Univerzita Karlova v Praze

Přírodovědecká fakulta / Faculty of Science

Studijní program: Organická chemie / Organic chemistry

Studijní obor: Organická chemie / Organic chemistry

ASWATHI CHAKRAPANI

Bioorthogonální reakce na DNA pro regulaci transkripce

Bioorthogonal reactions on DNA for regulation of transcription

Disertační práce

DISSERTATION

Školitel: Prof. Ing. Michal Hocek, CSc., DSc.

Praha 2022

Disertační práce byla vypracována na Ústavu organické chemie a biochemie, Akademie věd České Republiky v Praze v období září 2018 - červenec 2022.

Prohlášení

Prohlašuji, že jsem závěrečnou práci zpracovala samostatně a že jsem uvedla všechny použité informační zdroje a literaturu. Tato práce, ani její podstatná část, nebyla předložena k získání jiného nebo stejného akademického titulu.

V Praze, dne Podpis: M.S. Aswathi Chakrapani

Acknowledgment

First and foremost, my sincere gratitude goes to my supervisor Prof. Michal Hocek for the opportunity to work on such an interesting research topic, and for his guidance and trust. He is an amazing mentor and it would not have been possible without his support. Next, I would like to thank Dr. Libor Krásný for the productive collaboration, Dr. Viola Vaňková Hausnerová and Dr. Olatz Ruiz-Larrabeiti for the DNA transcription experiments. Many thanks to Dr. Radek Pohl for his help with the measurement and interpretation of NMR spectra, Ing. Kateřina Nováková (MALDI-TOF), and Ing. Martin Svoboda (LC-MS) for the measurement of mass spectra, to all the staff of IOCB who made a very constructive working environment.

I would like to thank all of my colleagues in the group for creating a positive working environment. First of all, to Dr. Zuzana Vaníková and Dr. Alessandro Panattoni who were my mentors at the beginning of my PhD and made the transition to an independent researcher quite easy for me; to my dear friends and colleagues Denise and Nemanja to whom I approached anytime and every time I had a doubt and who supported me immensely; to Filip, Ambra and David, with whom I started the PhD together and always helped each other out; to Dr. Michal Tichý and Dr. Veronika Sýkorová, for managing the lab so well and for their useful suggestions; to lab technician Tereza Schröpferová who made all our lives easy.

My sincere gratitude goes to my friends outside the group, firstly to my best friend and roommate Navyasree without whom I wouldn't be the same person I am today; Sarah who became my very close friend in a short period; to my friends outside work, Nithya, Balu who were family away from home. Together, you all helped to make the good days better and the terrible days tolerable.

I am forever grateful to my family for supporting me to do what I love. My Dad, who inspired me to pursue basic sciences, my Mom who always found the silver lining in every cloud, to my sister with whom I share a very special bond. Last but not least, to my then boyfriend and now husband Vaisakh who supported me and motivated me throughout the whole journey.

My PhD work was a part of a multidisciplinary project performed by Prof. Michal Hocek research group at the IOCB AS CR, partly in collaboration with Dr. Libor Krásný research group from the Institute of Microbiology of the Czech Academy of Sciences. All the chemical syntheses, enzymatic and chemical DNA incorporations, and irradiation experiments were performed by me

in person. The experiments mentioned in the thesis, which were performed by others are distinctively denoted at the beginning of each chapter and in the particular place of the thesis.

This work was supported by the Czech Science Foundation (20-00885X and 22-12023S) and by European Regional Development Fund; OP RDE (no. CZ.02.1.01/0.0/0.0/16_019/0000729).

Summary

This PhD thesis describes the design and synthesis of photocaged or glucosylated derivatives of epigenetic 5-(hydroxymethyl)pyrimidine-modified nucleotides and DNA using chemical and enzymatic methods and the studies on their regulation of gene expression in bacterial (*Escherichia coli* RNA polymerase) *in vitro* transcription level.

In the first part of the thesis, the design and syntheses of 5-(nitrobenzyloxymethyl)-2'-deoxyuridine (**dU**$^{\mathbf{NB}}$) and -cytidine (**dC**$^{\mathbf{NB}}$) phosphoramidites are described. These photocaged nucleoside phosphoramidite building blocks were used in the automated solid-phase synthesis of oligonucleotides (ONs) modified at specific positions. The ONs were used as forward primers in a polymerase chain reaction (PCR) to construct DNA templates modified at specific sites of the promoter region. The specific site photocaged DNA was then irradiated with light to result in the corresponding specific site 5-(hydroxymethyl)-modified DNA. Bacterial *in vitro* transcription studies of both the specific site photocaged and uncaged DNA were carried out. The incorporation of the photocaged epigenetic pyrimidine nucleotides at the -35 region of the promoter region of the template DNA inhibited transcription partially while the presence of the same outside the -35 region did not have any significant effect. Photochemical uncaging upon irradiation resulted in the corresponding 5-(hydroxymethyl)pyrimidine-modified epigenetic DNA, which restored the transcription to the level of natural, thus achieving a bioorthogonal switching ON of transcription. This way, specific site photoactivatable epigenetic labeling of DNA using a combination of chemical and enzymatic methods was demonstrated.

In the second part of the thesis, the syntheses of 5-(β-D-glucopyranosyloxymethyl)-2'-deoxyuridine (**dU**$^{\mathbf{Glc}}$) and -cytidine (**dC**$^{\mathbf{Glc}}$) 5'-*O*-triphosphates and phosphoramidites are described. The novel triphosphates were tested as substrates for several DNA polymerases in the enzymatic synthesis of major groove-modified DNA. In all cases (except for **dU**$^{\mathbf{Glc}}$**TP** with Pwo polymerase in PCR) the full-length amplified products were obtained showing that the glucosylated dNTPs are very good substrates for DNA polymerases. A series of modified DNAs with specific sequences were prepared by primer extension (PEX) and/or PCR for the studies of restriction endonuclease cleavage and bacterial *in vitro* transcription, respectively. The restriction endonuclease cleavage studies showed that the presence of any glucosylated pyrimidines fully protected DNA from cleavage by type II restriction endonucleases. But the transcription activity

of the glucosylated pyrimidines was found to be significantly different. While the presence of **dU^{Glc}** completely inhibited transcription by RNA polymerase (RNAP), DNA containing **dC^{Glc}**, although reduced the transcription to 50% of that of the 5hmC-modified DNA, still allowed transcription comparable to that of natural DNA. These results suggest the possible biological role of the glucosylation of hydroxymethylpyrimidine bases in DNA of some bacteriophages and protozoan parasites.

Consequently, enzymatic glucosylation of 5hmC-containing DNA using T4 phage *β*-glucosyltransferase (T4-βGT) was carried out to bioorthogonally tune the transcription activity. Finally, the glucosylated 5-(hydroxymethyl)-modified pyrimidine phosphoramidites were also synthesized, slightly optimizing the reported procedures, which can be used in the automated solid-phase synthesis of specific site glucosylated 5-(hydroxymethyl)-modified oligonucleotides.

Souhrn

Tato disertační práce popisuje návrh a syntézu derivátů nukleotidů a DNA nesoucích fotolabilní nebo glukosylované modifikace připojené k epigenetické značce 5-(hydroxymethyl)pyrimidinu, pomocí chemických a enzymatických metod a dále studium regulace genové exprese na bakteriální (*Escherichia coli* RNA polymeráza) *in vitro* transkripční úrovni.

V první části práce je popsán design a syntéza 5-nitrobenzyloxymethyl-2'-deoxyuridinových (**dU^{NB}**) a -cytidinových (**dC^{NB}**) fosforamiditů. Tyto nukleosidové fosforamiditové stavební bloky chráněné fotolabilní skupinou byly použity při automatizované syntéze oligonukleotidů (ONs) na pevné fázi modifikovaných ve specifických polohách. ONs byly následně použity jako přímé primery v polymerázové řetězové reakci (PCR) pro konstrukci **dU^{NB}**- a **dC^{NB}**-modifikovaných DNA templátů v konkrétních pozicích promotorové oblasti. Značená DNA byla poté v této oblasti ozářena světlem, které způsobilo uvolnění NB-skupin a vedlo ke vzniku odpovídajících specifických míst nesoucích volnou 5-(hydroxymethyl)ovou (5hm) modifikaci. Následně byly provedeny bakteriální *in vitro* transkripční studie modifikované i přirozené DNA. Inkorporace epigenetických pyrimidinových nukleotidů **dU^{NB}** a **dC^{NB}** nesoucích fotosenzitivní chránící skupiny do promotorové oblasti (-35) templátové DNA vedla k částečné inhibici transkripce, zatímco přítomnost stejných modifikací v pozici mimo promotorovou oblast nevykazovala žádný významný účinek na transkripci. Následnou fotochemickou reakcí došlo k odštěpení fotolabilní skupiny, která vedla ke vzniku DNA nesoucí odpovídající epigenetickou 5-(hydroxymethyl)pyrimidinovou modifikovaci, což způsobilo obnovení transkripce na úrovni přirozené DNA a tím dosažení bioortogonálního zapnutí transkripce. Pomocí kombinace chemických a enzymových metod tak byl potvrzen vznik specifických foto-aktivních míst DNA epigenetickým značením této DNA.

V druhé části práce jsou popsány syntézy 5-(β-D-glukopyranosyloxymethyl)-2'-deoxyuridin (**dU^{Glc}**) a -cytidin (**dC^{Glc}**) 5'-*O*-trifosfátů a fosforamiditů. Nové trifosfáty byly testovány jako substráty pro několik DNA polymeráz za použití enzymové syntézy DNA modifikované ve velkém žlábku. Ve všech případech (kromě **dU^{Glc}TP** s Pwo polymerázou v PCR) byly získány amplifikované produkty plné délky, ukazující, že glukosylované dNTPs jsou velmi dobré substráty pro DNA polymerázy. Série modifikovaných DNA se specifickými sekvencemi byla připravena metodou prodlužování primeru (PEX) a/nebo PCR a použita pro studium štěpení restrikčními

endonukleázami a následně i pro studium bakteriální *in vitro* transkripce. Experimenty zahrnující štěpení restrikčními endonukleázami ukázaly, že přítomnost jakýchkoli glukosylovaných pyrimidinů plně chránila DNA před štěpením restrikčními endonukleázami typu II. Zároveň však bylo zjištěno, že transkripční aktivita glukosylovaných pyrimidinů se významně liší. Zatímco přítomnost **dU^{Glc}** modifikace zcela inhibovala transkripci RNA polymerázou (RNAP), DNA obsahující **dC^{Glc}**, redukující transkripci na 50 % transkripce 5hmC-modifikované DNA, stále umožňovala transkripci srovnatelnou s přirozenou DNA. Následně byla provedena enzymová glukosylace DNA obsahující 5hmC pomocí T4 fágové glukosyltransferázy (T4-βGT) za účelem bioortogonálního přepínání transkripční aktivity. Tyto výsledky naznačují možné biologické role glukosylace hydroxypyrimidinových bází v DNA některých bakteriofágů a protozoových parazitů.

Nakonec byly také syntetizovány glukosylované 5-(hydroxymethyl)-modifikované pyrimidinové fosforamidity, s mírně optimalizovanými již popsanými postupy, které lze použít při automatizované syntéze na pevné fázi specifických glukosylovaných 5-(hydroxymethyl)-modifikovaných oligonukleotidů.

List of publications of the author related to the thesis

1. **A. Chakrapani**, V. Vanková Hausnerová, O. Ruiz-Larrabeiti, R. Pohl, L. Krásný, M. Hocek: Photocaged 5-(hydroxymethyl) pyrimidine Nucleoside Phosphoramidites for Specific Photoactivatable Epigenetic Labelling of DNA. *Org. Lett.* **2020**, *22*, 9081-9085.

2. **A. Chakrapani**, O. Ruiz-Larrabeiti, R. Pohl, M. Svoboda, L. Krásný, M. Hocek: Glucosylated 5-hydroxymethylpyrimidines as epigenetic DNA bases regulating transcription and restriction cleavage. *Chem. Eur. J.* **2022**, *28*, e202200911.

Table of Contents

1. Introduction

1.1. DNA - Discovery, Structure, and Function

DNA is the most basic molecule in life which serves as the genetic substance of practically all living organisms on the planet. Johann Friedrich Miescher, a Swiss scientist, was the first to discover DNA in 1869.[1] He discovered a unique molecule from the nucleus of leukocyte cells that functioned differently from proteins and named it "Nuclein". After entirely separating the non-protein component, Richard Altmann renamed Nuclein to Nucleic acid in 1889.[2] Later research by Albrecht Kossel led to the identification of the fundamental building blocks of the DNA- the nitrogen bases (adenine, guanine, cytosine, and thymine), deoxyribose sugar, and phosphoric acid, which gave DNA its current chemical designation of deoxyribonucleic acid (DNA).[3] In the following decade, Erwin Chargaff conducted a series of studies known as Chargaff's laws, proving that while DNA differs between species, the molar ratios of cytosine to guanine and adenine to thymine remain close to one.[4-6] Later works by Rosalind Franklin and Maurice Wilkins along with the other discoveries, contributed greatly to Watson and Crick's derivation of the double-helical structure of DNA[7] (Figure 1).

The DNA macromolecule exist as two antiparallel strands coiled around one another in the form of a double helix. These two DNA strands are known as polynucleotides, and they consist of monomeric units known as nucleotides. Each nucleotide consists of a sugar termed deoxyribose, a phosphate group, and one of the four nitrogen-containing nucleobases (cytosine (C), guanine (G), adenine (A), or thymine (T)). The nitrogen base and sugar together constitute the nucleoside. Covalent connections (referred to as the phosphodiester linkage) between the sugar of one nucleotide and the phosphate of the following nucleotide, link the nucleotides together, resulting in an alternating backbone of sugar and phosphate. Two of the nitrogenous bases are complementary to each other and are classified into two groups, pyrimidines, and purines. In DNA, thymine and cytosine constitute the pyrimidines; adenine and guanine constitute the purines. There exist hydrogen bonds between these complementary nucleobase pairs (A = T, C ≡ G) that hold the two DNA strands together. Other weaker bonding interactions, such as the π-π nucleobase stacking, also contribute to the overall structural stability.[8]

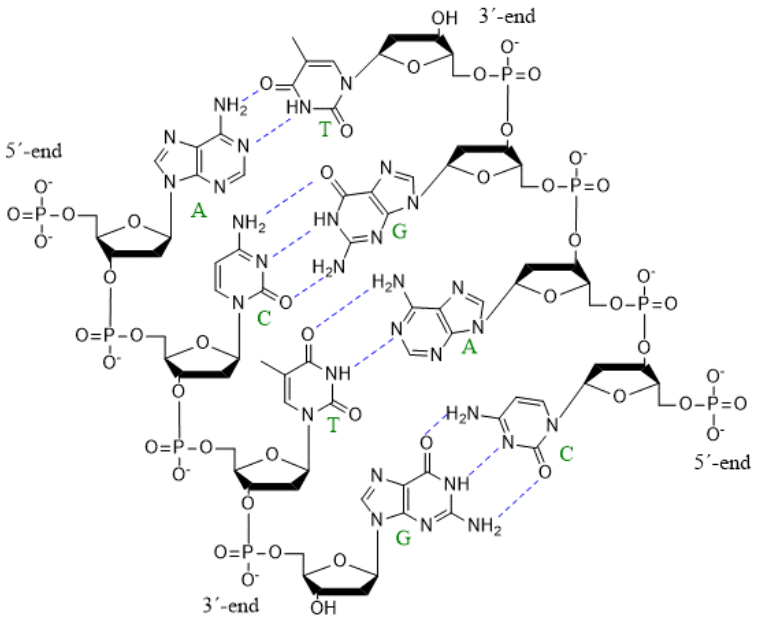


Figure 1: Schematic representation of Watson-Crick base pairing of nucleobases in DNA

The DNA double helix is found to exist in a variety of conformers. The three major conformers are A-DNA, B-DNA, and Z-DNA. The B-DNA helix which is the most prevalent confirmation in most living cells, twists in a right-hand direction.[9] It has two essential locations for the interactions with the proteins: a large major groove and a small minor grove. The A-DNA forms a right-handed duplex with closer spacing of base pairs.[10] The two strands of Z-DNA coil in left-handed helices in a distinct zig-zag pattern, forming a significantly different duplex structure[11] (Figure 2).

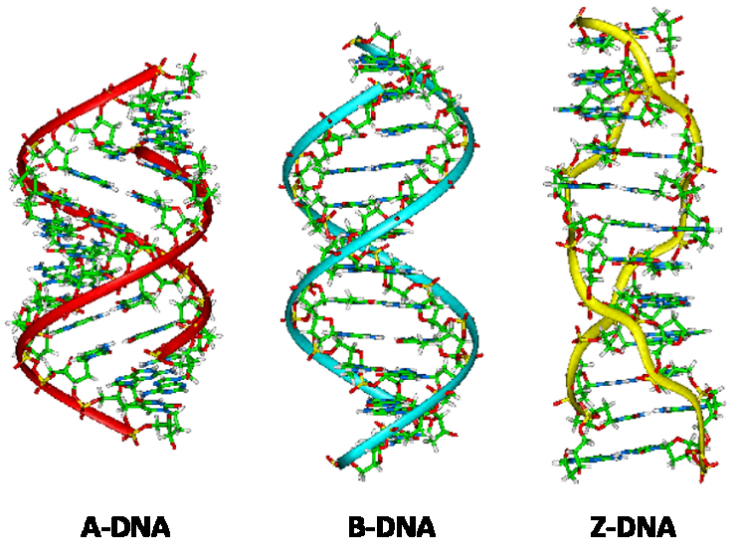


Figure 2: A-DNA, B-DNA and Z-DNA conformers of DNA (Taken from snapsolve.com)

DNA is organized into functional units called genes based on the nucleotide sequences. Each gene has unique biological information that can be duplicated through the replication process. The DNA template can be transcribed into RNA, which in turn can be translated into specific amino acid sequences that code for a certain protein. This transfer of genetic information within a biological system is illustrated as the central dogma of molecular biology [12] (Figure 3).

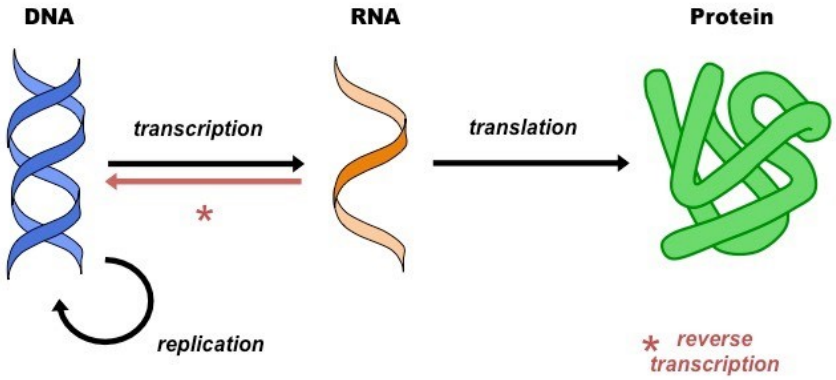


Figure 3: Schematic illustration of the central dogma of molecular biology
(Taken from https://ib.bioninja.com.au)

1.2. Non-canonical nucleobases

The four nucleobases A, G, C and T are the developing blocks that constitute the basic code contained in DNA and the unique linear arrangement of which comprises the genetic information layer. In addition to the four canonical nucleobases that make up the genomic DNA, there exists naturally occurring variation of these canonical nucleobases (non-canonical bases) that has the ability to affect the chromosomal shape and gene expression. These modifications can add chemical functionality to double helical major groove without interfering with the Watson–Crick pairing. Therefore, they do not change the specificity of base pairing but can interact with cellular and viral encoded proteins and can regulate a broad range of different biological pathways. These non-canonical nucleobases that can alter the phenotypic expression of the genome without altering the genotype, are referred to as epigenetic bases. The significant advancement in analytical tools and methodologies in recent years has enabled the genome-wide mapping of these bases and contributed to their discovery and functional studies in both prokaryotes and eukaryotes.[13-17]

Selected examples of some of the non-canonical pyrimidine nucleobases discovered so far are summarized in Figure 4.

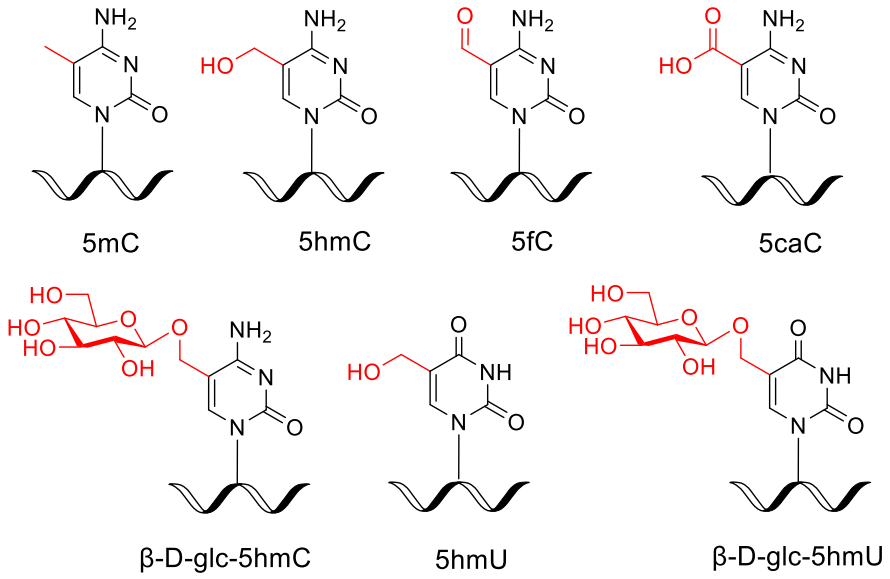


Figure 4: Chemical structures of selected non-canonical pyrimidine nucleobases discovered in DNA. 5mC, 5-methylcytosine; 5hmC, 5-(hydroxymethyl)cytosine; 5fC, 5-formylcytosine; 5caC, 5-carboxycytosine; β-D-glc-5hmC, β-D-glucosyl-5-(hydroxymethyl)cytosine; 5hmU, 5-(hydroxymethyl)uracil; β-D-glc-5hmU, β-D-glucosyl-5-(hydroxymethyl)uracil.

1.2.1. Modified cytosine nucleobases

5mC

5-Methylcytosine (5mC) was the first variant of a non-canonical nucleobase to be discovered and probably the most studied.[18, 19] Later, it was recognized to have regulatory roles.[20-23] Numerous locations in bacteria contain 5mC, which is frequently employed as a flag to prevent native methylation-sensitive restriction enzymes from cleaving DNA.[24, 25] It is discovered to be predominately present in CpG dinucleotide sequences in eukaryotes, where it, along with histone modifications, is essential for controlling gene expression. The majority of plants, animals, and fungi have 5mC, which significantly affects the integrity of the genome, gene expression, and development.[26-28] For instance, transcription is directly repressed when DNA methylation is present in the promoter region.[29] DNA methyltransferases (DNMTs) deposit methyl groups on cytosine's

position 5 while employing *S*-adenosyl-methionine as the methyl donor to create 5mC [30] (Figure 5).

The oxidized congeners of 5mC: 5hmC, 5fC, and 5caC

It is found that cytosine methylation in DNA is dynamic and reversible.[31, 32] It's been shown that the Ten-Eleven-Translocation (TET) enzymes[33] are responsible for the active demethylation of 5hmC by oxidizing it to 5-(hydroxymethyl)cytosine (5hmC)[34, 35], 5-formylcytosine (5fC)[36], and 5-carboxycytosine (5caC)[37]. The latter two can go through base excision repair (BER) with the help of glycosylases like thymine-DNA glycosylase (TDG), which would eventually recover the natural cytosine.[38-40] Additionally, the latter two may be passively diluted by cell division to the unmethylated stage [41] (Figure 5). Accumulated evidence currently suggests that these oxidative derivatives might also have significant roles in the regulation of gene expression rather than being just intermediates in active demethylation.[42-44] It was also found to have significant effects on DNA flexibility.[45-46]

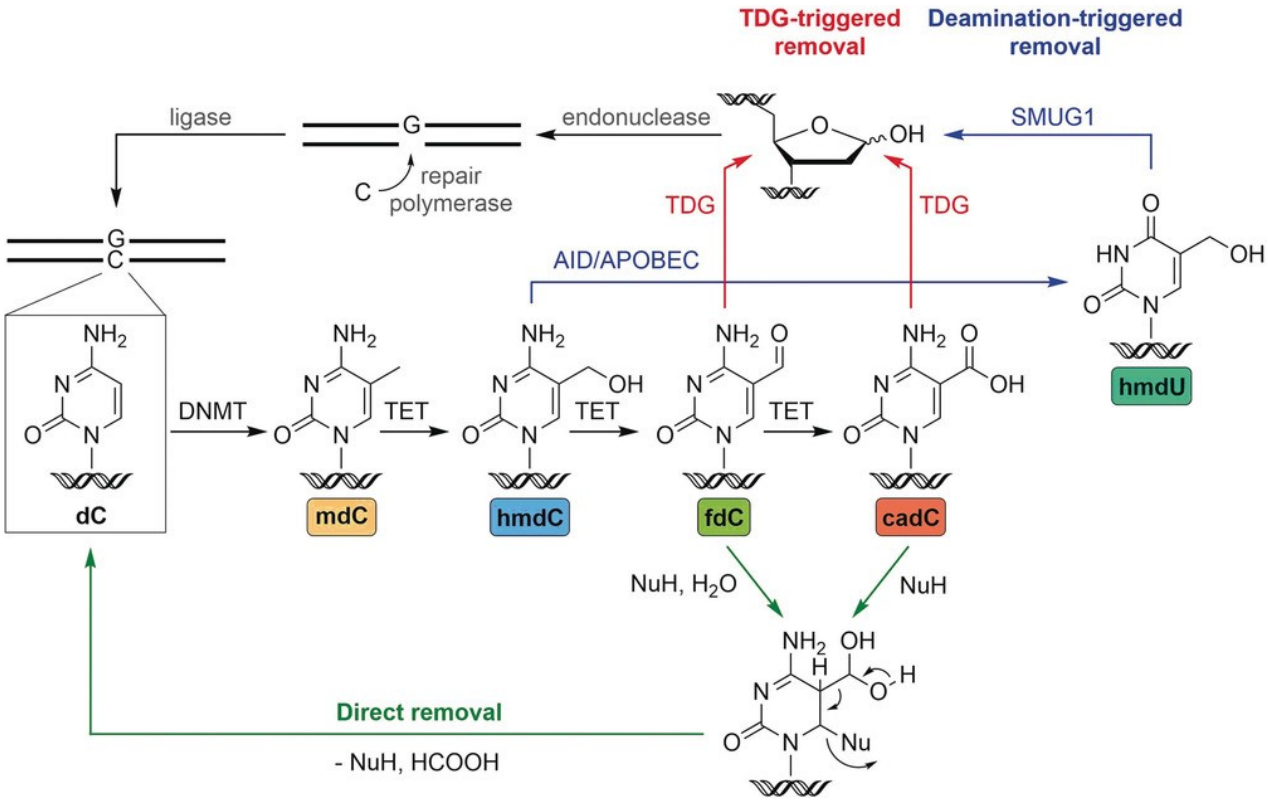


Figure 5: Schematic illustration of the proposed active methylation and demethylation pathway (Taken from *Angew. Chem. Int. Ed.* **2018**, *57*, 4296– 4312).

5hmC

Of the oxidized congeners of 5mC, the most prevalent and rather stable one was found to be 5hmC. It was first discovered in bacteriophages, where all the cytosine bases in the DNA of T-even bacteriophages, such as T4 bacteriophage, are substituted with 5hmC[47, 48], which may then be glucosylated to prevent the phage DNA from being attacked down by bacterial restriction enzymes.[49] It was later reported in mammalian tissues as well[50, 51]: human and mouse brains[52-54], embryonic stem cells and in high concentrations in growing neurons.[55, 56] It could be recognized as an epigenetic mark that controls the chromatin structure and transcription.[57, 58] In addition, many kinds of human cancer cells were found to have low levels of 5hmC.[59, 60] Recently, Hocek group reported that 5hmC could enhance the *in vitro* transcription by bacterial RNAP. [61]

β-D-glc-5hmC

The T-even phages which contain the 5hmC base can further glucosylate it to form the β-D-glc-5hmC.[49] They encode a DNA modifying enzyme called β-glucosyltransferase (β-GT) that catalyzes the glucose transfer from uridine diphosphoglucose (UDP-glc) to 5hmC in double-stranded DNA.[62-64] The infecting viral DNA is shielded from host restriction enzymes by this glucosylation. It also allows the phage to turn on viral endonucleases, which can digest the unaltered host DNA.[49] The glucosylation has been linked to the regulation of phage-specific gene expression in addition to its protective function, by affecting transcription both *in vivo* and *in vitro*.[65, 66] Even though β-GT does not show much sequence specificity, it specifically modify only 5hmC bases.

1.2.2. Modified thymine nucleobases

The homologous complementary nucleotide for adenine (A) in DNA is thymine (T) while it is uracil (U) in RNA. However, DNA has been found to contain trace amounts of 2'-deoxyuridine (dU), a nucleotide that lacks a methyl group at the C5 position. It is likely to be caused by cytosine deamination resulting in a U: G mispair. Several repair enzymes can remove uracil from DNA via the BER pathway. [67] Another alteration, 5-formyluracil (5fU), a byproduct of thymine oxidative damage also exists and can generate mutations in DNA as it can base pair with both A and G.[68, 69]

5hmU

The genomic DNA of several creatures, ranging from bacteriophages to mammals, contains another oxidized thymine modification, 5hmU.[70, 71] Thymine can be oxidized or hydroxylated by TET proteins[72] or reactive oxygen species (ROS) [73] to produce the 5hmU base (Figure 6).

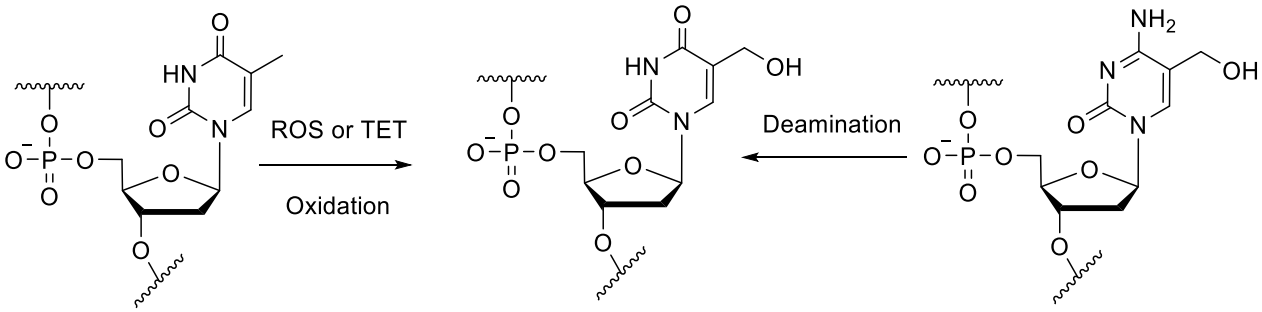


Figure 6: Schematic illustration of the proposed ways of 5hmU formation

However, it was recently reported that 5hmC could be deaminated to produce the 5hmU base in which case it is mispaired with guanine (hmU: G). The latter provides an alternate pathway for active demethylation by following a mismatch repair. Smug1 TDG has the ability to cleave the hmU: G mismatches from the DNA along with other oxidized nucleotides (5fC or 5caC) and restore them with natural deoxycytidine (Figure 5, 6).[74, 75] The presence of 5hmU in blood DNA was examined as a potential breast cancer marker.[76, 77] The binding of the chromatin remodeling proteins as well as the transcription factors can be affected by this base, indicating a unique function in stem cells.[72, 78] It was also found to alter the flexibility and hydrophilicity of DNA molecules. [79] Additionally, Hocek group reported that 5hmU can enhance the *in vitro* transcription by bacterial RNAP. [61]

β-D-glc-5hmU

β-D-glucopyranosyloxymethyluracil (base J), has been found in the kinetoplastid protozoan parasites, human pathogenic Trypanosoma[80] or Leishmania[81] where it replaces 1% of thymidine. It is the first hypermodified base found in eukaryotic DNA. It is mostly present in repetitive DNA sequences, with telomeric DNA repeats having the highest concentration.[82-84] J is biosynthesized in two stages: T is converted to 5hmU by the TET homologue enzymes JBP1 and JBP2, some of which are then glycosylated to produce base J by J glycosyltransferase (JGT). [85] Base J is crucial for the right termination of transcription in these organisms.[81, 86, 87]

1.3. Synthesis of functionalized or modified DNA

DNA is an interesting tool for a variety of applications since it is easily synthesized and has a wide range of capabilities. The functional range of these applications is constrained by the small chemical repertory of natural nucleic acids. To get around these limitations, many oligonucleotides have been developed with modifications incorporated to the nucleobase, ribose sugar, and/or phosphate backbone [88-90] (Figure 7).

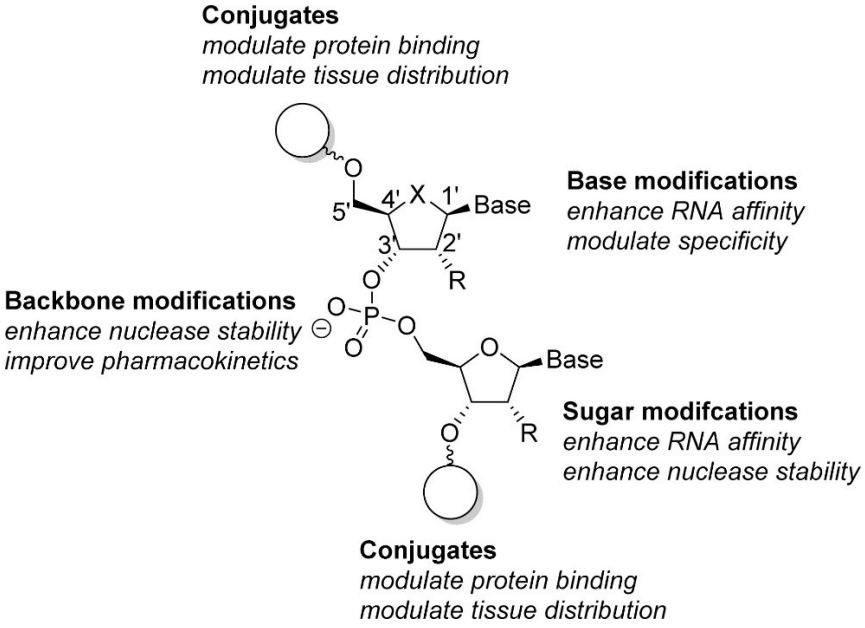


Figure 7: Schematic illustration of a dinucleotide unit and the possible sites for modification (Taken from *J. Med. Chem.* **2016**, *59*, 9645−9667)

Modifications of the sugar moiety bring added functionality and help in stabilizing secondary structures, boosting binding affinity for target sequences. It can improve nuclease resistance, bioavailability, and toxicity. Chemically modifying/replacing the phosphodiester bond between the nucleotides can improve the oligonucleotide in vivo stability. Such sugar and backbone modifications can be combined to produce non-native nucleic acids with increased functionality and resistance to nuclease-mediated degradation, opening up a larger range of uses.

1.4. Synthesis of base modified nucleosides and DNA

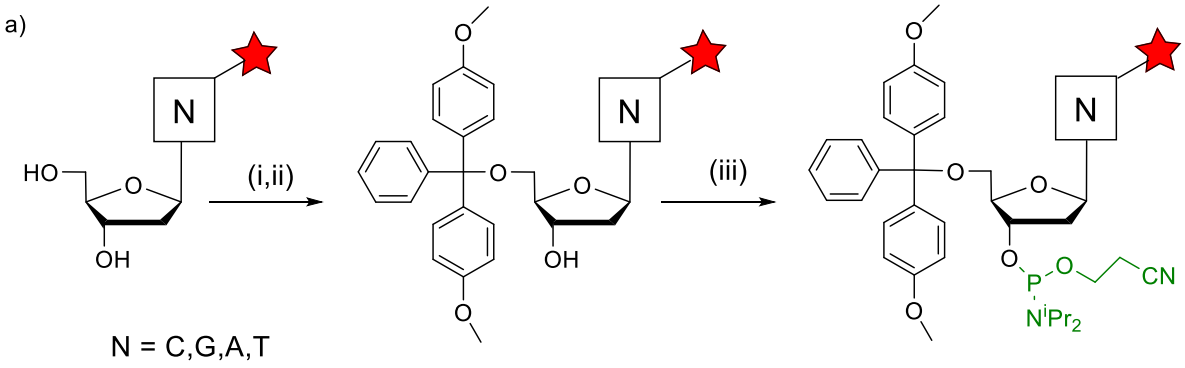
The nucleobase is a lot more favoured location for adding functional groups to nucleic acids compared to the sugar and backbone sites, because it doesn't considerably change the Watson-Crick pattern of recognition and doesn't interfere with the genetic data included in DNA sequences.[91] Furthermore, DNA polymerases, which enable the effective enzymatic synthesis of base modified DNA templates, generally tolerate them. They can be also converted to their corresponding developing blocks suitable for the chemical synthesis of short oligonucleotides. Gene silencing and duplex thermal stability can both be enhanced by such modified nucleobases. They also make it possible to add fluorescent or redox indicators, communicate extra reactivity, and expand the genetic alphabet.[92-94] They are typically made via enzymatic integration of functionalized 2'-deoxynucleoside 5'-*O*-triphosphates (enzymatic synthesis) or by solid phase synthesis of 2'-deoxynucleoside 3'-phosphoramidites (chemical synthesis).

1.5. Chemical synthesis of DNA

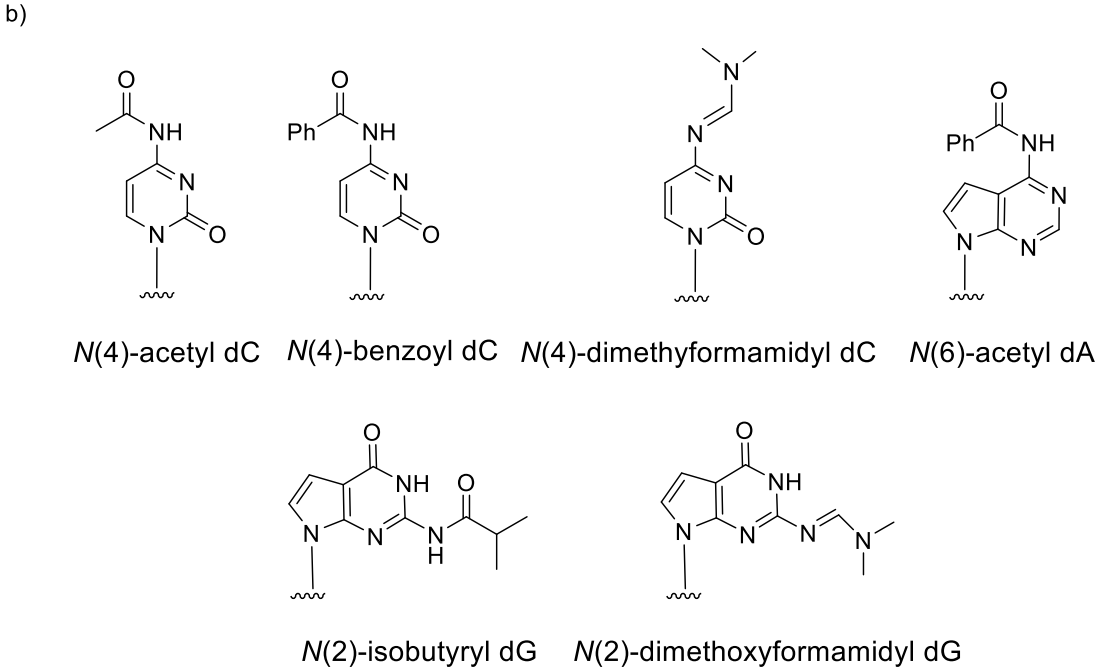
The most well-established method for producing short oligonucleotides in a large scale (microgram) is the solid phase DNA synthesis (up to 100 mer).[95] Multiple methods have been developed for the chemical synthesis that are H-phosphonate [96, 97], phosphodiester [98], phosphotriester [99, 100], and phosphite triester [101-103] methods, each with their own advantages and drawbacks. The phosphite triester method, which employs 3'-*O*-chlorophosphites as building blocks, was soon improved by M. Caruthers group, who used the more stable nucleoside 3'-*O*-phosphoramidites [104-106] as building blocks on solid phase. It remains the most reliable and favoured method for the automated synthesis of oligonucleotides. Unlike the enzymatic synthesis, the solid phase synthesis of oligonucleotides takes place in the 3'-5' direction.

1.5.1. Synthesis of base modified nucleoside phosphoramidites

The phosphoramidite approach utilizes the nucleoside phosphoramidites where the 5'-OH of the nucleoside is protected with a dimethoxytrityl (DMTr) group, the 3'-OH with a phosphoramidite moiety (Figure 8a). In addition, for the bases except thymidine, the reactive primary amines are protected with bioorthogonal protecting groups (Figure 8b).



(i) DMTrCl, Pyridine, r.t. (ii) Protection of amino group (except for T)
(iii) 2-Cyanoethyl *N,N*-diisopropylchlorophosphoramidite, DIPEA, DCM, r.t.



N(4)-acetyl dC *N*(4)-benzoyl dC *N*(4)-dimethyformamidyl dC *N*(6)-acetyl dA

N(2)-isobutyryl dG *N*(2)-dimethoxyformamidyl dG

Common protecting groups for the nucleobases

Figure 8: Synthetic scheme of 2'-deoxyribonucleoside phosphoramidites

1.5.2. Solid phase synthesis of oligonucleotides

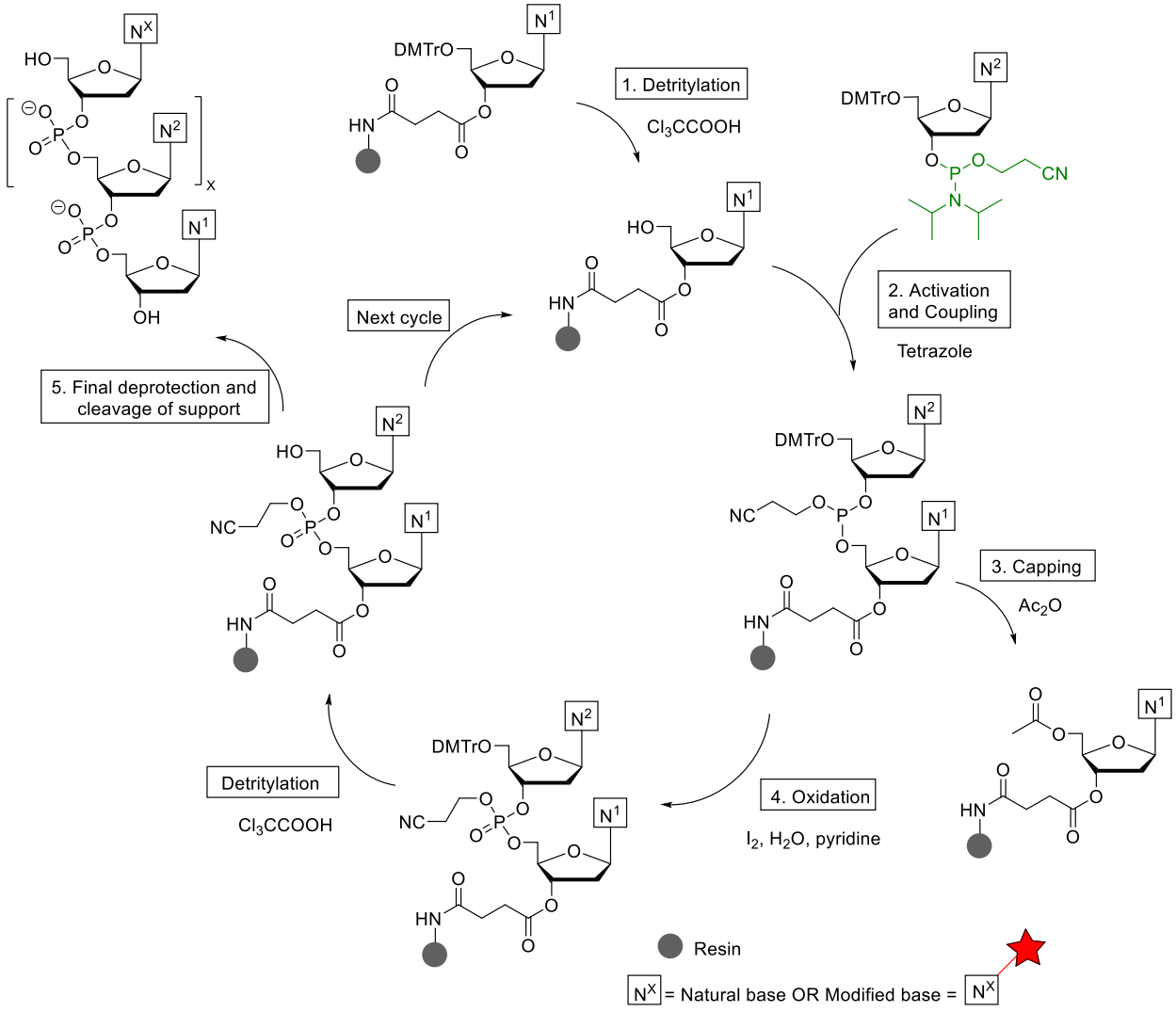


Figure 9: Scheme of the solid phase synthesis of oligonucleotides using phosphoramidite chemistry

The synthesis cycle consists of four steps:

First step is the deblocking/detritylation step where the deprotection of the DMTr group of the nucleoside attached to the solid support, take place. Next step is the activation/coupling. Here, the deprotected nucleoside reacts with the newly introduced protected nucleoside phosphoramidite. Tetrazole serves as an activator for this coupling. The next step is capping, where any unreacted deprotected nucleoside gets acetylated. This prevents the formation of any truncated oligonucleotides. Following cycles of reaction of these truncated oligonucleotides can produce near full-length oligonucleotides with internal deletions, unless they are inhibited. The next step is oxidation, where the oxidation of the unstable trivalent phosphite triester to the stable pentavalent phosphotriester, take place. Then the cycle is repeated, starting with the detritylation of the newly incorporated nucleoside. The final step of the entire synthesis involves the removal of the oligonucleotide from the solid support. In addition to the cleavage, the cyanoethyl groups, the amine protecting groups of the nucleobase are also removed (Figure 9).

This approach is ideal to introduce one or several modified bases at specific sites. However, the functional group range is frequently a limiting element for the chemistry of phosphoramidites. Moreover, it becomes ineffective for the synthesis of longer oligonucleotides (> 100 mer) due to overall low yield.

1.6. Enzymatic synthesis of DNA

The enzymatic synthesis is an alternate route to the synthesis of DNA especially for the synthesis of longer oligonucleotides (>100 mer) and also for cases where the modifications are not compatible with the phosphoramidite chemistry. It can be used for single or multiple incorporations of the modifications but the synthetic scale is usually in nanograms. The enzymatic synthesis involves the incorporation of natural/ functionalized nucleoside triphosphates (nucleotides) by various DNA polymerases in the 5'-3' direction.

1.6.1. Synthesis of base modified 2´-deoxyribonucleoside triphosphates (dN^{X}TPs)

For the manufacture of modified triphosphates, several approaches have been devised. Of them, two most commonly used ones are:

(1) **Cross coupling approach** which involves the reactions of halogenated nucleotides with a functional group (usually sensitive to phosphorylation conditions) catalyzed by palladium (Pd). [107, 108] The most utilized Pd catalyzed reactions for this cross-coupling approach are mentioned below. Here, the functional group/ modification is incorporated into the nucleobase with the help of a linker. For Sonagashira cross coupling, the linker is a triple bond, in case of Suzuki, it is an aryl group and for Heck reaction, it is a double bond (Figure 10). The major advantage of these reactions is the tolerance to a wide range of reactive functional groups.

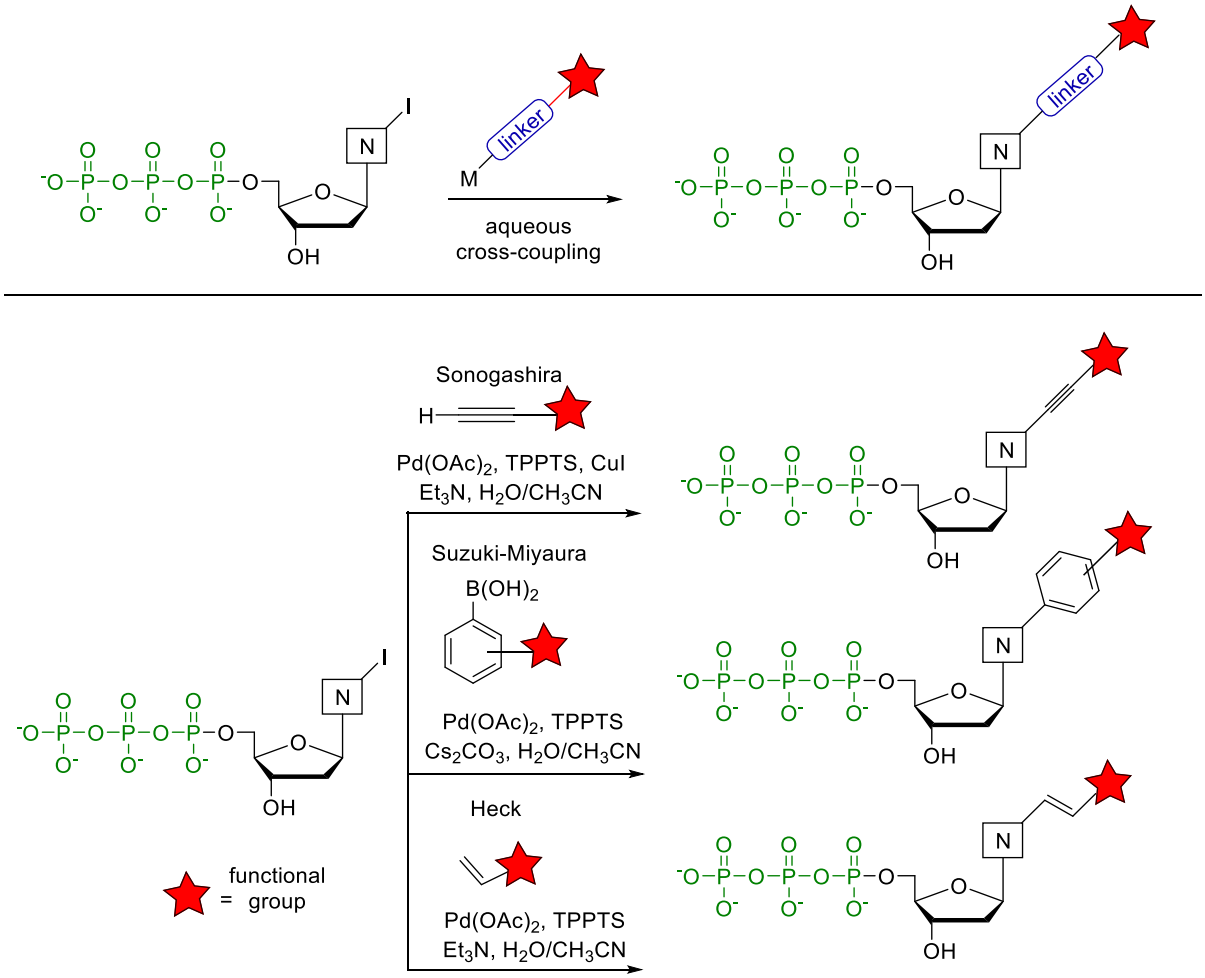


Figure 10: Cross-coupling approach for the synthesis of **dN^{X}TPs**

(2) **Phosphorylation approach** where chemically modified nucleosides (that are stable under the phosphorylation conditions) are phosphorylated directly. The modifications are introduced to the nucleobase either at the nucleobase level or nucleoside level and are useful in cases where it needs to be attached directly without any linker moiety. In contrast to the cross-coupling approach, the major limitation of this approach is the limited tolerance to various functional groups (Figure 11).

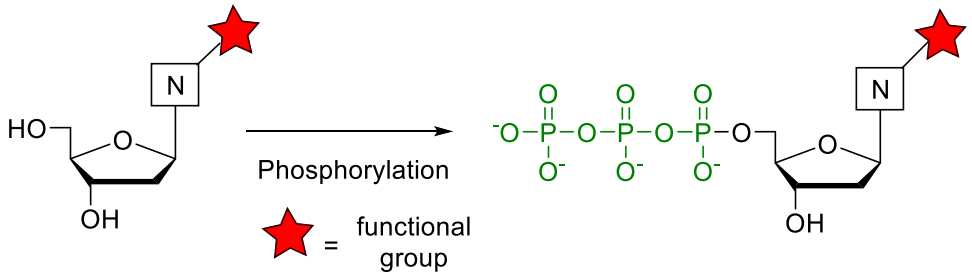


Figure 11: Phosphorylation approach for the synthesis of **dN^{X}TPs**

Two popular phosphorylation techniques that exist in the literature are:

(A) Yoshikawa method[109, 110]:

This approach is the most common and straightforward reaction for the preparation of (modified) dNTPs. Selective 5'-monophosphorylation of unprotected nucleosides using $POCl_3$ gives the corresponding phosphorodichlorate derivative. It is then possible to hydrolyze this intermediate to get the 2'-deoxynucleoside monophosphate (dNMP). It can also combine with the pyrophosphate in situ to form the cyclic triphosphate, which is then hydrolyzed to yield the target triphosphate moiety (dNTP). Both monophosphates and triphosphates can be synthesized by this approach (Figure 12).

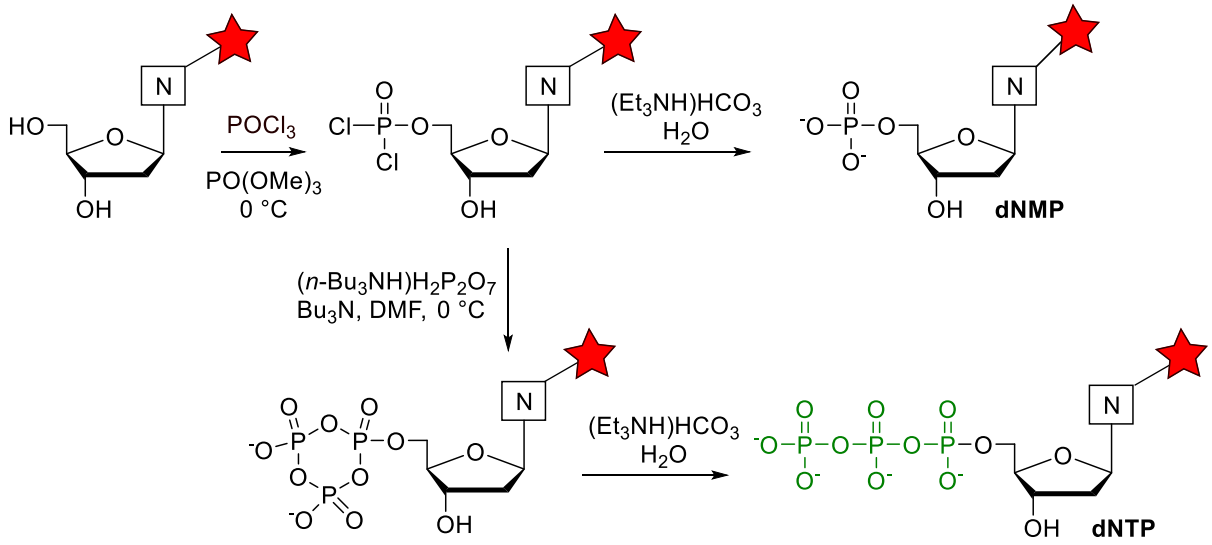


Figure 12: Yoshikawa method of phosphorylation

(B) Ludwig and Eckstein method [111]:

This methodology involves the creation of a 3'-*O*-acetylated nucleoside derivative, which reacts with salicyl phosphorochloride to produce an activated phosphite intermediate. The cyclic nucleoside triphosphate is generated after two nucleophilic substitution processes and is then oxidized to form the desired triphosphate (Figure 13).

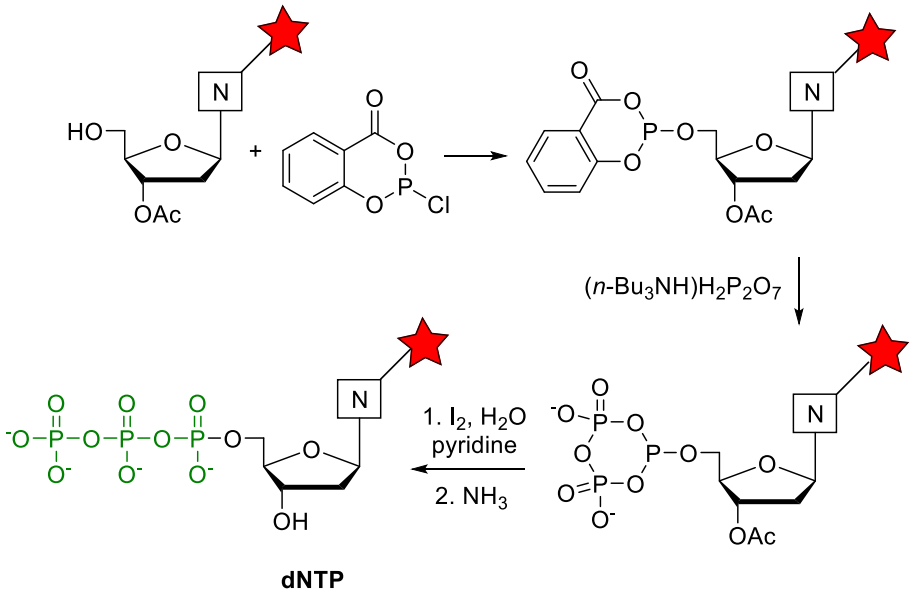


Figure 13: Ludwig and Eckstein method of phosphorylation

1.6.2. Modified nucleoside triphosphates as substrates for DNA polymerases

The position 5 of pyrimidines and position 7 of deazapurines (purines with the nitrogen at position 7 replaced by carbon) are the most suitable sites for modifications as the nucleotides modified this way serve as good substrates for DNA polymerases. This placement will result in the modifications projecting out of the DNA major groove enabling them for further interactions. In addition, it doesn't interrupt the successful Watson-Crick base pairing (Figure 14).

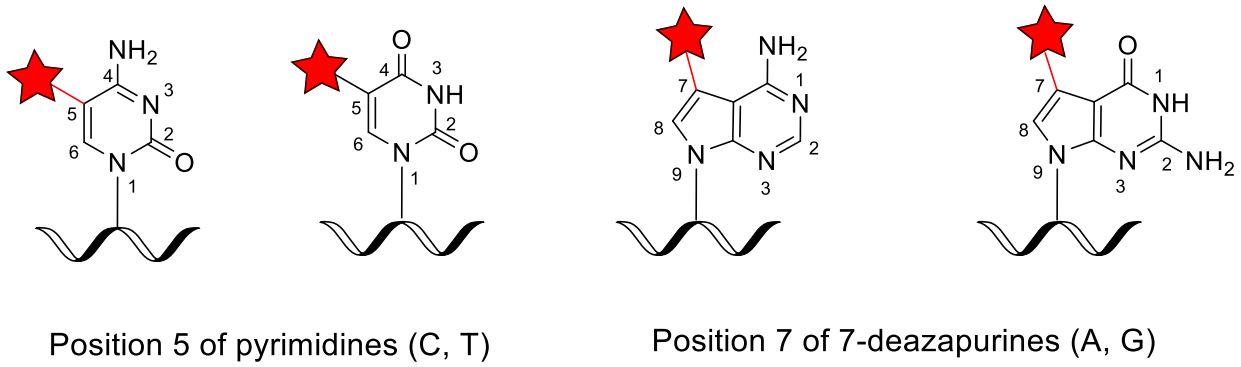


Figure 14: Nucleobase modification positions accepted by DNA polymerases in enzymatic incorporation

Representative examples of the modifications incorporated to DNA using the polymerase incorporation of such base modified nucleoside triphosphates reported from the Hocek group are shown below (Figure 15).

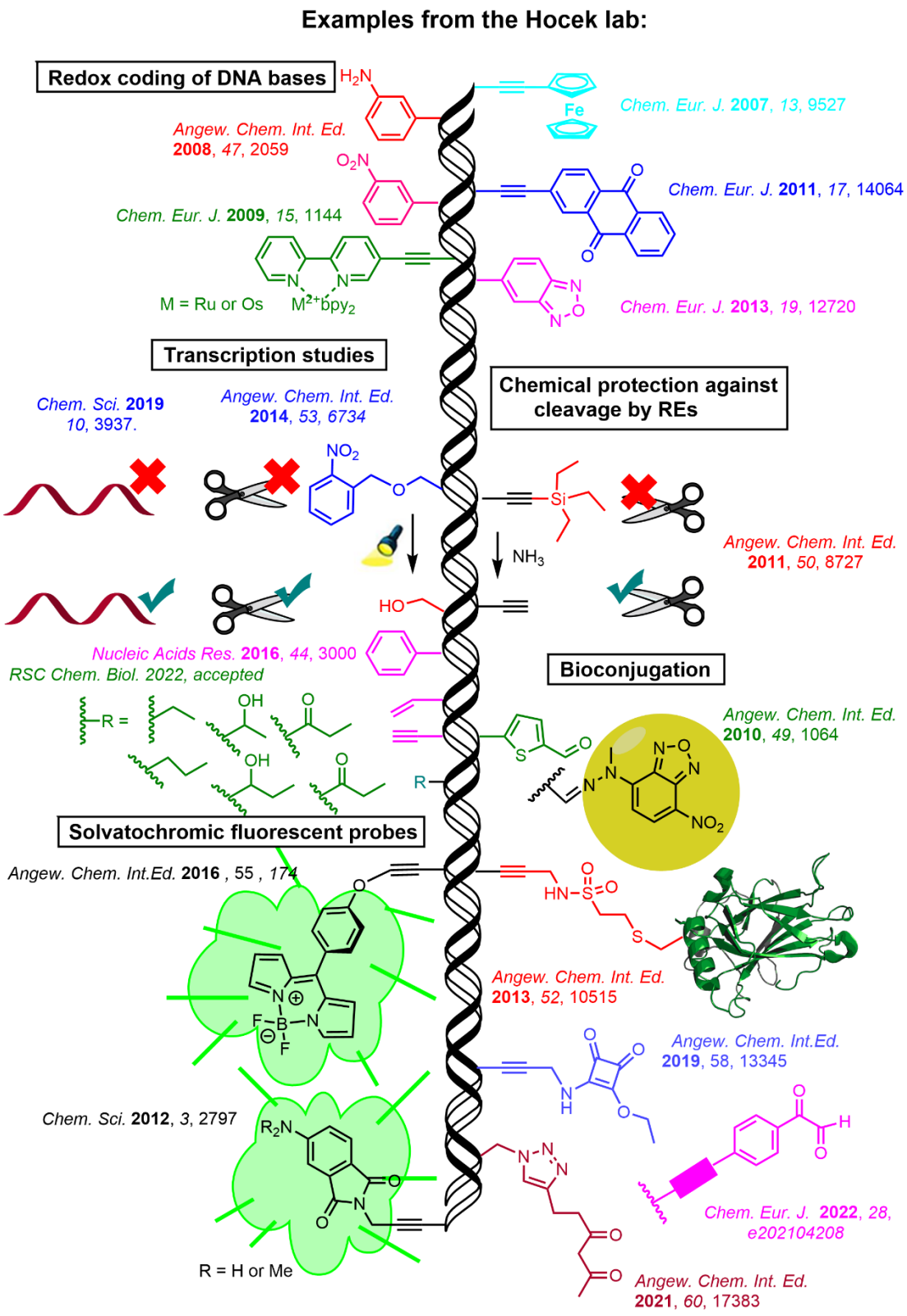


Figure 15: Schematic illustration of a selected set of nucleobase modifications enzymatically incorporated into DNA and their applications

1.6.3. Enzymatic synthetic methods

PEX and PCR are two common methods for the enzymatic DNA synthesis. The enzymatic synthesis involves a template, a DNA polymerase, primers (oligonucleotides that are complement to the target DNA sequences) and dNTPs. The primer is annealed to the corresponding template and the polymerase elongates the 3' end of the primer using the available dNTPs. The basic requirements for the enzymatic synthesis of modified DNA are the tolerance of the modified nucleotides by DNA polymerases and the successful subsequent incorporation of other nucleotides.

Primer Extension Reaction (PEX)

The PEX reaction is used to make small DNA fragments (up to 100 bp) with the template strand unmodified and the primer strand bearing one or several modifications. The 5' end of the primer is typically labelled with a ^{32}P-phosphate group or some fluorescent probe to visualize the extended product. The reaction takes place at a suitable temperature for the DNA polymerase in question. Polyacrylamide gel electrophoresis is used to determine the denatured DNA product (Figure 16).

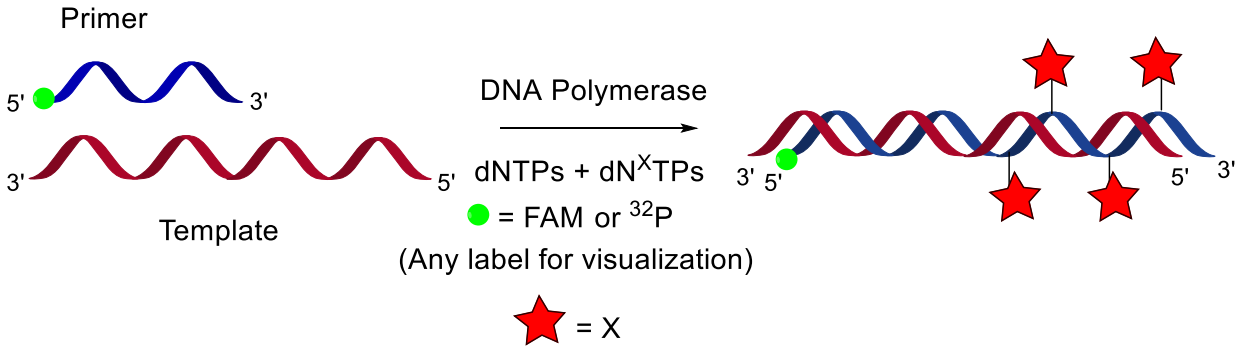


Figure 16: Schematic illustration of PEX

Polymerase Chain Reaction (PCR)

The PCR reaction is used to make longer DNA fragments (100 bp-1000 bp) with a higher frequency of modifications on both strands. [112] Thermal cycling is the basic principle behind the PCR. Reactants are repeatedly heated and cooled to accommodate the various temperature-dependent PCR operations. The DNA double helix is physically broken into its two strands at a high temperature during the first stage of a PCR, known as denaturation (usually 95 °C). The temperature is reduced in the second stage (known as annealing) depending on the melting point of the two primers (forward and reverse), allowing them to bind to the complementary DNA sequences. In the third step called elongation, the temperature is again altered resulting in the two DNA strands annealed to their complementary primers to act as templates for the DNA polymerase, creating a new DNA strand using the provided nucleotides as building blocks. The newly synthesized DNA now serves as a replication template for a chain reaction (30-40 cycles) resulting in an exponential amplification. The end of the PCR reaction is usually with a 72-75 °C final extension step. Agarose gel electrophoresis is generally used to analyze the amplified DNA. A fluorescent intercalating reagent (typically GelRed) is used for visualization. In addition to the preliminary requisites of an enzymatic DNA synthesis, the ability of the polymerase to replicate the existing modified DNA strands remain a main limitation for the PCR reaction (Figure 17).

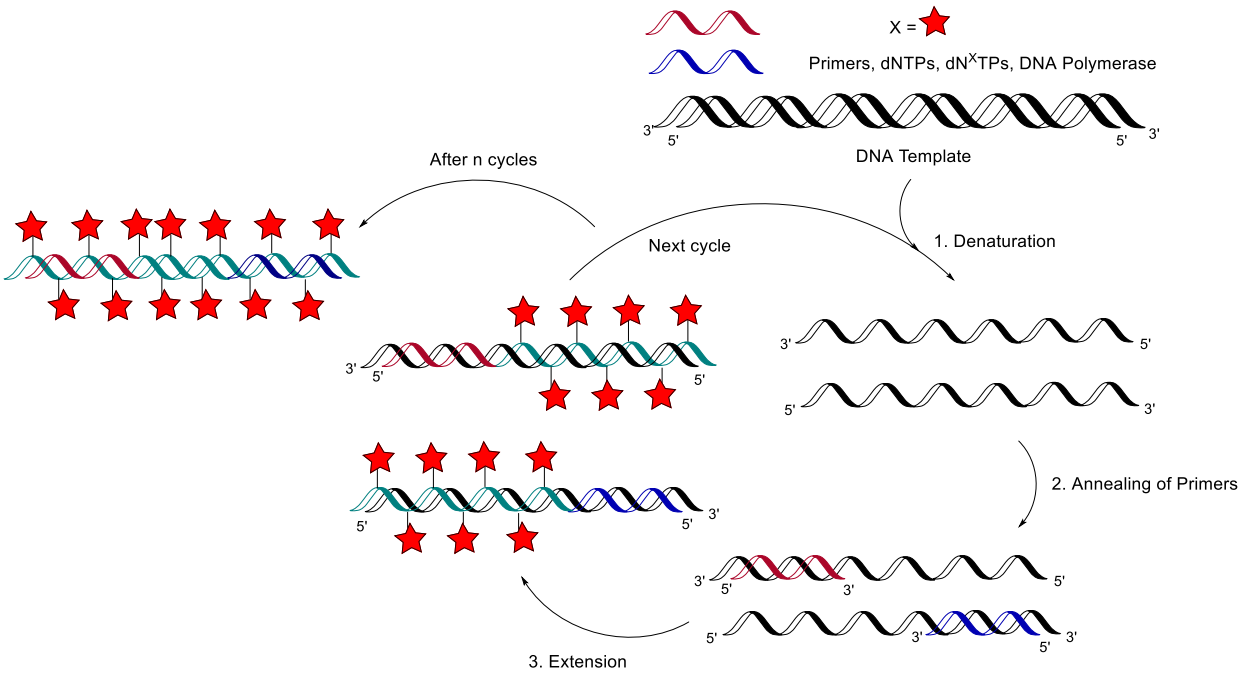


Figure 17: Schematic illustration of PCR

1.6.4. Enzymatic synthesis of modified single-stranded oligonucleotides

Unlike the solid phase DNA synthesis which generates oligonucleotides (single-stranded DNA), the enzymatic DNA synthesis usually generates double-stranded DNA. Many methods associated with PEX have been employed to generate the single-stranded modified oligonucleotides from the double-stranded DNA. One of the most convenient methods is magnetoseparation. In this approach, PEX is performed using a biotinylated template (5' or both 5' and 3') to synthesize the double-stranded DNA. It is then incubated with streptavidin magnetic beads upon which the biotin complexes with the streptavidin unit resulting in a DNA-streptavidin complex attached to the magnet. All unbound components are then washed away. Upon denaturation of the double-stranded DNA, the biotinylated template remains bound to the magnetic particles while the modified single strand is released into the solution (Figure 18).

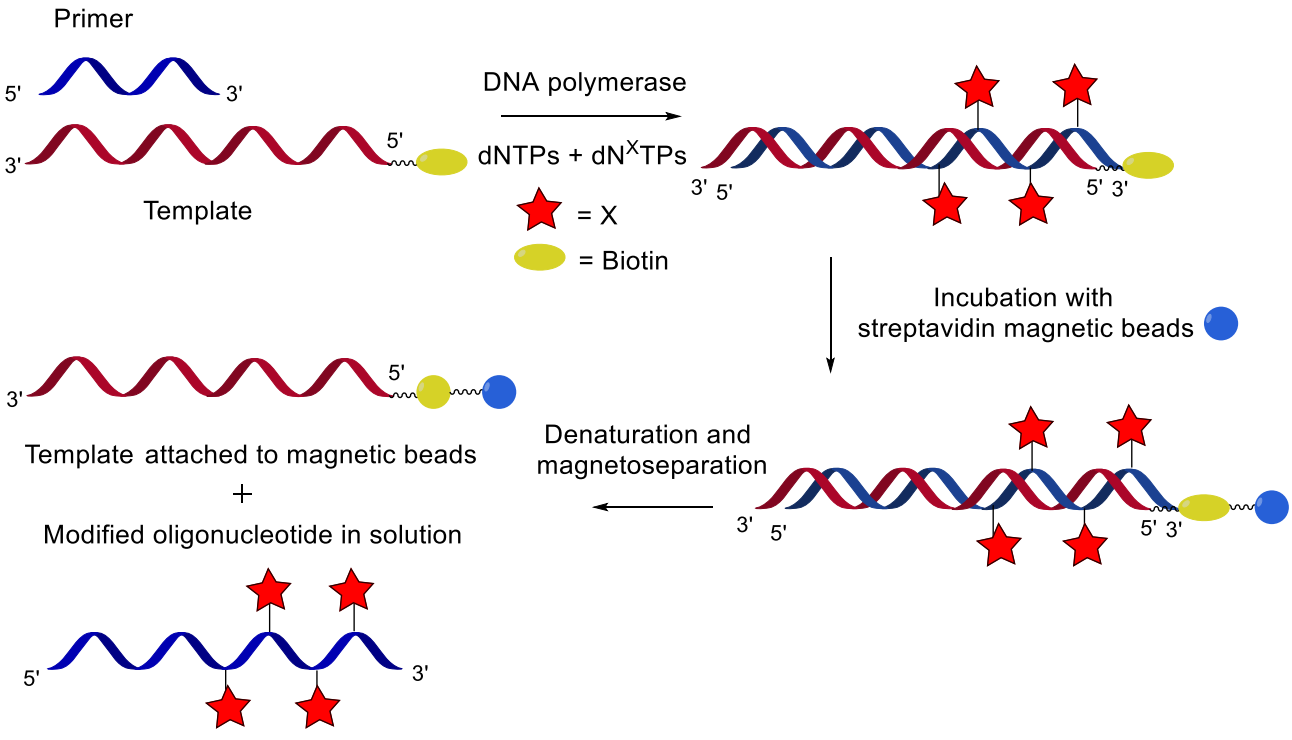


Figure 18: Schematic illustration of magnetoseparation

1.7. Restriction endonucleases

Restriction endonucleases are enzymes which recognize a single or a few target sequences on the DNA and cut the DNA twice (once through each sugar-phosphate backbone) at or near them.[113]

1.7.1. Different types of restriction endonucleases

Restriction enzymes are divided into four main groups based on their activity sites, required cofactors, and recognition sequences. [114] However, great variety among restriction enzymes has been uncovered showing that there are many more than four different types. The popular four classes are Type I, II, III, and IV.

Type I enzymes cleave the DNA at regions that are random and distant from their recognition sequences. They do not yield discrete restriction fragments and are considered to have not much practical use.

Type II enzymes cleave close to or within a short distance from the recognition site. [115] They can cleave near the middle of both strands to produce a blunt end, or they can cleave in a staggered pattern to produce sticky ends (Figure 19). They are the most often utilized and available class of restriction enzymes because they yield discrete restriction fragments.

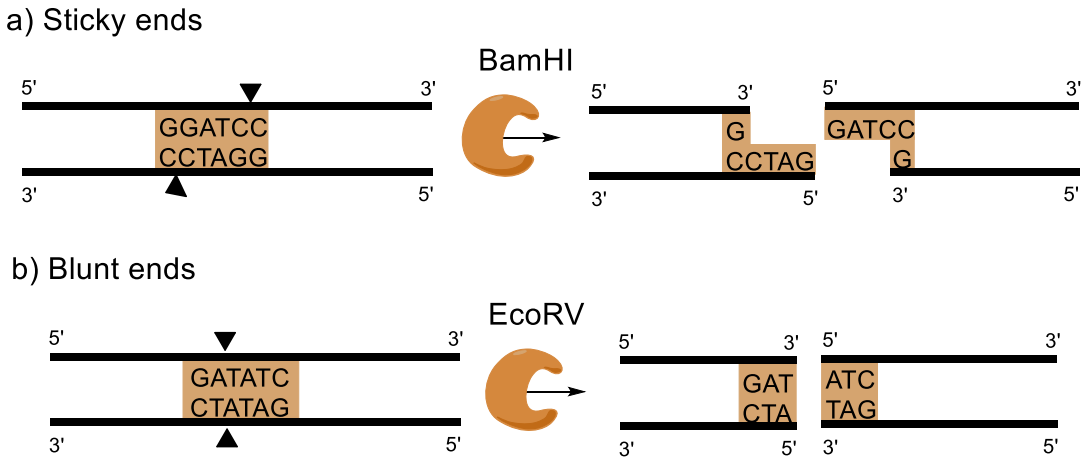


Figure 19: Schematic illustration of the two types of type II restriction endonuclease cleavage

Table 1 contains the list of selected type II restriction endonucleases with their cleavage sequences and type of cutting.

Table 1: Sequences of various type II restriction endonucleases

Restriction endonuclease	Sequence	Cut	Type of cut
EcoRI	GAATC CTTAG	5'---G AATTC---3' 3'---CTTAA G---5'	Sticky end
BamHI	GGATCC CCTAGG	5'---G GATCC---3' 3'---CCTAG G---5'	
HindIII	AAGCTT TTCGAA	5'---A AGCTT---3' 3'---TTCGA A---5'	
KpnI	GGTACC CCATGG	5'---GGTAC C---3' 3'---C CATGG---5'	
PstI	CTGCAG GACGTC	5'---CTGCA G---3' 3'---G ACGTC---5'	
SacI	GAGCTC CTCGAG	5'---GAGCT C---3' 3'---C TCGAG---5'	
SalI	GTCGAC CAGTG	5'---G TCGAC---3' 3'---CAGCT G---5'	
AflIII	CTTAAG GAATTC	5'---CTTA AG---3' 3'---GAAT TC---5'	
RsaI	GTAC CATG	5'---GT AC---3' 3'---CA TG---5'	

PvuII	CAGCTG GTCGAC	5’---CAG CTG---3’ 3’---GTC GAC---5’	Blunt end
SmaI	CCCGGG GGGCCC	5’---CCC GGG---3’ 3’---GGG CCC---5’	
AluI	AGCT TCGA	5’---AG CT---3’ 3’---TC GA---5’	
EcoRV	GATATC CTATAG	5’---GAT ATC---3’ 3’---CTA TAG---5’	
ScaI	AGTACT TCATGA	5’---AGT ACT---3’ 3’---TCA TGA---5’	

The X-ray crystal structure of two representative examples of type II restriction endonucleases, BamHI and EcoRV, are shown below (Figure 20, 21). The restriction endonuclease cleavage is carried out by the binding of the restriction endonuclease that recognizes and cleaves the specific palindromic sequence of 6 base pairs.

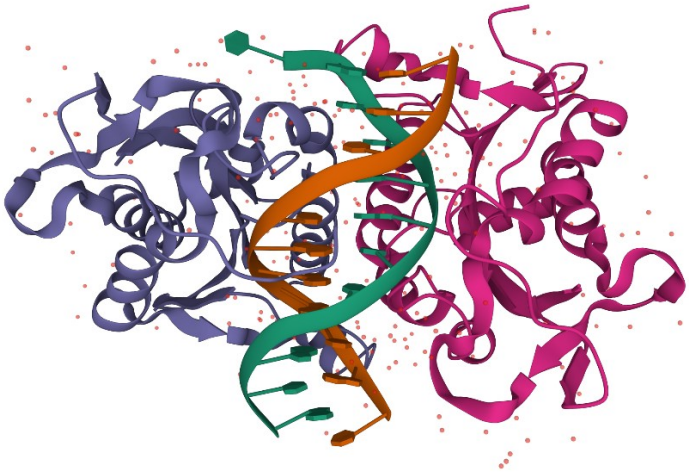


Figure 20: Crystal structure of BamHI complexed with double-stranded DNA (Taken from 10.2210/pdb1BHM/pdb; *Science* **1995**, *269*, 656- 663)

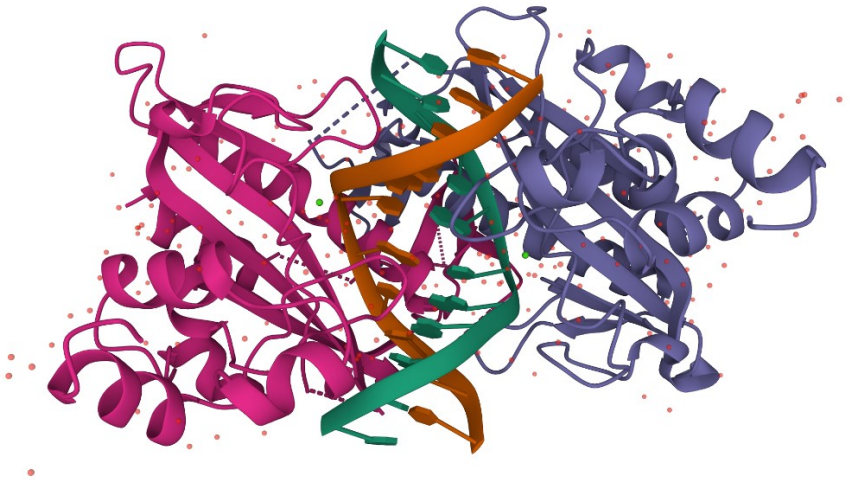


Figure 21: Crystal structure of EcoRV complexed with double-stranded DNA (Taken from 10.2210/pdb1AZ0/pdb; *J. Mol. Biol.* **1997**, *273*, 207-225)

Type III enzymes cleave at sites 20-30 bp after the recognition site and require two inversely oriented palindromic sequences to accomplish cleavage, often giving incomplete digests. Unlike the above-mentioned enzymes, modified DNA are recognized by type IV enzymes, for example, methylated DNA.

1.7.2. Tolerance of base modified DNA to restriction endonuclease cleavage

Restriction endonuclease cleavage of base modified DNA has been studied although not extensively. [116-119] Hocek group has investigated the effect of several major-groove modifications on type II restriction endonuclease cleavage. [120-124] Small modifications (H, vinyl, and ethynyl) attached to uracil and adenine bases were tolerated by restriction endonucleases whereas the bulky modifications (phenyl) inhibited cleavage. At the same time, no modifications attached to cytosine and guanine bases were tolerated by the same set of enzymes. Following this study, a transient chemical defense against RE cleavage was developed using (triethylsilyl)ethynyl-modified DNA. [124] The modified DNA synthesized using the enzymatic incorporation of 7-(triethylsilylethynyl)-7-deaza-2'-deoxyadenosine triphosphate (**dA^{TESE}TP**) provided complete protection against restriction endonuclease cleavage. The bulky silyl groups were deprotected by aqueous ammonia to produce the ethynyl-modified DNA, which was easily cleaved. Later, a bioorthogonal transient protection of DNA was developed using photocaging groups. [125-127] Photocaged DNA prepared using 5-nitrobenzyloxymethyl-2'-deoxyuridine (**dU^{NB}TP**) or –cytidine (**dC^{NB}TP**) or 5-nitroperonyloxymethyl-2'-deoxyuridine (**dU^{NP}TP**) or cytidine (**dC^{NP}TP**) triphosphates, inhibited any cleavage from restriction endonucleases. Cleavage of the bulky photocaging groups occurred upon irradiation with light, releasing the corresponding hydroxymethyl-modified DNA which is fully cleaved (Figure 22a, 22b). In another study, ethynyl-modified DNA was synthesized using the enzymatic incorporation of ethynyluracil-2'-deoxyribonucleoside triphosphate (**dU^{E}TP**). The CuAAC reaction of ethynyl-modified DNA in the presence of 3-azidopropane-1,2-diol or 3-azido-7-hydroxycoumarin, completely inhibited restriction endonuclease cleavage [128] (Figure 22c). It has also been observed that introducing acyl groups to DNA can shield it against RE cleavage.[129]

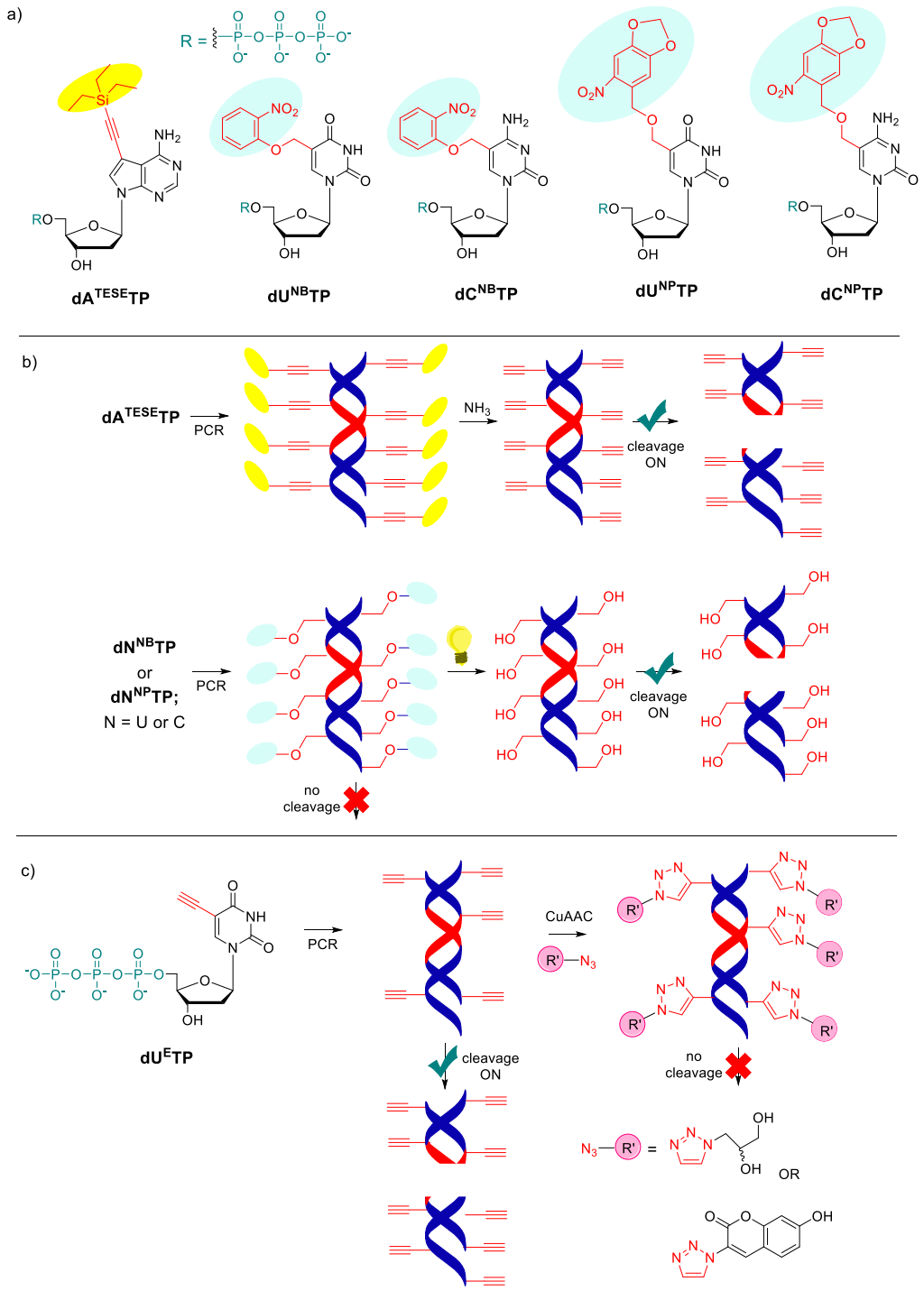


Figure 22: Schematic illustration of modified DNA that are protected against restriction endonuclease cleavage.[124-128] a) Structures of modified nucleoside triphosphates bearing bulky cleavable groups. b) Turn OFF of cleavage by REs using either bulky triethylsilyl group or photocleavable nitrobenzyl or nitroperonyl group and turn ON of the cleavage upon deprotection. c) Protection of ethynyl-modified DNA against cleavage by REs by CuAAC click reaction.

1.8. Transcription

Transcription is the process in which a segment of DNA is copied to form an RNA. It is done by an RNA polymerase that reads the DNA (using one of the DNA strands as the template) and generates a primary transcript, which is a complementary, antiparallel RNA strand. In addition to the RNA polymerase, several accessory proteins called transcription factors are also required for the transcription process. The process of transcription consists of three steps in general (Figure 23).

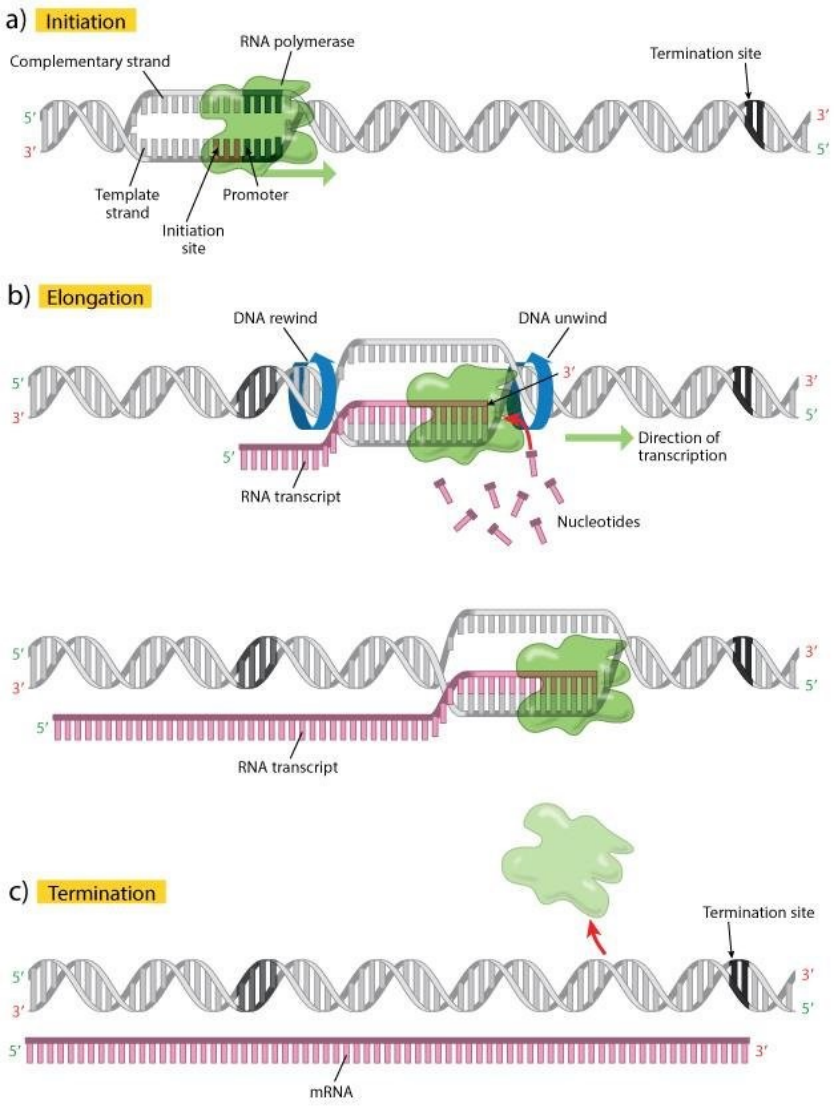


Figure 23: Schematic illustration of the three steps in transcription

(Taken from S. Clancy, *Nature Education* **2008**, *1*, 41)

The first step is the initiation, where the RNA polymerase complex binds to a specific DNA region at the start of the gene known as the promoter. In the next step called elongation, the template strand (the strand that's read in the 3'- 5' direction) of the DNA is employed. The RNA polymerase reads this template strand and builds an RNA molecule out of it, using the complementary ribonucleotides. The other strand is called the coding strand and it has the same sequence as that of the RNA except for the fact that thymine (T) is substituted by uracil (U). The third step is termination. The end of the transcription unit is recognized by the polymerase which ceases the transcription and releases the newly synthesized RNA molecule from the DNA template.

Transcription take place in both prokaryotic and eukaryotic genome. While the major steps in transcription are the same, there are striking differences between the prokaryotic and eukaryotic transcription.[130]

1.8.1. Bacterial transcription

The prokaryotes do not possess an organized nucleus. Therefore, the DNA is found in the cytoplasm where the transcription also occurs. In bacteria, there exists only a single type of RNA polymerase that transcribes all the genes. It consists of five polypeptide subunits. Four subunits denoted as α, α, β, and β' comprise the polymerase core enzyme. The fifth subunit σ, is involved only in the commencement of transcription and it helps the polymerase to locate the proper transcriptional initiation point. The sigma factors are distinct for a distinct set of genes and this adds a layer of regulation on transcription. The polymerase core enzyme along with the σ factor is called the holoenzyme.

Because DNA is unattached to any histones in bacteria, transcription starts right away. It starts at the promoter region and extends through the coding area. When the RNA polymerase reaches the terminator sequences, the transcription comes to an end. Bacteria have two different types of these sequences; Rho independent and Rho dependent terminator sequences. Rho-independent terminators generate inverted repeat sequences that has the ability to fold back on themselves to form hairpin loops, which will stop the RNA polymerase and release the transcript. Contrarily, Rho-dependent terminators depend on a protein called Rho, to deliberately disassemble the DNA-

RNA hybrid created during transcription and release the newly made RNA. Up until the RNA polymerase comes across such stop sequences, elongation continues. Once it reaches the terminator sequences, the RNA polymerase releases the DNA template and the transcription ceases. There is no post-transcriptional processing as the transcribed mRNA gets translated simultaneously in the cytoplasm. The transcription factors that can impact the stability of the holoenzyme structure at transcriptional start, provide additional regulation of transcription.[130]

1.8.2. Eukaryotic transcription

Transcription in eukaryotes is more complicated than in prokaryotes. DNA is largely found in the nucleus, where transcription also takes place. The nuclear membrane divides the nucleus from the cytoplasm, where translation takes place. In contrast to bacteria, eukaryotes have three distinct forms of RNA polymerases (RNA pol I-III). Ribosomal RNAs (rRNAs) are produced by RNA pol I, messenger RNAs (mRNAs) and tiny regulatory RNAs are transcribed by RNA pol II, and transfer RNAs are produced by RNA pol III (tRNAs). These polymerases differ from one another in terms of the class of RNAs they transcribe as well as the number and type of subunits they contain. Typically, they are made up of 10–17 subunits. Ancillary transcription factors are present that aids in the identification and binding of promoter sites, instead of the sigma factors that are present in bacteria. Only after the promoter area has been occupied by the transcription initiation factors, does the RNA polymerase identify it. The DNA separates from the histone proteins and unwinds close to the promoter region. The promoter area will also be bound by other transcription factors, such as enhancers.

Similar to bacteria, transcription starts at the initiation site and ends at the termination signal. It is still unclear what exact processes distinct RNA polymerases employ for termination. The newly produced mRNA is called pre-mRNA. During the post transcriptional processing, the non-coding regions are removed and the coding portions are reconnected. This results in the mature mRNA which is ready for translation.[130]

1.8.3. Bacterial *in vitro* transcription studies of base modified DNA

Hocek group reported an extensive study on the effect of several major groove-modified DNA on bacterial *in vitro* transcription by RNAP. By employing the appropriate base-modified dNTPs in PCR, several major groove-modified DNA templates were synthesized. The modifications included 5-substituted pyrimidines or 7-substituted 7-deazapurines carrying H, methyl, vinyl, ethynyl or phenyl groups [131] (Figure 24a). The *in vitro* transcription experiments with these modified templates utilized RNAP from *Bacillus subtilis (B. subtilis)* and *Escherichia coli (E. coli)*. Both enzymes fully tolerated nucleobases with smaller modifications (H, Me in 7-deazapurines), but bulkier modifications (phenyl at any nucleobase) and uracil, inhibited transcription. In addition, the *E. coli* enzyme partially tolerated certain medium-sized modifications-vinyl or ethynyl (Figure 24b).

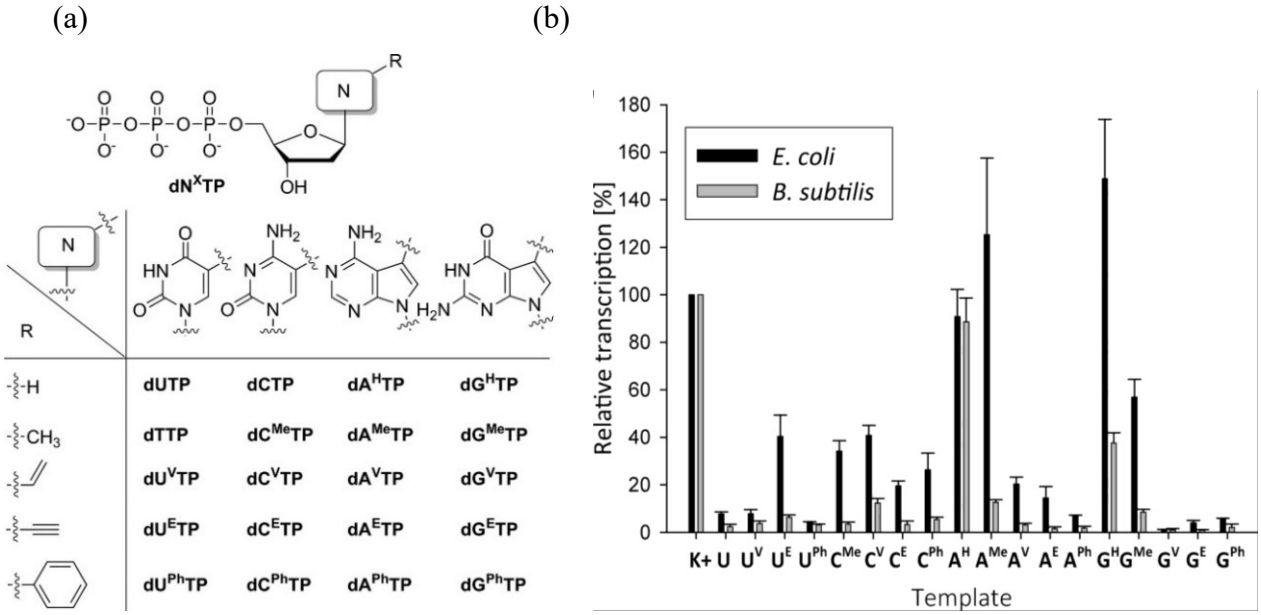


Figure 24: a) Structures of modified dNTPs b) Relative transcription of modified DNA templates compared to natural template using *E. coli* and *B. subtilis* RNAP. (Taken from *Nucleic Acids Res.* **2016**, *44*, 3000-3012).

In a follow up study, the bacterial transcription of modified DNA templates containing the epigenetic modifications 5mC or uracil (U) or 5hmU or 5hmC were studied.[61] DNA templates with the P*veg* promoter containing the 5hmU (350%) and 5hmC (250%) modifications supported a strong enhancement of transcription while the presence of U inhibited the transcription significantly (Figure 25).

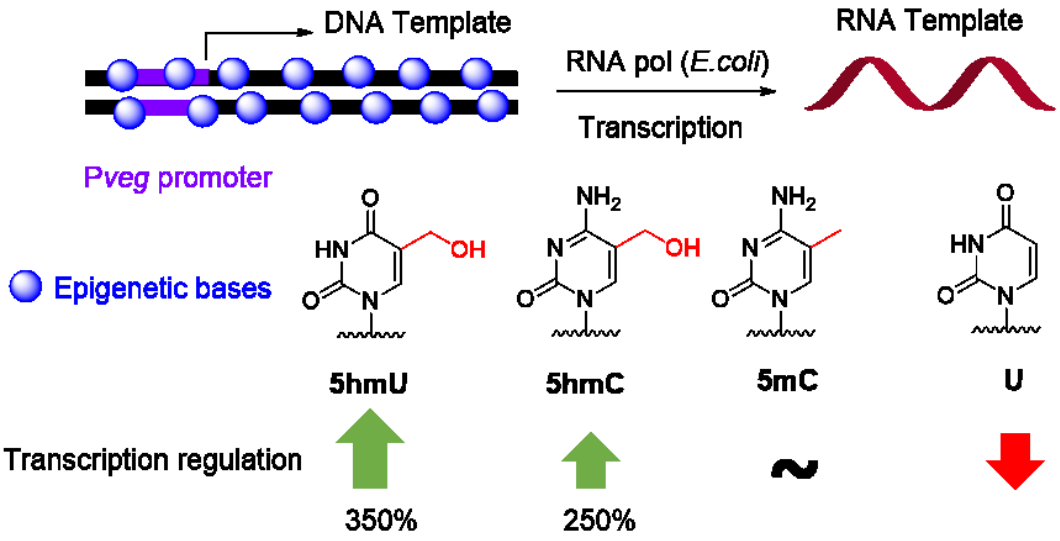


Figure 25: Schematic representation of the effect of epigenetic bases in the major groove of DNA on bacterial *in vitro* transcription by RNAP.[61]

This shed light to the possibility of developing a bioorthogonal chemical switch of transcription by employing a chemical transformation between bulkier and smaller substituents in the major groove of DNA. Soon after, a transcription turn OFF using a CuAAC click reaction of water soluble azides and 5-ethynyluracil-modified DNA (**U**$^{\mathbf{E}}$-modified DNA), was developed. [128] Compared to the natural DNA templates, the transcription rate of the **U**$^{\mathbf{E}}$-modified DNA template was about 43%, whereas the transcription rate of the same after the azide cycloaddition, was close to zero (Figure 26).

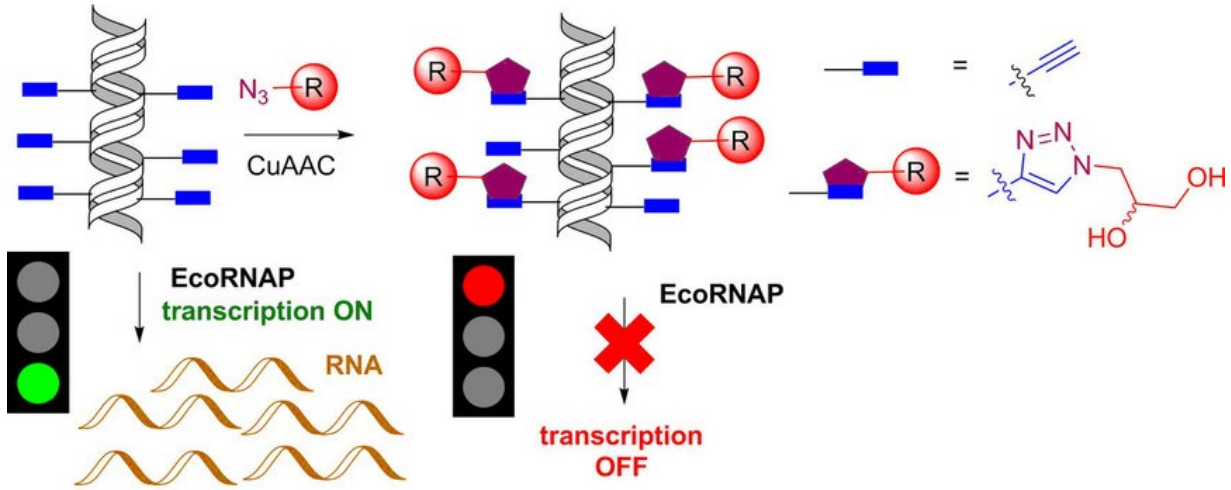


Figure 26: Schematic representation of transcription turn OFF with bacterial RNAP using CuAAC click reactions on the DNA. (Taken from *Chem. Eur. J.* **2018**, *24*, 8311-8314).

This study was further extended to the development of a photo induced bioorthogonal switching of transcription with a 0-1-0 logical gate.[132] Nitrobenzyl photocaged DNA templates were developed using the corresponding base modified dNTPs. No transcription using *E. coli* RNAP was supported by these photocaged DNA templates (OFF state). Irradiation with 400 nm light produced the uncaged hydroxymethylpyrimidine containing DNA templates which switched ON the transcription. Enzymatic phosphorylation of the templates containing 5hmU (but not 5hmC) switched the transcription OFF again (Figure 27).

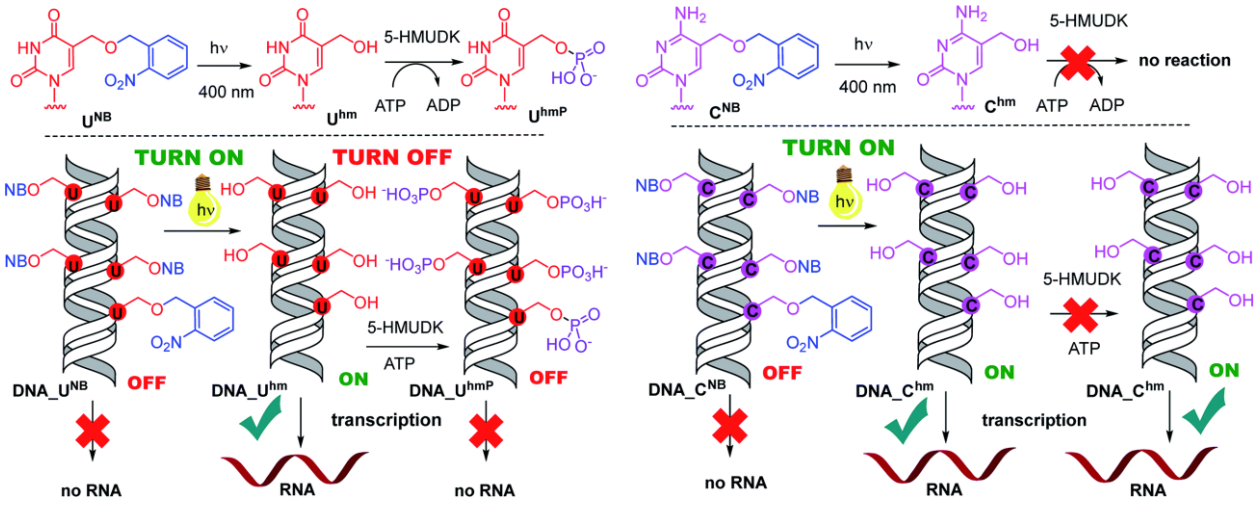


Figure 27: Schematic representation of transcription turn ON and turn OFF with bacterial RNAP through a series of bioorthogonal reactions in the major groove of DNA. (Taken from *Chem. Sci.* **2019**, *10*, 3937-3942).

2. Specific aims of the thesis

1. Synthesis of nitrobenzyl photocaged 5-(hydroxymethyl)pyrimidine-2'-deoxyribonucleoside phosphoramidites for specific photoactivatable epigenetic labeling of DNA.
   1.1. Synthesis of nitrobenzyl photocaged 5-(hydroxymethyl)pyrimidine-2'-deoxyribonucleoside phosphoramidites and their incorporation into oligonucleotides using solid phase synthesis.
   1.2. Development of specific site photoactivatable epigenetic modified DNA and it's bacterial *in vitro* transcription studies.

2. Synthesis of glucosylated 5-(hydroxymethyl)pyrimidine-2'-deoxyribonucleotides and their studies as epigenetic DNA bases regulating transcription and restriction endonuclease cleavage
   2.1. Synthesis of glucosylated 5-(hydroxymethyl)pyrimidine-2'-deoxyribonucleoside triphosphates and their enzymatic incorporation to DNA.
   2.2. Restriction endonuclease cleavage studies of glucosylated 5-(hydroxymethyl)-modified DNA
   2.3. PCR synthesis of glucosylated 5-(hydroxymethyl)-modified DNA for the bacterial *in vitro* transcription studies
   2.4. Synthesis of glucosylated 5-(hydroxymethyl)pyrimidine-2'-deoxyribonucleoside phosphoramidites for their incorporation into oligonucleotides using solid phase synthesis.

2.1. Rationale of the specific aims

The synthesis of nitrobenzyl photocaged 5-(hydroxymethyl)pyrimidine-2'-deoxyribonucleoside triphosphates for the enzymatic synthesis of major groove photocaged DNA and its restriction endonuclease cleavage and bacterial transcription studies have been previously reported from our group.[125, 132] Following the studies on this lane, my first goal was to design and synthesize nitrobenzyl photocaged 5-(hydroxymethyl)pyrimidine-2'-deoxyribonucleoside phosphoramidites for the construction of specific site photocaged oligonucleotides using chemical synthesis. The next step was to use these modified oligonucleotides as primers for the synthesis of specific site

photocaged DNA. These specific site photocaged DNA when irradiated with light would yield the specific site hydroxymethyl-modified DNA. Therefore, the next task was to study the bacterial *in vitro* transcription of these specific site photocaged and uncaged epigenetically modified DNA and thereby to develop a specific site photoactivatable epigenetic labelling of DNA. From the overall perspective of the unidentified biological function of these naturally occurring epigenetic bases, research into the impact of these photocaged or uncaged hydroxymethyl groups, on certain regions of the DNA promoter region, on bacterial transcription, is of special interest.

Attempts to further switch the transcription of the 5-(hydroxymethyl)pyrimidine-modified DNA by a bioorthogonal reaction led us to the glucosylation reaction on these epigenetic nucleobases. The glucosylated 5-(hydroxymethyl)pyrimidines are naturally found hypermodifications and their biological roles are still unclear. Therefore, the next part of the thesis was focused on the systematic study of the glucosylated 5-(hydroxymethyl)pyrimidine-modified DNA. The synthesis of glucosylated 5-(hydroxymethyl)pyrimidine-2'-deoxyribonucleoside phosphoramidites and its incorporation to oligonucleotides using solid phase synthesis have been previously reported. [132-134] My aim was to design and synthesize the corresponding triphosphate building blocks and their enzymatic incorporation to DNA by polymerases. The next task was to study the effect of these glucosylated 5-(hydroxymethyl)pyrimidines on restriction endonuclease cleavage and bacterial *in vitro* transcription by bacterial RNAP. The final goal was to develop an enzymatic glucosylation using T4-βGT glucosyl transferase for the bioorthogonal regulation of transcription.

The photochemical release or the enzymatic glucosylation of epigenetic 5-(hydroxymethyl)pyrimidine bases described in this thesis could be useful for the regulation of gene expression through bioorthogonal reactions in the DNA major groove, especially in the field of artificial chemical epigenetics.

3. Results and discussions

3.1. Photocaged 5-(hydroxymethyl)pyrimidine nucleoside phosphoramidites for specific photoactivatable epigenetic labeling of DNA

5-(Hydroxymethyl)uridine and -cytidine nucleosides, their triphosphates and their enzymatic incorporation to DNA have been previously reported in the literature. [125, 136-138] Photocaged 5-(hydroxymethyl)uridine and -cytidine nucleosides using nitrobenzyl protecting groups and their triphosphate building blocks have also been reported. [125, 126] The enzymatic incorporation of these nucleotides to form photocaged or bare 5-(hydroxymethyl)pyrimidine-modified DNA and their studies on restriction endonuclease cleavage and bacterial *in vitro* transcription have been studied. The nitrobenzyl photocaged DNA was found to be resistant to restriction endonuclease cleavage by several restriction endonucleases while the corresponding uncaged DNA (hydroxymethyl -modified) was fully cleaved by the same set of endonucleases. The nitrobenzyl photocaging of hydroxymethylated templates inhibits the *in vitro* transcription by bacterial RNAP but upon irradiation with light, uncaging occurs and the resultant hydroxymethylated DNA switches ON the transcription. Infact, the hydroxymethylated DNA was found to show enhanced transcription activity compared to the natural DNA (350% for 5hmU and 250% for 5hmC). [61] This way, a bioorthogonal chemical reaction in the major groove of DNA to turn ON the transcription with bacterial RNAP *in vitro* is depicted.[132]

It is challenging to use these base modified dNTP based enzymatic synthesis of DNA for the site-specific incorporation of one or more photocaging groups. For further in-depth structural and functional studies of the impact of these epigenetic alterations, such site-specifically photocaged DNAs might be helpful. There are several reports on the protected 5hmC [139-143] and 5hmU [138, 143] 2'-deoxyribonucleoside phosphoramidites suitable for automated solid-phase synthesis of oligonucleotides, but their photocaged analogues have not been described yet. Considering the specific importance of site-specific photocaging of a single epigenetic nucleobase in the major groove of DNA, we set out to design and synthesize photocaged 5hmU and 5hmC 2'-deoxyribonucleoside phosphoramidites and to study their use in solid-phase synthesis of photocaged oligonucleotides.

The syntheses of all compounds, solid phase synthesis of modified oligonucleotides, enzymatic DNA synthesis, and the irradiation experiments were carried out by me. The *in vitro* transcription experiments were carried out by Dr. Viola Vaňková Hausnerová and Dr. Olatz Ruiz-Larrabeiti (Institute of Microbiology). Structural aspects of RNAP were provided by Dr. I. Barvík (Charles University). Syntheses of the nitrobenzyl-modified 5-(hydroxymethyl)pyrimidine nucleosides were performed following the reported procedures.[126, 144, 145]

3.1.1. Synthesis of modified nucleosides and nucleoside phosphoramidites

3.1.1.1. Synthesis of nitrobenzyl photocaged 5-(hydroxymethyl)-modified nucleosides

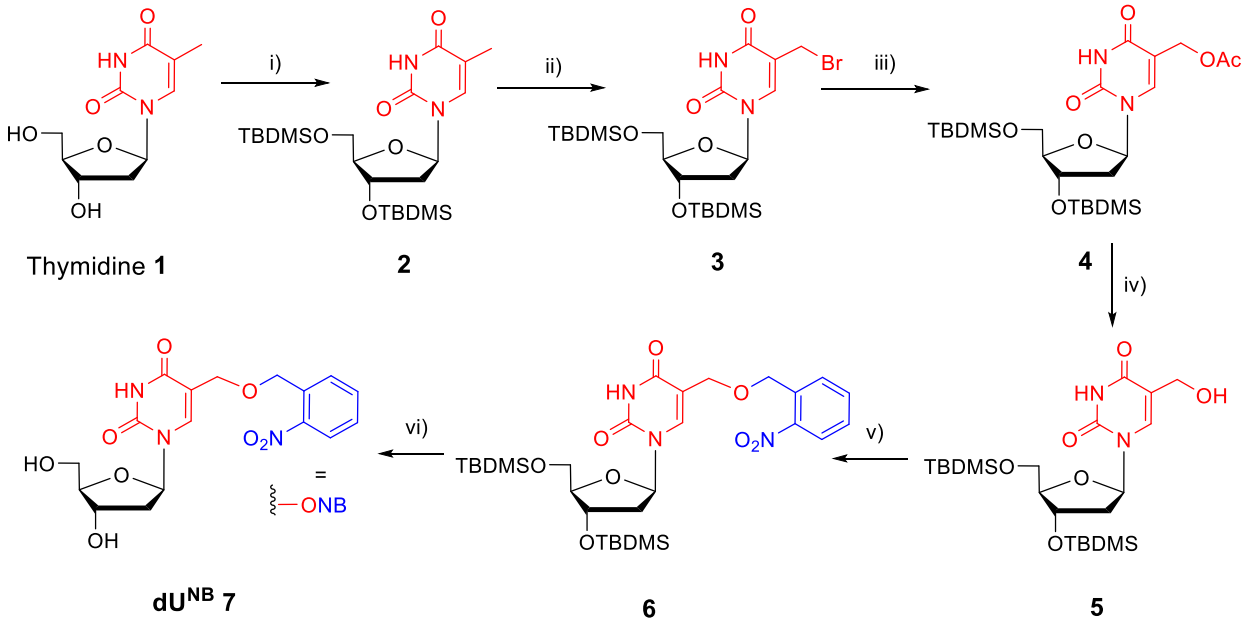


Figure 28: Synthetic scheme of nitrobenzyl photocaged 5-(hydroxymethyl)-2'-deoxyuridine (**7, dU^{NB}**). Reagents and conditions: i) TBDMSCl, imidazole, DMF, r.t., 12 h, 95-99 %; ii) NBS, AIBN, benzene, 85 °C, 30 min.; iii) CsOAc, DMF, r.t., 2 h, 53% over two steps; iv) K_2CO_3, MeOH, r.t., 2 h, 69%; v) 2-nitrobenzyl bromide, AgOTf, 2,6-di-*tert*-butylpyridine, DCM, r.t., 3 h, 40%; iv) $Et_3N{\cdot}3HF$, THF, r.t., 16 h, 40%.

The synthesis begins with the TBDMS group protection of 5'-hydroxyl and 3'-hydroxyl groups of thymidine which gives **2** in quantitative yield. Benzylic bromination of **2** using NBS and catalytic amount of AIBN gave **3**. Nucleophilic treatment of crude **3** with cesium acetate in DMF gave the acetyloxymethyl derivative **4** (53%) which upon deprotection gave the hydroxymethyl derivative **5** (69%). The hydroxymethyl derivative **5** undergoes nucleophilic reaction with 2-nitrobenzyl bromide, with the aid of 2,6-di-*tert*-butylpyridine and AgOTf, which yields silyl-protected 5-(2-nitrobenzyloxymethyl)- 2′-deoxyuridine **6** (40%). Final deprotection of the silyl groups of **6** was carried out using $Et_3N.3HF$ to yield 5-(2- nitrobenzyloxy)methyl-2′-deoxyuridine (**7**, **dU^{NB}**) in 40 % yield (Figure 28).

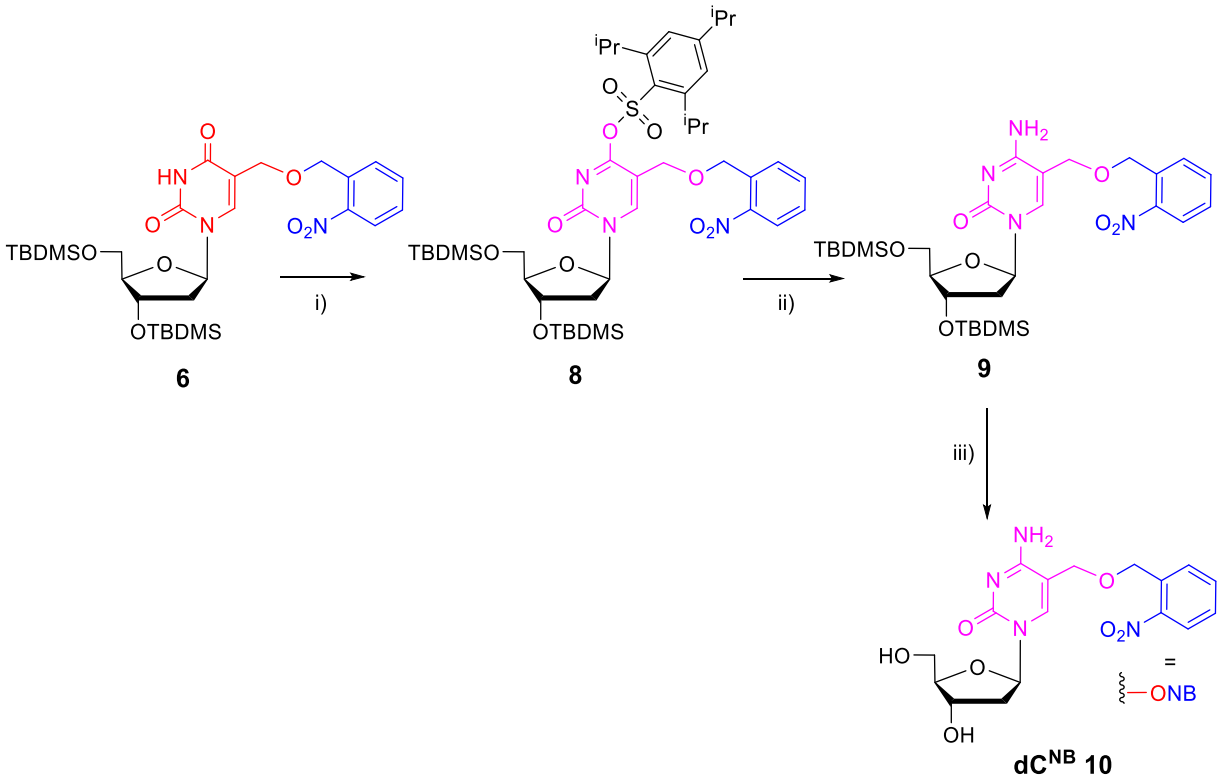


Figure 29: Synthetic scheme of nitrobenzyl photocaged 5-(hydroxymethyl)-2'-deoxycytidine (**10, dC^{NB}**). Reagents and conditions: i) TIPSCl, DMAP, Et_3N, DCM, r.t., 12 h; ii) NH_3 (g), dioxane, r.t., 3 h, 80% over 2 steps; iii) $Et_3N{\cdot}3HF$, THF, r.t., 16 h, 50%.

Silyl-protected 5-(2-nitrobenzyloxy)methyl-2′-deoxycytidine **9** was prepared by the conversion of silyl-protected 5-(2-nitrobenzyloxy)methyl-2′-deoxyuridine **6** in two steps. [127] The first step involves the reaction with 2, 4, 6-triisopropyl-benzenesulfonyl chloride in the presence of DMAP which activates the oxo group in position four of **6**. The benzene sulfonyl moiety of crude **8** then undergoes nucleophilic substitution with gaseous ammonia to give the silyl-protected 5-(2-nitrobenzyloxymethyl)-2′-deoxycytidine **9** in 80 % yield. Final deprotection of the silyl groups of **9** was carried out using $Et_3N.3HF$ to yield 5-(2-nitrobenzyloxymethyl)-2′-deoxycytidine (**10**, **dC^{NB}**) in 50 % yield (Figure 29).

3.1.1.2. Synthesis of nitrobenzyl photocaged 5-(hydroxymethyl)nucleoside phosphoramidites

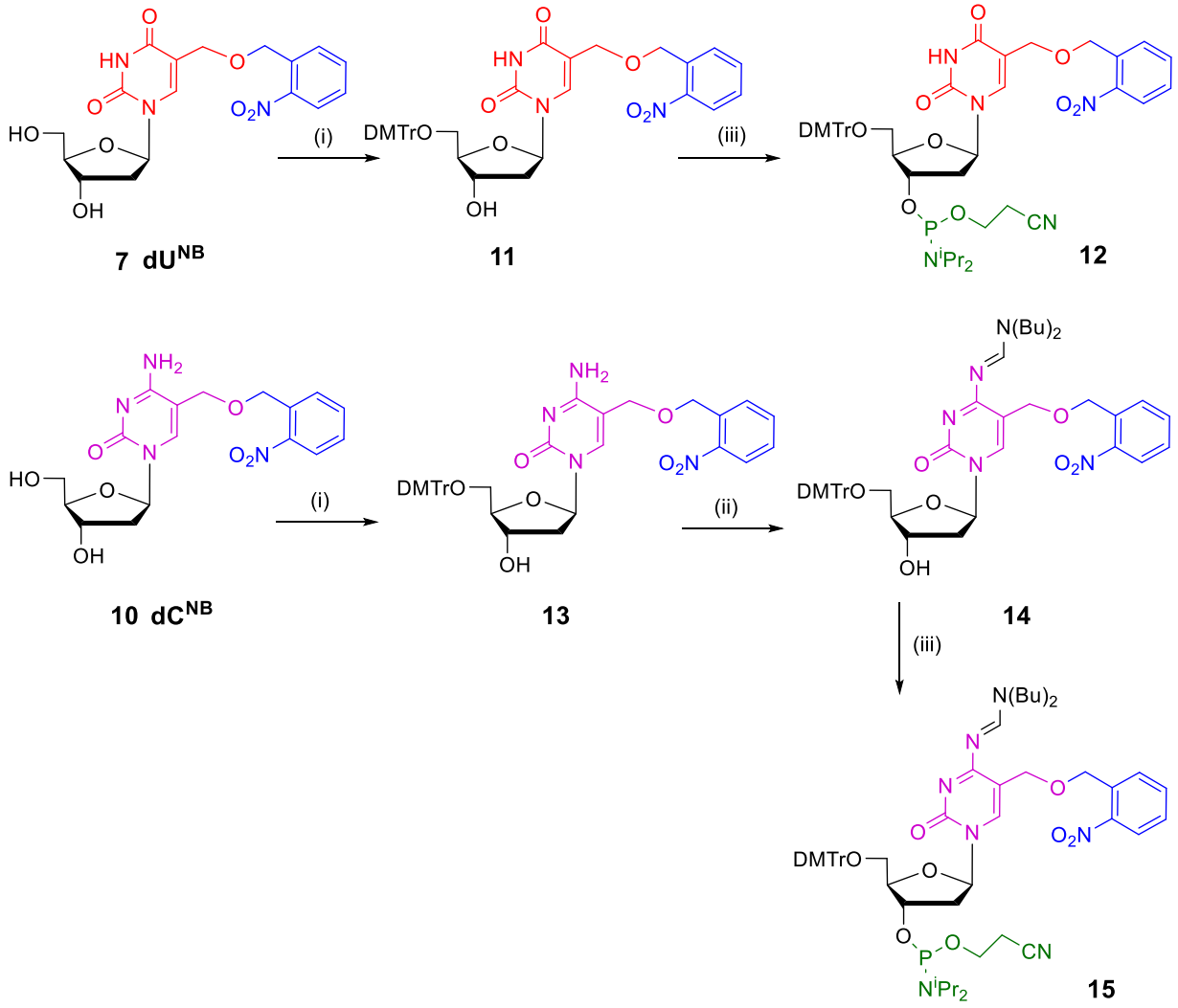


Figure 30: Synthetic scheme of the nitrobenzyl photocaged 5-(hydroxymethyl)-2'-deoxyuridine and -cytidine phosphoramidites **12** and **15**. Reagents and conditions: (i) DMTrCl, DMAP, pyridine, r.t., 6 h, 50-70% (ii) DBF-DMA, DMF, r.t., 12 h, 76% (iii) 2-Cyanoethyl *N,N*-diisopropylchlorophosphoramidite, DIPEA, DCM, r.t., 1 h, 70-85%.

Dimethoxytritylation of nucleosides **7** and **10** gave the 5'-*O*-DMTr-protected intermediates **11** and **13** in 50% and 68% yields, respectively. Various protecting group strategies were then attempted for the transient protection of the amino group of cytidine in compound **13** before the conversion to the corresponding phosphoramidite. First, benzoyl or acetyl groups were tried, but low yields and/or side acylation at 3'-OH was obtained. Therefore, the dibutylformamidine protecting group was attached through the reaction of **13** with dibutylformamide-dimethylacetal (DBF-DMA) in anhydrous DMF.[146] This gave the corresponding amidine derivative **14** in good yield, which was further converted to the phosphoramidite **15** (85% yield) by treatment with 2-cyanoethyl-*N,N*-diisopropylchlorophosphoramidite. In the same way, **11** was transformed to the desired phosphoramidite **12** in 74% yield (Figure 30).

3.1.2. Strategy and design of DNA templates

The photocaged building blocks **12**, **15** were used in the synthesis of oligonucleotides (ONs) aimed for the construction of specifically modified DNA templates for the study of bacterial transcription. The previous works [61, 132] used a 339 mer DNA template containing the 38 bp-long P*veg* promoter for the study of the influence of major-groove modifications on transcription with RNAP from *E. coli*. Those fully modified templates were constructed using PCR with modified dNTPs. For specific single-point modifications in the non-template strand of the promoter region, the synthesis was designed through PCR using modified forward primers (modified ONs). In order to access the important regions of the promoter, the 339 mer template was truncated to a 222mer DNA containing only a core promoter flanked by two nucleotides at the 5'-end. This shortened 222mer template has retained ~90% of the activity of the original 339 mer construct (Figure 31).

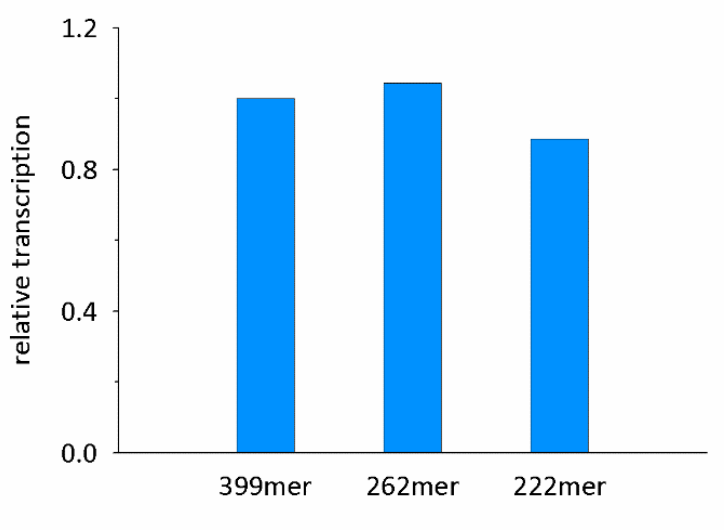


Figure 31: Relative transcription of different lengths of natural P*veg* template

3.1.3. Synthesis of modified oligonucleotides

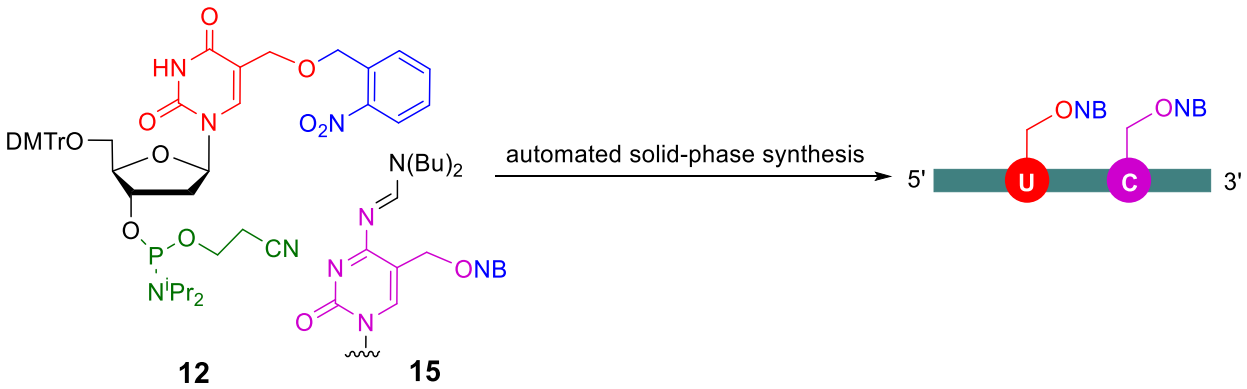


Figure 32: Schematic depiction of the solid phase synthesis of photocaged oligonucleotides

Based on a structural study[147] and on our previous experience[148], specific sites of the core promoter were identified, mainly in the −35 region of the nontemplate strand, as important for the interaction with the RNAP. Specific sites in the -10 region as well as in the transcription start site (TSS) which could play a crucial role were also identified. Therefore, photocaged ONs containing one or several photocaged 5hmU (**U^{hm}**) or 5hmC (**C^{hm}**) bases in the −35 region (**ON2–ON5**), -10 region (**ON9–ON12**) and in TSS (**ON13–ON16**) were designed. For comparison, ONs containing these modifications outside of these regions (**ON6**–**ON8**), were also designed.

ONs **ON2**–**ON16** were synthesized on an automated DNA synthesizer using the phosphoramidite building blocks **12** and **15** (Figure 32, Table 2).

Table 2: Sequences of the modified oligonucleotides. The regions -35, -10 and TSS are highlighted in yellow, blue, and green respectively.

ON	Sequence (5'-3')	nt	Region modified
1	TCTATTTGACAAAAATGGGC	20	-35 region
2	TCTATU*TGACAAAAATGGGC	20	
3	TCTATTU*GACAAAAATGGGC	20	
4	TCTATTTGAC*AAAAATGGGC	20	
5	TCTATU*U*GAC*AAAAATGGGC	20	
6	TCTATTTGACAAAAU*GGGC	20	Linker region
7	TCTATTTGACAAAATGGGC*T	21	
8	TCTATTTGACAAAAU*GGGC*T	21	
9	TCTATTTGACAAAATGGGCTCGTGTTGU*A	30	-10 region
10	TCTATTTGACAAAATGGGCTCGTGTTGTAC*A	32	
11	TCTATTTGACAAAATGGGCTCGTGTTGTACAAU*A	36	
12	TCTATTTGACAAAATGGGCTCGTGTTGU*AC*AAU*A	36	
13	TCTATTTGACAAAATGGGCTCGTGTTGTACAATAAAU*G	39	TSS
14	TCTATTTGACAAAATGGGCTCGTGTTGTACAATAAATGU*G	41	
15	TCTATTTGACAAAATGGGCTCGTGTTGTACAATAAATGTGU*C	43	
16	TCTATTTGACAAAATGGGCTCGTGTTGTACAATAAATGTGTCU*A	45	

The synthesis proceeded under standard conditions (with slightly increased coupling volume and time for the modified phosphoramidites compared to natural), giving the desired photocaged ONs with efficient incorporation of the modified phosphoramidites, compared to their natural counterparts. The incorporation of the intact 5-(2-nitrobenzyloxy)methyl nucleobases was confirmed by MALDI and ESI (Table 3, Figure 56-63 in Appendix 1).

Table 3: Overview of the ESI-MS and MALDI-TOF mass data of modified oligonucleotides (na: not achieved).

ON	M (calculated) [Da]	ESI-MS (found) [Da]	MALDI found [Da]
2	6291.13	6290.08 $[M-H]^+$	6292.8 $[M+H]^+$, 6156.8 $[M-NB]^+$
3	6291.13	6290.08 $[M-H]^+$	6292.5 $[M+H]^+$, 6157.1 $[M-NB+H]^+$
4	6305.15	6304.09 $[M-H]^+$	6306.8 $[M+H]^+$, 6171.6 $[M-NB+H]^+$
5	6607.21	6606.14 $[M-H]^+$	6473.6 $[M-NB+H]^+$, 6338.4 $[M-2NB+H]^+$, 6203.2 $[M-3NB+H]^+$
6	6291.13	6290.08 $[M-H]^+$	6157.1 $[M-NB+H]^+$
7	6609.35	6608.14 $[M-H]^+$	6610.3 $[M+H]^+$, 6474.8 $[M-NB+H]^+$
8	6760.38	6760.16 $[M]^+$	6626.2 $[M-NB+H]^+$, 6491.0 $[M-2NB+H]^+$
9	9402.6	9400.6 $[M-2H]^+$	9267.8 $[M-NB]^+$
10	10018.5	10016.5 $[M-2H]^+$	9884.3 $[M-NB+H]^+$
11	10935.4	10933.4 $[M-2H]^+$	10801.3 $[M-NB+H]^+$
12	11251.2	11250.2 $[M-H]^+$	10846.7 $[M-3NB]^+$
13	12195	na	12196.9 $[M+H]^+$, 12063.3 $[M-NB+3H]^+$
14	12828.4	na	12830.4 $[M+2H]^+$, 12694.7 $[M-NB+H]^+$
15	13421.7	na	13286.7 $[M-NB]^+$
16	14039.2	na	13904.6 $[M-NB]^+$, 13942.0 $[M-NB+K]^+$

3.1.4. Synthesis of specific site modified DNA

222 MER DNA

5'- TCTATTTGACAAAAATGGGCTCGTGTTGTACAATAAATGTGTCTAAGCTTGGGTCCC
ACCTGACCCCATGCCGAACTCAGAAGTGAAACGCCGTAGCGCCGATGGTAGTGTGG
GGTCTCCCCATGCGAGAGTAGGGAACTGCCAGGCATCAAATAAAACGAAAGGCTCAG
TCGAAAGACTGGGCCTTTCGTTTTATCTGTTGTTTGTCGGTGAACGCTCTCC-3'

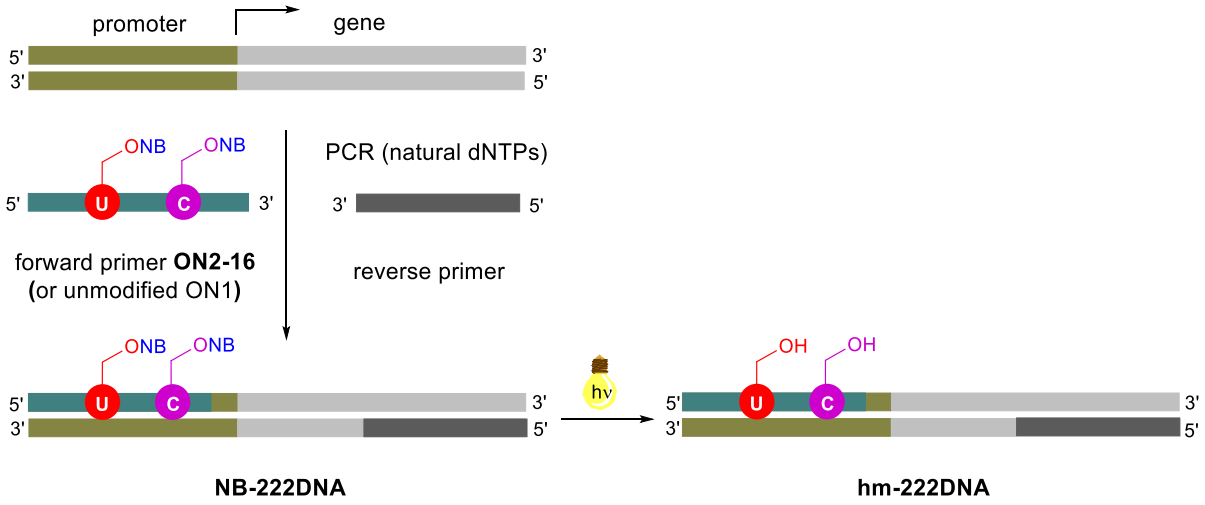


Figure 33: Schematic depiction of the synthesis of specific site modified DNA

The photocaged ONs (**ON2**–**ON16**), along with the natural ON1 (P*veg*_F_222), were used as forward primers for the PCR reaction on the P*veg* plasmid, along with a natural reverse primer (REV-PgII-R1/R3) and natural dNTPs (Figure 33, Table 4).

Table 4: List of oligonucleotides used for the PCR experiments and sequencing. [a] primer sequences in template are underlined. [b] core promoter sequence of the template is in bold italics. [c] 5′-end of the oligonucleotide is labelled by 6-carboxyfluorescein (6-FAM).

Oligonucleotide	**Sequence (5'-3')**	**Length**
PRIMERS		
P*veg*_F_222 (ON1)	TCTATTTGACAAAAATGGGC	20-mer
REV-PgII-R1/R3[c]	GGAGAGCGTTCACCGACA	18-mer
TEMPLATE		
Temp$^{P\mathit{veg}}$222 [a,b]	TC***TATTTGACAAAAATGGGCTCGTGTTGTACAAT AAATGT***GTCTAAGCTTGGGTCCCACCTGACCCC ATGCCGAACTCAGAAGTGAAACGCCGTAGCGC CGATGGTAGTGTGGGGTCTCCCCATGCGAGAGT AGGGAACTGCCAGGCATCAAATAAAACGAAAG GCTCAGTCGAAAGACTGGGCCTTTCGTTTTATC TGTTGTTTGTCGGTGAACGCTCTCC	222-mer

In all cases, full-length 222mer DNA amplicons 222DNA1 (unmodified) and **NB-222DNA2-16** (modified in the promoter region) were obtained efficiently (Figure 34). The photocaged DNA templates **NB-222DNA2-16** were then irradiated by a 3W 400-nm photodiode for 10 (for $\mathbf{C^{NB}}$) or 30 minutes (for $\mathbf{U^{NB}}$) in the presence of additives DTT and sodium azide (in analogy to previous works [132]), to release the uncaged DNA containing the 5-(hydroxymethyl)pyrimidine bases (**hm-222DNA2-16**). The integrity of the sequences was confirmed by Sanger sequencing.

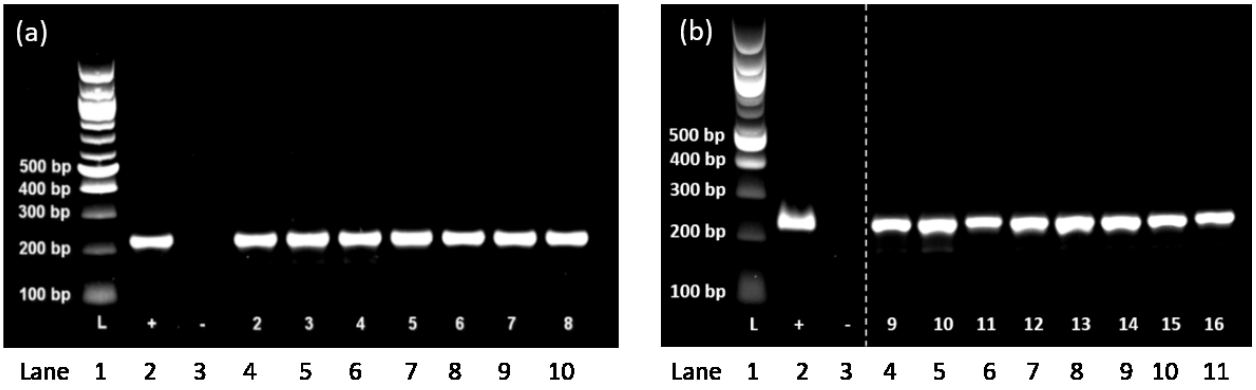


Figure 34: Agarose gel analysis of specific site nitrobenzyl photocaged DNA (a) **NB-222DNA2-8** and (b) **NB-222DNA9-16**. Lane 1 (L): 100 bp ladder (commercial mix of dsDNA of specific lengths); Lane 2 (+): PCR product with natural primers (222DNA1); Lane 3 (-): PCR product in the absence of forward primer. (a) Lane 4-10 (2-8): PCR products in the presence of modified forward primers (**ON2-8**) and natural reverse primer. (b) Lane 4-11 (9-16): PCR products in the presence of modified forward primers (**ON9-16**) and natural reverse primer.

3.1.5. *In vitro* transcription studies

The *in vitro* transcription studies were performed by Dr. Viola Vaňková Hausnerová and Dr. Olatz Ruiz-Larrabeiti (Institute of Microbiology). The irradiation experiments were carried out by me.

3.1.5.1.Transcription studies of specific site modified DNA

Specific site modified DNAs (photocaged as well as irradiated) prepared by PCR, were used as templates in the *in vitro* transcription studies by RNAP from *E.coli*. The natural DNA template (222DNA1) yielded about the same transcription before and after irradiation. DNA templates containing the photocaged bases in the −35 region (**NB-222DNA2**–**5**), which is critical for interaction with RNAP, displayed significantly lower transcription (14%–78%), whereas transcription of templates containing modifications outside the −35 region (**NB-222DNA6**–**8**) was comparable to the unmodified control, showing that they are not important for the interaction with RNAP. The lowest transcription was observed from the template containing $\mathbf{U^{NB}}$ at the −35 position (**NB-222DNA2**, 33%) and the template with a combination of three photocaged pyrimidines (**NB-**

222DNA5, 14%). After irradiation and uncaging of the hydroxymethylpyrimidine bases, the transcription of the out-of-region-modified DNA templates was unchanged (**hm-222DNA6**–**8**), but the transcription of the in-the-region-modified uncaged templates (**hm-222DNA2**–**5**) was restored to the level of the natural DNA. In other words, the transcription was switched ON (Figure 35, Table 5).

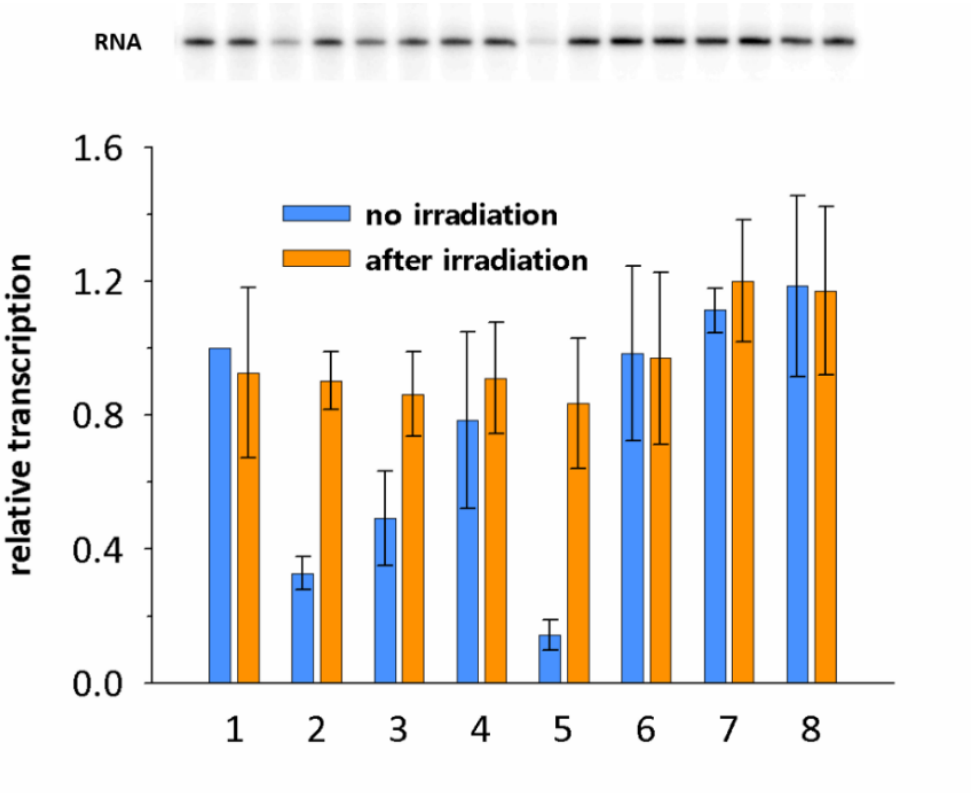


Figure 35: Relative transcription of DNA templates unmodified 222DNA1 (nonirradiated and irradiated), **NB-222DNA2-8**, and **hm-222DNA2-8**. Representative primary data (radiolabelled RNA transcripts) are shown above the graph. The bars show the average from at least five independent experiments, the error bars show ±SD.

Table 5: Average relative transcription (± SD; average from 5 experiments) of the DNA templates unmodified 222DNA1 (nonirradiated and irradiated), **NB-222DNA2-8** and **hm-222DNA2-8**.

DNA template	**Average relative transcription ± SD**	
	No irradiation (unmodified 222DNA1 or **NB-222DNA2-8**)	After irradiation (unmodified 222DNA1 or **hm-222DNA2-8**)
1 (Natural)	1	0.93 ± 0.26
2	0.33 ± 0.05	0.90 ± 0.09
3	0.49 ± 0.14	0.86 ± 0.13
4	0.78 ± 0.26	0.91 ± 0.17
5	0.14 ± 0.05	0.83 ± 0.19
6	0.98 ± 0.26	0.97 ± 0.26
7	1.11 ± 0.07	1.20 ± 0.18
8	1.19 ± 0.27	1.17 ± 0.25

DNA templates containing the photocaged and uncaged bases in the -10 region (**NB/hm-222DNA9**–**12**) did not show any reproducible specific trend in transcription showing that these positions might not be important for interaction with RNAP. Transcription of the DNA templates containing the photocaged and uncaged bases in the TSS region (**NB/hm-222DNA13**–**16**) were comparable to the unmodified templates proving that these positions are also not significant for interaction with RNAP. (The transcription data are not included in the thesis).

3.1.5.2. Transcription data for the kinetic study of deprotection of DNA template NB-222DNA-5

The kinetic study of deprotection was done on the **NB-222DNA-5** template since this template contains both uridine and cytidine modified bases. Approx. 480 ng of the stock DNA (FAM labelled DNA sample **NB-222DNA-5**) was irradiated by the UV lamp (400nm) and aliquots were taken at the following time intervals 10 minutes / 20 minutes / 30 minutes / 40 minutes/ and 60 minutes, following the reported procedures. [132] These progressively irradiated DNA templates were then directly used as templates in the *in vitro* transcription assay. Complete switch ON of transcription was observed from the sample after the first 10 minutes denoting that the uncaging is complete within the first 10 minutes (Figure 36).

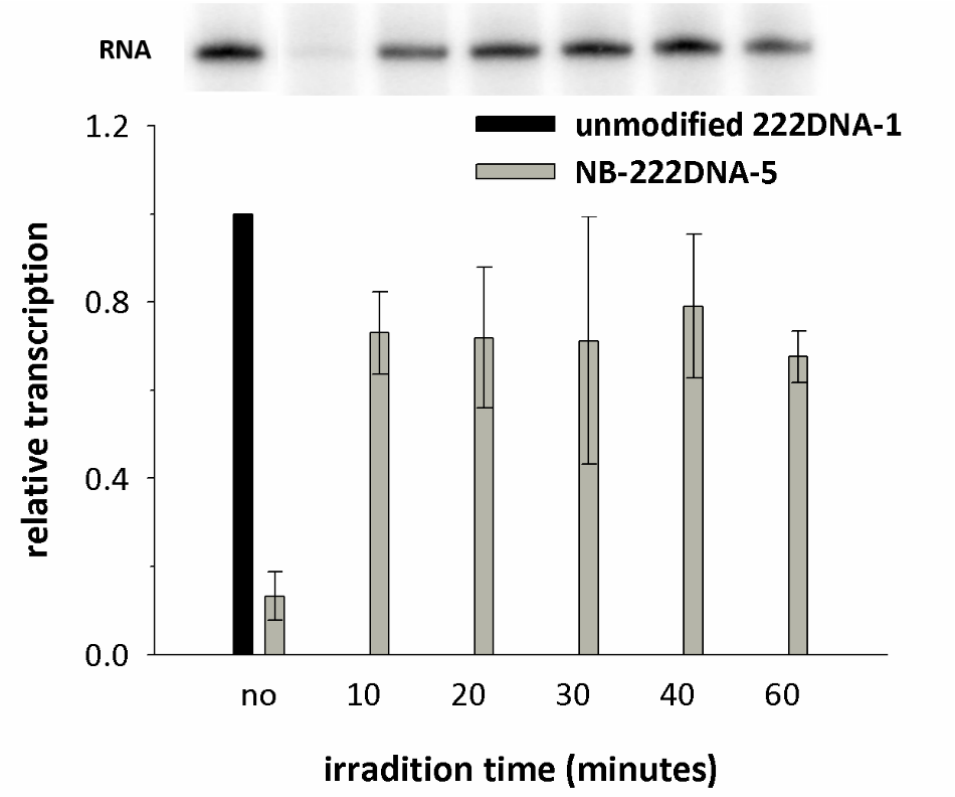


Figure 36: Relative transcription of DNA template **NB-222DNA-5** irradiated at different time intervals. Representative primary data (radiolabelled RNA transcripts) are shown above the graph. The bars show averages of two experiments. The error bars show the range.

3.2. Glucosylated 5-(hydroxymethyl)pyrimidines as epigenetic DNA bases regulating transcription and restriction cleavage

The *O*-glucosylated 5hmU and 5hmC are hypermodified bases that has been recently discovered. β-D-glucosyl-hydroxymethyluracil or Base J discovered in the nuclear DNA of some pathogenic protozoa, was found to have regulatory roles in transcription. [81, 86, 87] Whereas, β-D-glucosyl-hydroxymethylcytosine is found in bacteriophages formed through some bacteriophage-encoded biosynthetic pathway to circumvent the endonuclease-based defenses of their hosts. [50] It is known that this hypermodified base is formed by the selective transfer of the glucose moiety of UDP-glc to the 5hmC residues in double-stranded DNA. [62-64] Although these hypermodifications have been studied recently, no systematic study on their effect on *in vitro* bacterial transcription or restriction endonuclease cleavage have been done so far. The chemical syntheses of β-D-glucopyranosyloxymethyl uridine [133, 134] and cytidine [135] and their corresponding phosphoramidite analogues have been reported previously. Solid phase synthesis of oligonucleotides containing these glucosylated 5-(hydroxymethyl)pyrimidine modifications at specific positions, have also been described. [133-135] However, the synthesis of their triphosphate analogues has not yet reported. These triphosphate building blocks are essential for the synthesis of longer DNA templates that are suitable for transcription studies. Therefore, we chose to develop the synthesis of the β-D-glucopyranosyloxymethyl-2'-deoxyuridine and –cytidine 5'-*O*-triphosphates and their enzymatic incorporation to DNA for systematic studies on transcription and restriction endonuclease cleavage. The glucosylated 5-(hydroxymethyl)pyrimidine nucleoside phosphoramidites were also synthesized for the use in the automated solid phase synthesis of specific site glucosylated DNA.

The syntheses of all the compounds and enzymatic incorporations were carried out by me. The hydroxymethyl (**dU^{hm}TP 32** or **dC^{hm}TP 33**) triphosphates were synthesized by my colleagues Dr. Zuzana Vaníková and Filip Gracias. While some of the synthetic procedures were modified and optimized, other steps were performed mostly in accordance with the literature. The ESI-MS analysis was performed by my colleague Martin Svoboda. The *in vitro* transcription experiments were carried out by Dr. Olatz Ruiz-Larrabeiti (Institute of Microbiology).

3.2.1. Synthesis of modified nucleosides, nucleoside triphosphates, and phosphoramidites

3.2.1.1. Synthesis of glucosylated 5-(hydroxymethyl)-modified nucleosides

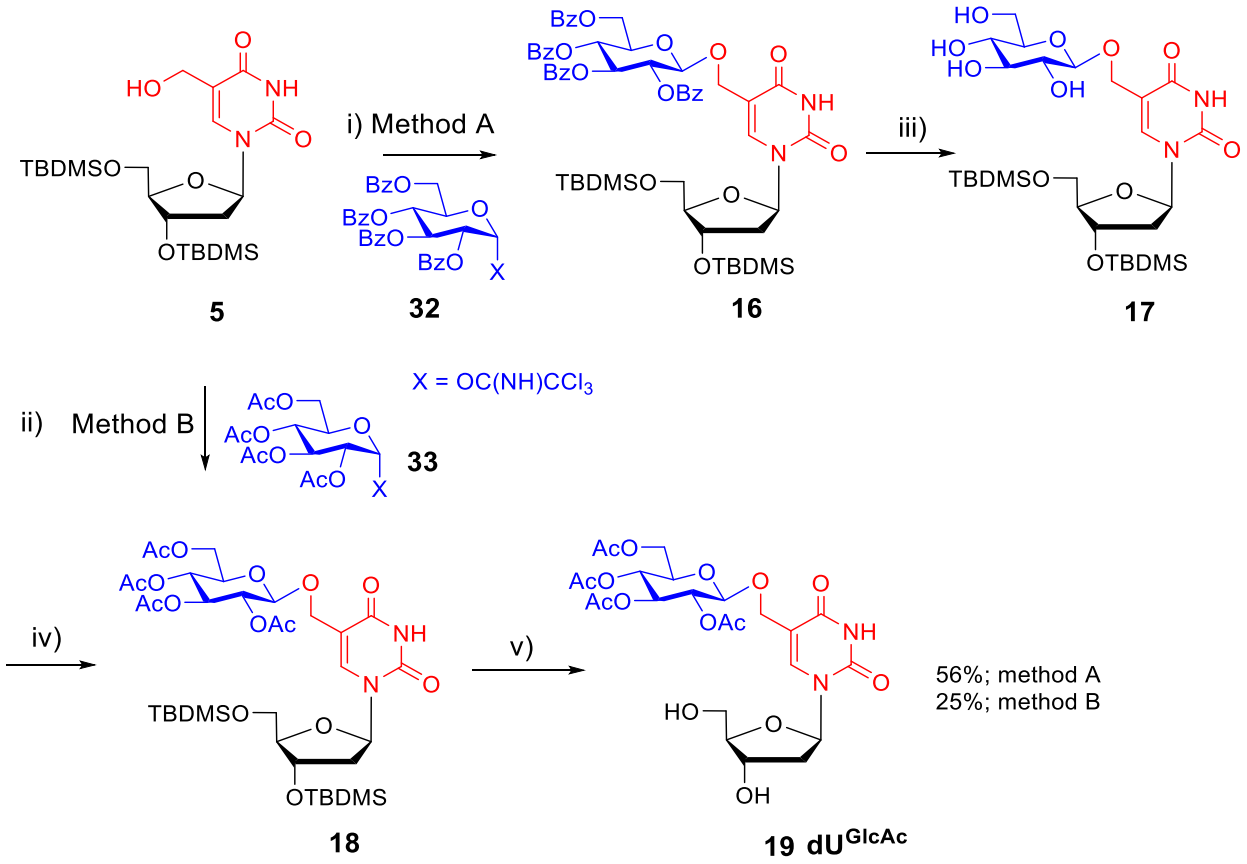


Figure 37: Synthetic scheme of glucosylated 5-(hydroxymethyl)-2'-deoxyuridine (**19, dU^{GlcAc}**). Reagents and conditions: i) perbenzoylated α-trichloroacetimidate ester sugar **32**, cat. TMSOTf, DCE, -25 °C, 1.5 h; ii) peracetylated α-trichloroacetimidate ester sugar **33**, cat. TMSOTf, DCE, -25 °C, 1.5 h; iii) K_2CO_3, MeOH, r.t., 4 h; iv) Ac_2O, pyridine, r.t., 16 h; v) $Et_3N{\cdot}3HF$, pyridine, r.t., 16 h. Method A: (i, iii, iv, v) 56% over four steps; Method B: (ii, v) 25% over two steps.

Synthesis of acetyl-protected 5-(glucopyranosyloxymethyl)-2'-deoxyuridine (**19, dU^{GlcAc}**) was performed as described in the literature. [134] The glucosylation reaction was achieved by means of Schmidt condensation of silyl-protected 5-(hydroxymethyl)uridine (acceptor) **5** in the presence of TMSOTf at -25 °C, with the sugars (donors) either as their corresponding perbenzoylated α-trichloroacetimidate esters **32** (Method A) or as their corresponding peracetylated α-trichloroacetimidate esters **33** (Method B). Method B was attempted to improvise and shorten the previously reported Method A but it resulted in a considerably lower yield and an inseparable reaction mixture of **18** with byproducts, hence was not followed further. Conversion of the resulting perbenzoylated sugar-protected 5-(hydroxymethyl)uridine **16** into their more ammonia-labile peracetylated analogue **18** was then achieved by treatment with K_2CO_3/MeOH followed by acetic anhydride (80% over three steps). Desilylation in the presence of $Et_3N.3HF$ in pyridine gave the final nucleoside (**19, dU^{GlcAc}**) in 70% yield (Figure 37).

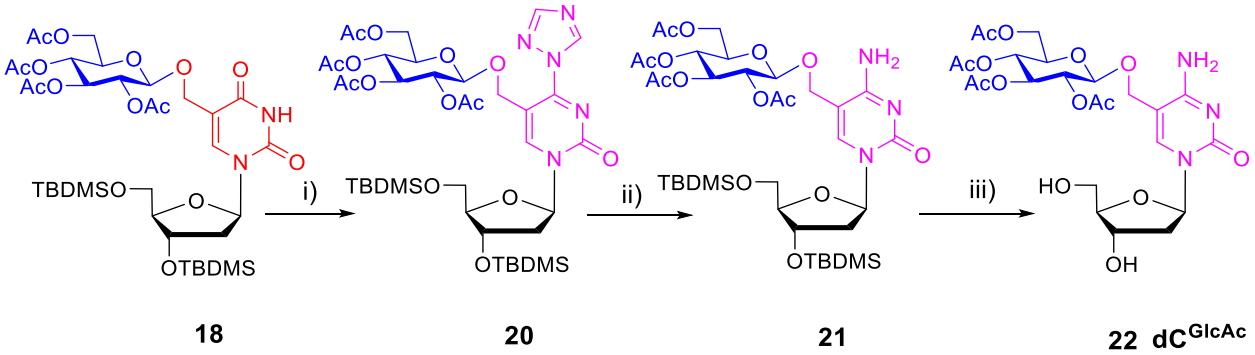


Figure 38: Synthetic scheme of glucosylated 5-(hydroxymethyl)-2'-deoxycytidine (**22, dC^{GlcAc}**). Reagents and conditions: i) $POCl_3$, 1,2,4-triazole, Et_3N, ACN, 16 h, 86%. (iii) 25% NH_4OH/1, 4-dioxane, 1:10, v/v, 30 min, 80%; (iii) $Et_3N{\cdot}3HF$, pyridine, 16 h, 76%.

Synthesis of acetyl-protected 5-(glucopyranosyloxymethyl)-2'-deoxycytidine (**22**, **dC^{GlcAc}**) was performed as described in the literature. [135] Treatment of **18** with phosphoryl chloride and 1,2,4-triazole gave the triazolide derivative **20**. Subsequent mild ammonolysis of **20** gave the glucosylated cytidine derivative **21** in 68% overall yield. Desilylation of **21** with $Et_3N{\cdot}3HF$ in pyridine gave the final nucleoside (**22**, **dC^{GlcAc}**) in 76% yield (Figure 38).

3.2.1.2. Synthesis of glucosylated 5-(hydroxymethyl)pyrimidine nucleoside triphosphates

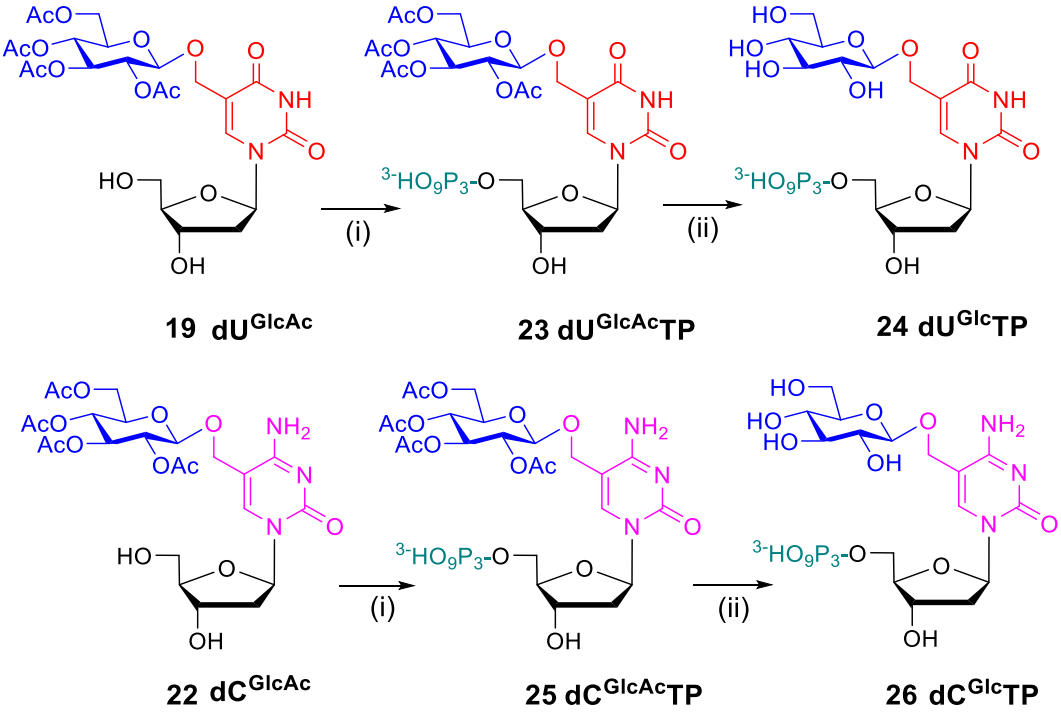


Figure 39: Synthetic scheme of the glucosylated 5-(hydroxymethyl)-2'-deoxyuridine and cytidine nucleoside triphosphates **dU^{GlcAc}TP (24)** or **dC^{GlcAc}TP (26)**. Reagents and conditions: i) 1. $POCl_3$, proton sponge, $PO(OMe)_3$, 0 °C, 1 h; 2. $(NHBu_3)_2H_2P_2O_7$, Bu_3N, DMF, 0 °C, 1 h; 3. 2 M TEAB; 20-30%; ii) NH_4OH, H_2O, 0 °C, 1 h; 78-80%.

The standard Yoshikawa triphoshorylation reaction on the modified nucleosides gave the corresponding modified triphosphates bearing acetyl-protected glucose (**dU^{GlcAc}TP 23** and **dU^{GlcAc}TP 25**), in moderate yields (20-30%). Further acetyl deprotection on these triphosphates produced the final triphosphates (**dU^{Glc}TP 24** and **dU^{Glc}TP 26**), in 78-80% yield (Figure 39).

3.2.1.3. Synthesis of glucosylated 5-(hydroxymethyl)nucleoside phosphoramidites

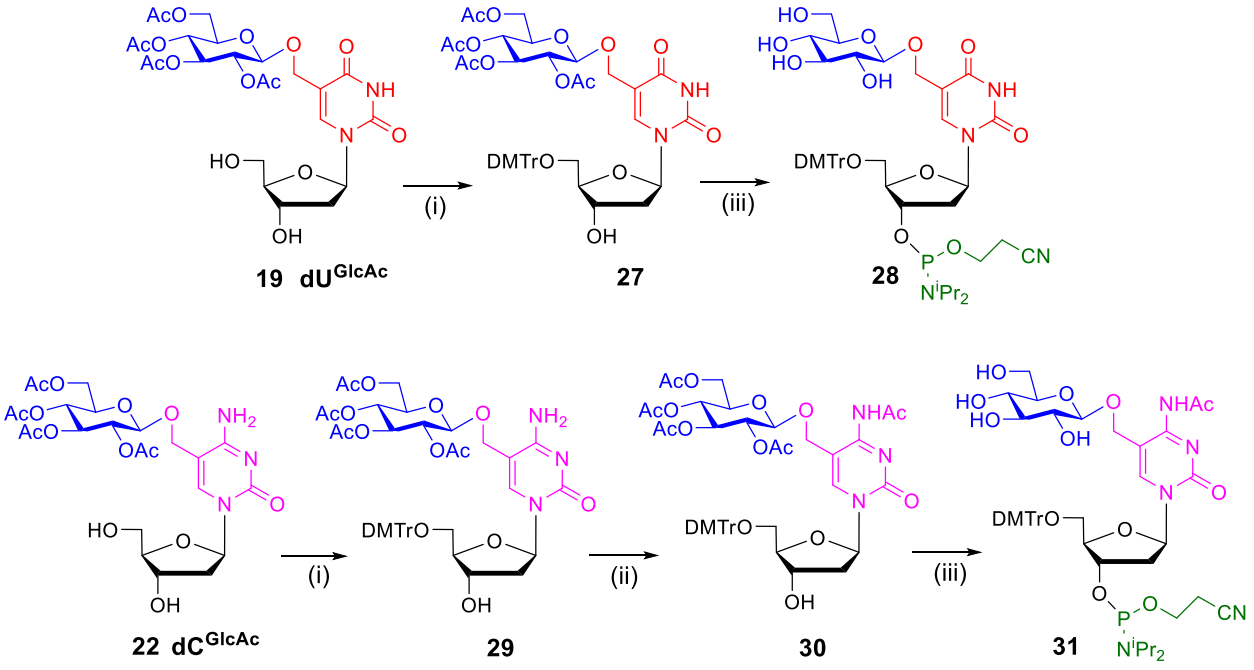


Figure 40: Synthetic scheme of the glucosylated 5-(hydroxymethyl)-2'-deoxyuridine and -cytidine phosphoramidites **28** or **31**. Reagents and conditions: i) DMTrCl, pyridine, 4 h, 60%. (ii) Ac_2O, DMF, r.t., 20 h, 64% (iii) 2-Cyanoethyl *N,N*-diisopropylchlorophosphoramidite, DIPEA, DCM, 1 h, 89-90%.

The standard trityl protection on the modified nucleosides gave the corresponding trityl-protected nucleosides bearing acetyl-protected glucose (**27** or **29**), in 60% yield. In the previous report [135], benzoyl group was used for the protection of the amino group of cytidine which is not suitable for the optimized final deprotection conditions that avoid potential byproducts (Deprotection conditions of the TAC-protected nucleoside phosphoramidites–2 hours at room temperature using NH_4OH). [133] Therefore, acetyl group was chosen as a suitable protecting group which can undergo deprotection at room temperature in 2 hours. The acetyl protection yielded the amino group-protected nucleoside **30** in 64% yield. Final phosphoramidite reaction on these protected nucleosides produced the final phosphoramidite building blocks **28** and **31** in 89-90% yield (Figure 40).

3.2.2. Enzymatic synthesis of modified DNA

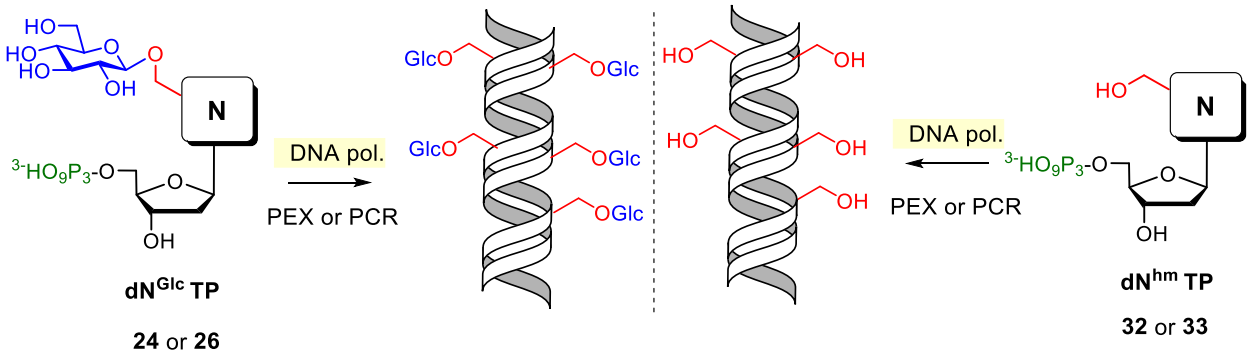


Figure 41: Schematic depiction of the enzymatic incorporation of glucose-modified and hydroxymethyl modified nucleoside triphosphates to DNA

3.2.2.1. Enzymatic incorporation of modified triphosphates by PEX

The enzymatic incorporation of glucosylated 5-(hydroxymethyl)-modified deoxyribonucleotides (Figure 41) was studied by PEX using three different DNA polymerases: Vent (exo-), Pwo and KOD XL. The design of the templates and primers were in such a way to accommodate one or four modifications. Non-labelled templates and FAM-labelled primers were used for the study (Table 6). The obtained products were verified by denaturing PAGE gel analysis and visualized using FAM.

Table 6: Sequences of oligonucleotides used for the PEX. [a] 5'-end of oligonucleotide is labelled by 6-carboxyfluorescein (6-FAM). [b] primer sequences in template are underlined.

List of primers and templates used in the study:		
Oligonucleotide	**Sequence (5'-3')**	**Length (nt)**
Primers		
Prim^{PEX} [a]	CATGGGCGGCATGGG	15
Templates		
Temp^{19}_^{1T} [b]	CCCACCCATGCCGCCCATG	19
Temp^{19}_^{1C} [b]	CCCGCCCATGCCGCCCATG	19
Temp^{31}_^{4X} [b]	CTAGCATGAGCTCAGTCCCATGCCGCCCATG	31

The incorporation of the modified nucleotides **24** and **26** were tested for both single and multiple nucleotide extension. Both the modified triphosphates **24, 26** were successfully integrated into DNA bearing one modification as well as four modifications using all the three DNA polymerases. These novel triphosphates served as good substrates for all the three polymerases (Figure 42, 43).

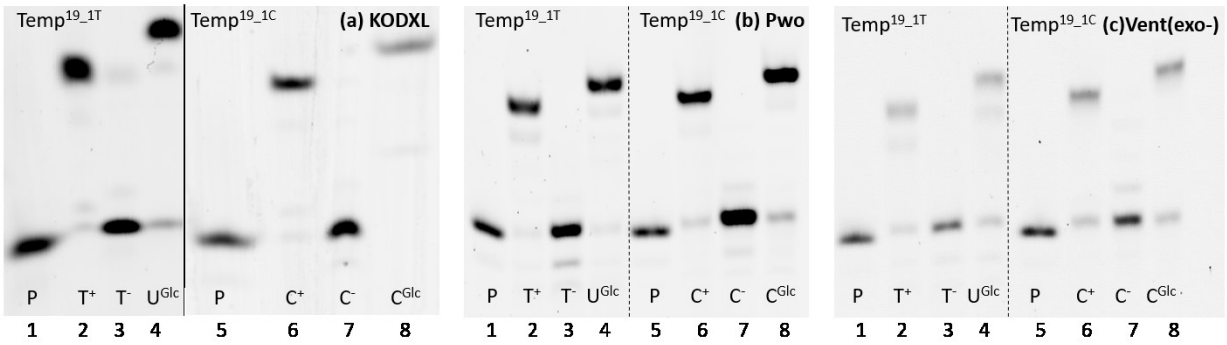


Figure 42: Denaturing PAGE analysis of PEX products prepared using (a) KOD XL, (b) Pwo and (c) Vent (exo-) DNA polymerases and templates Temp^{19_1X}. Temp^{19_1T}; **19DNA_1U^{X}** (lanes 2-4), Temp^{19_1C}; **19DNA_1C^{X}** (lanes 6-8). Lane 1, 5 (P): Primer Prim^{PEX}; Lane 2 (T^+): PEX product with dATP, dCTP, dGTP, dTTP; Lane 3 (T^-): PEX product with dATP, dCTP, dGTP; Lane 4 (**U^{Glc}**): PEX product with **dU^{Glc}TP**, dGTP, dCTP, dATP; Lane 6 (C^+): PEX product with dATP, dCTP, dGTP, dTTP; Lane 7 (C^-): PEX product with dATP, dTTP, dGTP; Lane 8 (**C^{Glc}**): PEX product with **dC^{Glc}TP**, dGTP, dTTP, dATP.

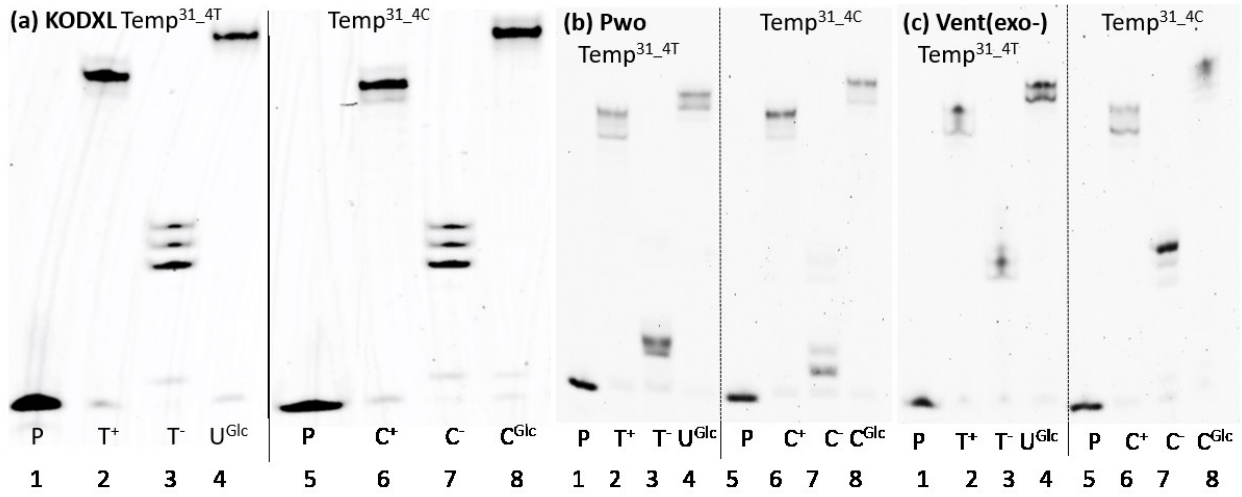


Figure 43: Denaturing PAGE analysis of PEX products prepared using (a) KOD XL, (b) Pwo and (c) Vent (exo-) DNA polymerases and templates Temp^{31_4X}. Temp^{31_4T}; **31DNA_4U^{X}** (lanes 2-4), Temp^{31_4C}; **31DNA_4C^{X}** (lanes 6-8). Lane 1, 5 (P): Primer Prim^{PEX}; Lane 2 (T^+): PEX product with dATP, dCTP, dGTP, dTTP; Lane 3 (T^-): PEX product with dATP, dCTP, dGTP; Lane 4 (**U^{Glc}**): PEX product with **dU^{Glc} TP**, dGTP, dCTP, dATP; Lane 6 (C^+): PEX product with dATP, dCTP, dGTP, dTTP; Lane 7 (C^-): PEX product with dATP, dTTP, dGTP; Lane 8 (**C^{Glc}**): PEX product with **dC^{Glc}TP**, dGTP, dTTP, dATP.

3.2.2.2. Enzymatic glucosylation of C^{hm}-modified PEX DNA by glucosyl transferase

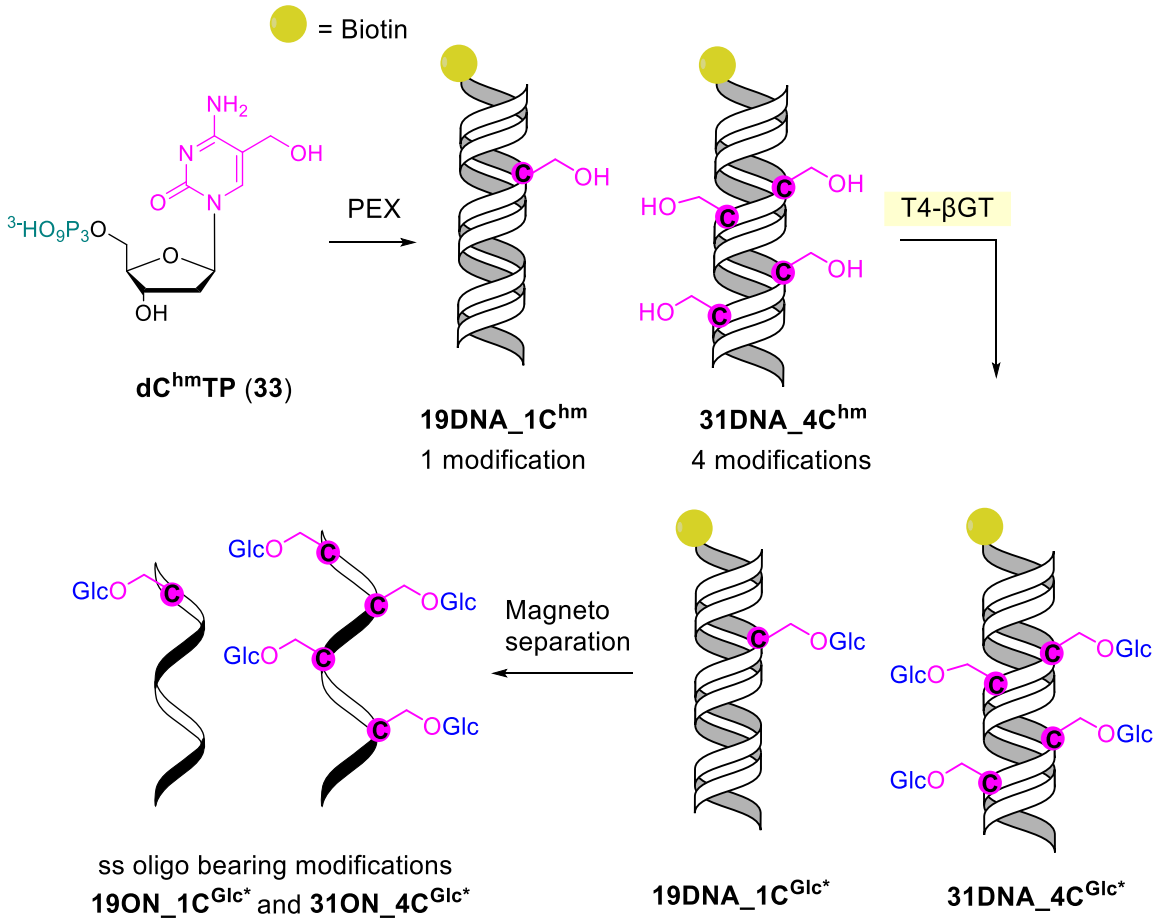


Figure 44: Scheme of the synthesis of ss oligonucleotides **19ON_1C^{Glc*}** and **31ON_4C^{Glc*}**

To develop and confirm the complete enzymatic glucosylation on the hydroxymethylcytidine-modified DNA using T4-βGT glucosyl transferase, it was initially attempted on shorter DNA PEX products. PEX products containing one or four modified hydroxymethyl cytidine bases were prepared in semi-preparative scale using biotinylated templates and non-labelled primers in the presence of KODXL following the reported protocols. [126] The enzymatic glucosylation was carried out on the purified PEX product using UDP-glc and T4-βGT glucosyl transferase. [149] MALDI of the ssON (**19ON_1C^{Glc*}** and **31ON_4C^{Glc*}**) obtained after the enzymatic glucosylation of **C^{hm}**-modified shorter DNA (**19DNA_1C^{hm}** and **31DNA_4C^{hm}**) followed by magnetoseparation, verified full conversion (Figure 44, Table 7).

3.2.2.3. Synthesis of modified single-stranded oligonucleotides

To confirm the incorporation of the modified triphosphates **24**, **26** and exclude the possibility of any misincorporation of other 2'-deoxyribonucleotides, PEX synthesis of oligonucleotides containing one and four modifications were carried out with biotinylated template and non-labelled primers using KOD XL polymerase on a semi-preparative scale. By using magnetoseparation, modified ssDNA was extracted, and the mass of the isolated product was analyzed using MALDI-TOF mass spectrometry. Data of the MALDI results are summarized in Table 7.

Table 7: Overview of the modified oligonucleotides and MALDI-TOF mass data.

ON	M (calculated) [Da]	M (found) [Da]	Δ
19ON_1U^{Glc}	6143.9	6145.0	1.1
19ON_1C^{Glc}	6142.9	6146.4	3.5
31ON_4U^{Glc}	10329.3	10331.9	2.6
31ON_4C^{Glc}	10385.3	10387.6	2.3
19ON_1C^{Glc*}	6142.9	6145.6	2.7
31ON_4C^{Glc*}	10385.3	10383.8	1.5

3.2.2.4. Enzymatic incorporation of modified triphosphates by PCR

In order to study the incorporation of glucosylated 5-(hydroxymethyl)-modified triphosphates into longer DNA and the amplification of glucosylated DNA, modified DNA templates were prepared by PCR using three different DNA polymerases (98 bp or 235 bp long DNA). Three natural dNTPs and an excess of modified one were used in the PCR reactions, together with non-labelled primers, templates, and DNA polymerase (Table 8). Agarose gel analysis was used to confirm the products that were obtained. In all cases (except for **dU^{Glc}TP** with Pwo polymerase) full-length amplified products were obtained showing that the glucosylated dNTPs are very good substrates in general (Figure 45, 46, 47). Sanger sequencing confirmed the fidelity of the DNA sequences (only 235DNA).

Table 8: Sequences of oligonucleotides used for the PCR. [a] 5'-end of oligonucleotide is labelled by 6-carboxyfluorescein (6-FAM). [b] primer sequences in template are underlined. [c] core promoter sequence of the template is in bold italics.

List of primers and templates used in the study:		
Oligonucleotide	**Sequence (5'-3')**	**Length (nt)**
Primers		
Prim^{F_98} [a]	GACATCATGAGAGACATCGC	20
Prim^{R_98} [a]	CAAGGACAAAATACCTGTATTCCTT	25
Prim^{F_235} [a]	CGTCTTCAAGAATTCTAT	18
Prim^{R_235} [a]	GGAGAGCGTTCACCGACA	18
Templates		
Temp^{98DNA} [b]	<u>GACATCATGAGAGACATCGCCTCTG</u>GGCTAATAGG ACTACTTCTAATCTGTAAGAGCAGATCCCTGGACA GGCAAGGA<u>ATACAGGTATTTTGTCCTTG</u>	98
Temp^{235DNA} [b,c]	<u>CGTCTTCAAGAATTC***TATTTGACAAAAATGGGCTCG TGTTGTACAATAAATGT***</u>GTCTAAGCTTGGGTCCCAC CTGACCCCATGCCGAACTCAGAAGTGAAACGCCGT AGCGCCGATGGTAGTGTGGGGTCTCCCCATGCGAG AGTAGGGAACTGCCAGGCATCAAATAAAACGAAA GGCTCAGTCGAAAGACTGGGCCTTTCGTTTTATCTG TTGTT<u>TGTCGGTGAACGCTCTCC</u>	235

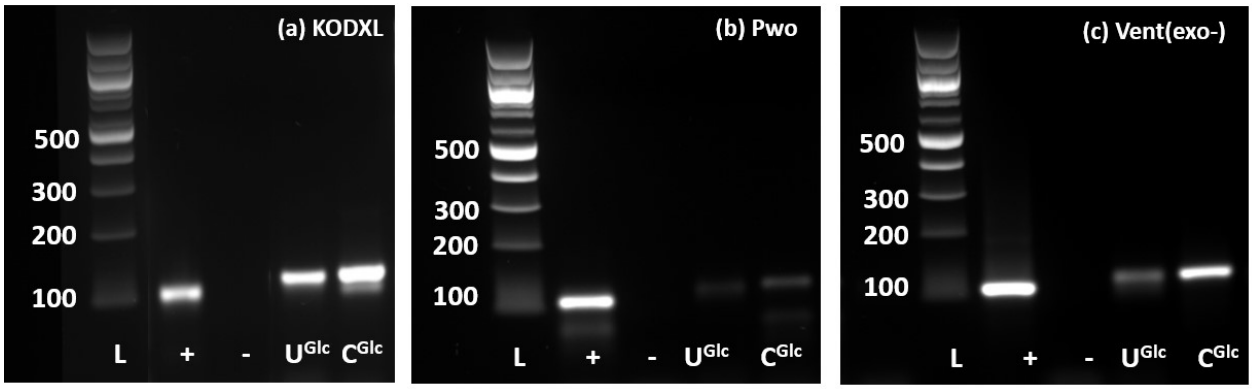


Figure 45: Agarose gel analysis of natural and modified PCR template products (98DNA) using (a) KODXL (b) Pwo and (c) Vent (exo-) DNA polymerase. Lane 1 (L): 100 bp ladder (commercial mix of dsDNA of specific lengths); Lane 2 (+): PCR product with dATP, dCTP, dGTP, dTTP; Lane 3: (-): PCR product with d(GAC)TP or d(ATG)TP; Lane 4 (**U^{Glc}**): PCR product with **dU^{Glc}TP**, dATP, dCTP, dGTP; Lane 5 (**C^{Glc}**): PCR product with **dC^{Glc}TP**, dATP, dTTP, dGTP.

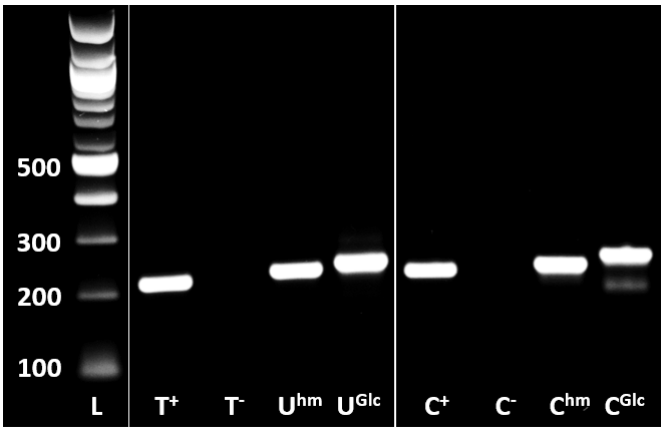


Figure 46: Agarose gel analysis of natural and modified PCR template products using KODXL DNA Polymerase (235DNA). Lane 1 (L): 100 bp ladder (commercial mix of dsDNA of specific lengths); Lane 2 (T^+): PCR product with dATP, dCTP, dGTP, dTTP; Lane 3: (T^-): PCR product with dATP, dCTP, dGTP; Lane 4 (**U^{hm}**): PCR product with **dU^{hm}TP**, dATP, dCTP, dGTP; Lane 5 (**U^{Glc}**): PCR product with **dU^{Glc}TP**, dATP, dCTP, dGTP. Lane 6 (C^+): PCR product with dATP, dCTP, dGTP, dTTP; Lane 7: (C^-): PCR product with dATP, dGTP, dTTP; Lane 8 (**C^{hm}**): PCR product with **dC^{hm} TP**, dATP, dGTP, dTTP; Lane 9 (**C^{Glc}**): PCR product with **dC^{Glc}TP**, dATP, dGTP, dTTP.

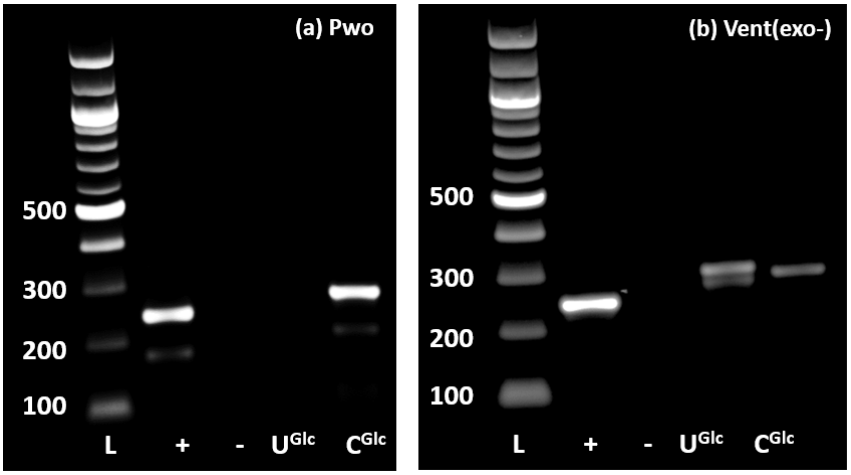


Figure 47: Agarose gel analysis of natural and modified PCR template products using (a) Pwo and (b) Vent (exo-) DNA polymerase (235DNA). Lane 1 (L): 100 bp ladder (commercial mix of dsDNA of specific lengths); Lane 2 (+): PCR product with dATP, dCTP, dGTP, dTTP; Lane 3: (-): PCR product with d(GAC)TP or d(ATG)TP; Lane 4 (**U^{Glc}**): PCR product with **dU^{Glc}TP**, dATP, dCTP, dGTP; Lane 5 (**C^{Glc}**): PCR product with **dC^{Glc}TP**, dTTP, dATP, dGTP.

3.2.2.5. Enzymatic glucosylation of C^{hm}-modified DNA

Next, the enzymatic glucosylation on the **235DNA_C^{hm}** template using T4-βGT glucosyl transferase was developed. In order to confirm the full conversion of glucosylation using glucosyl transferase, enzymatic digestion[150] of **235DNA_C^{Glc*}** to nucleoside level was performed which was followed by LC-MS analysis (Figure 48, 49). The retention time of the nucleosides in interest was analyzed using the standard nucleosides as references (Figure 50-51, Table 9). The results of the LC-MS analysis showed that the glucosylation was complete (no traces of **dC^{hm}** nucleoside peak were observed). Fidelity of the sequence of **235DNA_C^{Glc*}** was also confirmed by Sanger sequencing.

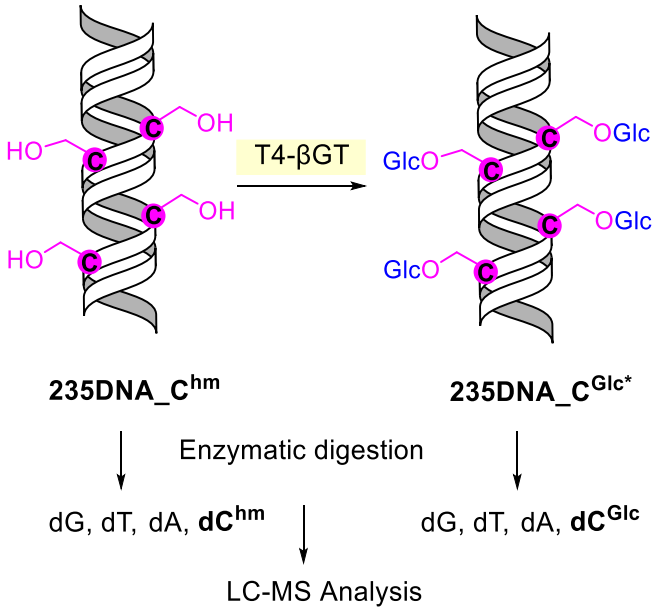


Figure 48: Schematic depiction of the enzymatic glucosylation of **235DNA_C^{hm}** and its enzymatic digestion

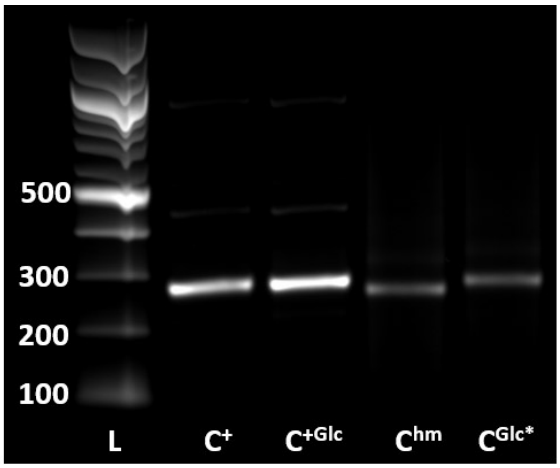


Figure 49: Agarose gel analysis of natural and **C^{hm}**-modified PCR template products before and after enzymatic glucosylation. Lane 1 (L): 100 bp ladder (commercial mix of dsDNA of specific lengths); Lane 2 (C^{+}): PCR product with dATP, dCTP, dGTP, dTTP; Lane 3: (**C^{+Glc}**): PCR product with dATP, dCTP, dGTP, dTTP followed by enzymatic glucosylation; Lane 4 (**C^{hm}**): PCR product with **dC^{hm}TP**, dATP, dTTP, dGTP; Lane 5 (**C^{Glc*}**): PCR product with **dC^{hm}TP**, dATP, dTTP, dGTP, followed by enzymatic glucosylation.

Table 9: Retention time and the mass of the extracted nucleoside peaks

Nucleoside	Retention time (minutes)	Mass fragments
dC^{Glc}	10.8	420.0, 304.1
dC^{hm}	6.6	258.0, 142.1

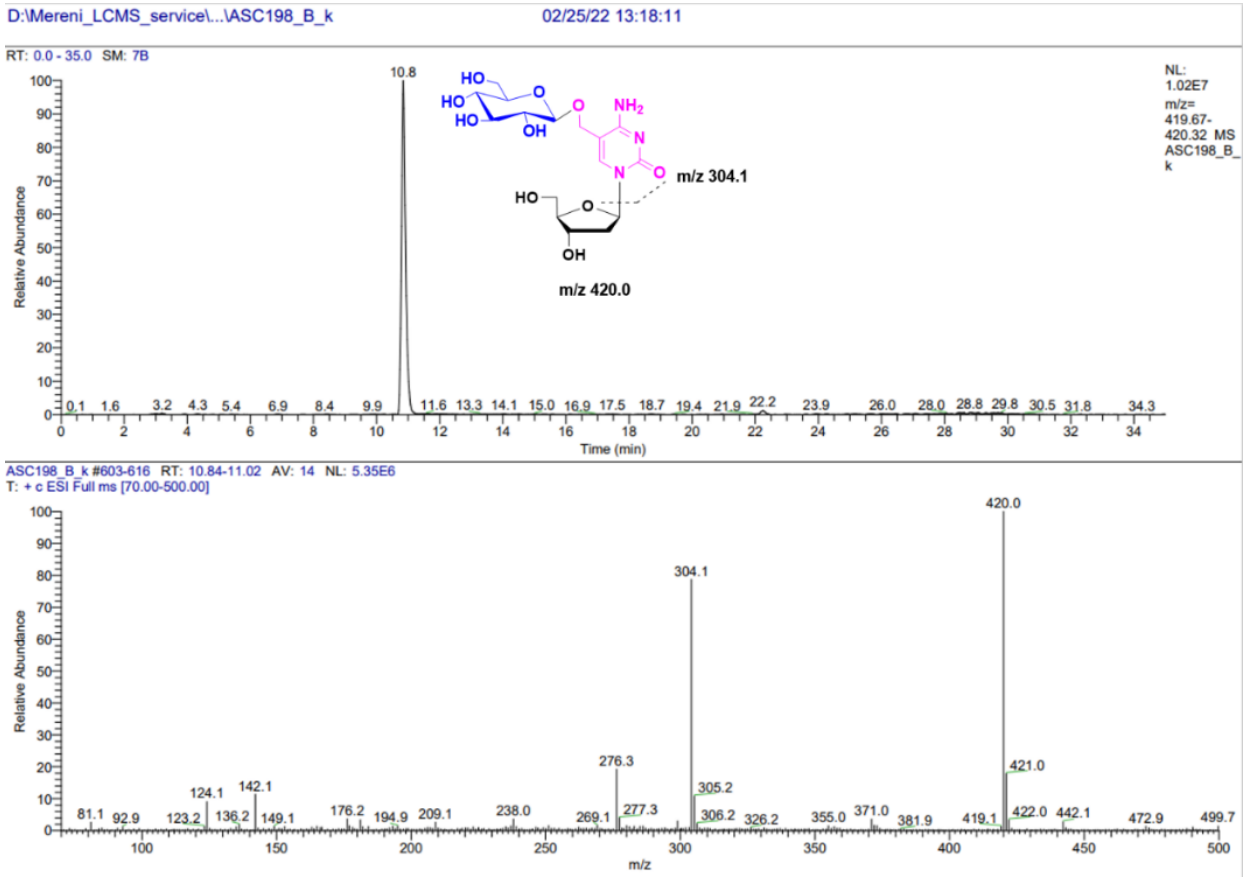


Figure 50: Extracted ion chromatogram of the fragment **dC^{Glc}** from the digested DNA mix of **235DNA_C^{Glc*}** and the mass spectrum of the peak

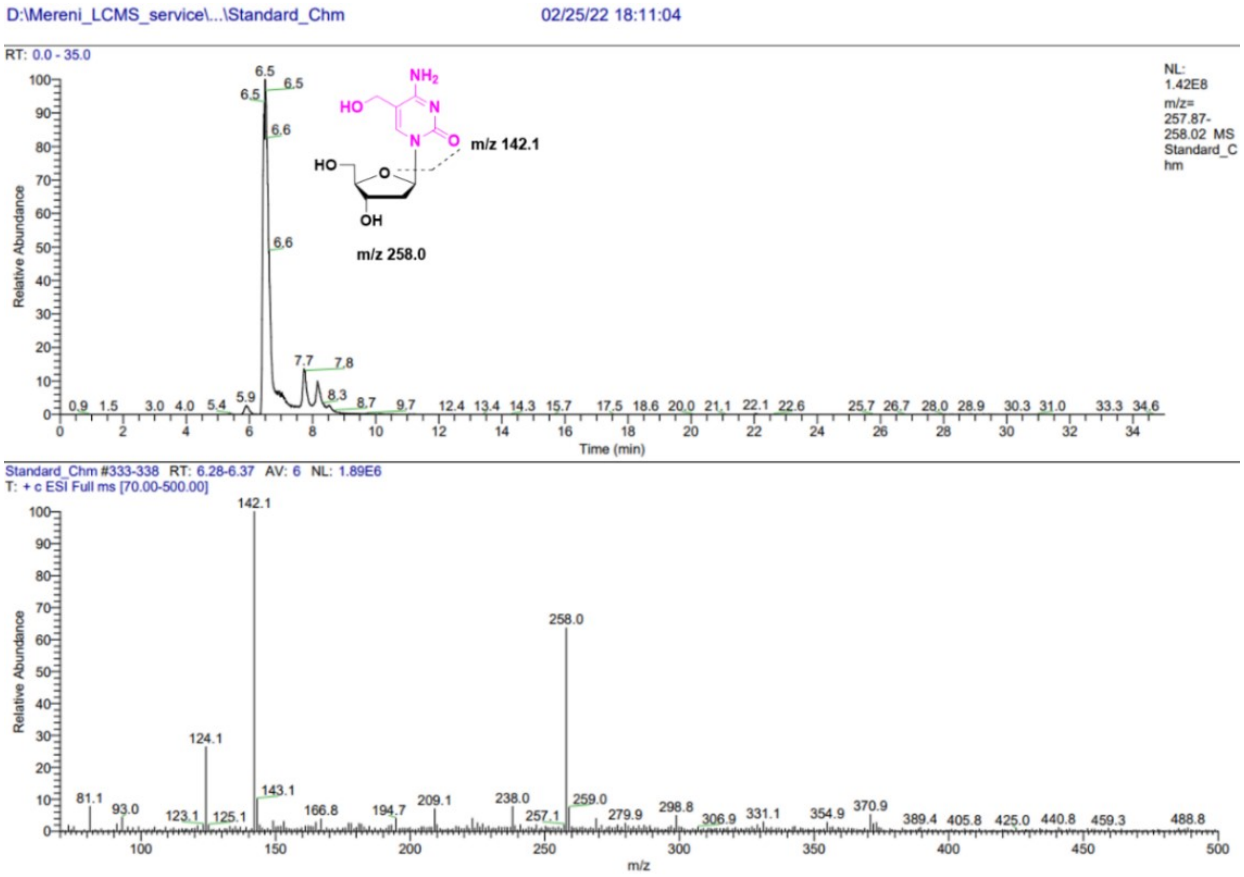


Figure 51: Extracted ion chromatogram of the standard **dC^{hm}** and the mass spectrum of the peak

3.2.3. Study of cleavage of modified DNA with REs

DNA templates for the restriction endonuclease cleavage investigations were chosen in accordance with the previously published studies from the group.[125, 126]. The design of the DNA templates was in such a way as to accommodate the modifications within the recognition sequences of the four REs (AflII, KpnI, PvuII, and RsaI). These templates containing the specific recognition sequences for the selected REs are previously shown to cleave DNA containing **U^{hm}** or **C^{hm}** (Table 10).

Table 10: Sequences of oligonucleotides used for the RE studies. [a] 5'-end of oligonucleotide is labelled by 6-FAM. [b] primer sequences in template are underlined. [c] specific sequence for restriction endonuclease is in bold and position of modified nucleobases in product strand is underlined.

List of primers and templates used in the study:		
Oligonucleotide	**Sequence (5'-3')**	**Length (nt)**
Primers		
Prim^{PEX} [a]	CATGGGCGGCATGGG	15
Templates		
Temp^{RsT} [b,c]	TTCGTCGTCG**GTA**CGCCCATGCCGCCCATG	30
Temp^{KpT} [b,c]	TTCGTCGTC**GGTA**CCCCCATGCCGCCCATG	30
Temp^{AflT} [b,c]	TTCGTCGTC**CTTAAG**CCCATGCCGCCCATG	30
Temp^{PvT} [b,c]	TTCGTCGTC**CAGCTG**CCCATGCCGCCCATG	30
Temp^{RsC} [b,c]	AACTACTACT**GTAC**ACCCATGCCGCCCATG	30
Temp^{KpC} [b,c]	AACTACTAC**GGTACC**CCCATGCCGCCCATG	30

To test the influence of the rather bulky glucose modification on RE cleavage, **dN^{X}**-modified (X = hm or Glc, N = U or C) DNA PEX products were prepared using KODXL. For all templates and modifications, the PEX reaction went quite smoothly and produced full-length extended products that were examined using denaturing PAGE. In the next step, all PEX reaction products (natural DNA, hm-modified DNA, or Glc-modified DNA) were then exposed to the appropriate RE for the cleavage reaction. As expected, the presence of the bulky glucose modification completely prevented the cleavage by the REs in all cases, whereas the sequences containing the corresponding hydroxymethyl modifications were fully cleaved by the same enzymes (Figure 52, 53).

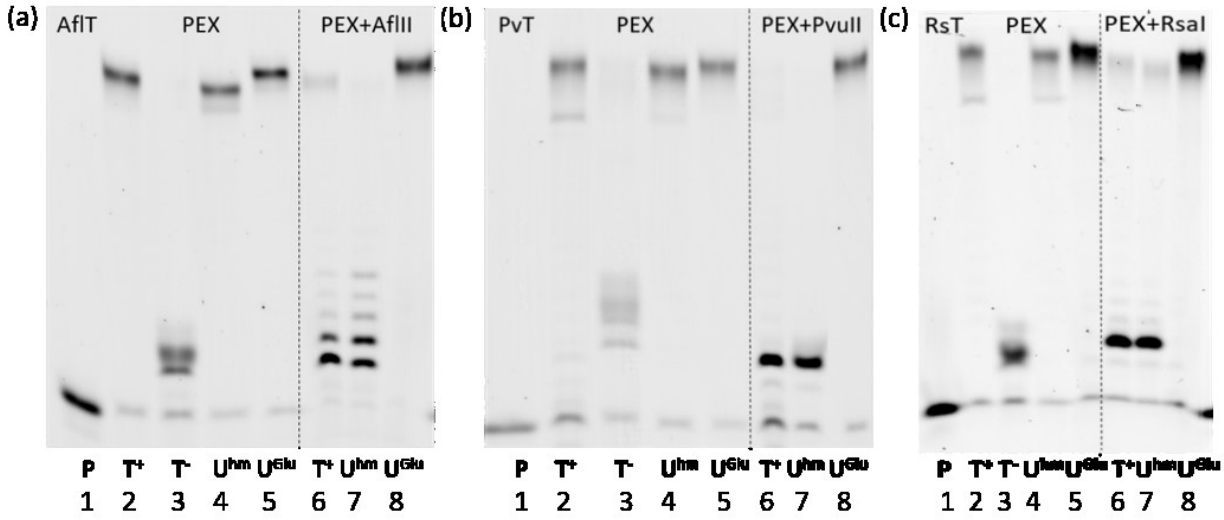


Figure 52: Denaturing PAGE analysis of PEX with templates (a) AflT (b) PvT (c) RsT; Lane 1 (P) : primer; lane 2 (T^+) : PEX product with dATP, dCTP, dGTP, dTTP; lane 3 (T^-) : PEX product with dATP, dCTP, dGTP; lane 4 (**U^{hm}**) : PEX product with **dU^{hm}TP**, dCTP, dGTP, dATP; lane 5 (**U^{Glc}**): PEX product with **dU^{Glc}TP**, dCTP, dGTP, dATP; lane 6 (T^+) : PEX product with dATP, dCTP, dGTP, dTTP, followed by the reaction with corresponding RE; lane 7 (**U^{hm}**) : PEX product with **dU^{hm}TP**, dCTP, dGTP, dATP, followed by the reaction with corresponding RE; lane 8 (**U^{Glc}**) : PEX product with **dU^{Glc}TP**, dCTP, dGTP, dATP, followed by the reaction with corresponding RE.

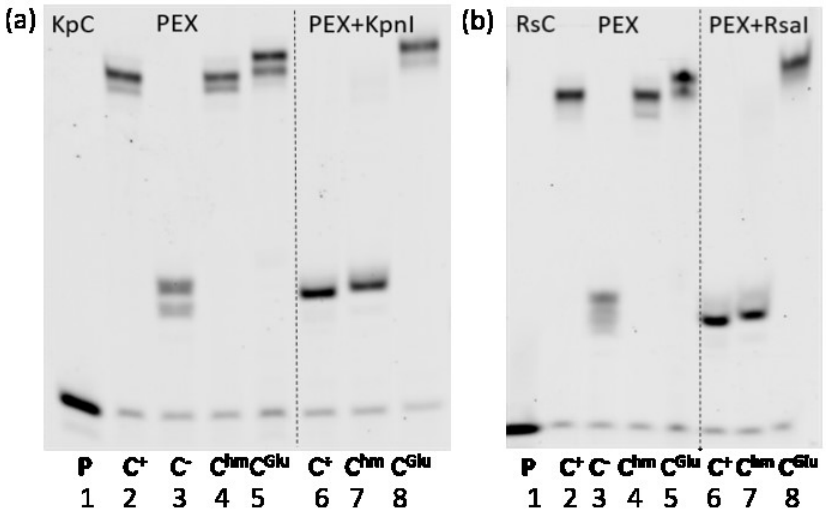


Figure 53: Denaturing PAGE analysis of PEX with templates (a) KpC (b) RsC; Lane 1 (P) : primer; lane 2 (C^+) PEX product with dTTP, dCTP, dGTP, dATP; lane 3 (C^-) : PEX product with dTTP, dGTP, dATP; lane 4 (**C^{hm}**) : PEX product with **dC^{hm}TP**, dTTP, dGTP, dATP; lane 5 (**C^{Glc}**): PEX product with **dC^{Glc}TP**, dTTP, dGTP, dATP; lane 6 (C^+) : PEX product with dTTP, dCTP, dGTP, dATP, followed by the reaction with corresponding RE; lane 7 (**C^{hm}**) : PEX product with **dC^{hm}TP**, dTTP, dGTP, dATP, followed by the reaction with corresponding RE; lane 8 (**C^{Glc}**) : PEX product with **dC^{Glc}TP**, dTTP, dGTP, dATP, followed by the reaction with corresponding RE.

3.2.4. *In vitro* transcription studies

The *in vitro* transcription studies were performed by Dr. Olatz Ruiz-Larrabeiti (Institute of Microbiology). From previous reports [61, 132], we know that the 5hmU or 5hmC-modified DNA support increased level of transcription whereas the hydroxymethyl DNA further modified by bulky protecting groups like nitrobenzyl inhibits transcription. Similar to the bulky nitrobenzyl photocaging groups, the glucosylated DNA also did not tolerate cleavage by restriction endonuclease. These findings moved our investigation further to study the transcription levels of the glucosylated 5-(hydroxymethyl)-modified DNA. For the study of the influence of the major-groove modifications on transcription with RNAP from *E. coli*, the 235-mer DNA template containing the 38 bp-long P*veg* promoter was chosen, which was also used in our previous studies. [61] The templates were prepared using natural and modified **dN^{X}TPs** (X = hm or Glc, N = U or C)

by PCR using FAM-labelled forward and reverse primers with KODXL polymerase. In all cases, full-length amplicons were obtained efficiently and after isolation and fluorescence quantification, were used directly for *in vitro* transcription experiments. In contrast to the previous studies (where radioactive labelling is used), FAM labelling was used as the labelling and quantification method. Therefore, the transcription of the hydroxymethyl-modified DNA were once again studied with the quantification by FAM labelling. Fully modified DNAs synthesized in the presence of **dU^{Glc}TP /dU^{hm}TP** or **dC^{Glc}TP /dC^{hm}TP** were used as templates in the *in vitro* transcription reactions by RNAP from *E.coli*.

3.2.4.1. Relative transcription of modified DNA templates prepared by the PCR synthesis of modified triphosphates

In comparison to the natural DNA template, there was a considerable augmentation of transcription in the case of the 5hmU and 5hmC-modified DNA templates, which was in accordance with the previous reports. [61] When compared to the hydroxymethyl-modified DNA, the templates containing the corresponding glucosylated bases (**235DNA_U^{Glc}** or **235DNA_C^{Glc}**) displayed decreased transcription. While **235DNA_U^{Glc}** did not allow any significant transcription (< 4 %) compared to natural DNA, the **235DNA_C^{Glc}** supported a comparable level of transcription as the natural DNA but reduced transcription (to ca. 50%) compared to **235DNA_C^{hm}**. In other words, glucosylation of 5hmU completely inhibits transcription whereas glucosylation of 5hmC reduces transcription to the level of natural DNA. Fluorescence quantification despite being a facile and convenient tool was somewhat less accurate compared to radioactive labelling and hence the standard deviations were higher than in the previous reports. Notwithstanding this issue, the trends in transcription activities were clear (Figure 54, Table 11).

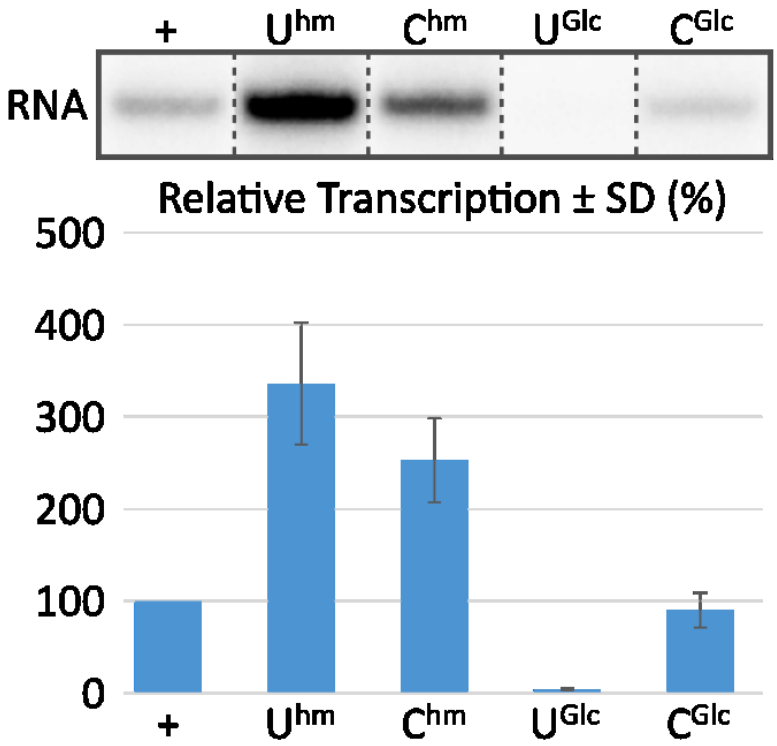


Figure 54: Relative transcription of modified DNA templates (235DNA). Representative primary data (radio-labelled RNA transcripts) are shown above the graph. The bars show the average from at least four independent experiments using three different batches, the error bars show ±SD.

Table 11: Average relative transcription (± SD; average from 4 experiments) of the modified DNA templates.

No.	235DNA	Average relative transcription ± SD
1	+ (Natural)	1
2	U^{hm}	3.36 ± 0.66
3	C^{hm}	2.53 ± 0.46
4	U^{Glc}	0.04 ± 0.01
5	C^{Glc}	0.96± 0.32

3.2.4.2. Relative transcription of modified DNA templates before and after enzymatic glycosylation using glucosyl transferase.

To achieve the bioorthogonal fine-tuning of the transcription, enzymatic glucosylation was done on the **235C^{hm}**-modified DNA (and on natural DNA as a control) using UDP-glc and T4-βGT. [149] The glucosylated DNA after purification and quantification correction, were directly used as templates for the *in vitro* transcriptions. As expected, the glucosylated product **235DNA_C^{Glc}*** supported a similar level of transcription as the **235DNA_C^{Glc}** template prepared by PCR using **dC^{Glc}TP** (Figure 55, Table 12).

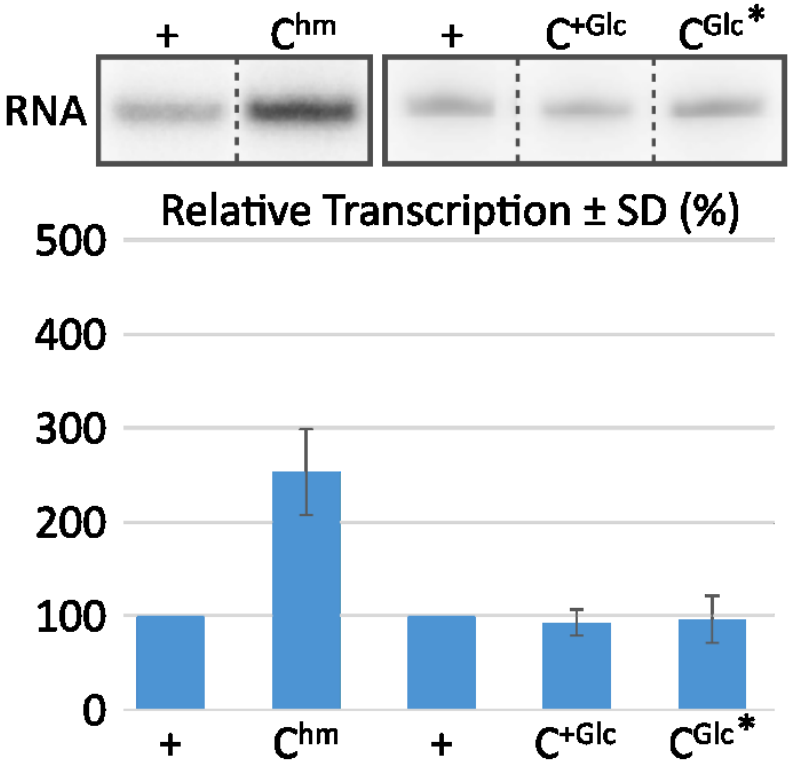


Figure 55: Relative transcription of natural and **C^{hm}**-modified DNA templates (235DNA) before and after enzymatic glucosylation. Representative primary data (radio-labelled RNA transcripts) are shown above the graph. The bars show the average from at least four independent experiments using two batches, the error bars show ±SD.

Table 12: Average relative transcription (± SD; average from 4 experiments) of the modified DNA templates.

No.	235DNA	Average relative transcription ± SD
1	+	1
2	$\mathbf{C^{hm}}$	2.53 ± 0.46
3	+	1
4	$\mathbf{C^{+Glc}}$	0.93 ± 0.14
5	$\mathbf{C^{Glc*}}$	0.94± 0.10

4. Conclusions

In this thesis work, we have synthesized photocaged or glucosylated derivatives of epigenetic 5-(hydroxymethyl)pyrimidine nucleotides and DNA combining the chemical and enzymatic techniques and have studied the bacterial *in vitro* transcription of the modified DNA templates.

In the first part of the thesis, we extended the previous studies done on the polymerase constructed nitrobenzyl photocaged 5-(hydroxymethyl)pyrimidine-modified DNA[132] to the corresponding specific site nitrobenzyl photocaged DNA. For this, we synthesized the nitrobenzyl photocaged 5-(hydroxymethyl)pyrimidine nucleoside phosphoramidites and these were used to create oligonucleotides containing photocaged pyrimidine bases at specific positions. These photocaged oligonucleotides served as primers for the PCR synthesis of specific site photocaged DNA templates. Photochemical uncaging of these specific site photocaged DNA upon irradiation resulted in the corresponding specific site 5-(hydroxymethyl)pyrimidine-modified epigenetic DNA. This way we developed the specific site photoactivatable epigenetic labelling of DNA. Bacterial *in vitro* transcription studies of the specific site photocaged DNA revealed that a combination of photocaged 5-(hydroxymethyl)pyrimidines in the -35 region of the promoter is required to achieve a significant inhibition in transcription showing that these sites are crucial for the interaction with bacterial RNAP. However we did not observe any enhancement in transcription due to the presence of the uncaged 5-(hydroxymethyl)-modifications at any specific sites of the promoter, compared to the natural template. Since we have demonstrated that fully modified templates with 5hmU or 5hmC bases improve transcription by up to 350% or 250%, respectively [61], it appears that a combination of several or many hydroxymethylpyrimidines at various sites in the promoter is required for such an observed enhancement. These newly developed photocaged phosphoramidites are useful for in-depth research into the effects of pyrimidines with epigenetic modifications on transcription using both bacterial and eukaryotic RNAP. They also find applications in the identification of the biological functions of these epigenetic modifications and for the development of transcription optical switches.

The second half of the thesis was focused on the glucosylated 5-(hydroxymethyl)pyrimidine-modified DNA. We have developed the synthesis of glucosylated 5-(hydroxymethyl)pyrimidine nucleoside triphosphates and shown that they are good substrates for DNA polymerases. We have shown that they can be used to efficiently synthesize glucosylated DNA both by PEX (for short sequences with many modifications) and by PCR (for long DNA sequences with a high frequency of modifications). The enzymatic technique is useful to create long fully modified DNA sequences suitable for transcription experiments, as opposed to prior works [133-135] that used glucosylated 5-(hydroxymethyl)pyrimidine nucleoside phosphoramidites and their incorporation to short oligonucleotides. We have then confirmed that glucosylation of DNA provides complete protection against RE cleavage. The impact on transcription, on the other hand, was found to be substantially different for each nucleobase. While the presence of **dU^{Glc}** totally suppressed RNAP transcription, DNA containing **dC^{Glc}** allowed transcription that is comparable to natural DNA but roughly 50% lower than DNA carrying 5hmC. This suggests that the glucosylation by T4-βGT serves to protect viral DNA from bacterial restriction enzymes while simultaneously fine-tuning viral gene expression by influencing transcription. We also developed enzymatic glucosylation using T4-βGT to bioorthogonally fine-tune the transcription. Glucosylation of 5hmU, on the other hand, fully turned OFF bacterial transcription, but the situation in kinetoplastid parasites could be very different, and would require further study. To the end, glucosylated 5-(hydroxymethyl) pyrimidine nucleoside phosphoramidites were also synthesized for the use in the automated solid phase synthesis of specific site glucosylated DNA. Further detailed studies on this could shed light into the unknown biological roles of these hypermodified nucleobases.

5. Experimental section

5.1. General remarks

Synthesis Part:

Unless otherwise stated, all materials were obtained from commercial providers and used without additional purification. $POCl_3$ and $PO(OMe)_3$ were distilled prior to use. DIPEA was dried by distillation at reduced pressure over CaH_2. HPLC quality water was used in the synthetic part. Chemicals were of analytical grade. Unless otherwise stated, all reactions were carried out using the usual, septa, cannula, and syringe procedures in a positive environment of argon. ^{1}H, ^{13}C and ^{31}P NMR spectra were acquired on a Bruker AVANCE IIIHD 500 (^{1}H at 500.0 MHz, ^{13}C at 125.7 MHz, ^{31}P at 202.4 MHz) and JEOL ECZR 500 MHz (^{1}H at 500.2 MHz, ^{13}C at 125.8 MHz and ^{31}P at 202.5 MHz) in $CDCl_3$ or CD_3CN or D_2O (referenced to the residual solvent signal) at 25 °C. ^{1}H and ^{13}C resonances were fully assigned using H,H-COSY, H,C-HSQC and H,C-HMBC techniques. All chemical shifts are quoted on the δ scale in ppm. For samples measured in D_2O, signals are referenced to the signal of ^{t}BuOH (10% v/v solution in D_2O, 1 drop) as an external standard (1.24 ppm in ^{1}H, 32.43 ppm in ^{13}C). Coupling constants (J) are reported in Hz with the following splitting abbreviations: s = singlet, d= doublet, t = triplet, q = quartet, m = multiplet. MS and HRMS were measured on a LTQ Orbitrap XL spectrometer (Thermo Fisher Scientific) by ESI ionization technique. The MALDI-TOF spectra were measured on an UltrafleXtreme MALDI-TOF/TOF (Bruker) mass spectrometer. LC-MS measurements were done on a RP C18 column connected to a LC-MS system which consists of TSP quaternary pump P4000, TSP autosampler AS3000 and ion-trap mass spectrometer Advantage (Thermo). All mass spectra were acquired by the MS service at IOCB. Reactions were monitored by thin layer chromatography (TLC) on TLC Silica gel 60 F254 (Merck) and detected by UV (254 nm) or by a solution of 4-anisaldehyde 3.6 v/v in ethanol and 10% v/v of sulphuric acid. Reactions were monitored by Advion Expression Compact Mass Spectrometer connected with Plate Express® TLC Plate Reader using ESI. Nucleosides were purified using an ISCO Combiflash Rf system on disposable RediSep Rf Silica Gel columns. Purification of nucleoside phosphoramidites were performed using manual column chromatography using silica gel (40–63 μm). Purification of nucleoside triphosphates was

performed using HPLC (Waters modular HPLC system) on a Kinetex column (RP 18, 5µm EVO C18 100 LC-Column 250 x 21.2 mm) or on 10 µm C18 reverse-phase (RP) column (Phenomenex, Luna C18 (2) 100 Å). Triphosphorylation reactions were analyzed by TLC using IPAV (Isopropyl alcohol: NH_4OH: H_2O = 11: 7: 2) as mobile phase. Melting points were determined on a Stuart SMP40 melting point apparatus and are uncorrected. The purity of all final compounds was determined by NMR spectra.

Solid phase synthesis part:

Solvents and reagents for the solid-phase synthesis of oligonucleotides were purchased from Link Technologies, Sigma-Aldrich and Thermo Fisher-Scientifics. Natural phosphoramidites were purchased from Sigma-Aldrich. Modified oligonucleotides were synthesized through standard phosphoramidite chemistry with an automated DNA synthesizer (Mermade 8, Bioautomation). Purification of the prepared oligonucleotides was performed using semi-preparative HPLC (Waters modular HPLC system) on a RP column packed with 10 µm C18 (Phenomenex, Luna C18 (2) 100 Å). The products were analyzed by MALDI-TOF MS and ESI-MS. Concentrations of DNA solutions were calculated using extinction coefficients obtained from the on-line tool at https://www.atdbio.com/tools/oligo-calculator and A260 values were measured on a Cary100 Bio UV-Vis spectrophotometer (Varian).

Biochemistry part:

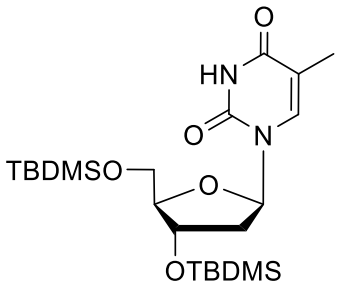
Synthetic oligonucleotides (primers, FAM-labelled primers, templates, and biotinylated templates) were purchased from Generi Biotech. Natural dNTPs (dATP, dCTP, dGTP, dTTP) were purchased from New England Biolabs. DNA polymerases and its corresponding polymerase reaction buffer were purchased from Merck (KOD XL), VWR Life Science (Peqlab; Pwo) and from New England Biolabs (Taq, Vent exo-). Restriction endonucleases and its corresponding reaction buffer were from New England Biolabs. T4-βGT along with Epi buffer and UDP-glc were purchased from Thermofisher scientific. Nucleoside digestion mix and its reaction buffer was from New England Biolabs. PAGE stop solution used after PEX reactions contains: 95% [v/v] formamide, 0.5 mM EDTA, 0.025% [w/v] bromophenol blue, 0.025% [w/v] xylene cyanol, 0.025% SDS. Samples after

PEX reaction were analyzed by 12.5% PAGE (Acrylamide/bisacrylamide 19:1) gel under denaturing conditions (1 hour, 50 °C, 1x TBE buffer). Samples were concentrated on a CentriVap vacuum concentrator system (Labconco). All PCR products were purified with Agencourt AMPure XP magnetic particles (Beckman Coulter Life Science - GE Healthcare). Streptavidin magnetic particles were purchased from Roche. QIAquick Nucleotide removal kit was purchased from QIAGEN (Biotech, Czech Republic). Microcon-10kDa centrifugal filters were purchases from Merck. The concentration of DNA solutions was calculated using A260 values measured on a Nanodrop 1000 Spectrophotometer (Thermo Scientific). PAGE and Agarose gels were analyzed with fluorescence scanner Typhoon FLA 9500 (GE Healthcare, USA). Image J quantifier was used for the quantification analysis of the agarose gel. All solutions for biochemistry experiments were prepared in Milli-Q water. The PCR products were sequenced by Sanger sequencing (Eurofins genomics) and analyzed by Geneious prime.

5.2. Synthesis of nitrobenzyl photocaged 5-(hydroxymethyl)pyrimidine nucleoside phosphoramidites and their use in chemical synthesis of specific site photocaged DNA.

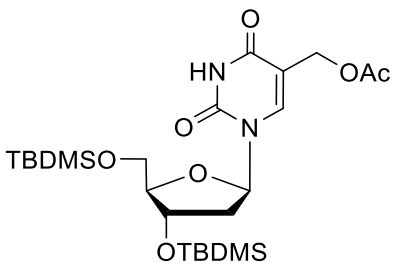
5.2.1. Synthesis of modified nucleosides

3', 5'-Bis-*O*-(*tert*-butyldimethylsilyl)thymidine (2)



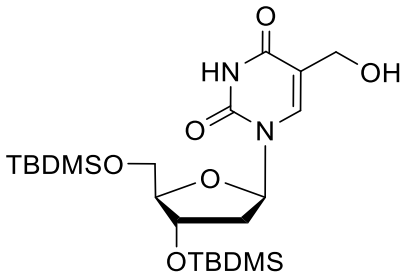
The title compound **2** was prepared from thymidine (**1**) following the reported procedures. The spectral data were consistent with the literature.[144, 145] Yield: 95-99 %.

3', 5'-Bis-*O*-(*tert*-butyldimethylsilyl)-5-(acetyloxymethyl)-2'-deoxyuridine (4)



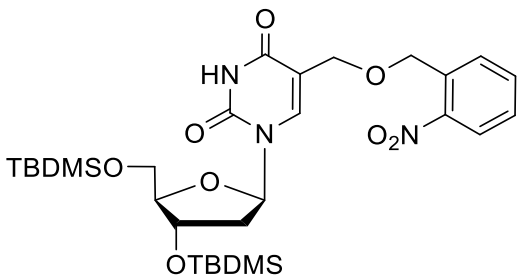
The title compound **4** was prepared in two steps from compound **2** following the reported procedures. The spectral data were consistent with the literature.[144, 145] Yield: 53 % in two steps.

3', 5'-Bis-*O*-(*tert*-butyldimethylsilyl)-5-(hydroxymethyl)-2'-deoxyuridine (5)



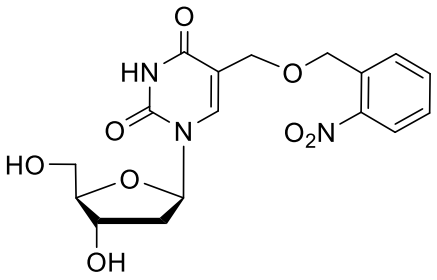
The title compound **5** was prepared from **4** following the reported procedures. The spectral data were consistent with the literature. [144, 145] Yield: 69 %.

3', 5'-Bis-*O*-(*tert*-butyldimethylsilyl)-5-(2-nitrobenzyloxymethyl)-2'-deoxyuridine (6)



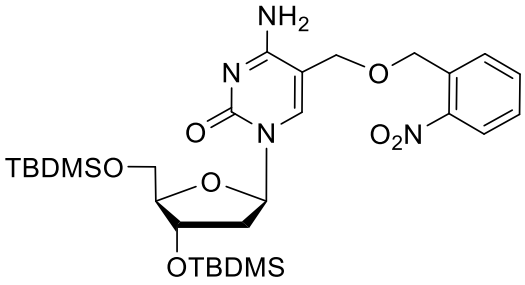
The title compound **6** was prepared by the reaction of **5** with 2-nitrobenzyl bromide following the reported procedures. The spectral data were consistent with the literature. [144, 145] Yield: 40 %.

5-(2-Nitrobenzyloxymethyl)-2'-deoxyuridine (7, dU^{NB})



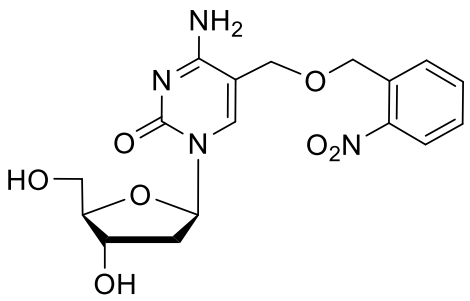
The silyl group deprotection of **6** gave the title compound **7**. The spectral data were consistent with the literature. [144, 145] Yield: 40 %.

3', 5'-Bis-*O*-(*tert*-butyldimethylsilyl)-5-(2-nitrobenzyloxymethyl)-2'-deoxycytidine (9)



The title compound **9** was prepared from **6** in two steps through its oxo form (**8**) following the reported procedures. The spectral data were consistent with the literature.[126] Yield: 80% in two steps.

5-(2-Nitrobenzyloxymethyl)-2'-deoxycytidine (10, dC^{NB})



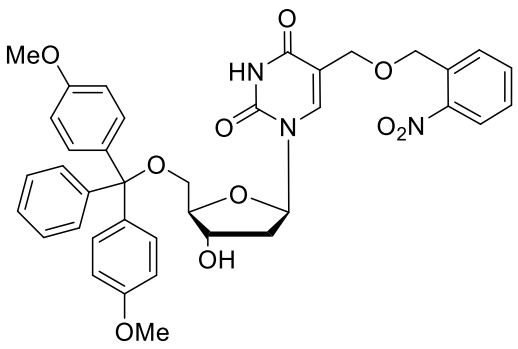
The title compound **10** was prepared by the silyl group deprotection of **9** following the reported procedures. The spectral data were consistent with the literature. [126] Yield: 50 %.

5.2.2. Synthesis of modified nucleoside phosphoramidites

General procedure A: Synthesis of trityl-protected modified nucleosides

The nucleosides were dried by repeated co-evaporations with anhydrous pyridine (3 × 5 ml) and finally dissolved in 10 ml of anhydrous pyridine along with DMAP (0.10 equiv.). DMTrCl (1.2 equiv.) in 5 ml of anhydrous pyridine, was added in 4 portions over one hour and the reaction mixture was stirred at room temperature for 6 hours. Pyridine was removed under reduced pressure at a rotary evaporator and the crude reaction mixture was re-dissolved in CH_2Cl_2 (50 ml), washed with 10% aqueous $NaHCO_3$ (50 ml), H_2O (100 ml) and brine (50 ml), and dried over Na_2SO_4. Purification by flash chromatography (CH_2Cl_2: MeOH, with 0.5% Et_3N) afforded the desired compounds.

5'-*O*-[Bis(4-methoxyphenyl)phenylmethyl]-5-(2-nitrobenzyloxymethyl)-2'-deoxyuridine (11)



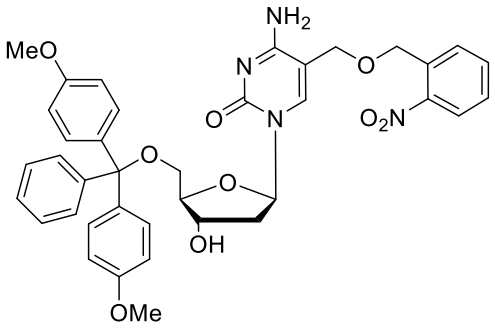
The title compound **11** was obtained from **7** (1 g, 2.5 mmol), DMAP (31 mg, 0.25 mmol, 0.10 equivalent) and DMTrCl (1 g, 3.0 mmol, 1.2 equiv.) following the general procedure A. The product was obtained as a yellowish foam (885 mg, 50%) after purification by flash chromatography (CH_2Cl_2: MeOH 99:1, Et_3N 0.5%). m.p. 58-65 °C.

^{1}H NMR (500.0 MHz, $CDCl_3$): 2.35 (ddd, 1H, J_{gem} = 13.6, $J_{2'b,1'}$ = 7.5, $J_{2'b,3'}$ = 6.2, H-2′b); 2.45 (ddd, 1H, J_{gem} = 13.6, $J_{2'a,1'}$ = 6.0, $J_{2'a,3'}$ = 3.6, H-2′a); 3.38 (dd, 1H, J_{gem} = 10.6, $J_{5'b,4'}$ = 3.1, H-5′b); 3.46 (dd, 1H, J_{gem} = 10.6, $J_{5'a,4'}$ = 3.4, H-5′a); 3.57 (d, 1H, J_{gem} = 11.0, $CH_a\mathbf{H_b}O$-5); 3.71, 3.72 (2 × s, 2 × 3H, CH_3O); 3.89 (d, J_{gem} = 11.0, $C\mathbf{H_a}H_bO$-5); 4.06 (ddd, 1H, $J_{4',5'}$ = 3.4, 3.1, $J_{4',3'}$ = 2.9, H-4′); 4.55, 4.59 (2 × d, J_{gem} = 15.3, CH_2O-1″); 4.60 (ddd, 1H, $J_{3',2'}$ = 6.2, 3.6, $J_{3',4'}$ = 2.9, H-3′); 6.40 (dd, 1H, $J_{1',2'}$ = 7.5, 6.0, H-1′); 6.76 – 6.80 (m, 4H, H-*m*-C_6H_4-DMTr); 7.17 (m, 1H, H-*p*-C_6H_5-DMTr); 7.23 – 7.31 (m, 6H, H-*o*-C_6H_4-DMTr, H-*m*-C_6H_5-DMTr); 7.36 7.40 (m, 3H, H-4″, H-*o*-C_6H_5-DMTr); 7.54 (td, 1H, $J_{5'',4''} = J_{5'',6''}$ = 7.7, $J_{5'',3''}$ = 1.4, H-5″); 7.71 (dq, 1H, $J_{6'',5''}$ = 7.7, $J_{6'',4''} = J_{6'',CH2}$ = 1.2, H-6″); 7.91 (s, 1H, H-6); 8.01 (dd, 1H, $J_{3'',4''}$ = 8.2, $J_{3'',5''}$ = 1.4, H-3″).

^{13}C NMR (125.7 MHz, $CDCl_3$): 41.01 (CH_2-2′); 55.15 (CH_3O); 63.39 (CH_2-5′); 64.87 (CH_2O-5); 69.13 (CH_2O-1″); 72.10 (CH-3′); 84.83 (CH-1′); 86.06 (CH-4′); 86.90 (C-DMTr); 111.46 (C-5); 113.22, 113.23 (CH-*m*-C_6H_4-DMTr); 124.47 (CH-3″); 127.11 (CH-*p*-C_6H_5-DMTr); 127.78 (CH-4″); 128.00 (CH-*m*-C_6H_5-DMTr); 128.08 (CH-*o*-C_6H_5-DMTr); 128.67 (CH-6″); 130.07, 130.08 (CH-*o*-C_6H_4-DMTr); 133.71 (CH-5″); 134.87 (C-1″); 135.16, 135.30 (C-*i*-C_6H_4-DMTr); 139.57 (CH-6); 144.25 (C-*i*-C_6H_5-DMTr); 146.96 (C-2″); 149.97 (C-2); 158.62, 158.63 (C-*p*-C_6H_4-DMTr); 162.36 (C-4).

HR MS (ESI+) m/z : $[M+Na]^+$ Calcd for $C_{38}H_{37}O_{10}N_3Na$ 718.2371; found 718.2363.

5'-*O*-[Bis(4-methoxyphenyl)phenylmethyl]-5-(2-nitrobenzyloxymethyl)-2'-deoxycytidine (13)



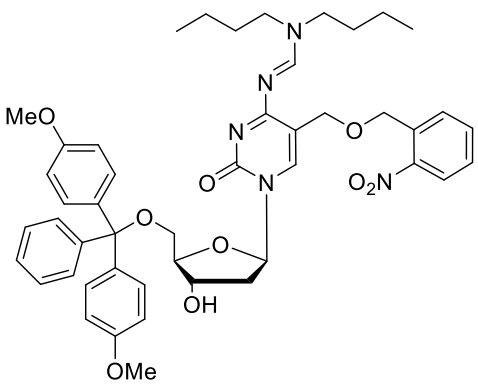
The title compound **13** was obtained from **10** (1.1 g, 2.8 mmol), DMAP (34 mg, 0.28 mmol, 0.10 equivalent) and DMTrCl (1.15 g, 3.4 mmol, 1.2 equiv.) following the general procedure A. The product was obtained as a yellowish foam (1.32 g, 68 %) after purification by flash chromatography (CH_2Cl_2: MeOH 99:1, Et_3N 0.5%). m.p. 148-150 °C.

^{1}H NMR (500.0 MHz, $CDCl_3$): 2.28 (ddd, 1H, J_{gem} = 13.6, $J_{2'b,1'}$ = 6.7, $J_{2'b,3'}$ = 6.2, H-2′b); 2.66 (ddd, 1H, J_{gem} = 13.6, $J_{2'a,1'}$ = 6.1, $J_{2'a,3'}$ = 3.8, H-2′a); 3.32, 3.51 (2 × dd, 2 × 1H, J_{gem} = 10.6, $J_{5',4'}$ = 3.0, H-5′); 3.62 (d, 1H, J_{gem} = 11.9, $CH_a\mathbf{H_b}$O-5); 3.72, 3.73 (2 × s, 2 × 3H, CH_3O); 3.83 (d, J_{gem} = 11.9, $C\mathbf{H_a}H_b$O-5); 4.11 (q, 1H, $J_{4',3'}$ = $J_{4',5'}$ = 3.0, H-4′); 4.47 (s, 2H, CH_2O-1″); 4.57 (ddd, 1H, $J_{3',2'}$ = 6.4, 3.8, $J_{3',4'}$ = 3.0, H-3′); 6.44 (dd, 1H, $J_{1',2'}$ = 6.7, 6.1, H-1′); 6.78 – 6.82 (m, 4H, H-*m*-C_6H_4-DMTr); 7.18 (m, 1H, H-*p*-C_6H_5-DMTr); 7.24 – 7.29 (m, 6H, H-*o*-C_6H_4-DMTr, H-*m*-C_6H_5-DMTr); 7.36 - 7.40 (m, 2H, H-*o*-C_6H_5-DMTr); 7.43 (ddd, 1H, $J_{4'',3''}$ = 8.2, $J_{4'',5''}$ = 7.3, $J_{4'',6''}$ = 1.6, H-4″); 7.50 (dd, 1H, $J_{6'',5''}$ = 7.8, $J_{6'',4''}$ = 1.6, H-6″); 7.55 (ddd, 1H, $J_{5'',6''}$ = 7.8, $J_{5'',4''}$ = 7.3, $J_{5'',3''}$ = 1.3, H-5″); 8.02 (s, 1H, H-6); 8.04 (dd, 1H, $J_{3'',4''}$ = 8.2, $J_{3'',5''}$ = 1.3, H-3″).

^{13}C NMR (125.7 MHz, $CDCl_3$): 42.18 (CH_2-2′); 55.18 (CH_3O); 63.30 (CH_2-5′); 67.76 (CH_2O-5); 68.87 (CH_2O-1″); 71.86 (CH-3′); 86.14 (CH-4′); 86.17 (CH-1′); 86.77 (C-DMTr); 102.19 (C-5); 113.24, 113.25 (CH-*m*-C_6H_4-DMTr); 124.80 (CH-3″); 127.12 (CH-*p*-C_6H_5-DMTr); 127.99 (CH-*m*-C_6H_5-DMTr); 128.24 (CH-*o*-C_6H_5-DMTr); 128.50 (CH-4″); 129.17 (CH-6″); 130.12 (CH-*o*-C_6H_4-DMTr); 133.41 (C-1″); 133.77 (CH-5″); 135.38, 135.44 (C-*i*-C_6H_4-DMTr); 140.82 (CH-6); 144.33 (C-*i*-C_6H_5-DMTr); 147.30 (C-2″); 155.86 (C-2); 158.66 (C-*p*-C_6H_4-DMTr); 165.05 (C-4).

HR MS (ESI+) m/z : $[M+Na]^+$ Calcd for $C_{38}H_{38}O_9N_4Na$ 717.2531; found 717.2525.

N^4-*N*,*N*-dibutylformimidamide-5'-*O*-[bis(4-methoxyphenyl)phenylmethyl]-5-(2-nitrobenzyl oxymethyl)-2'-deoxycytidine (14).



The title compound **14** was obtained from **13** (1.36 g, 1.9 mmol). To the precursor **13** dissolved in anhydrous DMF (10 ml) under inert conditions, DBF-DMA (2.2 ml, 9.5 mmol, 5 equiv.) was added and the reaction was stirred overnight at room temperature. The solvent was evaporated under vacuum and the crude reaction mixture was re-dissolved in CH_2Cl_2 (50 ml), washed twice with saturated $NaHCO_3$ (50 ml) and dried over Na_2SO_4. The crude was then purified by flash chromatography (CH_2Cl_2: MeOH 98:2, Et_3N 0.5%) and obtained as a yellowish foam (1.24 g, 76 %). m.p. 79-82 °C.

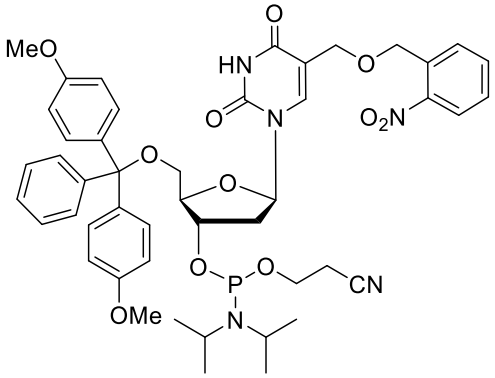
^{1}H NMR (500.0 MHz, $CDCl_3$): 0.80, 0.92 (2 × t, 2 × 3H, J_{vic} = 7.4, C**H₃**$CH_2CH_2CH_2N$); 1.17 – 1.34 (m, 4H, CH_3C**H₂**CH_2CH_2N); 1.49 – 1.62 (m, 4H, CH_3CH_2C**H₂**CH_2N); 2.24 (ddd, 1H, J_{gem} = 13.5, $J_{2'b,1'}$ = 7.2, $J_{2'b,3'}$ = 6.2, H-2′b); 2.70 (ddd, 1H, J_{gem} = 13.5, $J_{2'a,1'}$ = 5.9, $J_{2'a,3'}$ = 3.5, H-2′a); 3.29 – 3.33 (m, 2H, $CH_3CH_2CH_2C$**H₂**N); 3.36, 3.43 (2 × dd, 2 × 1H, J_{gem} = 10.4, $J_{5',4'}$ = 3.5, H-5′); 3.45 – 3.49 (m, 2H, $CH_3CH_2CH_2C$**H₂**N); 3.70, 3.71 (2 × s, 2 × 3H, CH_3O); 3.99 (d, 1H, J_{gem} = 11.1, CH_a**H$_b$**O-5); 4.15 (q, 1H, $J_{4',3'}$ = $J_{4',5'}$ = 3.5, H-4′); 4.24 (d, J_{gem} = 11.1, C**H$_a$**H_bO-5); 4.55 (dt, 1H, $J_{3',2'}$ = 6.2, 3.5, $J_{3',4'}$ = 3.5, H-3′); 4.69 (s, 2H, CH_2O-1″); 6.45 (dd, 1H, $J_{1',2'}$ = 7.2, 5.8, H-1′); 6.75 – 6.78 (m, 4H, H-*m*-C_6H_4-DMTr); 7.14 (m, 1H, H-*p*-C_6H_5-DMTr); 7.28 – 7.31 (m, 6H, H-*o*-C_6H_4-DMTr, H-*m*-C_6H_5-DMTr); 7.33 (ddd, 1H, $J_{4'',3''}$ = 8.0, $J_{4'',5''}$ = 7.5, $J_{4'',6''}$ = 1.6, H-4″); 7.38 (td, 1H, $J_{5'',4''}$ = $J_{5'',6''}$ = 7.5, $J_{5'',3''}$ = 1.5, H-5″); 7.40 - 7.43 (m, 2H, H-*o*-C_6H_5-DMTr); 7.59 (ddt, 1H, $J_{6'',5''}$ = 7.5, $J_{6'',4''}$ = 1.6, $J_{6'',3''}$ = $J_{6'',CH2O}$ = 1.4, H-6″); 8.00 (dd, 1H, $J_{3'',4''}$ = 8.0, $J_{3'',5''}$ = 1.5, H-3″); 8.09 (s, 1H, H-6); 8.79 (s, 1H, CH=N).

^{13}C NMR (125.7 MHz, $CDCl_3$): 13.60, 13.64 (**C**$H_3CH_2CH_2CH_2N$); 19.71, 20.01 (CH_3**C**$H_2CH_2CH_2N$); 28.93, 30.93 (CH_3CH_2**C**H_2CH_2N); 42.12 (CH_2-2′); 45.68, 52.26 ($CH_3CH_2CH_2$**C**H_2N); 55.06 (CH_3O); 63.52 (CH_2-5′); 66.63 (CH_2O-5); 69.00 (CH_2O-1″); 72.07 (CH-3′); 85.96 (CH-4′); 86.37 (CH-1′); 86.53 (C-DMTr); 111.21 (C-5); 113.11 (CH-*m*-C_6H_4-DMTr); 124.30 (CH-3″); 126.78 (CH-*p*-C_6H_5-DMTr); 127.42 (CH-4″); 127.83 (CH-*m*-C_6H_5-DMTr); 128.07 (CH-*o*-C_6H_5-DMTr); 128.49 (CH-6″); 130.01 (CH-*o*-C_6H_4-DMTr); 133.46 (CH-5″); 135.57, 135.60, 135.68 (C-1″, C-*i*-C_6H_4-DMTr); 141.12 (CH-6); 144.56 (C-*i*-C_6H_5-DMTr); 146.80 (C-2″); 156.36 (C-2); 157.69 (CH=N); 158.41 (C-*p*-C_6H_4-DMTr); 170.14 (C-4).
HR MS (ESI+) m/z: $[M+H]^+$ Calcd for $C_{47}H_{56}O_9N_5$ 834.4072; found 834.4071, $[M+Na]^+$ Calcd for $C_{47}H_{55}O_9N_5Na$ 856.3892; found 856.3894.

General procedure B: Synthesis of modified nucleoside phosphoramidites

5'-*O*-DMTr-substituted nucleosides **11** or **14** were dried by repeated co-evaporations with anhydrous pyridine (3 × 5 ml) and subsequent co-evaporations with anhydrous CH_2Cl_2 (3 × 5 ml), and finally dissolved in 10 ml of anhydrous CH_2Cl_2. Freshly distilled DIPEA (2.5 equiv.) was added, followed by the addition of 2-cyanoethyl-*N*,*N*-diisopropylchlorophosphoramidite (1.2 equiv.). The reaction mixture was stirred at room temperature until a complete conversion was observed by TLC analysis (approx. 1 hour). The reaction mixture was diluted with anhydrous CH_2Cl_2 (50 ml), quickly washed under an argon atmosphere with saturated aqueous KCl (20 ml) and dried over Na_2SO_4. Purification by flash chromatography (Cy: EtOAc, with 0.5% Et_3N) under argon atmosphere provided each desired compounds as two pure diastereomers.

5'-*O*-[Bis(4-methoxyphenyl)phenylmethyl]-5-(2-nitrobenzyloxymethyl)-2'-deoxyuridine-3'-[2-cyanoethyl bis(1-methylethyl)phosphoramidite] (12).



The title compound **12** was obtained from the precursor **11** (500 mg, 0.72 mmol), DIPEA (313 µl, 1.8 mmol, 2.5 equiv.) and 2-cyanoethyl-*N,N*-diisopropylchlorophosphoramidite (192 µl, 0.86 mmol, 1.2 equiv.) following the general procedure B. The product was purified by flash chromatography (Cy: EtOAc 50: 50, Et_3N 0.5%) and obtained as a yellowish foam (476 mg, 74%).

^{1}H NMR (500.2 MHz, CD_3CN): 1.15, 1.17 (2 × d, 2 × 6H, J_{vic} = 6.8, (C**H$_3$**)$_2$CH); 2.34 – 2.46 (m, 2H, H-2′); 2.52 (t, 2H, J_{vic} = 6.0, OCH$_2$C**H$_2$**CN); 3.31 – 3.38 (m, 2H, H-5′); 3.55 – 3.78 (m, 11H, CH_3O-DMTr, $(CH_3)_2$C**H**, OC**H$_2$**CH_2CN, CH_a**H$_b$**O-5); 4.01 (d, J_{gem} = 11.4, C**H$_a$**H$_b$O-5); 4.11 (qd, 1H, $J_{4',3'} = J_{4',5'}$ = 3.5, $J_{H,P}$ = 1.0, H-4′); 4.59, 4.62 (2 × d, 2 × 1H, J_{gem} = 15.1, CH_2O-1″); 4.66 (dddd, 1H, $J_{H,P}$ = 10.5, $J_{3',2'}$ = 6.4, 4.5, $J_{3',4'}$ = 3.5, H-3′); 6.23 (t, 1H, $J_{1',2'}$ = 6.6, H-1′); 6.79 – 6.84 (m, 4H, H-*m*-C_6H_4-DMTr); 7.19 (m, 1H, H-*p*-C_6H_5-DMTr); 7.25 – 7.34 (m, 6H, H-*o*-C_6H_4-DMTr, H-*m*-C_6H_5-DMTr); 7.42 – 7.49 (m, 3H, H-4″, H-*o*-C_6H_5-DMTr); 7.58 (ddd, 1H, $J_{5'',6''}$ = 7.8, $J_{5'',4''}$ = 7.2, $J_{5'',3''}$ = 1.4, H-5″); 7.62 (ddt, 1H, $J_{6'',5''}$ = 7.8, $J_{6'',4''}$ = 1.5, $J_{5'',CH2}$ = 1.0, H-6″); 7.79 (s, 1H, H-6); 7.98 (dd, 1H, $J_{3'',4''}$ = 8.2, $J_{3'',5''}$ = 1.4, H-3″); 9.27 (bs, 1H, NH-3).

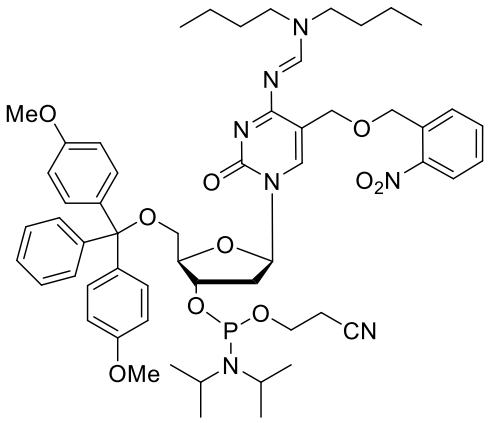
^{13}C NMR (125.8 MHz, CD_3CN): 20.97 (d, $J_{C,P}$ = 7.1, OCH$_2$**C**H$_2$CN); 24.88, 24.91 (2 × d, $J_{C,P}$ = 7.3, (**C**H$_3$)$_2$CH); 40.18 (d, $J_{C,P}$ = 4.4, CH_2-2′); 44.04 (d, $J_{C,P}$ = 12.4, $(CH_3)_2$**C**H); 55.89 (CH_3O-DMTr); 59.50 (d, $J_{C,P}$ = 19.1, O**C**H$_2$$CH_2CN$); 63.99 ($CH_2$-5′); 65.92 ($CH_2O$-5); 69.56 ($CH_2O$-1″); 73.78 (d, $J_{C,P}$ = 16.7, CH-3′); 85.75 (CH-1′); 86.16 (d, $J_{C,P}$ = 4.4, CH-4′); 87.52 (C-DMTr); 111.92 (C-5); 114.13 (CH-*m*-C_6H_4-DMTr); 119.40 (CN); 125.45 (CH-3″); 127.98 (CH-*p*-C_6H_5-DMTr); 128.95 (CH-*m*-C_6H_5-DMTr); 129.09 (CH-*o*-C_6H_5-DMTr); 129.16 (CH-4″); 129.75 (CH-6″);

131.11, 131.14 (CH-*o*-C_6H_4-DMTr); 134.64 (CH-5″); 135.58 (C-1″); 136.52, 136.75 (C-*i*-C_6H_4-DMTr); 140.43 (CH-6); 145.86 (C-*i*-C_6H_5-DMTr); 148.50 (C-2″); 151.22 (C-2); 159.76 (C-*p*-C_6H_4-DMTr); 163.72 (C-4).

$^{31}P\{^1H\}$ NMR (202.5 MHz, CD_3CN): 148.71.

HR MS (ESI+) m/z : $[M+H]^+$ Calcd for $C_{47}H_{55}O_{11}N_5P$ 896.3630; found 896.3632.

N^4-*N,N*-dibutylformimidamide-5'-*O*-[bis(4-methoxyphenyl)phenylmethyl]-5-(2-nitrobenzyl oxymethyl)-2'-deoxycytidine-3'-[2-cyanoethyl bis(1-methylethyl)phosphoramidite] (15).



The title compound **15** was obtained from the precursor **14** (300 mg, 0.36 mmol), DIPEA (157 µl, 0.9 mmol, 2.5 equiv.) and 2-cyanoethyl-*N,N*-diisopropylchlorophosphoramidite (96 µl, 0.43 mmol, 1.2 equiv.) following the general procedure B. The product was purified by flash chromatography (Cy: EtOAc 40:60, Et_3N 0.5%) and obtained as a yellowish foam (318 mg, 85%).

^1H NMR (500.0 MHz, CD_3CN): 0.80, 0.93 (2 × t, 2 × 3H, J_{vic} = 7.4, (C**H₃**$CH_2CH_2CH_2)_2N$); 1.15, 1.16 (2 × d, 2 × 6H, J_{vic} = 6.8, (C**H₃**$)_2CH$); 1.18 – 1.35 (m, 4H, (CH_3C**H₂**$CH_2CH_2)_2N$); 1.50 – 1.62 (m, 4H, (CH_3CH_2C**H₂**$CH_2)_2N$); 2.31 (dt, 1H, J_{gem} = 13.6, $J_{2'b,1'} = J_{2'b,3'}$ = 6.5, H-2′b); 2.49 (ddd, 1H, 1H, J_{gem} = 13.6, $J_{2'a,1'}$ = 6.5, $J_{2'b,3'}$ = 4.3, H-2′a); 2.50 – 2.54 (m, 2H, OCH_2C**H₂**CN); 3.33 (dd, 1H, J_{gem} = 10.7, $J_{5'b,4'}$ = 3.9, H-5′b); 3.37 (dd, 1H, J_{gem} = 10.7, $J_{5'a,4'}$ = 3.1, H-5′a); 3.37 – 3.41, 3.42 – 3.52 (2 × m, 2 × 2H, ($CH_3CH_2CH_2C$**H₂**$)_2N$); 3.55 – 3.69 (m, 4H, OC**H₂**CH_2CN, $(CH_3)_2C$**H**); 3.697, 3.699 (2 × s, 2 × 3H, CH_3O); 3.93 (d, 1H, J_{gem} = 11.0, U-CH_a**H**$_b$O); 4.14 (dddd, 1H, $J_{4',3'}$ = 4.3, $J_{4',5'}$ = 3.9, 3.1, $J_{H,P}$ = 1.0, H-4′); 4.27 (d, 1H, J_{gem} = 11.0, U-C**H**$_a$$H_b$O); 4.61 (m, 1H, H-3′); 4.62, 4.66 (2 × d, 2 × 1H, J_{gem} = 14.8, $C_6H_4NO_2$-C**H₂**O); 6.23 (t, 1H, $J_{1',2'}$ = 6.5, H-1′); 6.79 – 6.83 (m, 4H, H-

m-C_6H_4OMe-DMTr); 7.18 (m, 1H, H-*p*-C_6H_5-DMTr); 7.24 – 7.29 (m, 2H, H-*m*-C_6H_5-DMTr); 7.30 – 7.34 (m, 4H, H-*o*-C_6H_4OMe-DMTr); 7.42 – 7.46 (m, 3H, H-4-$C_6H_4NO_2$, H-*o*-C_6H_5-DMTr); 7.52 (m, 1H, H-5-$C_6H_4NO_2$); 7.57 (m, 1H, H-6-$C_6H_4NO_2$); 7.96 (s, 1H, H-6); 7.97 (dd, 1H, $J_{3,4}$ = 8.1, $J_{3,5}$ = 1.4, H-3-$C_6H_4NO_2$); 8.67 (s, 1H, CH=N).

^{13}C NMR (125.7 MHz, CD_3CN): 13.96, 14.03 (($\mathbf{C}H_3CH_2CH_2CH_2)_2N$); 20.35, 20.80 (($CH_3\mathbf{C}H_2CH_2CH_2)_2N$); 20.96 (d, $J_{C,P}$ = 7.2, $OCH_2\mathbf{C}H_2CN$); 24.89 (d, $J_{C,P}$ = 7.3, $(\mathbf{C}H_3)_2CH$); 29.77, 31.58 (($CH_3CH_2\mathbf{C}H_2CH_2)_2N$); 41.06 (d, $J_{C,P}$ = 4.3, CH_2-2′); 44.00 (d, $J_{C,P}$ = 2.4, $(CH_3)_2\mathbf{C}H$); 46.52, 52.89 (($CH_3CH_2CH_2\mathbf{C}H_2)_2N$); 55.86 ($CH_3O$); 59.47 (d, $J_{C,P}$ = 19.1, $O\mathbf{C}H_2CH_2CN$); 63.84 (CH_2-5′); 67.35 ($C_6H_4NO_2$-$\mathbf{C}H_2O$); 69.54 (U-$\mathbf{C}H_2O$); 73.64 (d, $J_{C,P}$ = 16.5, CH-3′); 86.16 (d, $J_{C,P}$ = 4.3, CH-4′); 86.94 (CH-1′); 87.44 (C-DMTr); 111.31 (C-5); 114.08 (CH-*m*-C_6H_4OMe-DMTr); 119.39 (CN); 125.37 (CH-3-$C_6H_4NO_2$); 127.89 (CH-*p*-C_6H_5-DMTr); 128.91 (CH-*m*-C_6H_5-DMTr); 129.01 (CH-4-$C_6H_4NO_2$); 129.09 (CH-*o*-C_6H_5-DMTr); 129.69 (CH-6-$C_6H_4NO_2$); 131.10, 131.13 (CH-*o*-C_6H_4OMe-DMTr); 134.54 (CH-5-$C_6H_4NO_2$); 135.98 (C-1-$C_6H_4NO_2$); 136.63, 136.82 (C-*i*-C_6H_4OMe-DMTr); 142.40 (CH-6); 145.91 (C-*i*-C_6H_5-DMTr); 148.42 (C-2-$C_6H_4NO_2$); 156.47 (C-2); 158.23 (CH=N); 159.69 (C-*p*-C_6H_4OMe-DMTr); 171.21 (C-4).

^{31}P{^{1}H} NMR (202.4 MHz, CD_3CN): 147.95.

HR MS (ESI+) m/z : $[M+H]^+$ Calcd for $C_{56}H_{73}O_{10}N_7P$ 1034.5151; found 1034.5148, $[M+Na]^+$ Calcd for $C_{56}H_{72}O_{10}N_7PNa$ 1056.4970; found 1056.4969.

5.2.3. Solid phase synthesis of modified oligonucleotides

Synthesis of oligonucleotides **ON2-16** were performed in 1 µM scale with the end trityl OFF mode. The amidites were diluted to a 0.1 M solution. Standard cycle procedures provided by Bioautomation were applied for the unmodified and modified phosphoramidites. The coupling volume and duration for the natural phosphoramidites were 220 µl and 1 minute 30 seconds whereas that for the modified phosphoramidites were increased to 300 µl and 6 minutes. The reaction products were deprotected by incubating them in 30% aqueous NH_3 at 55 °C for 6 hours. HPLC purification of the oligonucleotides was performed using a linear gradiant of ACN (0 to 30%) in 0.1 M TEAB buffer (pH 7.6) in 30 minutes.

5.2.4. Synthesis of specific site modified DNA

Synthesis of the unmodified 222DNA1 and modified DNA NB-222DNA2-16 templates.

To prepare the natural and modified templates for transcription, PCR reactions were performed in a total final volume of 30 µl. Natural DNAs (positive controls on agarose gels), which were used as controls for the DNA transcription experiments, were synthesized according to the procedure for modified DNA without any additives and in the presence of natural forward primer (P*veg*_F_222). All modified DNA templates were prepared by PCR in the presence of modified forward primers (**ON2-16**) and either non-labelled or FAM-labelled natural reverse primers (REV-PgII-R1/R3) from the P*veg* plasmid directly. One PCR reaction mixture (30 µl) contained Taq DNA polymerase for ThermoPol buffer (New England Biolabs; 5000 U/ml; 1.8 µl) with ThermoPol buffer (3 µl), natural dNTPs (4 mM; 1.125 µL), primers (20 µM; 4.5 µl; **ON2-16** and 20 µM; 4.5 µl; REV-PgII-R1/R3) and appropriate plasmid template (70 ng). The reaction was performed in a total volume 30 µl (3x10 µl). Forty PCR cycles were run in the thermal cycler under the following conditions: preheating for 3 minutes at 95 °C, denaturation for 1 minute at 94 °C, annealing for 1 minute at 52 °C, extension for 1.5 minutes at 75 °C, followed by final extension step of 5 minutes at 75 °C. 30 µl of PCR reaction mixture was combined into one eppendorf vial and the PCR products were purified using Agencourt AMPure XP magnetic particles. In the last step of purification, the product was eluted with 20 µl of Milli-Q water and the concentration was measured by Nanodrop. For non-labelled PCR products, approximately 100 ng of DNA was loaded on a control 1.3% agarose gel stained with GelRed (Biotium) and analyzed in 0.5xTBE buffer. The sequence of the templates were confirmed by Sanger DNA sequencing.

5.2.5. FAM Quantification of DNA templates

Quantification using radioactive labelling could not be used here due to the practical difficulties involved in distinguishing the template and the transcript which are of similar length. Since the modifications were only one or a few in the 5' end of the DNA template, we assumed that it cannot quench the FAM at the 3' end. Therefore, FAM labelling at the 3' end was chosen as the method to quantify the PCR products for transcription. To verify the concentrations of modified DNAs for the *in vitro* transcription assay, FAM-labelled DNA templates were prepared (using FAM-labelled reverse primer (FAM-REV-PgII-R1/R3)) in 30 µl reactions. The templates were then purified and finally diluted in 20 µl H_2O. 1 µl of the DNA (along with 4 µl H_2O and 1 µl 6X DNA loading dye) was loaded on a 1.3% agarose gel without Gelred. The relative intensities of the DNA bands in the gels were analyzed using the ImageJ software. The relative concentrations of the modified DNA templates were calculated using these values.

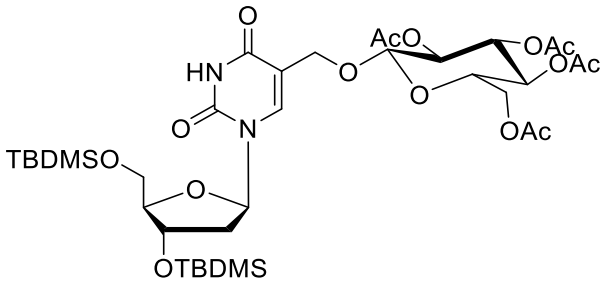
5.2.6. Deprotection of DNA by irradiation

The FAM-labelled DNA templates after purification and quantification was used for the deprotection experiments. Two batches of each sample were prepared to serve as the non-irradiated and irradiated samples. Approximately 240 ng (according to Nanodrop) of the natural DNA was taken and diluted to the final concentration of approximately 20 ng/µl either with H_2O (for non-irradiated samples) or with H_2O and additives (1 µl of 1 mM NaN_3 + 1 µl of 50 mM DTT, for irradiated samples). The modified DNA was also diluted in the same manner, based on quantification correction. Irradiation was performed on samples diluted in the presence of additives, with UV lamp 400 nm (3W) at room temperature. The natural and modified samples containing either $\mathbf{U^{NB}}$ or both $\mathbf{U^{NB}}$ and $\mathbf{C^{NB}}$ (unmodified 222DNA1 and **NB-222DNA-2,3,5,6,8,9,11,12,13,14,15,16**) were irradiated for 30 minutes and the samples containing only $\mathbf{C^{NB}}$ (**NB-222DNA-4,7,10**) for 10 minutes, following the optimized protocols. [132] We then used both the non-irradiated and irradiated DNA samples directly as templates for the *in vitro* transcription assay.

5.3. Synthesis of glucosylated 5-(hydroxymethyl)pyrimidine nucleoside triphosphates and phosphoramidites and their use in the synthesis of modified DNA

5.3.1. Synthesis of modified nucleosides

3', 5'-Bis-*O*-(*tert*-butyldimethylsilyl)-5-(2, 3, 4, 6-tetra-*O*-acetyl-β-glucopyranosyloxymeth yl) -2'-deoxyuridine (18)



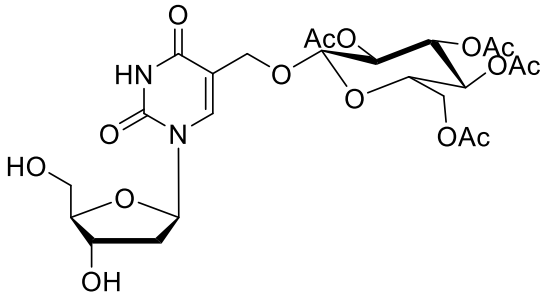
Method A:

The title compound **18** was prepared from **5** in three steps (**5** → **16** →**17**→**18**) following the reported procedures. [134] The spectral data were consistent with the literature. [133] Yield: 80 % in three steps.

Method B:

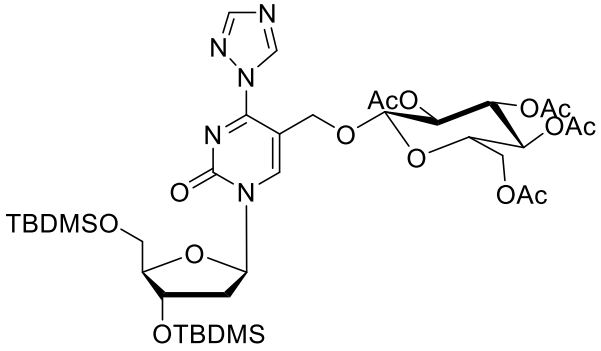
A mixture of **5** (500 mg, 1 mmol) and **33** (556 mg, 1.1 mmol, 1.1 equiv.) was dried by co-evaporation with dry DCE (2* 10 ml) and kept under vaccum overnight. The mixture was dissolved in dry DCE (10 ml) and stirred for 15 minutes with powdered molecular sieves (4 A°) (activated by heating overnight under vaccum) at room temperature. Then it is cooled down to -25 °C. TMSOTf (28 µl, 0.154 mmol, 0.15 equiv.) was added dropwise in 4 portions over 30 minutes and the reaction mixture was allowed to stir for 1.5 hour under an atmosphere of Argon. The mixture was quenched with Et_3N (1 ml) and allowed to come to room temperature. It is then filtered through celite. The filtrate was washed with H_2O and brine, dried over $MgSO_4$ and concentrated. Repeated purification by column chromatography (100% CH_2Cl_2 to CH_2Cl_2: MeOH 99:1) gave the desired compound. The spectral data were in accordance with literature. [133] Yield: 25 %.

5-(2, 3, 4, 6-Tetra-O-benzoyl-β-D-glucopyranosyloxymethyl)-2'-deoxyuridine (19, dU^{GlcAc})



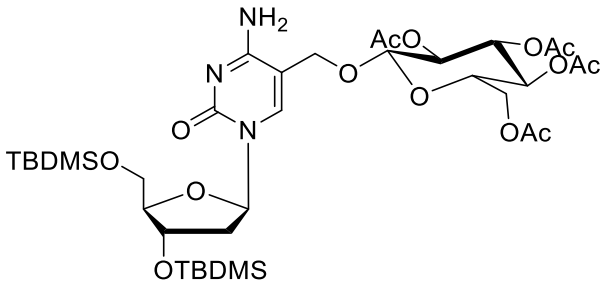
The title compound **19** was prepared from **18** following the reported procedures. [135] The spectral data were consistent with the literature. [133] Yield: 70%.

***N^4*-(*N^1*-1, 2, 4-triazolyl)-3', 5'-Bis(*O-tert*-butyldimethylsilyl)-5-(2, 3, 4, 6-tetra-*O*-acetyl-β-D-glucopyranosyloxymethyl)-2'-deoxyuridine (20)**



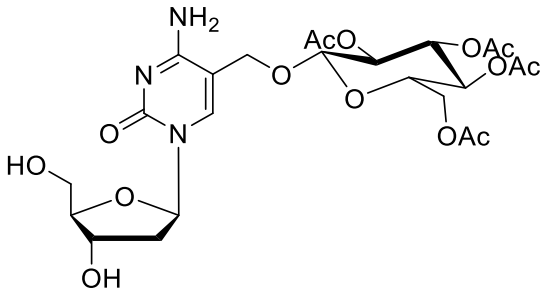
The title compound **20** was prepared from **18** following the reported procedures. [134] The spectral data were consistent with the literature. [135] Yield: 86 %.

3', 5'-bis(*O-tert*-butyldimethylsilyl)-5-(2, 3, 4, 6-tetra-*O*-acetyl-β-D-glucopyranosyloxy methyl)-2'-deoxycytidine (21)



The title compound **21** was prepared from **20** following the reported procedures. [134] The spectral data were consistent with the literature. [135] Yield: 80 %.

5-(2, 3, 4, 6-Tetra-O-benzoyl-β-D-glucopyranosyloxymethyl)-2'-deoxycytidine (22, dC^{GlcAc})



The title compound **22** was prepared from **21** following the reported procedures. [134] The spectral data were consistent with the literature. [135] Yield: 76%.

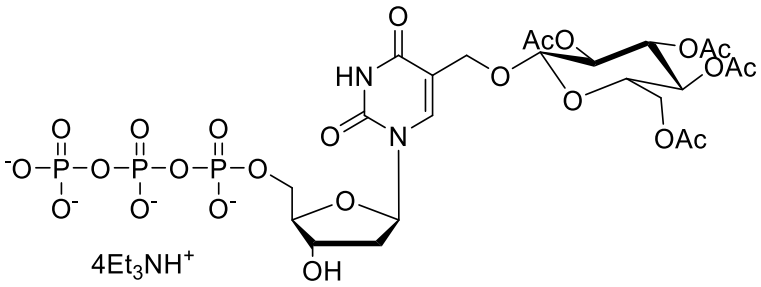
5.3.2. Synthesis of modified nucleoside triphosphates

General procedure C: Triphosphorylation reaction of modified nucleosides

Nucleosides (**dU^{GlcAc}** or **dC^{GlcOAc}**) along with proton sponge (1 equiv.) were dried under reduced pressure overnight, dissolved in dry $PO(OMe)_3$ (600 µl) and cooled at 0 °C. Freshly distilled $POCl_3$ (1 equiv.) was added dropwise and the reaction mixture was stirred at 0 °C for 2 hours. Then, an ice-cold solution of $(n\text{-}Bu_3NH)_2H_2P_2O_7$ (5 equiv.) and $n\text{-}Bu_3N$ (4.2 equiv.) in anhydrous ACN (500

µl) was added dropwise to the reaction mixture at 0 °C. The reaction mixture was stirred for 1 hour at 0 °C and then quenched by dropwise addition of cold 2M TEAB (1 ml). The mixture was concentrated on a rotary evaporator and the residue was co-evaporated several times with water. The crude product was dissolved in water (5 ml), filtered and purified by semi-preparative HPLC using a linear gradient of methanol (0 → 100%) in 0.1 M TEAB buffer. The appropriate fractions were combined and evaporated on a rotary evaporator. The viscous oil was co-evaporated three times with Milli-Q water and the product was freeze-dried either in the triethylammonium form itself or in the sodium form (by passing it through Na^+ dowex).

5-(2, 3, 4, 6-Tetra-*O*-acetyl-β-D-glucopyranosyloxymethyl)-2'-deoxyuridine-5'-*O*-triphosphate (23, dU^{GlcAc}TP)



dU^{GlcAc}TP (**23**) was obtained from the starting nucleoside **dU^{GlcAc}** (**19,** 30 mg, 0.051 mmol) following the general procedure C. RP HPLC, using a linear gradient of MeOH (0 → 100%) in 0.1 M TEAB buffer over 2 hours, afforded the desired compound in triethylammonium form as a white solid (12.5 mg, 20%).

^{1}H NMR (500.0 MHz, D_2O, ref (^{t}BuOH) = 1.24 ppm): 1.23 (t, 36H, J_{vic} = 7.3, CH_3CH_2N); 2.022 (s, 3H, CH_3CO-3-Glc); 2.024 (s, 3H, CH_3CO-2-Glc); 2.08 (s, 3H, CH_3CO-4-Glc); 2.11 (s, 3H, CH_3CO-6-Glc); 2.31 – 2.38 (m, 2H, H-2′); 3.09 (q, 24H, J_{vic} = 7.3, CH_3CH_2N); 4.07 (ddd, 1H, $J_{5,4}$ = 10.2, $J_{5,6}$ = 3.2, 2.3, H-5-Glc); 4.12 – 4.21 (m, 3H, H-4′,5′b, H-6b-Glc); 4.25 (ddd, 1H, J_{gem} = 11.3, $J_{H,P}$ = 6.3, $J_{5'a,4'}$ = 4.0, H-5′a); 4.40 (dd, 1H, J_{gem} = 12.7, $J_{6a,5}$ = 3.3, H-6a-Glc); 4.51, 4.60 (2 × d, 2 × 1H, J_{gem} = 12.3, CH_2O); 4.66 (dt, 1H, $J_{3',2'}$ = 5.3, 3.6, $J_{3',4'}$ = 3.6, H-3′); 4.87 (dd, 1H, $J_{2,3}$ = 9.4, $J_{2,1}$ = 8.1, H-2-Glc); 4.98 (d, 1H, $J_{1,2}$ = 8.1, H-1-Glc); 5.07 (dd, 1H, $J_{4,5}$ = 10.2, $J_{4,3}$ = 9.4, H-4-Glc); 5.35 (t, 1H, $J_{3,2}$ = $J_{3,4}$ = 9.4, H-3-Glc); 6.30 (t, 1H, $J_{1',2'}$ = 6.8, H-1′); 7.88 (s, 1H, H-6).

Integration of Et_3NH^+ corresponds to more number of ions, possibly from the TEAB from HPLC, which was not removed even after a couple of freeze dryings.

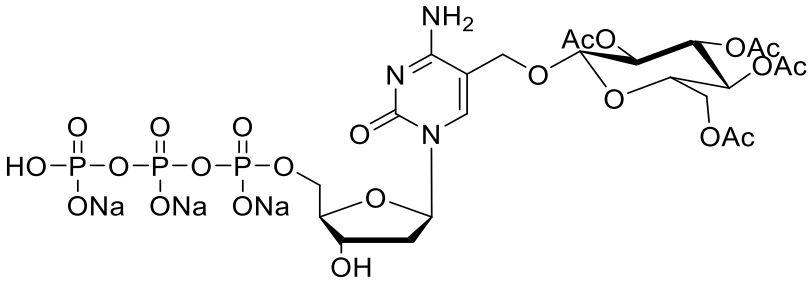
^{13}C NMR (125.7 MHz, D_2O, ref (^{t}BuOH) = 32.43 ppm): 11.20 (CH_3CH_2N); 22.88, 22.91, 22.97 (CH_3CO-2,3,4,6-Glc); 41.48 (CH-2′); 49.36 (CH_3CH_2N); 64.46 (CH_2-6-Glc); 67.95 (b, CH_2-5′); 69.44 (b, CH_2O); 70.94 (CH-4-Glc); 73.05 (CH-3′); 73.77 (CH-5-Glc); 74.46 (CH-2-Glc); 76.07 (CH-3-Glc); 87.97 (CH-1′); 88.14 (b, CH-4′); 102.82 (CH-1-Glc); 114.06 (C-5); 143.67 (b, CH-6); 158.10 (C-2); 172.50 (C-4); 175.55 (CH_3CO-2-Glc); 175.73 (CH_3CO-4-Glc); 175.91 (CH_3CO-3-Glc); 176.55 (CH_3CO-6-Glc).

^{31}P {^1H} NMR (202.4 MHz, D_2O): -21.95 (dd, J = 21.1, 20.1, P_β); -10.79 (d, J = 20.1, P_α); -5.72 (bd, J = 21.1, Pγ).

MS (ESI): m/z: 413.0 (18 %, M-2H/2^{2-}), 667.1 (24, M-$H_3P_2O_6^-$), 747.1 (100, M-$H_2PO_3^-$), 827.1 (1, M-H^-), 849.1 (11, M-2H+Na^-), 871 (10, M-3H+2Na^-).

HRMS (ESI): m/z: Calcd for $C_{24}H_{34}O_{24}N_2P_3$ [M-H]$^-$: 827.07198; found: 827.07124. Calcd for $C_{24}H_{33}O_{24}N_2P_3Na$ [M-2H+Na]$^-$: 849.05393; found: 849.05240. Calcd for $C_{24}H_{32}O_{24}N_2P_3Na_2$ [M-3H+2Na]$^-$: 871.03587; found: 871.03440.

5-(2, 3, 4, 6-Tetra-*O*-acetyl-β-D-glucopyranosyloxymethyl)-2'-deoxycytidine-5'-*O*-triphosphate (25, dC^{GlcAc}TP)



dC^{GlcAc}TP (**25**) was obtained from the starting nucleoside **dC^{GlcAc}** (**22,** 26 mg, 0.044 mmol) following the general procedure C. RP HPLC, using a linear gradient of methanol (0→100%) in 0.1 M TEAB buffer over 2 hours, followed by conversion to sodium form in Na^+ dowex, afforded the desired compound as a white solid (13.6 mg, 30%).

^1H NMR (500.0 MHz, D_2O, ref (^{t}BuOH) = 1.24 ppm): 2.03 (s, 3H, CH_3CO-3-Glc); 2.07 (s, 3H, CH_3CO-4-Glc); 2.08 (s, 3H, CH_3CO-2-Glc); 2.11 (s, 3H, CH_3CO-6-Glc); 2.34 (dt, 1H, J_{gem} = 13.9,

$J_{2'b,1'} = J_{2'b,3'} = 6.6$, H-2′b); 2.42 (ddd, 1H, $J_{gem} = 13.9$, $J_{2'a,1'} = 6.6$, $J_{2'a,3'} = 4.4$, H-2′a); 4.11 (ddd, 1H, $J_{5,4} = 10.0$, $J_{5,6} = 3.3$, 2.5, H-5-Glc); 4.17 – 4.29 (m, 4H, H-4′,5′, H-6b-Glc); 4.39 (dd, 1H, J_{gem} = 12.7, $J_{6a,5} = 3.3$, H-6a-Glc); 4.67 (dt, 1H, $J_{3',2'} = 6.6$, 4.4, $J_{3',4'} = 4.4$, H-3′); 4.68, 4.75 (2 × d, 2 × 1H, $J_{gem} = 13.1$, CH_2O); 4.92 – 4.99 (m, 2H, H-1,2-Glc); 5.09 (dd, 1H, $J_{4,5} = 10.0$, $J_{4,3} = 9.5$, H-4-Glc); 5.37 (m, 1H, H-3-Glc); 6.30 (t, 1H, $J_{1',2'} = 6.6$, H-1′); 8.03 (s, 1H, H-6).

^{13}C NMR (125.7 MHz, D_2O, ref (^{t}BuOH) = 32.43 ppm): 22.87, 22.91, 22.95, 23.00 (CH_3CO-2,3,4,6-Glc); 42.03 (CH-2′); 64.45 (CH_2-6-Glc); 67.79 (d, $J_{C,P} = 5.2$, CH_2-5′); 69.31 (CH_2O); 70.89 (CH-4-Glc); 72.88 (CH-3′); 73.78 (CH-5-Glc); 74.33 (CH-2-Glc); 76.00 (CH-3-Glc); 88.36 (d, $J_{C,P}$ = 9.0, CH-4′); 88.66 (CH-1′); 102.21 (CH-1-Glc); 107.03 (C-5); 144.71 (CH-6); 160.14 (C-2); 167.86 (C-4); 175.64 (CH_3CO-2-Glc); 175.74 (CH_3CO-4-Glc); 175.89 (CH_3CO-3-Glc); 176.54 (CH_3CO-6-Glc).

$^{31}P\{^1H\}$ NMR (202.4 MHz, D_2O): -21.32 (bt, $J = 19.4$, P_β); -10.56 (d, $J = 19.4$, P_α); -5.49 (bd, J = 19.4, Pγ).

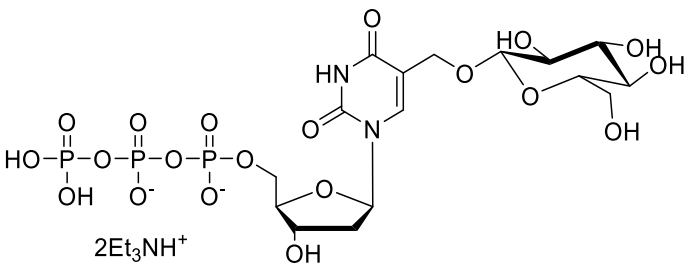
MS (ESI): m/z: 412.5 (79%, M-3H/2^{3-}) ; 666.2 [1, M-$H_3P_2O_6^-$) ; 746.1 (100, M-$H_2PO_3^-$), 826.1 (72, M-H^-),848.0(7,M-2H+Na^-).

HRMS (ESI): m/z: Calcd for $C_{24}H_{35}O_{23}N_3P_3$ $[M-H]^-$: 826.08797; found: 826.08788. Calcd for $C_{24}H_{33}O_{24}N_2P_3Na$ $[M-3H+Na]^-$: 849.05393; found: 849.05240. Calcd for $C_{24}H_{32}O_{24}N_2P_3Na_2$ $[M-4H+2Na]^-$: 871.03587; found: 871.03440.

General procedure D: Acetyl deprotection of the modified triphosphates

To the purified triphosphates **dU^{GlcAc}TP** and **dC^{GlcAc}TP**, 30% solution of NH_3 in H_2O (1 ml) was added and the resulting mixture was stirred at 0 °C for 1 hour, before evaporating near dryness. The crude product was dissolved in water (3 ml), filtered and purified by semi-preparative HPLC using a short linear gradient of methanol (0 → 100%) in 0.1 M TEAB buffer over 30 minutes. The appropriate fractions were combined and evaporated on a rotary evaporator. The viscous oil was co-evaporated three times with Milli-Q water and the product was freeze-dried.

5-(β-D-glucopyranosyloxymethyl)-2'-deoxyuridine-5'-*O*-triphosphate (24, dU^{Glc}TP)



dU^{Glc}TP (**24**) was obtained from the corresponding protected triphosphate **dU^{GlcAc}TP** (**23**, 5.5 mg, 0.004 mmol) following the general procedure D. RP HPLC, using a linear gradient of methanol (0→100%) in 0.1 M TEAB buffer over 30 minutes, afforded the desired compound in triethylammonium form as a white solid (2.7 mg, 78%).

^{1}H NMR (500.0 MHz, D_2O, ref(^{t}BuOH ext.) = 1.24 ppm): 1.28 (t, 18H, J_{vic} = 7.3, CH_3CH_2N); 2.38 (ddd, 1H, J_{gem} = 14.1, $J_{2'b,1'}$ = 6.5, $J_{2'b,3'}$ = 3.9, H-2′b); 2.43 (ddd, 1H, J_{gem} = 14.1, $J_{2'a,1'}$ = 7.4, $J_{2'a,3'}$ = 5.9 H-2′a); 3.20 (q, 12H, J_{vic} = 7.3, CH_3CH_2N); 3.25 (dd, 1H, $J_{2,3}$ = 9.4, $J_{2,1}$ = 8.0, H-2-Glc); 3.36 (dd, 1H, $J_{4,5}$ = 9.9, $J_{4,3}$ = 9.0, H-4-Glc); 3.44 (ddd, 1H, $J_{5,4}$ = 9.9, $J_{5,6}$ = 5.6, 2.3, H-5-Glc); 3.52 (dd, 1H, $J_{3,2}$ = 9.4, $J_{3,4}$ = 9.0, H-3-Glc); 3.70 (dd, 1H, J_{gem} = 12.4, $J_{6b,5}$ = 5.6, H-6b-Glc); 3.86 (dd, 1H, J_{gem} = 12.4, $J_{6a,5}$ = 2.3, H-6a-Glc); 4.16 – 4.27 (m, 3H, H-4′,5′); 4.56 (d, 1H, $J_{1,2}$ = 8.0, H-1-Glc); 4.60, 4.64 (2 × d, 2 × 1H, J_{gem} = 12.1, CH_2O); 4.67 (m, 1H, H-3′); 6.34 (dd, 1H, $J_{1',2'}$ = 7.4, 6.5, H-1′); 8.06 (s, 1H, H-6).

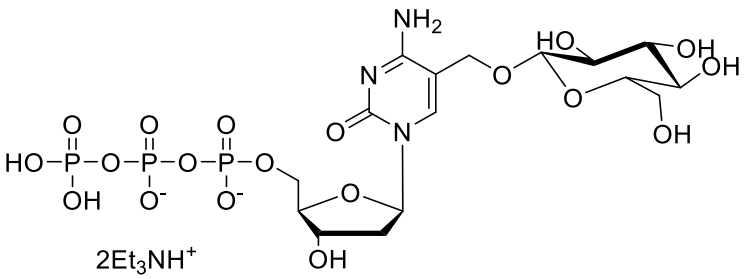
^{13}C NMR (125.7 MHz, D_2O, ref(^{t}BuOH ext.) = 32.43 ppm): 11.06 (CH_3CH_2N); 41.56 (CH-2′); 49.50 (CH_3CH_2N); 63.49 (CH_2-6-Glc); 67.58 (CH_2O); 68.23 (d, $J_{C,P}$ = 5.7, CH_2-5′); 72.47 (CH-4-Glc); 73.60 (CH-3′); 75.96 (CH-2-Glc); 78.57 (CH-3-Glc); 78.67 (CH-5-Glc); 88.17 (CH-1′); 88.43 (d, $J_{C,P}$ = 9.1, CH-4′); 104.70 (CH-1-Glc); 113.92 (C-5); 144.42 (CH-6); 154.45 (C-2); 168.03 (C-4).

^{31}P{^{1}H} NMR (202.4 MHz, D_2O): -22.60 (dd, J = 20.1, 19.1, P$_\beta$); -10.95 (d, J = 20.1, P$_\alpha$); -10.21 (d, J = 19.1, Pγ).

MS (ESI): m/z: 329.0 (100%, $(M+H)/2^{2-}$), 499.1 (8, $M-H_3P_2O_6^{3-}$), 579.1 (42, $M-H_2PO_3^-$), 659.0 (5, $M-H^-$), 681.0 (13, $M-2H+Na^-$).

HRMS (ESI): m/z: Calcd for $C_{16}H_{26}O_{20}N_2P_3$ $[M-H]^-$: 659.02972; found: 659.02881. Calcd for $C_{16}H_{25}O_{20}N_2NaP_3$ $[M-2H+Na]^-$: 681.01167; found: 681.01068.

5-(β-D-glucopyranosyloxymethyl)-2'-deoxycytidine-5'-O-triphosphate (26, dC^{Glc}TP)



dC^{Glc}TP (**26**) was obtained from the starting nucleoside **dC^{GlcAc}TP** (**25**, 4.6 mg, 0.005 mmol) following the general procedure D. RP HPLC, using a short linear gradient of methanol (0 →100%) in 0.1 M TEAB buffer over 30 minutes, afforded the desired compound in triethylammonium form as a white solid (3.5 mg, 80%).

^{1}H NMR (500.0 MHz, D_2O, ref(^{t}BuOH ext.) = 1.24 ppm): 1.28 (t, 18H, J_{vic} = 7.3, CH_3CH_2N); 2.34 (ddd, 1H, J_{gem} = 14.0, $J_{2'b,1'}$ = 7.0, $J_{2'b,3'}$ = 6.3, H-2′b); 2.43 (ddd, 1H, J_{gem} = 14.0, $J_{2'a,1'}$ = 6.3, $J_{2'a,3'}$ = 3.9 H-2′a); 3.20 (q, 12H, J_{vic} = 7.3, CH_3CH_2N); 3.28 (dd, 1H, $J_{2,3}$ = 9.4, $J_{2,1}$ = 8.0, H-2-Glc); 3.38 (dd, 1H, $J_{4,5}$ = 10.0, $J_{4,3}$ = 9.4, H-4-Glc); 3.48 (ddd, 1H, $J_{5,4}$ = 10.0, $J_{5,6}$ = 5.7, 2.3, H-5-Glc); 3.52 (t, 1H, $J_{3,2}$ = $J_{3,4}$ = 9.4, H-3-Glc); 3.72 (dd, 1H, J_{gem} = 12.4, $J_{6b,5}$ = 5.7, H-6b-Glc); 3.88 (dd, 1H, J_{gem} = 12.4, $J_{6a,5}$ = 2.3, H-6a-Glc); 4.18 – 4.24 (m, 3H, H-4′,5′); 4.58 (d, 1H, $J_{1,2}$ = 8.0, H-1-Glc); 4.64 (m, 1H, H-3′); 4.70, 4.78 (2 × d, 2 × 1H, J_{gem} = 12.6, CH_2O); 6.32 (dd, 1H, $J_{1',2'}$ = 7.0, 6.3, H-1′); 8.07 (s, 1H, H-6). ^{13}C NMR (125.7 MHz, D_2O, ref(^{t}BuOH ext.) = 32.43 ppm): 11.06 (CH_3CH_2N); 42.23 (CH-2′); 49.50 (CH_3CH_2N); 63.41 (CH_2-6-Glc); 68.07 (d, $J_{C,P}$ = 5.5, CH_2-5′); 68.54 (CH_2O); 72.43 (CH-4-Glc); 73.38 (CH-3′); 75.84 (CH-2-Glc); 78.57 (CH-3-Glc); 78.76 (CH-5-Glc); 88.38 (d, $J_{C,P}$ = 9.1, CH-4′); 88.80 (CH-1′); 104.19 (CH-1-Glc); 106.91 (C-5); 144.93 (CH-6); 159.81 (C-2); 167.71 (C-4).

^{31}P{^{1}H} NMR (202.4 MHz, D_2O): -22.54 (dd, J = 20.1, 19.3, P$_\beta$); -10.92 (d, J = 20.1, P$_\alpha$); -10.05 (d, J = 19.3, Pγ).

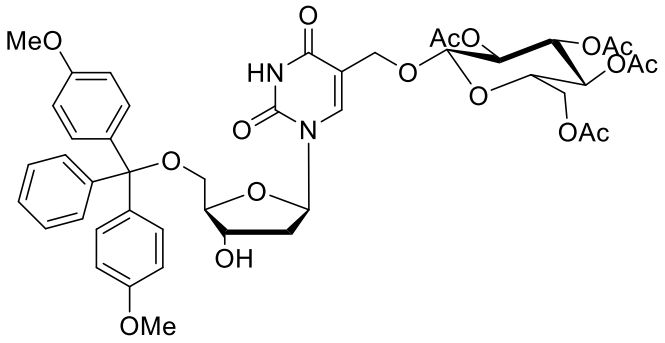
MS (ESI): m/z: 328.5 (46 %, M-2H/2^{2-}), 498.1 (15, M-$H_3P_2O_6^-$), 578.1 (100, M-$H_2PO_3^-$), 658.0 (15, M-H$^-$), 680.0 (5, M-2H+Na$^-$).

HRMS (ESI): m/z: Calcd for $C_{16}H_{27}O_{19}N_3P_3$ [M-H]$^-$: 658.04571; found: 658.04553. Calcd for $C_{16}H_{26}O_{19}N_3NaP_3$ [M-2H+Na]$^-$: 680.02765; found: 680.02728.

5.3.3. Synthesis of modified nucleoside phosphoramidites

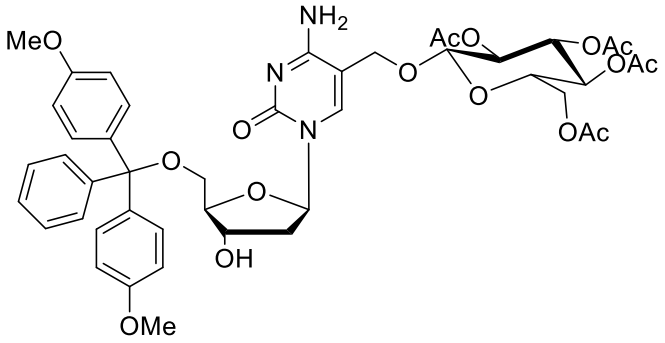
Synthesis of trityl-protected modified nucleosides

5'-*O*-[bis (4-methoxyphenyl)phenylmethyl]-5-(2, 3, 4, 6-tetra-*O*-benzoyl-β-D-glucopyranosyl oxymethyl)-2'-deoxyuridine (27)



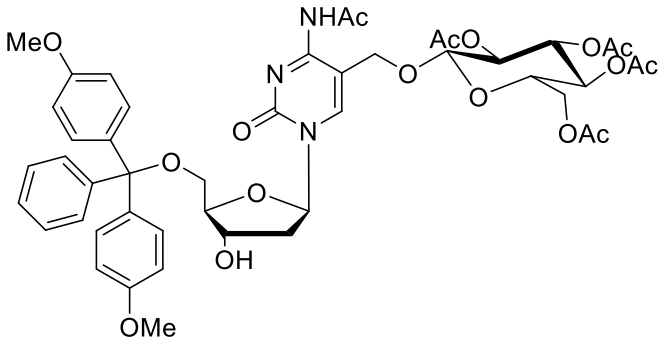
The title compound **27** was prepared from **19** following the reported procedures. The spectral data were consistent with the literature. [133] Yield: 60%.

5'-*O*-[bis (4-methoxyphenyl)phenylmethyl]-5-(2, 3, 4, 6-tetra-*O*-benzoyl-β-D-glucopyranosyl oxymethyl)-2'-deoxycytidine (29)



The title compound **29** was prepared from **22** following the reported procedures. The spectral data were consistent with the literature. [135] Yield: 60%.

***N*^{4}-acetyl-5'-*O*-[bis(4-methoxyphenyl)phenylmethyl]-5-(2, 3, 4, 6-tetra-*O*-benzoyl-β-D-gluco pyranosyl) oxymethyl-2'-deoxycytidine (30)**



The title compound **30** was obtained from **29** (216 mg, 0.24 mmol). To the precursor **29** dissolved in anhydrous DMF (2 ml) under inert conditions, acetic anhydride (27 µl, 0.29 mmol, 1.2 equiv.) was added and the reaction was stirred for 20 hours at room temperature. The solvent was evaporated under vacuum and the crude reaction mixture was re-dissolved in CH_2Cl_2 (50 ml), washed twice with saturated $NaHCO_3$ (50 ml) and dried over Na_2SO_4. The crude was then purified by flash chromatography (CH_2Cl_2: MeOH 99:1, Et_3N 0.5%) to obtain the pure product (145 mg, 64 %).

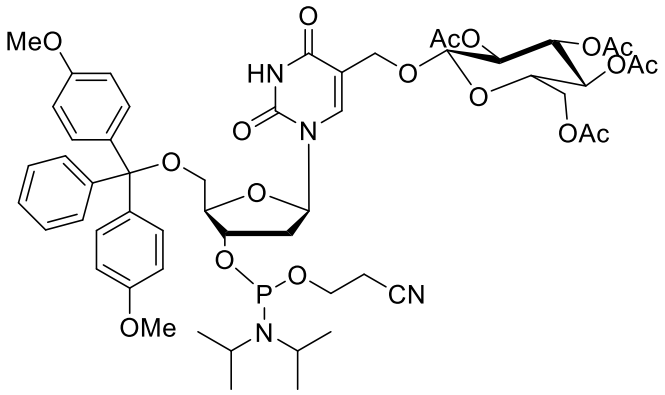
^{1}H NMR (500.0 MHz, CD_3CN): 1.90, 1.94, 1.95, 1.97 (s, 12 H, CH_3CO-2,3,4,6-Glc); 2.24 (dt, 1H, J_{gem} = 13.7, $J_{2'b,1'} = J_{2'b,3'}$ = 6.3, H-2′b); 2.41 (bs, 3H, CH3CON); 2.43 (ddd, 1H, J_{gem} = 13.7, $J_{2'a,1'}$ = 6.3, $J_{2'a,3'}$ = 4.9, H-2′a); 3.28 (dd, 1H, J_{gem} = 10.7, $J_{5'b,4'}$ = 3.4, H-5′b); 3.32 (dd, 1H, J_{gem} = 10.7, $J_{5'a,4'}$ = 4.2, H-5′a); 3.70 (ddd, 1H, $J_{5,4}$ = 10.0, $J_{5,6}$ = 4.5, 2.6, H-5-Glc); 3.78 (s, 6 H, CH_3O-DMTr); 3.90 (bd, 1H, J_{gem} = 12.4, H-6b-Glc); 3.99 (m, 1H, H-4′); 4.09 (bm, 1H, $CH_a\mathbf{H_b}O$); 4.16 (dd, 1H, J_{gem} = 12.4, $J_{6a,5}$ = 4.5, H-6a-Glc); 4.19 (bm, 1H, $CH_a\mathbf{H_b}O$); 4.41 (bm, 1H, H-3′); 4.60 (bd, 1H, $J_{1,2}$ = 8.1, H-1-Glc); 4.88 (dd, 1H, $J_{2,3}$ = 9.8, $J_{2,1}$ = 8.1, H-2-Glc); 5.03 (dd, 1H, $J_{4,5}$ = 10.0, $J_{4,3}$ = 9.4, H-4-Glc); 5.21 (dd, 1H, $J_{3,2}$ = 9.8, $J_{3,4}$ = 9.4, H-3-Glc); 6.10 (t, 1H, $J_{1',2'}$ = 6.3, H-1′); 6.85 – 6.90 (m, 4H, H-*m*-C_6H_4OMe-DMTr); 7.23 – 7.34 (m, 7H, H-*o*-C_6H_4OMe-DMTr, H-*m,p*-C_6H_5-DMTr); 7.41 – 7.44 (m, 2H, H-*o*-C_6H_5-DMTr); 7.94 (bs, 1H, H-6); 8.14 (bs, 1H, NH).

^{13}C NMR (125.7 MHz, CD_3CN): 20.84, 22.86, 20.88 (**C**H_3CO-2,3,4,6-Glc); 26.45 (**C**H_3CONH); 41.78 (CH-2′); 55.95 (CH_3O-DMTr); 62.57 (CH_2-6-Glc); 64.09 (CH_2-5′); 66.05 (b, CH_2O); 69.08

(CH-4-Glc); 71.45 (CH-3′); 71.85 (CH-2-Glc); 72.63 (CH-5-Glc); 73.29 (CH-3-Glc); 87.15 (CH-4′); 87.39 (C-DMTr); 87.57 (CH-1′); 99.70 (CH-1-Glc); 103.66 (C-5); 114.19 (CH-*m*-C_6H_4OMe-DMTr); 128.02 (CH-*p*-C_6H_5-DMTr); 128.97, 129.01 (CH-*o,m*-C_6H_5-DMTr); 131.05 (CH-*o*-C_6H_4OMe-DMTr); 136.55, 136.75 (C-*i*-C_6H_4OMe-DMTr); 144.30 (CH-6); 146.02 (C-*i*-C_6H_5-DMTr); 155.21 (C-2); 159.74, 159.75 (C-*p*-C_6H_4OMe-DMTr); 161.69 (C-4); 170.43, 170.48 (CH_3**C**O-2,4-Glc); 170.89 (CH_3**C**O-3-Glc); 171.29 (CH_3**C**O-6-Glc); 171.90 (CH_3**C**ONH).
MS (ESI+): m/z: 954.3 (100%, (M+Na), 932.3 (23%, M+H).
HR MS (ESI+) m/z: $[M+H]^+$ Calcd for $C_{47}H_{53}O_{17}N_3$ 931.34466; found 932.34477, $[M+Na]^+$ Calcd for $C_{47}H_{53}O_{17}N_3Na$ 954.32618; found 954.32677.

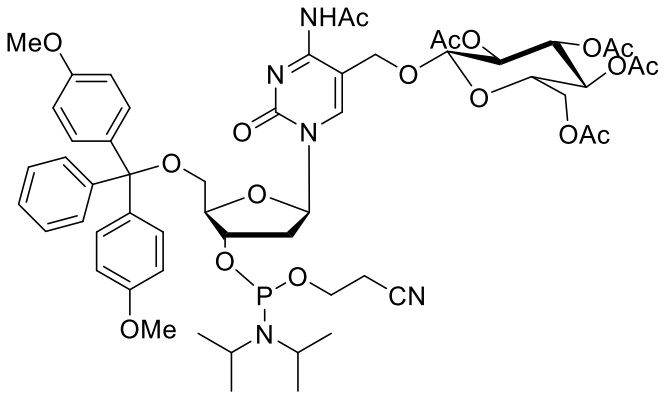
Synthesis of modified nucleoside phosphoramidites

5'-*O*-[Bis (4-methoxyphenyl)phenylmethyl]-5-(2, 3, 4, 6-tetra-*O*-benzoyl-β-D-glucopyranosyl)oxymethyl-2'-deoxyuridine-3'-[2-cyanoethyl bis(1-methylethyl)phosphoramidite] (28)



The title compound **28** was prepared from **27** following the reported procedures. The spectral data were consistent with the literature. [133] Yield: 89%.

N^4-Acetyl-5'-*O*-[bis(4-methoxyphenyl)phenylmethyl]-5-(2, 3, 4, 6-tetra-*O*-benzoyl-β-D-gluco pyranosyl)oxymethyl-2'-deoxycytidine-3'-[2-cyanoethylbis(1-methylethyl)phosphoramidite] (31)



The title compound **31** was obtained from the precursor **30** (133 mg, 0.14 mmol), DIPEA (62 µl, 0.36 mmol, 2.5 equiv.) and 2-cyanoethyl-*N,N*-diisopropylchlorophosphoramidite (38 µl, 0.17 mmol, 1.2 equiv.) following the general procedure B. The crude was purified by flash chromatography (EtOAc: Et_3N, 100: 0.5) to obtain the product as a yellowish foam as a 1:1 mixture of diastereomers (148 mg, 90%).

^{1}H NMR (500.0 MHz, CD_3CN): 1.05, 1.146, 1.149, 1.17 (4 × d, 4 × 3H, J_{vic} = 6.8, $(CH_3)_2CHN$); 1.93, 1.94, 1.945, 1.950, 1.953, 1.964, 1.966, 1.97 (8 × s, 8 × 3H, CH_3CO-2,3,4,6-Glc); 2.31 – 2.39 (m, 2H, H-2′b); 2.41 (bs, 6H, CH_3CON); 2.52 (t, 2H, J_{vic} = 6.0, $OCH_2C\mathbf{H_2}CN$); 2.54 – 2.63 (m, 2H, H-2′a); 2.64 (t, 2H, J_{vic} = 5.9, $OCH_2C\mathbf{H_2}CN$); 3.30 – 3.38 (m, 4H, H-5′); 3.53 – 3.90 (m, 24H, CH_3O-DMTr, $OC\mathbf{H_2}CH_2CN$, $(CH_3)_2C\mathbf{H}N$, H-5,6b-Glc); 4.07 – 4.29 (m, 8H, H-4′); CH_2OH-6a-Glc); 4.51 – 4.67 (m, 4H, H-3′, H-1-Glc); 4.84 – 4.90 (m, 2H, H-2-Glc); 5.02 (t, 2H, $J_{4,3} = J_{4,5}$ = 9.7, H-4-Glc); 5.20 (t, 2H, $J_{3,2} = J_{3,4}$ = 9.7, H-3-Glc); 6.10, 6.13 (2 × t, 2 × 1H, $J_{1',2'}$ = 6.3, H-1′); 6.85 – 6.90 (m, 8H, H-*m*-C_6H_4OMe-DMTr); 7.23 – 7.34 (m, 14H, H-*o*-C_6H_4OMe-DMTr, H-*m,p*-C_6H_5-DMTr); 7.41 – 7.44 (m, 4H, H-*o*-C_6H_5-DMTr); 7.93, 7.97 (2 × bs, 1H, H-6); 8.14 (bs, 2H, NH).

^{13}C NMR (125.7 MHz, CD_3CN): 20.83, 22.86, 20.88 ($\mathbf{C}H_3CO$-2,3,4,6-Glc); 20.97, 21.04 (2 × d, $J_{C,P}$ = 7.4, $OCH_2\mathbf{C}H_2CN$); 24.78, 24.83, 24.85, 24.87, 24.91, 24.92, 24.93 (($\mathbf{C}H_3)_2CHN$); 26.40 (b, $\mathbf{C}H_3CONH$); 40.80, 40.98 (CH-2′); 43.97, 44.06 (2 × d, $J_{C,P}$ = 2.6, $(CH_3)_2\mathbf{C}HN$); 55.93, 55.95

(CH_3O-DMTr); 59.31, 59.46 (2 × d, $J_{C,P}$ = 12.6, O**C**H_2CH_2CN); 62.52 (CH_2-6-Glc); 63.55, 63.80 (CH_2-5′); 65.99 (b, CH_2O); 69.05 (CH-4-Glc); 71.83 (CH-2-Glc); 72.62 (CH-5-Glc); 73.28 (CH-3-Glc); 73.90 (d, $J_{C,P}$ = 17.4, CH-3′); 86.17 (d, $J_{C,P}$ = 6.2, CH-4′); 86.43 (d, $J_{C,P}$ = 4.3, CH-4′); 87.44, 87.47 (C-DMTr); 87.60, 87.67 (b, CH-1′); 99.56 (b, CH-1-Glc); 114.19 (CH-*m*-C_6H_4OMe-DMTr); 119.42, 119.59 (OCH_2CH_2**C**N); 128.04, 128.07 (CH-*p*-C_6H_5-DMTr); 128.95, 128.99, 129.01, 129.02 (CH-*o,m*-C_6H_5-DMTr); 131.07, 131.08, 131.10, 131.12 (CH-*o*-C_6H_4OMe-DMTr); 136.45, 136.47, 136.57, 136.59 (C-*i*-C_6H_4OMe-DMTr); 144.25 (b, CH-6); 145.95, 145.96 (C-*i*-C_6H_5-DMTr); 155.20 (b, C-2); 159.77 (C-*p*-C_6H_4OMe-DMTr); 161.75 (b, C-4); 170.41, 170.46 (CH_3**C**O-2,4-Glc); 170.88 (CH_3**C**O-3-Glc); 171.26 (CH_3**C**O-6-Glc); 172.00 (CH_3**C**ONH); (C-5 not detected).

$^{31}P\{^1H\}$ NMR (202.4 MHz, CD_3CN): 148.03, 148.10.

MS (ESI+): m/z: 1132.5 (100%, M).

HR MS (ESI+) m/z: $[M+H]^+$ Calcd for $C_{56}H_{71}O_{18}N_5P$ 1132.45290; found 1132.45262, $[M+Na]^+$ Calcd for $C_{56}H_{70}O_{18}N_5PNa$ 1154.43437; found 1154.43457.

5.3.4. Enzymatic incorporation of modified triphosphates by PEX

5.3.4.1. Single incorporation of modified dN^{X}TP by PEX - Analytical scale

Single incorporation of modified dU^{X}TP by PEX using three DNA polymerases (KOD XL, Pwo, Vent (exo-)) – 19DNA_1U^X

Three reaction mixtures - for KOD XL polymerase, Vent (exo-) polymerase or Pwo polymerase - (20 µl each) contained appropriate DNA polymerase (KOD XL - Merc4biosciences, Novagen, 2.5 U/µl, 0.02 µl; Vent (exo-) - New England Biolabs, 2 U/µl, 0.05 µl; or Pwo - Peqlab, 1 U/µl, 0.08 µl), dGTP (1 mM, 0.4 µl), dTTP (1 mM, 0.5 µl) or **dU^{Glc}TP** (1 mM, 0.8 µl), 6-FAM-labelled primer Prim$^{PEX\,a}$ (10 µM, 1 µl), 19-mer template Temp^{19_1T} (10 µM, 1.5 µl) in appropriate reaction buffer (10x KOD XL reaction buffer/ 10x Thermopol reaction buffer for Vent(exo-)/ 10x Pwo reaction buffer, 2 µl) supplied by the manufacturer. Reaction mixtures were incubated for 40 minutes at 60 °C in a thermal cycler, subsequently denatured by addition of stop solution (20 µl) and heated 5 minutes at 95 °C. Reaction mixtures were analyzed using 12.5 % denaturing PAGE.

Single incorporation of modified dC^{X}TP by PEX using three DNA polymerases (KOD XL, Pwo, Vent (exo-)) – 19DNA_1C^{X}

Three reaction mixtures - for KOD XL polymerase, Vent (exo-) polymerase or Pwo polymerase - (20 µl each) contained appropriate DNA polymerase (KOD XL - Merc4biosciences, Novagen, 2.5 U/µl, 0.02 µl; Vent (exo-) - New England Biolabs, 2 U/µl, 0.05 µl; or Pwo - Peqlab, 1 U/µl, 0.08 µl), dGTP (1 mM, 0.6 µl), dCTP (1 mM, 0.6 µl) or **dC^{Glc}TP** (1 mM, 0.8 µl), 6-FAM-labelled primer Prim$^{PEX\,a}$ (3 µM, 1 µl), 19-mer template Temp^{19_1C} (3 µM, 1.5 µl) in appropriate reaction buffer (10x KOD XL reaction buffer/ 10x Thermopol reaction buffer for Vent(exo-)/ 10x Pwo reaction buffer, 2 µl) supplied by the manufacturer. Reaction mixtures were incubated either for 40 minutes at 60 °C in a thermal cycler. After reaction, samples were denatured by the addition of the stop solution (20 µl) followed by heating for 5 minutes at 95 °C. Reaction mixtures were analyzed using 12.5 % denaturing PAGE.

5.3.4.2. Multiple incorporation of modified dN^{X}TP by PEX - Analytical scale

Multiple incorporation of modified dU^{X}TP by PEX using three DNA Polymerases (KOD XL, Pwo, Vent (exo-)) – 31DNA_4U^{X}

Three reaction mixtures - for KOD XL polymerase, Vent (exo-) polymerase or Pwo polymerase - (20 µl each) contained appropriate DNA polymerase (KOD XL - Merc4biosciences, Novagen, 2.5 U/µl, 0.02 µl; Vent (exo-) - New England Biolabs, 2 U/µl, 0.05 µl; or Pwo - Peqlab, 1 U/µl, 0.08 µl), d(GAC)TP (1 mM, 1.5 µl), dTTP (1 mM, 1.5 µl) or **dU^{Glc}TP** (1 mM, 3 µl), 6-FAM-labelled primer Prim$^{PEX\,a}$ (3 µM, 1 µl), 31-mer template Temp^{31_4X} (3 µM, 1.5 µl) in appropriate reaction buffer (10x KOD XL reaction buffer/ 10x Thermopol reaction buffer for Vent(exo-)/10x Pwo reaction buffer, 2 µl) supplied by the manufacturer. In case of Pwo and Vent (exo-) polymerase, all the reaction mixtures were incubated for 40 minutes at 60 °C and in case of KODXL polymerase, reaction mixtures were incubated either for 40 minutes at 60 °C (dTTP) or 3 minutes at 72 °C (**dU^{Glc}TP**), in a thermal cycler. After reaction, samples were denatured by the addition of the stop solution (20 µl) followed by heating for 5 minutes at 95 °C. Reaction mixtures were analyzed using 12.5 % denaturing PAGE.

Multiple incorporation of modified dC^{X}TP by PEX using three DNA Polymerases (KOD XL, Pwo, Vent (exo-)) – 31DNA_4C^{X}

Three reaction mixtures - for KOD XL polymerase, Vent (exo-) polymerase or Pwo polymerase - (20 µl each) contained appropriate DNA polymerase (KOD XL - Merc4biosciences, Novagen, 2.5 U/µl, 0.02 µl; Vent (exo-) - New England Biolabs, 2 U/µl, 0.05 µl; or Pwo - Peqlab, 1 U/µl, 0.08 µl), d(ATG)TP (1 mM, 1.5 µl), dCTP (1 mM, 1.5 µl) or **dC^{Glc}TP** (1 mM, 3 µl), 6-FAM- labelled primer Prim$^{PEX\ a}$ (3 µM, 1 µl), 31-mer template Temp^{31_4X} (3 µM, 1.5 µl in appropriate reaction buffer (10x KOD XL reaction buffer/ 10x Thermopol reaction buffer for Vent(exo-)/ 10x Pwo reaction buffer, 2 µl) supplied by the manufacturer. In case of Pwo and Vent (exo-) polymerase, reaction mixtures were incubated for 40 minutes at 60 °C and in case of KOD XL polymerase, reaction mixtures were incubated either for 40 minutes at 60 °C (dTTP) or 3 minutes at 72 °C (**dU^{Glc}TP**), in a thermal cycler. After reaction, samples were denatured by the addition of the stop solution (20 µl) followed by heating for 5 minutes at 95 °C. Reaction mixtures were analyzed using 12.5 % denaturing PAGE.

5.3.5. Enzymatic glucosylation of C^{hm}-modified PEX DNA by glucosyl transferase

5.3.5.1. PEX Synthesis of C^{hm}-modified DNA- Semi-preparative scale

Single incorporation of modified dC^{hm}TP by PEX - 19DNA_1C^{hm}

The reaction mixture (50 µl) contained KOD XL DNA polymerase (2.5 U/µl, 0.1 µl), dGTP (4 mM, 2.6 µl), **dC^{hm}TP** (4 mM, 2.6 µl), primer Prim^{PEX} (100 µM, 1.6 µl), 19-mer template Temp$^{19_1C\ b}$ (100 µM, 1.6 µl) and 10x buffer for KOD XL DNA polymerase (5 µl) supplied by the manufacturer. Reaction mixtures were incubated for 40 minutes at 60 °C in a thermal cycler. After the reaction, it was purified by Qiagen nucleotide removal kit and eluted in a final of 30 µl water.

Multiple incorporation of modified dC^{hm}TP by PEX - 31DNA_4C^{hm}

The reaction mixture (50 µl) contained KOD XL DNA polymerase (2.5 U/µl, 0.1 µl), d(GAT)TP (4 mM, 2.6 µl), **dC^{hm}TP** (4 mM, 2.6 µl), primer Prim^{PEX} (100 µM, 1.6 µl), 31-mer template Temp$^{31_4C\ b}$ (100 µM, 1.6 µl) and 10x buffer for KOD XL DNA polymerase (5 µl) supplied by the manufacturer. Reaction mixtures were incubated for 40 minutes at 60 °C in a thermal cycler. After the reaction, it was purified by Qiagen nucleotide removal kit and eluted in a final of 30 µl water.

5.3.5.2. Enzymatic glucosylation of C^{hm}-modified PEX DNA by glucosyl transferase

To ~ 500 ng of purified PEX Product (**19DNA_1C^{hm}** and **31DNA_4C^{hm}**) in 10x Epi buffer (5 µl, 1x final concentration) 10x UDP-glc (5 µl, 1x final concentration) and T4-βGT (5U/ µl, 1 µl) was added and the solution was diluted to 50 µl by water. It was incubated at 37 ℃ for 1 hour. Products were then purified using the DB strepatividin magnetoseparation procedure to get **19ON_1C^{Glc*}** and **31ON_4C^{Glc*}** which were then submitted for MALDI-TOF analysis.

5.3.6. Synthesis of modified single-stranded oligonucleotides

5.3.6.1. PEX synthesis of DNA – Semi-preparative scale

The PEX reaction mixture (KODXL; 20 µl) is scaled up to 5 times (5 x 20 µl = 100 µl) following the mentioned protocols (5.3.4) using the corresponding biotinylated templates (Temp$^{19_1T\ b}$ *or* Temp$^{19_1C\ b}$ *or* Temp$^{31_4T\ b}$ *or* Temp$^{31_4c\ b}$) and non-labelled primer Prim^{PEX} and rest of the reagents same as that used in the analytical scale. Products were purified using the DB streptavidin magnetoseparation procedure to get **19ON_1N^{X}** and **31ON_4N^{X}** which were then submitted for MALDI-TOF analysis.

5.3.6.2. DB streptavidin magnetoseparation procedure

Streptavidin particles (Roche; 100 µl) were washed with binding buffer (3 × 200 µl; 10 mM Tris, 1 mM EDTA, 100 mM NaCl, pH 7.5). The PEX solution (100 µl; 5 x 20 µl MM) was mixed with binding buffer (200 µl) and incubated for 30 minutes at 15 °C in 1400 rpm. The magnetic beads were collected on a magnet (DynaMagTM-2, Invitrogen), and washed with wash buffer (3 × 400 µl; 10 mM Tris, 1 mM EDTA, 500 mM NaCl, pH 7.5) and water (4 × 400 µl). Then water (50 µl) was added and the sample was denatured for 2 minutes at 900 rpm and 72 °C. The beads were collected on a magnet and the solution was transferred into a clean vial. The product was evaporated to dryness, then dissolved in water and analyzed by MALDI-TOF mass spectrometry.

5.3.7. Enzymatic incorporation of modified triphosphates by PCR

5.3.7.1. Synthesis of 98 mer DNA modified templates

PCR reactions were performed in a total final volume of 20 µl. Natural DNAs (positive controls on gels) were in all cases synthesized according to the reported procedures for modified DNA without any additives and in the presence of natural dNTPs. Three reaction mixtures - for KOD XL polymerase, Vent (exo-) polymerase or Pwo polymerase - (20 µl each) contained appropriate DNA polymerase (KOD XL - Merc4biosciences, Novagen, 2.5 U/µl, 1 µl; Vent (exo-) - New England Biolabs, 2 U/µl, 2 µl; or Pwo - Peqlab, 1 U/µl, 2 µl), natural (d(GACT) TP) or functionalized dNTPs (d(GAC)TP along with **dU^{Glc}TP** or d(GAT)TP along with **dC^{Glc}TP**; 4 mM, 0.5 µl), primers Prim^{F_98} (10µM, 4 µl) and Prim^{R_98} (10 µM, 4 µl) and 98- mer template Temp^{98DNA} (1µM, 0.5 µl) in appropriate reaction buffer (KOD XL reaction buffer/ Thermopol reaction buffer for Vent(exo-)/ Pwo reaction buffer, 2 µl) supplied by the manufacturer. 30 PCR cycles were run in PCR cycler, preheated to 80 °C, under the following conditions: preheating for 3 minutes at 94 °C, denaturation for 1 minute at 95 °C, annealing for 1 minute at 53 °C, extension for 1 minute at 72 °C, followed by the final elongation step of 2 minutes at 75 °C. The PCR products were purified by Agencourt AmPure XP magnetic particles and finally diluted in 20 µl H_2O. PCR products were analyzed on a 2% agarose gel stained with GelRed (Biotium, 10 000X in H_2O) in 0.5x TBE buffer.

5.3.7.2. Synthesis of 235 mer P*veg* DNA template

Synthesis of 235 mer P*veg* DNA template for transcription studies using KOD XL DNA polymerase:

Natural 235 mer P*veg* DNA template was synthesized from the plasmid following the reported procedure.[61] To prepare modified templates for transcription, PCR reactions were performed in a total final volume of 20 µl. Natural DNAs (positive controls on gels) were in all cases synthesized according to the reported procedures for **U^{hm}**-modified or **C^{hm}**-modified DNA without any additives and in the presence of natural dNTPs. All samples were prepared in the presence of either non-labelled (Prim^{F_235} and Prim^{R_235}) or FAM-labelled (Prim$^{F_235\,a}$ and Prim$^{R_235\,a}$) primers under slightly different PCR conditions. PCR products of each type were then purified by Agencourt AMPure XP magnetic particles. The concentration of purified samples was measured by the nano drop spectrophotometer. Control 1.3% agarose gels were stained with GelRed (Biotium) in the case of non-labelled templates and without any staining for FAM-labelled templates.

PCR synthesis of dU^{hm}-modified template (235DNA_U^{hm})

The PCR reaction mixture (20 µl) contained KOD XL (2.5 U/µl, 1.2 µl), d(GAC)TP (4 mM, 0.75 µl), **dU^{hm}TP** (4 mM, 1.5 µl), either FAM-labelled (Prim$^{F_235\,a}$) or non-labelled (Prim^{F_235}) forward primers (20 µM, 3 µl), either FAM-labelled (Prim$^{R_235\,a}$) or non-labelled (Prim^{R_235}) reverse primers (20 µM, 3 µl)) and 235 mer template Temp^{235DNA} (40 ng) in KOD XL reaction buffer (2 µl) supplied by the manufacturer. 40 PCR cycles were run in PCR cycler under the following conditions: preheating for 3 minutes at 94 °C, denaturation for 1 minute at 94 °C, annealing for 1 minute at 52 °C, extension for 1.5 minutes at 75 °C, followed by the final elongation step of 2 minutes at 75 °C. The PCR products were purified by Agencourt AmPure XP magnetic particles and finally diluted in 20 µl H_2O. PCR products were analyzed on a 1.3% agarose gel stained with or without GelRed (Biotium, 10 000X in H_2O) in 0.5x TBE buffer.

PCR synthesis of dC^{hm}-modified template (235DNA_C^{hm})

The PCR reaction mixture (20 µl) contained KOD XL (2.5 U/µl, 1.2 µl), d(ATG)TP (4 mM, 0.75 µl), **dC^{hm}TP** (4 mM, 1.5 µl), either FAM-labelled (Prim$^{F_235\,a}$) or non-labelled (Prim^{F_235}) forward primers (20 µM, 3 µl), either FAM-labelled (Prim$^{R_235\,a}$) or non-labelled (Prim^{R_235}) reverse primers

(20 µM, 3 µl)) and 235- mer template $Temp^{235DNA}$ (40 ng) in KOD XL reaction buffer (2 µl) supplied by the manufacturer. 40 PCR cycles were run in PCR cycler under the following conditions: preheating for 3 minutes at 94 °C, denaturation for 1 minute at 94 °C, annealing for 1 minute at 52 °C, extension for 1.5 minutes at 75 °C, followed by the final elongation step of 2 minutes at 75 °C. The PCR products were purified by Agencourt AmPure XP magnetic particles and finally diluted in 20 µl H_2O. PCR products were analyzed on a 1.3% agarose gel stained with or without GelRed (Biotium, 10 000X in H_2O) in 0.5x TBE buffer.

PCR synthesis of dU^{Glc}-modified template (235DNA_U^{Glc})

The PCR reaction mixture (20 µl) contained KOD XL (2.5 U/µl, 2 µl), d(GAC)TP (4 mM, 0.5 µl), **dU^{Glc}TP** (4 mM, 2 µl), either FAM-labelled ($Prim^{F\_235\ a}$) or non-labelled ($Prim^{F\_235}$) forward primers (20 µM, 3 µl), either FAM-labelled ($Prim^{R\_235\ a}$) or non-labelled ($Prim^{R\_235}$) reverse primers (20 µM, 3 µl)) and 235- mer template $Temp^{235DNA}$ (40 ng) in KOD XL reaction buffer (2 µl) supplied by the manufacturer. 40 PCR cycles were run in PCR cycler under the following conditions: preheating for 3 minutes at 94 °C, denaturation for 1 minutes at 94 °C, annealing for 1 minute at 52 °C, extension for 1 minute at 75 °C, followed by the final elongation step of 2 minutes at 75 °C. The PCR products were purified by Agencourt AmPure XP magnetic particles and finally diluted in 20 µl H_2O. PCR products were analyzed on a 1.3% agarose gel stained with or without GelRed (Biotium, 10 000X in H_2O) in 0.5x TBE buffer.

PCR synthesis of dC^{Glc}-modified template (235DNA_C^{Glc})

The PCR reaction mixture (20 µl) contained KOD XL (2.5 U/µl, 3 µl), d(ATG)TP (4 mM, 0.5 µl), **dC^{Glc}TP** (4 mM, 2 µl), either FAM-labelled ($Prim^{F\_235\ a}$) or non-labelled ($Prim^{F\_235}$) forward primers (20 µM, 3 µl), either FAM-labelled ($Prim^{R\_235\ a}$) or non-labelled ($Prim^{R\_235}$) reverse primers (20 µM, 3 µl)) and 235- mer template $Temp^{235DNA}$ (40 ng) in KOD XL reaction buffer (2 µl) supplied by the manufacturer. 40 PCR cycles were run in PCR cycler under the following conditions: preheating for 3 minutes at 94 °C, denaturation for 1 minute at 94 °C, annealing for 1 minute at 52 °C, extension for 1 minute at 75 °C, followed by the final elongation step of 2 minute at 75 °C. The PCR products were purified by Agencourt AmPure XP magnetic particles and finally

diluted in 20 µl H_2O. PCR products were analyzed on a 1.3% agarose gel stained with or without GelRed (Biotium, 10 000X in H_2O) in 0.5x TBE buffer.

Synthesis of 235 mer P*veg* DNA templates using Pwo and Vent (exo-) polymerases.

To study the incorporation of the modifications in PCR by other DNA Polymerases, optimized PCR conditions were carried out using Pwo, Vent (exo-) DNA polymerases in a total final volume of 20 µl. All samples were prepared in the presence of non-labelled ($Prim^{F\_235}$ and $Prim^{R\_235}$) primers. PCR products of each type were then purified by Agencourt AMPure XP magnetic particles. The concentration of purified samples was measured by the nano drop spectrophotometer. Control 1.3% agarose gels were stained with GelRed (Biotium).

Synthesis of dU^{Glc}-modified template

The PCR reaction mixture (20 µl) contained the corresponding polymerase (Vent (exo-) polymerase; New England Biolabs, 2 U/µl, 2 µl or Pwo polymerase; Peqlab, 1 U/µl, 3 µl), d(GAC)TP (4 mM, 0.5 µl), **dU^{Glc}TP** (4 mM, 2 µl), non-labelled ($Prim^{F\_235}$) forward primers (20 µM, 3 µl), non-labelled ($Prim^{R\_235}$) reverse primers (20 µM, 3 µl)) and 235-mer template $Temp^{235DNA}$ (40 ng) in KOD XL reaction buffer (2 µl) supplied by the manufacturer. 40 PCR cycles were run in PCR cycler under the following conditions: preheating for 3 minutes at 94 °C, denaturation for 1 minute at 94 °C, annealing for 1 minute at 52 °C, extension for 1 minute at 75 °C, followed by the final elongation step of 2 minutes at 75 °C. The PCR products were purified by Agencourt AmPure XP magnetic particles and finally diluted in 20 µl H_2O. PCR products were analyzed on a 1.3% agarose gel stained with GelRed (Biotium, 10 000X in H_2O) in 0.5x TBE buffer.

Synthesis of dC^{Glc}-modified template

The PCR reaction mixture (20 µl) contained the corresponding polymerase (Vent (exo-) polymerase; New England Biolabs, 2 U/µl, 3 µl or Pwo polymerase; Peqlab, 1 U/µl, 3 µl), d(ATG)TP (4 mM, 0.5 µl), **dC^{Glc}TP** (4 mM, 2 µl), non-labelled ($Prim^{F\_235}$) forward primers (20 µM, 3 µl), non-labelled ($Prim^{R\_235}$) reverse primers (20 µM, 3 µl)) and 235- mer template

$Temp^{235DNA}$ (40 ng) in KOD XL reaction buffer (2 µl) supplied by the manufacturer. 40 PCR cycles were run in PCR cycler under the following conditions: preheating for 3 minutes at 94 °C, denaturation for 1 minute at 94 °C, annealing for 1 minute at 52 °C, extension for 1 minute at 75 °C, followed by the final elongation step of 2 minutes at 75 °C. The PCR products were purified by Agencourt AmPure XP magnetic particles and finally diluted in 20 µl H_2O. PCR products were analyzed on a 1.3% agarose gel stained with GelRed (Biotium, 10 000X in H_2O) in 0.5x TBE buffer.

5.3.8. Enzymatic glucosylation of C^{hm}-modified PCR DNA by glucosyl transferase

Enzymatic glucosylation of 235DNA_C^{hm} to form 235DNA_C^{Glc}

To ~ 1000 ng of natural (235DNA_C^{+}) and **C^{hm}**-modified (**235DNA_C^{hm}**) PCR DNA ((non-labelled (for enzymatic digestion studies) or FAM-labelled (for *in vitro* transcription studies)) in 10x Epi buffer (5 µl, 1x final concentration), 10x UDP-glc (5 µl, 1x final concentration) and T4-β-GT (5U/µl, 1 µl) was added and the solution was diluted to 50 µl by water. It was incubated at 37^0C for 1 hour. Products were purified using Agencourt AmPure XP magnetic particles and eluted by a final volume of 10 µl. Control 1.3% agarose gels were stained with GelRed (Biotium) in the case of non-labelled templates and without any staining for FAM-labelled templates.

Enzymatic digestion of DNA

In order to confirm the full conversion of glucosylation reaction on 235DNA by glucosyl transferase, we used enzymatic digestion of modified DNA (**235DNA_C^{hm}** and **235DNA_C^{Glc*}**; non-labelled) and LC-MS analysis. The purified (non-labelled) **235DNA_C^{hm}** and **235DNA_C^{Glc*}** DNA (~ 1 µg) was fully digested by Nucleoside digestion mix (NEB) for 1 hour at 37 °C. Digested DNA samples were diluted to 200 µl and purified over Microcon-10kDa centrifugal filters (Merck). The flow through was concentrated on a Speedvac system to the volume of 20 µl for LC-MS analysis.

5.3.9. FAM quantification of DNA templates

To verify the concentrations of modified DNAs for the *in vitro* transcriptions, FAM-labelled DNA templates were used in all cases. 1 µl of the purified DNA (along with 4 µl H_2O and 1 µl 6X DNA loading dye) was loaded on a 1.3% agarose gel without Gelred. The relative intensities of the fluorescence from the DNA bands in the gels were analyzed using the ImageJ software. The relative concentrations of the modified DNA templates were calculated using these values. These quantified DNA templates were diluted to a final concentration of 5 ng/µl and used straight away for the transcription studies.

5.3.10. Study of cleavage of modified DNA with REs

Restriction endonuclease cleavage studies were done on DNA synthesized by PEX (with templates containing specific sequences corresponding to restriction endonucleases) using KOD XL DNA polymerase in analytical scale. The PEX conditions for each template are described below. For each sample, PEX reaction was of 40 µl which is divided into two portions of 20 µl each. First portion is the PEX reaction control which was directly used for gel analysis. Second portion was used for the restriction endonuclease cleavage studies.

5.3.10.1. Study of cleavage by restriction endonuclease for $dU^{X}TPs$

Incorporation of modified $dU^{X}TPs$ within the recognition sequence of AflII restriction endonuclease

The reaction mixture (40 µl) contained KOD XL DNA polymerase (2.5 U/µl, 0.05 µl), d(GAC)TP (1 mM, 1 µl), dTTP (1 mM, 2 µl) or modified dUTP (1 mM, 2 µl), 6-FAM-labelled primer $Prim^{PEX\ a}$ (3 µM, 2 µl), 30-mer template $Temp^{AflT}$ (3 µM, 3 µl) and 10x buffer for KOD XL DNA polymerase (4 µl) supplied by the manufacturer. Reaction mixtures were incubated either for 40 minutes at 60 °C (for dTTP and **$dU^{hm}TP$**) or 3 minutes at 72 °C (**$dU^{Glc}TP$**) in a thermal cycler. The reaction mixtures were then divided into two portions. The first portion (20 µl) was directly used for PAGE analysis and the second portion (20 µl) was employed for DNA cleavage studies. Restriction endonuclease AflII (20U/µl, 1.6 L), BSA (100 mg/ml, 0.2 µl), and CutSmart 10x buffer

(2 µl) were added to the second portion of the PEX reaction. The mixture was then incubated for 1 hour at 37 °C. Samples were denatured after the reaction by adding the stop solution (1:1 to the reaction volume) and heating for 5 minutes at 95 °C. 12.5 % denaturing PAGE was used to analyze the reaction mixtures.

Incorporation of modified dU^{X}TPs within the recognition sequence of PvuII restriction endonuclease

The reaction mixture (40 µl) contained KOD XL DNA polymerase (2.5 U/µl, 0.05 µl), d(GAC)TP (1 mM, 1 µl), dTTP (1 mM, 2 µl) or modified dUTP (1 mM, 2 µl), 6-FAM-labelled primer Prim^{PEX} ^{b} (3 µM, 2 µl), 30-mer template Temp^{PvT} (3 µM, 3 µl) and 10x buffer for KOD XL DNA polymerase (4 µl) supplied by the manufacturer. Reaction mixtures were incubated either for 40 minutes at 60 °C (for dTTP and **dU^{hm}TP**) or 3 minutes at 72^{0}C (**dU^{Glc}TP**) in a thermal cycler. The reaction mixtures were then divided into two portions. The first portion (20 µl) was directly used for PAGE analysis and the second portion (20 µl) was used for DNA cleavage studies. Restriction endonuclease PvuII (10U/µl, 1.6 µl) and NE buffer 3.1 (2 µl) were added to the second portion of the PEX reaction. The mixture was then incubated for 1 hour at 37 °C. Samples were denatured after the reaction by adding the stop solution (1:1 to the reaction volume) and heating for 5 minutes at 95 °C. 12.5 % denaturing PAGE was used to analyze the reaction mixtures.

Incorporation of modified dU^{X}TPs within the recognition sequence of RsaI restriction endonuclease

The reaction mixture (40 µl) contained KOD XL DNA polymerase (2.5 U/µL, 0.05 µl), d(GAC)TP (1 mM, 1 µl), dTTP (1 mM, 2 µl) or modified dUTP (1 mM, 2 µl), 6-FAM-labelled primer Prim^{PEX} ^{b} (3 µM, 2 µl), 30-mer template Temp^{RsT} (3 µM, 3 µl) and 10x buffer for KOD XL DNA polymerase (4 µl) supplied by the manufacturer. Reaction mixtures were incubated either for 40 minutes at 60 °C (for dTTP and **dU^{hm}TP**) or 3 minutes at 72^{0}C (**dU^{Glc}TP**) in a thermal cycler. The reaction mixtures were then divided into two portions. The first portion (20 µl) was directly used for PAGE analysis and the second portion (20 µl) was used for DNA cleavage studies. Restriction endonuclease RsaI (10U/µl, 0.8 µl) and Cutsmart buffer (2.2 µl) were added to the second portion

of the PEX reaction. The mixture was then incubated for 1 hour at 37 °C. Samples were denatured after the reaction by adding the stop solution (1:1 to the reaction volume) and heating for 5 minutes at 95 °C. 12.5 % denaturing PAGE was used to analyze the reaction mixtures.

5.3.10.2. Study of cleavage by restriction endonuclease for dC^{X}TPs

Incorporation of modified dC^{X}TPs within the recognition sequence of KpnI restriction endonuclease

The reaction mixture (40 µl) contained KOD XL DNA polymerase (2.5 U/µl, 0.04 µl), d(ATG)TP (1 mM, 1 µl), dCTP (1 mM, 2 µl) or modified dCTP (1 mM, 2 µl), 6-FAM-labelled primer Prim$^{PEX\ b}$ (3 µM, 2 µl), 30-mer template Temp^{KpC} (3 µM, 3 µl) and 10x buffer for KOD XL DNA polymerase (4 µl) supplied by the manufacturer. Reaction mixtures were incubated either for 40 minutes at 60 °C (for dCTP and **dC^{hm}TP**) or 3 minutes at 72°C (**dC^{Glc}TP**) in a thermal cycler. The reaction mixtures were then divided into two portions. The first portion (20 µl) was directly used for PAGE analysis and the second portion (20 µl) was used for DNA cleavage studies. Restriction endonuclease KpnI (10U/µl, 3.2 µl) and NEB 1 reaction buffer (10x, 2.2 µl) were added to the second portion of the PEX reaction. The mixture was then incubated for 1 hour at 37 °C. Samples were denatured after the reaction by adding the stop solution (1:1 to the reaction volume) and heating for 5 minutes at 95 °C. 12.5 % denaturing PAGE was used to analyze the reaction mixtures.

Incorporation of modified dC^{X}TPs within the recognition sequence of RsaI restriction endonuclease

The reaction mixture (40 µl) contained KOD XL DNA polymerase (2.5 U/µl, 0.04 µl), d(ATG)TP (1 mM, 0.4 µl), dCTP (1 mM, 0.5 µl) or modified dCTP (1 mM, 0.5 µl), 6-FAM- labelled primer Prim$^{PEX\ b}$ (3 µM, 2 µl), 30-mer template Temp^{RsC} (3 µM, 3 µl) and 10x buffer for KOD XL DNA polymerase (4 µl) supplied by the manufacturer. Reaction mixtures were incubated either for 40 minutes at 60 °C (for dCTP and **dC^{hm}TP**) or 3 minutes at 72 °C (**dC^{Glc}TP**) in a thermal cycler. The reaction mixtures were then divided into two portions. The first portion (20 µl) was directly used for PAGE analysis and the second portion (20 µl) was used for DNA cleavage studies. Restriction endonuclease RsaI (10U/µl, 3.2 µl) and Cutsmart reaction buffer (10x, 2.2 µl) were

added to the second portion of the PEX reaction. The mixture was then incubated for 1 hour at 37 °C. Samples were denatured after the reaction by adding the stop solution (1:1 to the reaction volume) and heating for 5 minutes at 95 °C. 12.5 % denaturing PAGE was used to analyze the reaction mixtures.

5.4. Transcription studies

5.4.1. Multiple round transcriptions of modified DNA

Multiple round *in vitro* transcription assays were performed in a volume of 10 µl containing 5 ng of DNA template (1 µl of DNA [quantification corrected and diluted to 5 ng/ µl concentration]) in transcription buffer 40 mM Tris-Cl pH 8, 10 mM $MgCl_2$, 1 mM DTT, 90 mM KCl, 10 µg/ml BSA, 30 nM RNAP holoenzyme (core RNAP saturated with σ^{70}) from *E. coli* (New England Biolabs) and NTPs (200 µM ATP, 1000 µM GTP, 200 µM CTP, 10 µM UTP plus 37 kBq [α-^{32}P] UTP). After allowing open complex formation between RNAP and the template DNA for 10 minutes at 37 °C, the transcription reactions were initiated by adding the NTP mix and run for 10 minutes at 37 °C. The reaction was stopped by adding 10 µl of stop solution (95% formamide, 20 mM EDTA, 0.05% bromophenol blue, and 0.05% xylene cyanol). The samples were then incubated at 95 °C for 5 minutes and chilled on ice. 10 µl of the reaction were analyzed on a 7% polyacrylamide urea gel. Dried gels were exposed to Fuji MS phosphor storage screens and scanned with Amersham typhoon scanner (Cytiva). The data are from at least four independent experiments performed using three different PCR batches.

5.4.2. Quantification of relative transcriptions

The signals from the gel were analyzed using Quantity One software (Biorad). To quantify the transcription intensity, background was subtracted and the bands were normalized to the value of non-modified DNA template which was set to 100%.

6. List of abbreviations and symbols

Ac	Acetyl
ACN	Acetonitrile
AgOTf	Silver trifluoromethanesulfonate
AIBN	Azobisisobutyronitrile
ATP	Adenosine triphosphate
BER	Base Excision Repair
bp	Base pair
Bu_3N	Tributylamine
BSA	Bovine Serum Albumin
COSY	Correlated spectroscopy
CpG	5'-C-phosphate-G-3'
CuAAC	Copper catalyzed Azide Alkyne Cycloaddition
CsOAc	Cesium acetate
CTP	Cytosine triphosphate
Cy	Cyclohexane
Da	Dalton
dATP	2'-deoxyadenosine triphosphate
DBF-DMA	*N,N*-dimethylformamide dimethyl acetal
DCE	Dichloroethane
DCM	Dichloromethane
dCTP	2'-deoxycytidine triphosphate
dGTP	2'-deoxyguanosine triphosphate
DIPEA	*N,N*-diisopropylethylamine
DMAP	4-Dimethylaminopyridine
DMF	*N,N*-dimethylformamide
DMTr	Dimethoxytrityl
DNA	Deoxyribonucleic acid

DNMTs	DNA methyltransferases
dNMP	2'-deoxyribonucleoside 5'-*O*-monophosphate
dNTP	2'-deoxyribonucleoside 5'-*O*-triphosphate
ds	double-stranded
DTT	Dithiothreitol
dTTP	2'-deoxythymidine triphosphate
dUTP	2'-deoxyuridine triphosphate
EDTA	Ethylenediaminetetraacetic acid
equiv.	Equivalents
ESI	Electrospray ionization
EtOAc	Ethyl acetate
FAM	Fluorescein amidite
Glc	Glucose
GlcAc	Acetyl-protected glucose
GTP	Guanosine triphosphate
h	hour
HF	Hydrogen fluoride
hm	hydroxymethyl
HMBC	Heteronuclear Multiple Bond Correlation
HPLC	High Performance Liquid Chromatography
HRMS	High Resolution Mass Spectroscopy
HSQC	Heteronuclear Single Quantum Coherence
LC	Liquid Chromatography
MALDI	Matrix Assisted Laser Desorption/Ionization
MeOH	Methanol
MHz	Mega Hertz
min	minutes
m.p.	melting point
MS	Mass Spectrometry
NB	Nitrobenzyl

NBS	N-bromosuccinimide
NMR	Nuclear Magnetic Resonance
nt	nucleotide
NTP	Ribonucleoside 5'-*O*-triphosphate
ON	Oligonucleotide
PAGE	Polyacrylamide Gel Electrophoresis
PCR	Polymerase Chain Reaction
PEX	Primer Extension
Pol	Polymerase
RE	Restriction Endonuclease
Rf	Retention factor
RNA	Ribonucleic acid
RNAP	RNA Polymerase
RP	Reverse-Phase
rpm	revolutions per minute
r.t.	room temperature
SD	Standard Deviation
SDS	Sodium Dodecyl Sulfate
ssDNA	single-stranded DNA
T4- βGT	T4 phage *β*-glucosyltransferase
TAC	*Tert*-butylphenoxyacetyl
TBDMS	*Tert*-butyldimethylsilyl
TBE	Tris-Borate-EDTA
TDG	Thymine DNA glycosylase
TEAB	Tetraethylammonium bicarbonate
TET	Ten-Eleven-Translocation enzyme
THF	Tetrahydrofuran
TIPS	Triisopropylsilyl
TLC	Thin Layer Chromatography
TMSOTf	Trimethylsilyl trifluoromethanesulfonate

TOF	Time Of Flight
TSS	Transcription Start Site
UDP	Uridine diphosphate
UTP	Uridine triphosphate
UV	Ultraviolet
X	Modification

7. Appendices

Appendix 1. Copies of MALDI spectra of modified oligonucleotides **ON9-16**

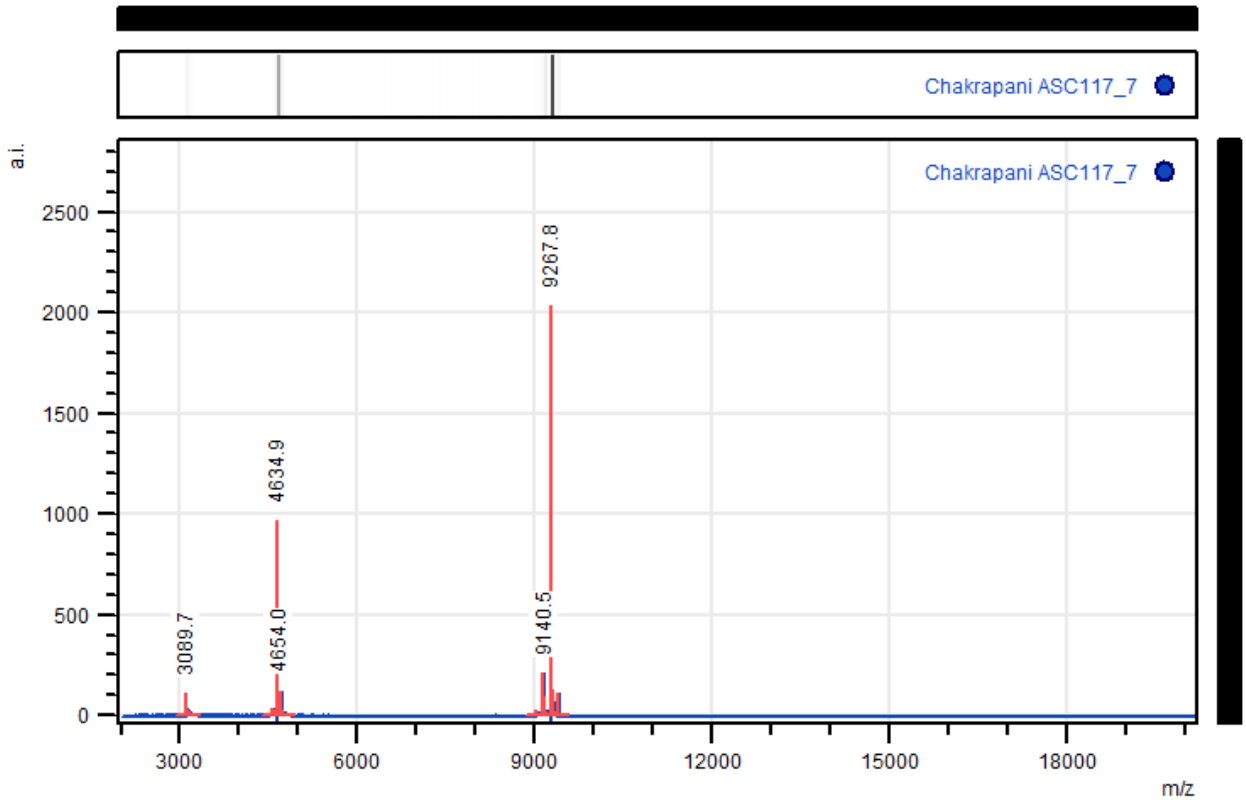


Figure 56: MALDI-TOF MS spectrum of **ON9**. Calculated for $[M]^+$ = 9402.6 Da, $[M\text{-}NB]^+$ = 9267.6 Da; found: $[M\text{-}NB]^+$ = 9267.8 Da.

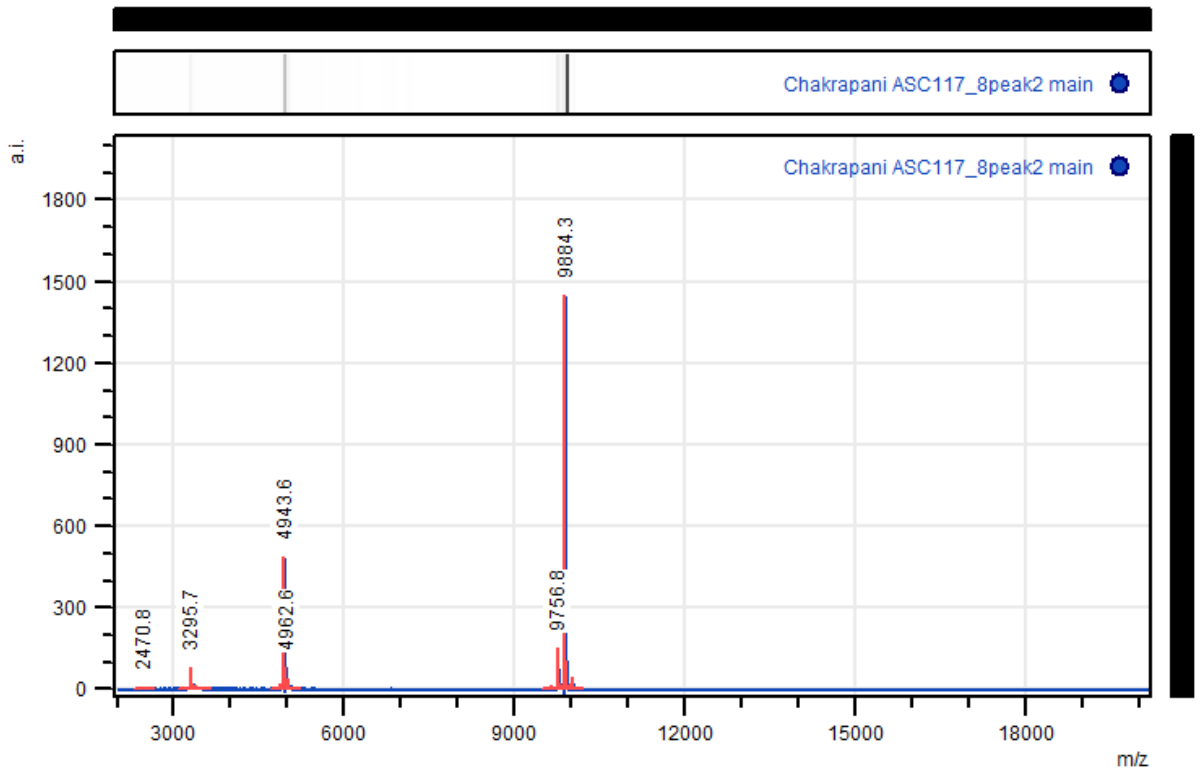


Figure 57: MALDI-TOF MS spectrum of **ON10**. Calculated for $[M]^+$ = 10018.5 Da, $[M\text{-}NB]^+$ = 9883.5 Da; found: $[M\text{-}NB+H]^+$ = 9884.3 Da.

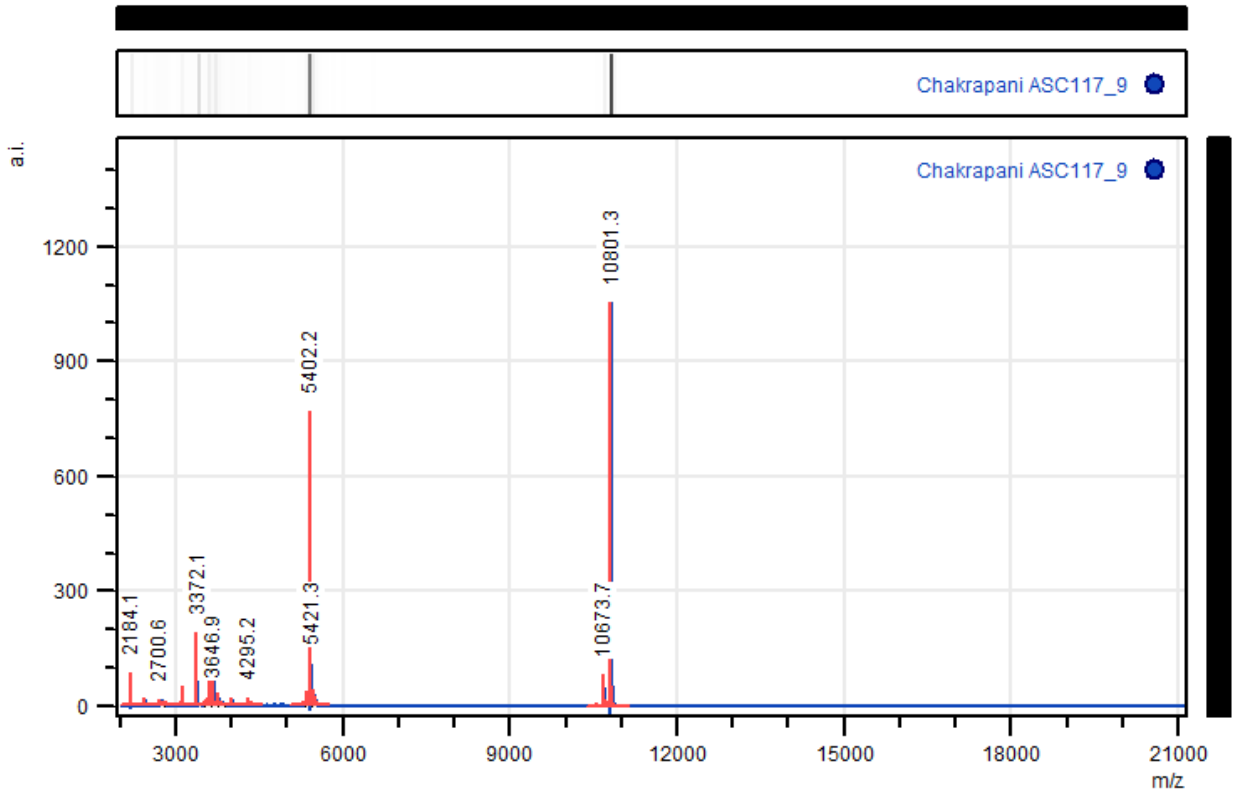


Figure 58: MALDI-TOF MS spectrum of **ON11**. Calculated for $[M]^+$ = 10935.4 Da, $[M\text{-}NB]^+$ = 10800.4 Da; found: $[M\text{-}NB+H]^+$ = 10801.3 Da.

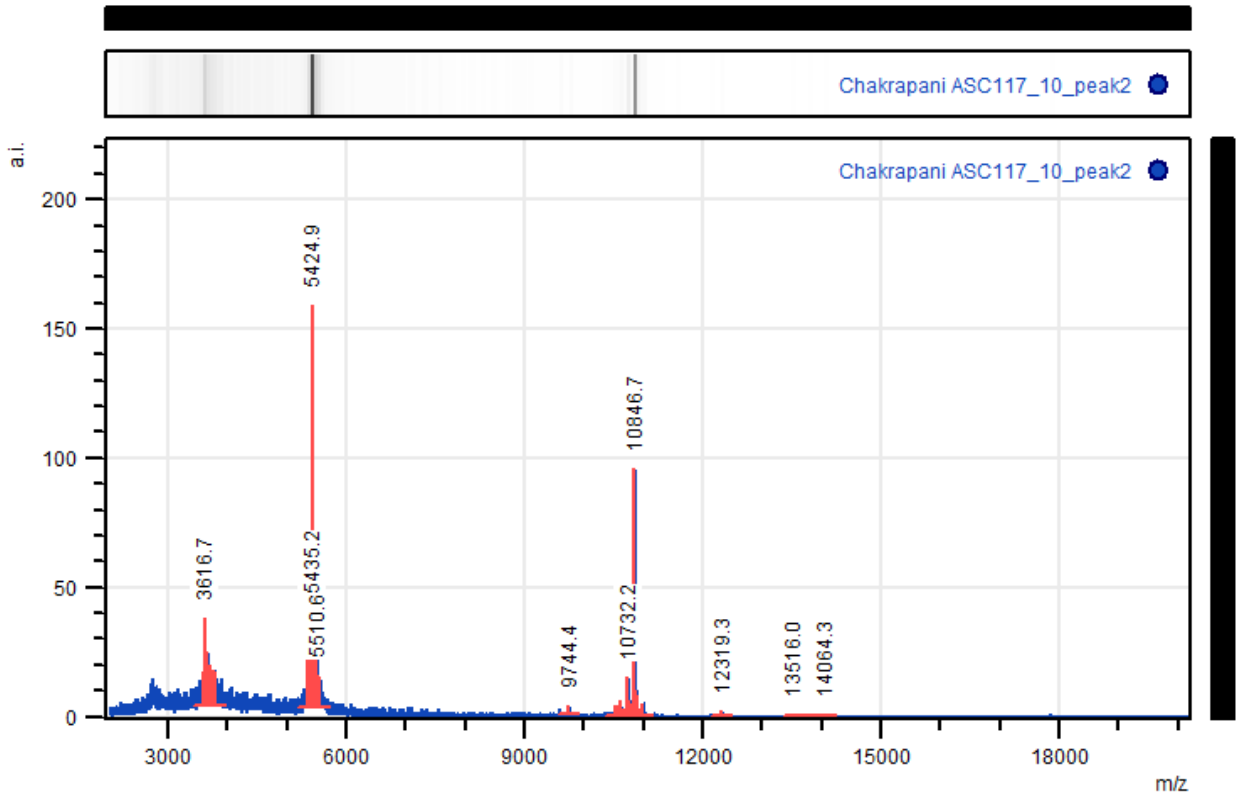


Figure 59: MALDI-TOF MS spectrum of **ON12**. Calculated for $[M]^+$ = 11251.2 Da, $[M\text{-}3NB]^+$ = 10846.2 Da; found: $[M\text{-}3NB]^+$ = 10846.7 Da.

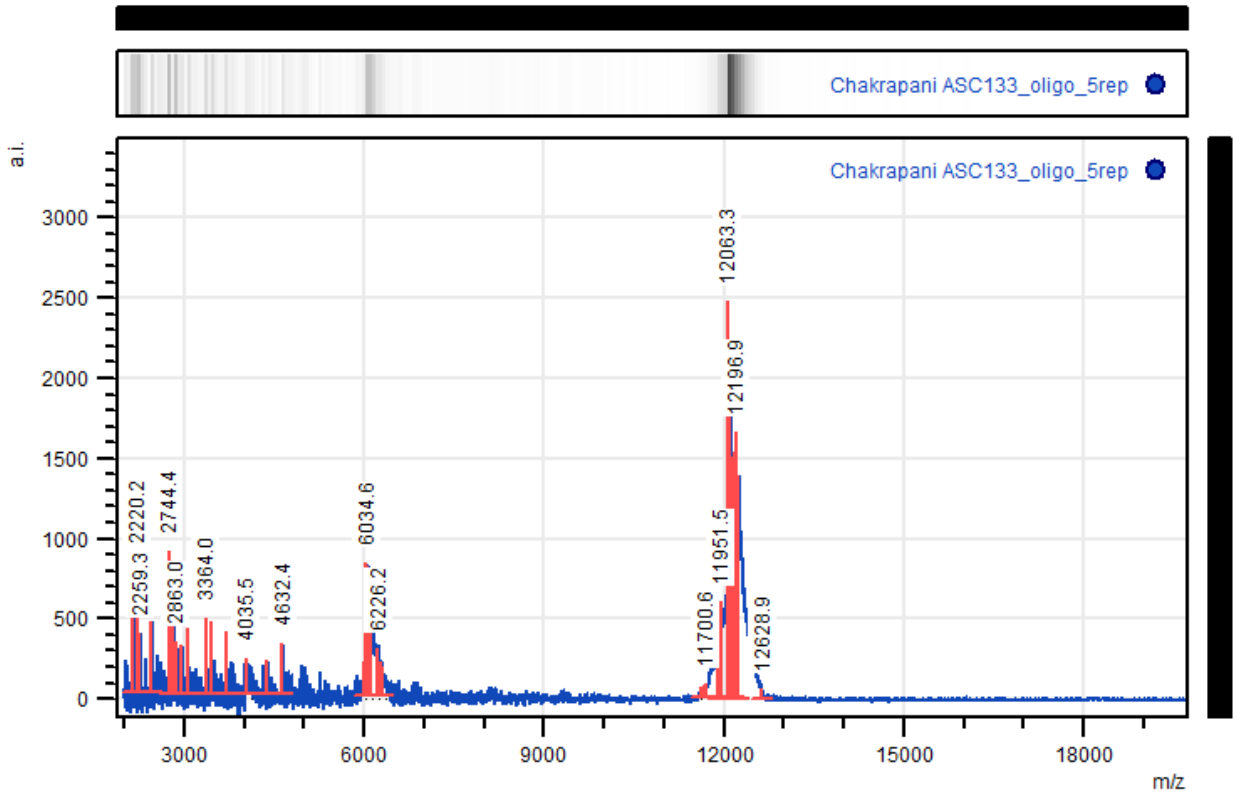


Figure 60: MALDI-TOF MS spectrum of **ON13**. Calculated for $[M]^+$ = 12195 Da, $[M\text{-}NB]^+$ = 12060 Da; found: $[M+H]^+$ = 12196.9 Da, $[M\text{-}NB+3H]^+$ = 12063.3 Da.

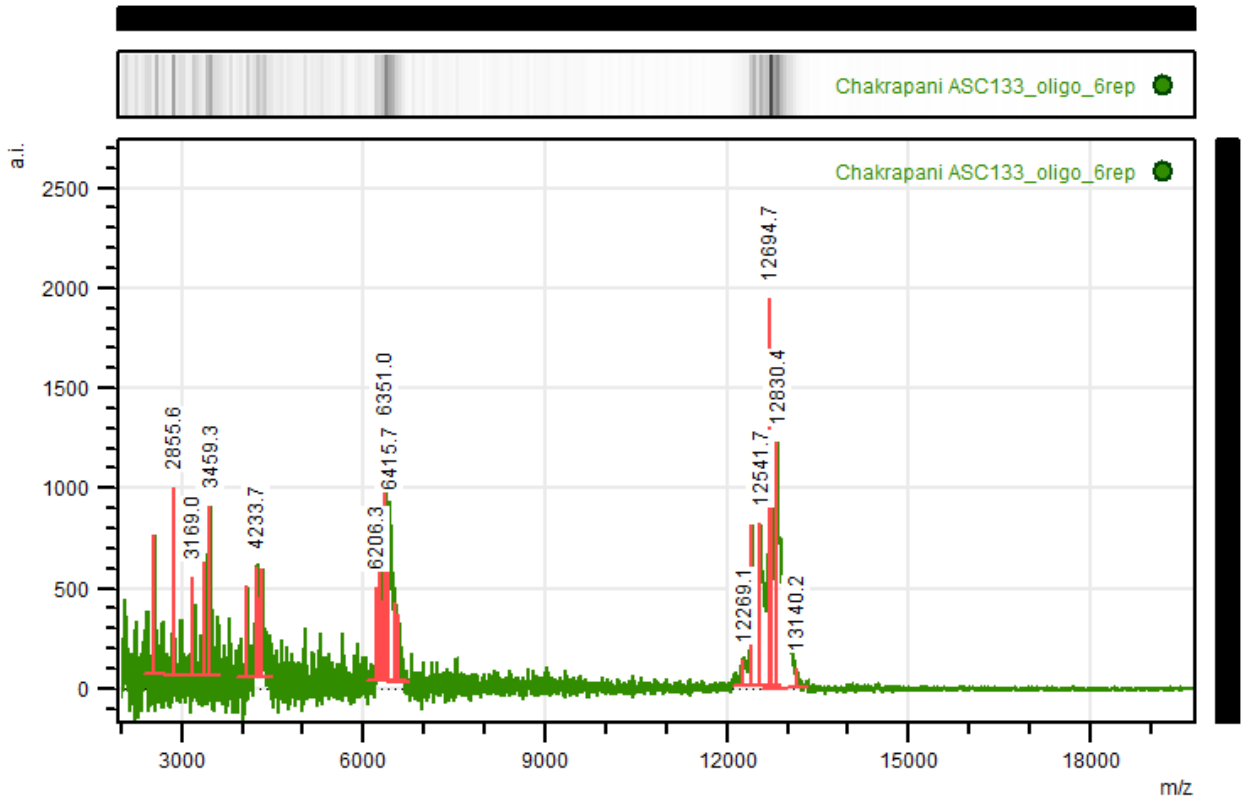


Figure 61: MALDI-TOF MS spectrum of **ON14**. Calculated for $[M]^+$ = 12828.4 Da, $[M\text{-}NB]^+$ = 12693.4 Da; found: $[M+2H]^+$ = 12830.4 Da, $[M\text{-}NB+H]^+$ = 12694.7 Da.

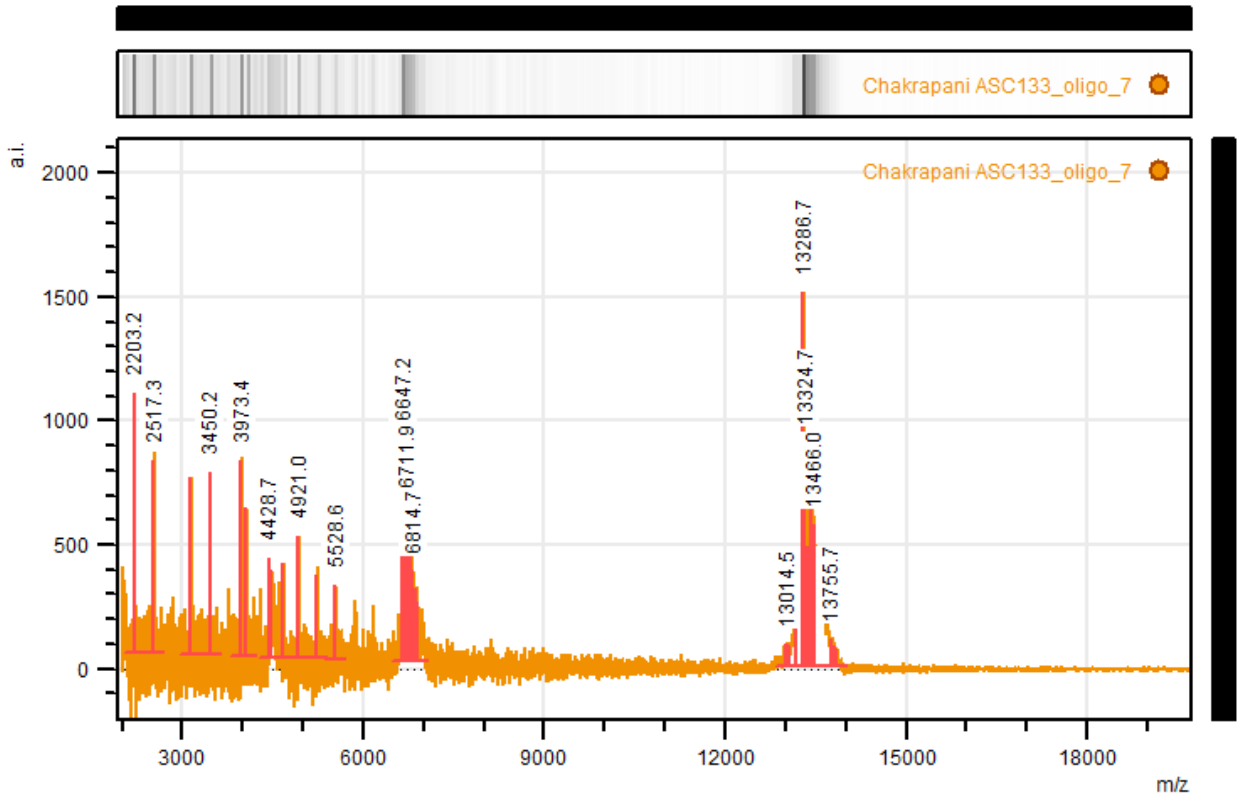


Figure 62: MALDI-TOF MS spectrum of **ON15**. Calculated for $[M]^+$ = 13421.7 Da, $[M\text{-}NB]^+$ = 13286.7 Da; found: $[M\text{-}NB]^+$ = 13286.7 Da.

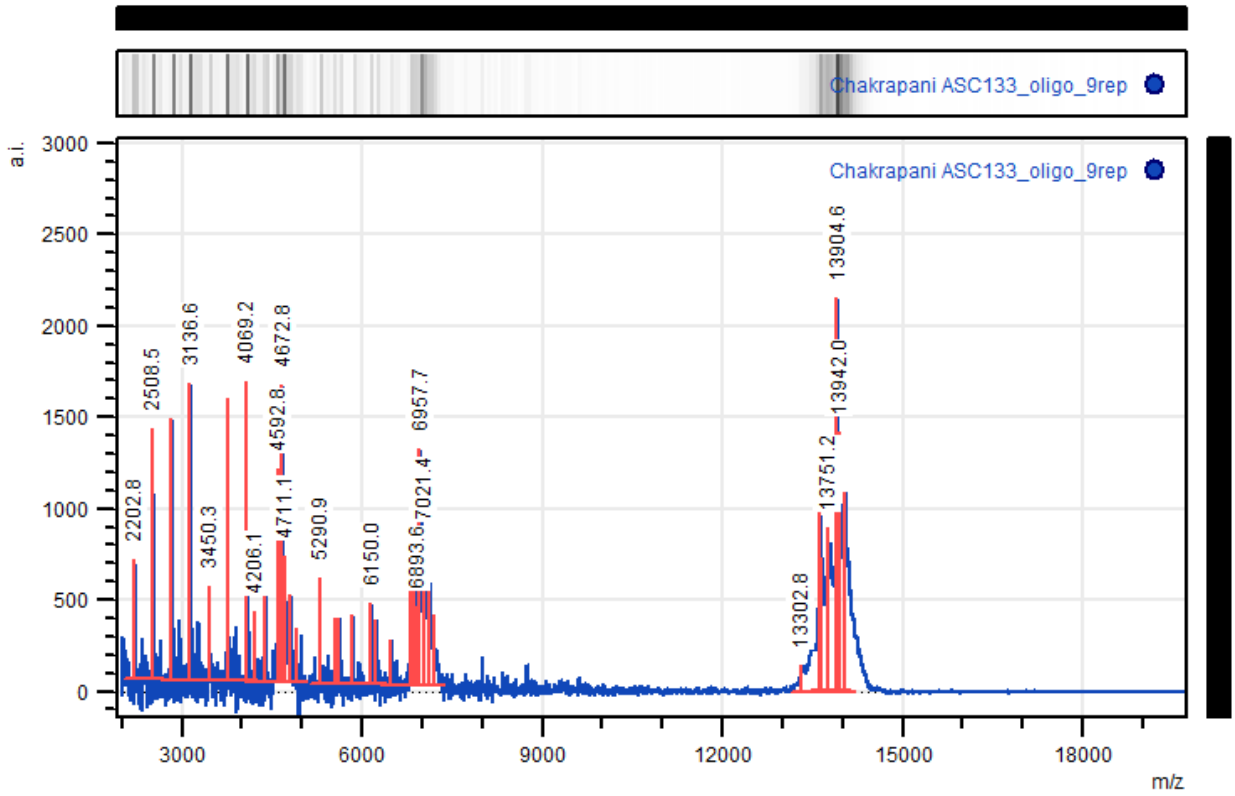
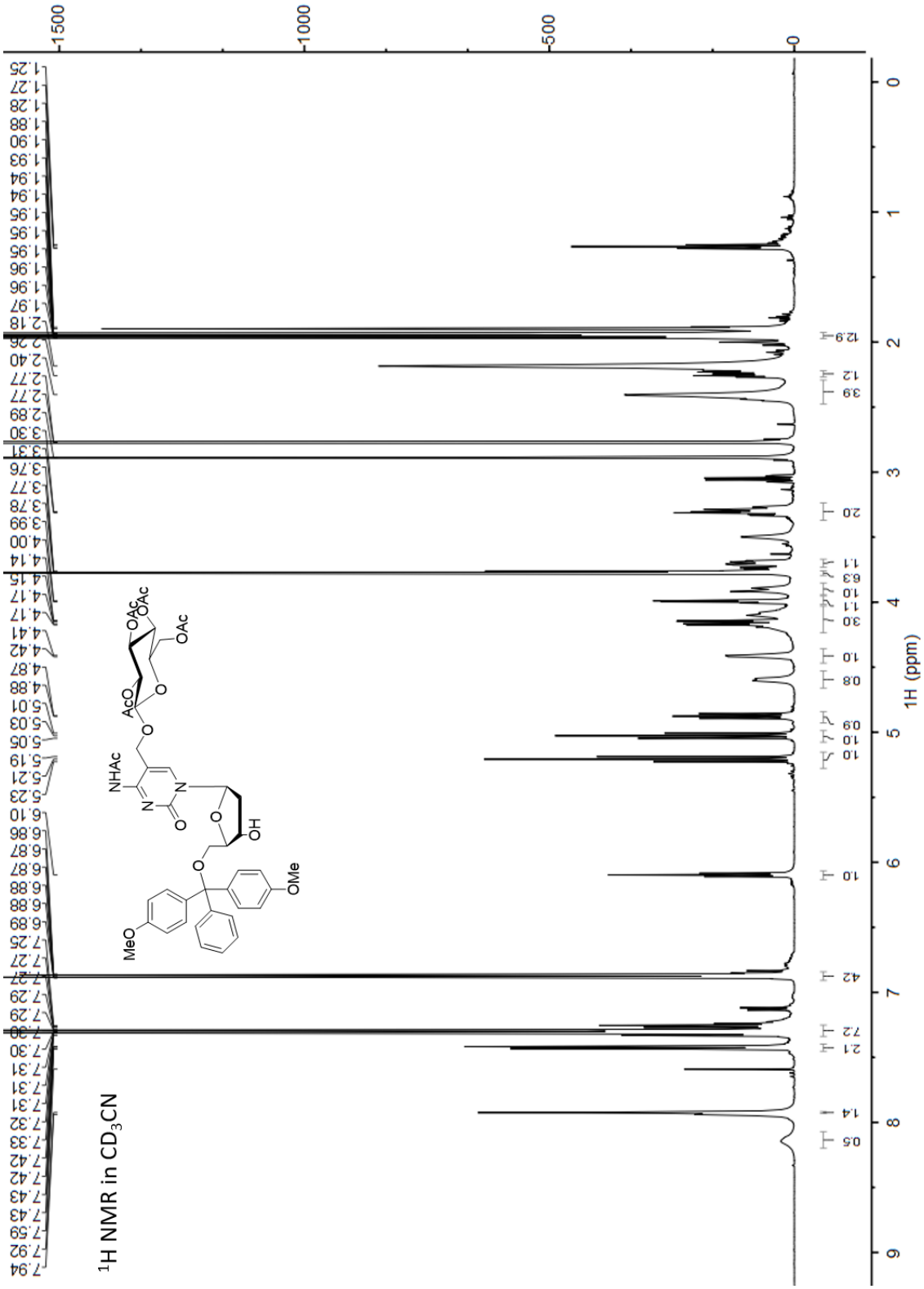
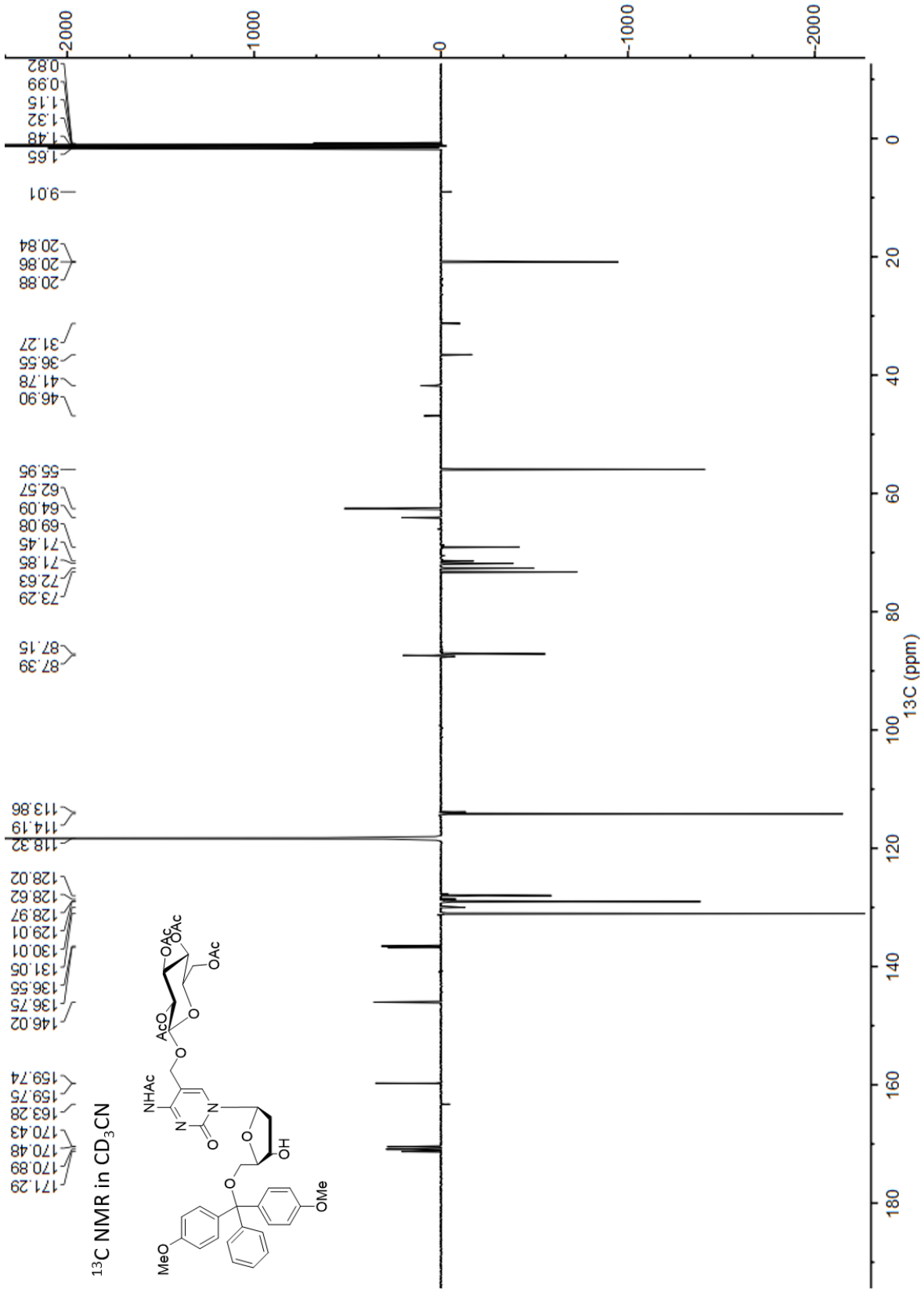


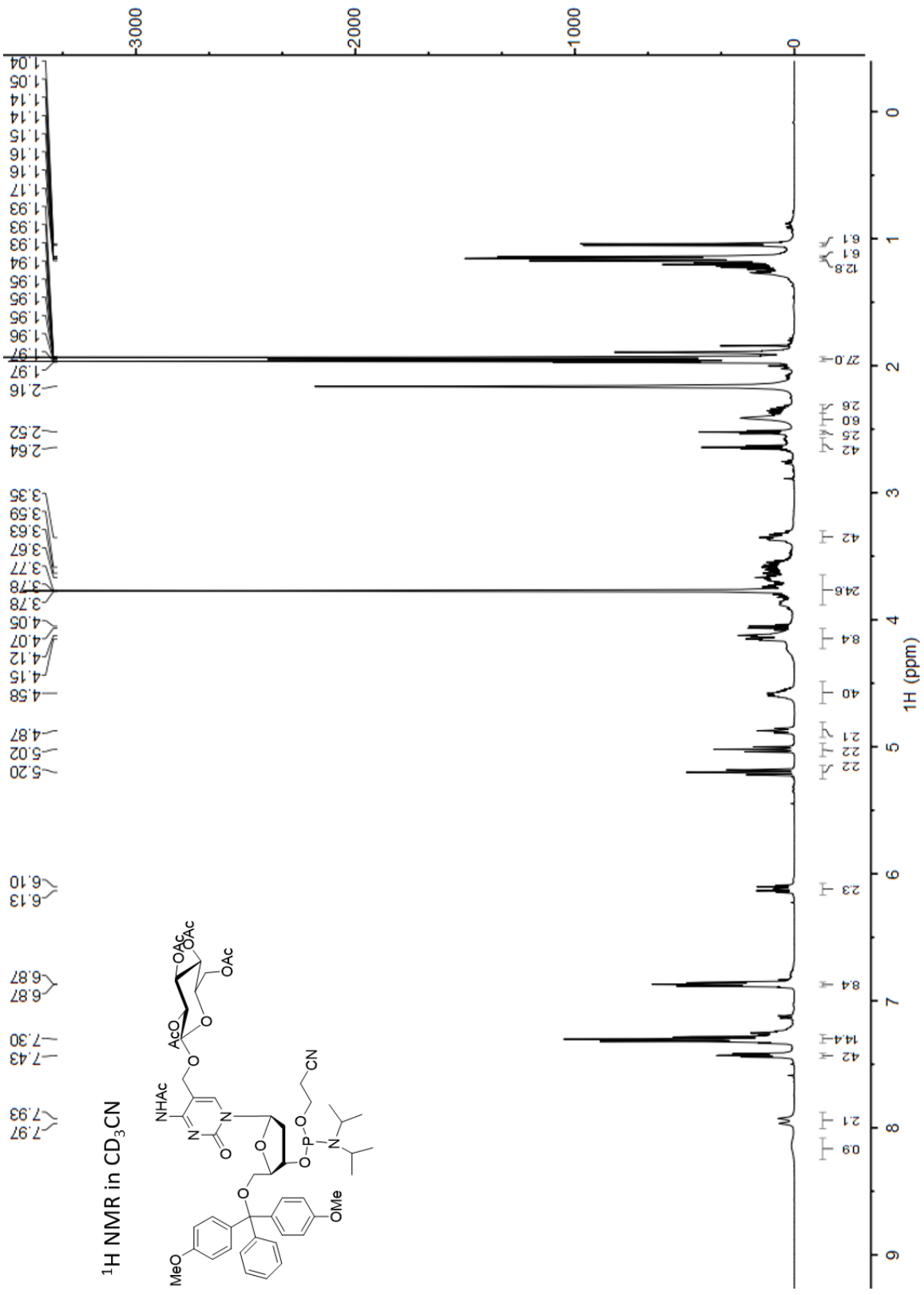
Figure 63: MALDI-TOF MS spectrum of **ON16**. Calculated for $[M]^+$ = 14039.2 Da, $[M\text{-}NB]^+$ = 13904.2 Da; found: $[M\text{-}NB]^+$ = 13904.6 Da.

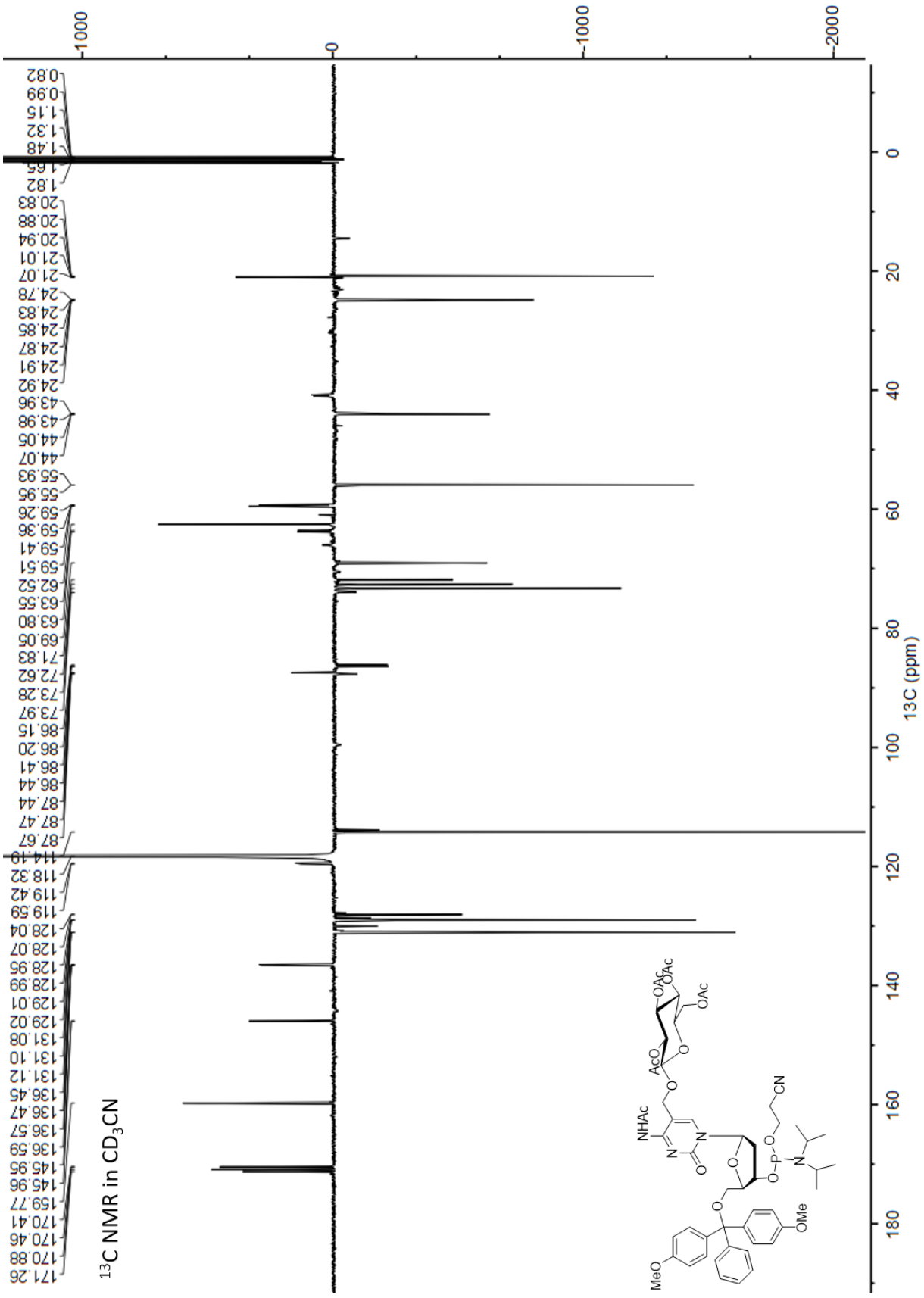
Appendix 2. Copies of ^{1}H, ^{13}C NMR spectra of **30**.

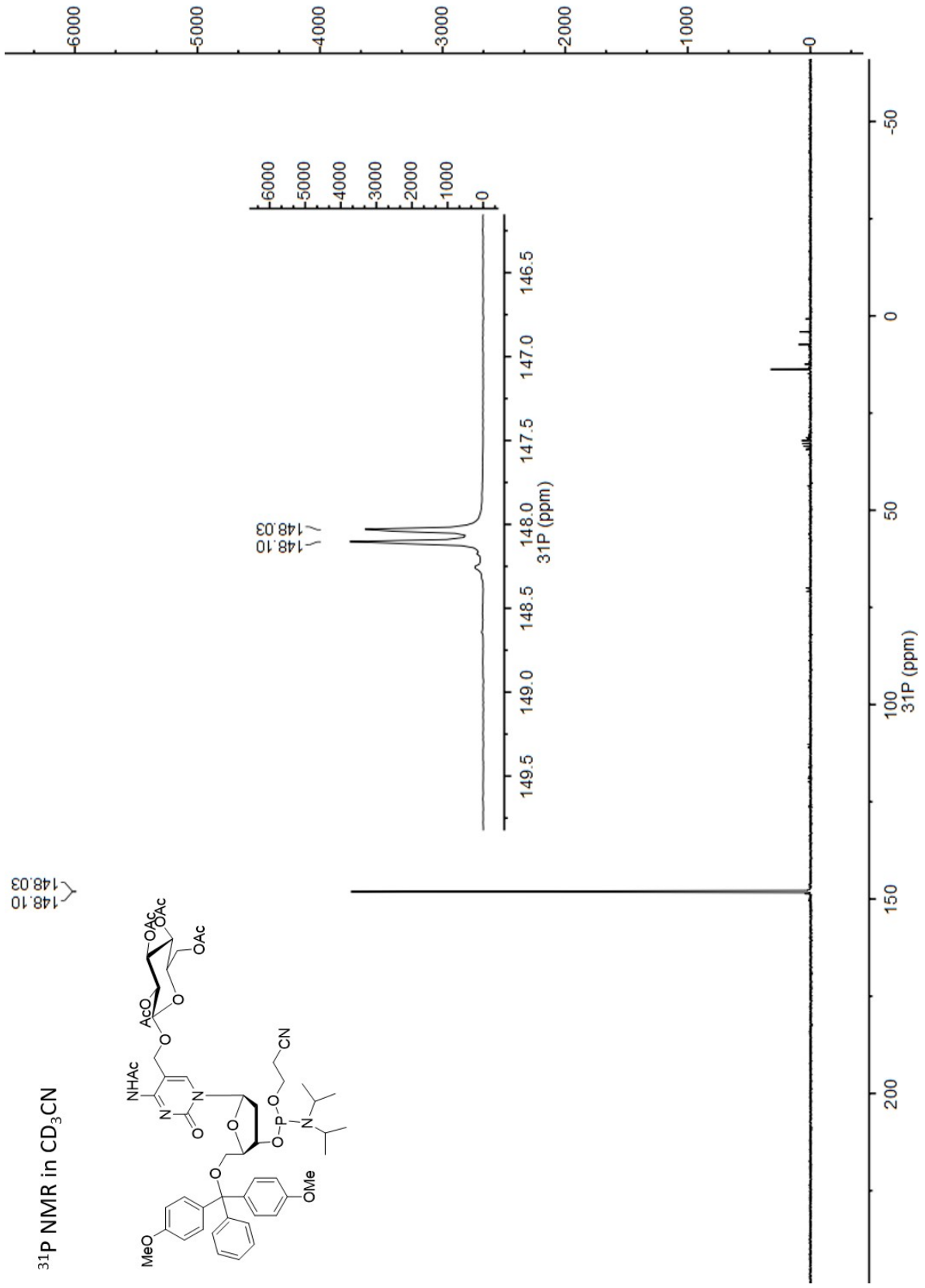



^{13}C NMR in CD_3CN
13C (ppm)
0.82
0.99
1.15
1.32
1.48
1.65
9.01
20.84
20.86
20.88
31.27
36.55
41.78
46.90
55.95
62.57
64.09
69.08
71.45
71.85
72.63
73.29
87.15
87.39
113.86
114.19
118.32
128.02
128.62
128.97
129.01
130.01
131.05
136.55
136.75
146.02
159.74
159.75
163.28
170.43
170.48
170.89
171.29
NHAc
AcO
OAc
OAc
OAc
O
O
N
N
O
O
OH
O
MeO
OMe

Appendix 3. Copies of ^{1}H, ^{13}C, ^{31}P NMR spectra of **31**.




13C (ppm)
13C NMR in CD3CN


^{31}P NMR in CD_3CN
148.10
148.03
NHAc
AcO
OAc
OAc
OAc
CN
OMe
MeO
31P (ppm)

8. References

1) R. Dahm, *Dev. Biol.* **2005**, *278*, 274–288.
2) R. Dahm, *Hum. Genet.* **2008**, *122*, 565–581.
3) M. E. Jones, *Yale J Biol Med*. **1953**, *26*, 80-97.
4) E. Chargaff, *Fed Proc*. **1951**, *10*, 654-659.
5) D. Elson, E. Chargaff, *Experientia* **1952**, *8*, 143–145.
6) E. Chargaff, R. Lipshitz, C. Green, *J. Biol. Chem.* **1952**, *195*, 155–160.
7) J. D. Watson, F. H. Crick, *Nature* **1953**, *171*, 737–738.
8) C. F. Matta, N. Castillo, R. J. Boyd, *J. Phys. Chem. B.* **2006**, *110*, 563–578.
9) G. W. Leslie, S. Arnott, R. Chandrasekaran, R. L. Ratliff, *J. Mol. Biol.* **1980**, *143*, 49–72.
10) M. C. Wahl, M. Sundaralingam, *Biopolymers* **1997**, *44*, 45–63.
11) A. H. J. Wang, G. J. Quigley, F. J. Kolpak, J. L. Crawford, J. H. van Boom, G. van der Marel, A. Rich, *Nature* **1979**, *282*, 680–686.
12) F. Crick, *Nature* **1970**, *227*, 561–563.
13) E. A. Raiber, R. Hardisty, P. van Delft, S. Balasubramanian, *Nat. Chem. Rev.* **2017**, *1*, 1-13.
14) T. Carell, M. Q. Kurz, M. Müller, M. Rossa, F. Spada, *Angew. Chem. Int. Ed.* **2018**, *57*, 4296–4312.
15) A. Hofer, Z. J. Liu, S. Balasubramanian, *J. Am. Chem. Soc.* **2019**, *141*, 6420– 6429.
16) M. K. Bilyard, S. Becker, S. Balasubramanian, *Curr. Opin. Chem. Biol.* **2020**, *57*, 1– 7.
17) B. F. Yuan, *Chem. Res. Toxicol.* **2020**, *33*, 695– 708.
18) G. R. Wyatt, *Nature* **1950**, *166*, 237 – 238.
19) G. R. Wyatt, *Biochem. J.* **1951**, *48*, 581–584.
20) R. Jaenisch, A. Bird, *Nat. Genet.* **2003**, *33*, 245 – 254.
21) K. Chen, B. S. Zhao, C. He, *Cell Chem. Biol*. **2016**, *23*, 74−85.
22) C. Luo, P. Hajkova, J. R. Ecker, *Science* **2018**, *361*, 1336− 1340.
23) T. Liu, C. J. Ma, B. F. Yuan, Y. Q. Feng, *Sci. China: Chem.* **2018**, *61*, 381−392.
24) K. Flodman, I. R. Corrêa, N. Dai, P. Weigele, S. Y. Xu, *Front. Microbiol.* **2020**, *11*, 604618.
25) L.H. Huang, C. M. Farnet, K. C. Ehrlich, M. Ehrlich, *Nucleic Acids Res.* **1982**, *10*, 1579-1591.
26) H. Leonhardt, H. P. Rahn, M. C. Cardoso, *Crit. Rev. Eukaryot. Gene Expr.* **1999**, *9*, 345–351.
27) J. A. Law, S. E. Jacobsen, *Nat. Rev. Genet.* **2010**, *11*, 204–220.

28) Z. D. Smith, A. Meissner, *Nat. Rev. Genet.* **2013**, *14*, 204−220.
29) P. A. Jones, *Nat. Rev. Genet.* **2012**, *13*, 484–492.
30) X. Cheng, R. J. Roberts, *Nucleic Acids Res.* **2001**, *29*, 3784−3795.
31) L. Shen, C. X. Song, C. He, Y. Zhang, *Annu. Rev. Biochem*. **2014**, *83*, 585−614.
32) H. Wu, Y. Zhang, *Cell* **2014**, *156*, 45–68.
33) X. Lu, B. S. Zhao, C. He, *Chem. Rev.* **2015**, *115*, 2225– 2239.
34) M. Tahiliani, K. P. Koh, Y. Shen, W. A. Pastor, H. Bandukwala, Y. Brudno, S. Agarwal, L. M. Iyer, D. R. Liu, L. Aravind, A. Rao, *Science* **2009**, *324*, 930.
35) S. Ito, A. C. D'Alessio, O. V. Taranova, K. Hong, L. C. Sowers, Y. Zhang, *Nature* **2010**, *466*, 1129–1151.
36) T. Pfaffeneder, B. Hackner, M. Truss, M. Munzel, M. Muller, C. A. Deiml, C. Hagemeier, T. Luo, *Angew. Chem. Int. Ed.* **2011**, *50*, 7008–7012.
37) S. Ito, L. Shen, Q. Dai, S. C. Wu, L. B. Collins, J. A. Swenberg, C. He, Y. Zhang Y, *Science* **2011**, *333*, 1300–1303.
38) A. Schön, E. Kaminska, F. Schelter, E. Ponkkonen, E. Korytiaková, S. Schiffers, T. Carell, *Angew. Chem. Int. Ed.* **2020**, *59*, 5591– 5594.
39) Y. F. He, B.-Z. Li, Z. Li, P. Liu, Y. Wang, Q. Tang, J. Ding, Y. Jia, Z. Chen, L. Li, Q. Dai, C. X. Song, K. Zhang, C. He, G.-L. Xu, *Science* **2011**, *333*, 1303– 1307
40) R. M. Kohli, Y. Zhang, *Nature* **2013**, *502*, 472–479.
41) A. Inoue, Y. Zhang, *Science* **2011**, *334*, 194-194.
42) C. You, D. Ji, X. Dai, Y. Wang, *Sci. Rep.* **2014**, *4*, 1-6.
43) L. Wang, Y. Zhou, L. Xu, R. Xiao, X. Lu, L. Chen, J. Chong, H. Li, C. He, X.-D. Fu, D. Wang, *Nature* **2015**, *523*, 621– 625.
44) N. Kitsera, J. Allgayer, E. Parsa, N. Geier, M. Rossa, T. Carell, A. Khobta, *Nucleic Acids Res.* **2017**, *45*, 11033– 11042.
45) E. A. Raiber, P. Murat, D. Y. Chirgadze, D. Beraldi, B. F. Luisi, S. Balasubramanian, *Nat. Struct. Mol. Biol.* **2015**, *22*, 44-49.
46) T. T. M. Ngo, J. Yoo, Q. Dai, Q. Zhang, C. He, A. Aksimentiev, T. Ha, *Nat. Commun.* **2016**, *7*, 10813.
47) G. R. Wyatt, S. S. Cohen, *Nature* **1952**, *170*, 1072–1073.
48) G. R. Wyatt, S. S. Cohen, *Biochem. J.* **1953**, *55*, 774–782.

49) P. Weigele, E. A. Raleigh, *Chem. Rev.* **2016**, *116*, 12655 – 12687.

50) N. W. Penn, R. Suwalski, C. O'Riley, K. Bojanowski, R. Yura, *Biochem. J.* **1972**, *126*, 781–790.

51) D. Globisch, M. Münzel, M. Muller, S. Michalakis, M. Wagne, S. Koch, T. Bruckl, M. Biel, T. Carell, *PLoS One* **2010**, *5*, e15367.

52) M. Münzel, D. Globisch, T. Bruckl, M. Wagner, V. Welzmiller, S. Michalakis, M. Muller, M. Biel, T. Carell, *Angew. Chem. Int. Ed.* **2010**, *49*, 5375–5377.

53) S. G. Jin, X. Wu, A. X. Li, G. P. Pfeifer, *Nucleic Acids Res.* **2011**, *39*, 5015–5024.

54) M. Wagner, J. Steinbacher, T. F. Kraus, S. Michalakis, B. Hackner, T. Pfaffeneder, A. Perera, M. Müller, A. Giese, H. A. Kretzschmar, T. Carell, *Angew. Chem. Int. Ed.* **2015**, *54*, 12511-12514.

55) C. X. Song, K. E. Szulwach , Y. Fu, Q. Dai, C. Yi, X. Li, Y. Li, C. H. Chen, W. Zhang, X. Jian, J. Wang, L. Zhang, T. J. Looney, B. Zhang, L. A. Godley, L. M. Hicks, B. T. Lahn, P. Jin, C. He, *Nat. Biotechnol.* **2011**, *29*, 68–72.

56) M. A. Hahn, R. Qiu, X. Wu, A. X. Li, H. Zhang, J. Wang, J. Jui, S. G. Jin, Y. Jiang, G. P. Pfeifer, Q. Lu, *Cell Rep.* **2013**, *3*, 291–300.

57) L. Lercher, M. A. McDonough, A. H. El-Sagheer, A. Thalhammer, S. Kriaucionis, T. Brown, C. J. Schofield, *Chem. Commun.* **2014***, 50*, 1794-1796.

58) A. Perera, D. Eisen, M. Wagner, S. K. Laube, A. F. Künzel, S. Koch, J. Steinbacher, E. Schulze, V. Splith, N. Mittermeier, M. Müller, M. Biel, T. Carell, S. Michalakis, *Cell Rep.* **2015**, *11*, 283– 294.

59) S. G. Jin, Y. R. Qiu, T. A. Rauch, Y. Wang, G. Schackert, D. Krex, Q. Lu, G. P. Pfeifer, *Cancer Res.* **2011**, *71*, 7360–7365.

60) L. Scourzic, E. Mouly, O. A. Bernard, *Genome Med.* **2015**, *7,* 9.

61) M. Janoušková, Z. Vaníková, F. Nici, S. Boháčová, D. Vítovská, H. Šanderová, M. Hocek, L. Krásný, *Chem. Comm*. **2017**, *53*, 13253-13255.

62) S. R. Kornberg, S. B. Zimmerman, A. Kornberg, *J. Biol. Chem.* **1961**, *236*, 1487-1493.

63) J. Tomaschewski, H. Gram, J. W. Crabb, W. Rüger, *Nucleic Acids Res.* **1985**, *13*, 7551-7566.

64) S. Moréra, A. Imberty, U. Aschke-Sonnenborn, W. Rüger, P. S. Freemont, *J. Mol. Biol.* **1991**, *292*, 717-730.

65) G. S. Cox, T. W. Conway, *J. Virol.* **1973**, *12*, 1279-1287.

66) K. Dharmalingam, E. B. Goldberg, *Virology* **1979**, *96*, 393-403.

67) H. E. Krokan, F. Drabløs, G. Slupphaug, *Oncogene* **2002***, 21*, 8935–8948.

68) K. Fujikawa , H. Kamiya, H. Kasai , *Nucleic Acids Res.* **1998**, **26** , 4582 —4587.

69) D. K. Rogstad , J. Heo , N. Vaidehi , W. A. Goddard , A. Burdzy, L. C. Sowers , *Biochemistry* **2004**, *43* , 5688 —5697.

70) R. G. Kallen, M. Simon, J. Marmur, *J. Mol. Biol.* **1962**, *5*, 248-250.

71) Y. J. Lee, N. Dai, S. E. Walsh, S. Müller, M. E. Fraser, K. M. Kauffman, C. Guan, I. R. Corrêa, P. R. Weigele, *Proc. Natl. Acad. Sci. USA.* **2018**, *115*, E3116– E3125

72) T. Pfaffeneder, F. Spada, M. Wagner, C. Brandmayr, S. K. Laube, D. Eisen, M. Truss, J. Steinbacher, B. Hackner, O. Kotljarova, D. Schuermann, *Nat. Chem. Biol.* **2014**, *10*, 574-581.

73) J. F. Mouret, M. Polverelli, F. Sarrazini, J. Cadet, *Chem. Biol. Interact.* **1991**, *77*, 187–201.

74) S. Cortellino, J. Xu, M. Sannai, R. Moore, E. Caretti, A. Cigliano, M. Le Coz, K. Devarajan, A. Wessels, D. Soprano, L. K. Abramowitz, *Cell* **2011**, *146*, 67-79.

75) A. L. Jacobs, P. Schär, *Chromosoma* **2012**, *121*, 1–20.

76) Z. Djuric, L. K. Heilbrun, M. S. Simon, D. Smith, D. A. Luongo, P. M. LoRusso, S. Martino, *Cancer* **1996**, *77*, 691-696.

77) Z. Djuric, L. K. Heilbrun, S. Lababidi, E. Berzinkas, M. S. Simon, M. A. Kosir, *Cancer Epidemiol Biomarkers Prev.* **2001**, *10*, 147−149.

78) R. Olinski, M. Starczak, D. Gackowski, *Mutat. Res.* **2016**, *767*, 59– 66.

79) S. Carson, J. Wilson, A. Aksimentiev, P. R. Weigele, M. Wanunu, *Nucleic Acids Res.* **2016**, *44*, 2085−2092.

80) J. H. Gommers-Ampt, F. Van Leeuwen, A. L. de Beer, J. F. Vliegenthart, M. Dizdaroglu, J. A. Kowalak, P. F. Crain, P. Borst, *Cell* **1993**, *75*, 1129−1136.

81) H. G. A. M. Van Luenen, C. Farris, S. Jan, P. A. Genest, P. Tripathi, A. Velds, R. M. Kerkhoven, M. Nieuwland, A. Haydock, G. Ramasamy, S. Vainio, T. Heidebrecht, A. Perrakis, L. Pagie, B. van Steensel, P. J. Myler, P. Borst, *Cell* **2012**, *150*, 909–921.

82) F. van Leeuwen, E. R. Wijsman, E. Kuyl-Yeheskiely, G. A. van der Marel, J. H. van Boom, P. Borst, *Nucleic Acids Res.* **1996**, *24*, 2476–2482.

83) F. van Leeuwen, E. R. Wijsman, R. Kieft, G. A. van der Marel, J. H. van Boom, P. Borst, *Genes Dev.* **1997**, *11*, 3232–3241.

84) F. van Leeuwen, M. C. Taylor, A. Mondragon, H. Moreau, W. Gibson, R. Kieft, P. Borst, *Proc. Natl. Acad. Sci. U. S. A.* **1998**, *95*, 2366−2371.

85) P. Borst, F. van Leeuwen, *Mol. Biochem. Parasitol.* **1997**, *90*, 1–8.

86) L. J. Cliffe, T. N. Siegel, M. Marshall, G. A. Cross, R. Sabatini, *Nucleic Acids Res.* **2010**, *38*, 3923−3935.

87) R. Kieft, Y. Zhang, A. P. Marand, J. D. Moran, R. Bridger, L. Wells, R. J. Schmitz, R. Sabatini, *PLoS Genet.* **2020**, *16*, e1008390.

88) W. B. Wan, P. P. Seth, *J. Med. Chem.* **2016**, *59*, 9645-9667.

89) T. C. Roberts, R. Langer, M. J. Wood, *Nat. Rev. Drug Discov.* **2020**, *19*, 673-694.

90) S. Ochoa, V. T. Milam, *Molecules* **2020**, *25*, 4659.

91) L. K. McKenzie, R. El-Khoury, J. D. Thorpe, M. J. Damha, M. Hollenstein, *Chem. Soc. Rev.* **2021**, *50*, 5126-5164.

92) M. Hocek, *J. Org. Chem.* **2014**, *79*, 21, 9914–9921

93) M. Hocek, *Acc. Chem. Res.* **2019**, *52*, 1730-1737.

94) I. Ivancová, M. Hocek, *Curr. Opin. Chem. Biol.* **2019**, *52*, 136-144.

95) C. B. Reese, *Org. Biomol. Chem.* **2005**, *3*, 3851–3868.

96) B. C. Froehler, P. G. Ng, M. D. Matteucci, *Nucleic Acids Res.* **1986**, *14*, 5399–5407.

97) P. J. Garegg, I. Lindh, T. Regberg, J. Stawinski, R. Strömberg, *Tetrahedron Lett.* **1986**, *27*, 4051.

98) P. T. Gilham, H. G. Khorana, *J. Am. Chem. Soc.* **1958**, *80*, 6212-6222.

99) R. L. Letsinger, K. K. Ogilvie, *J. Am. Chem. Soc.* **1969**, *91*, 3350-3355.

100) C. B. Reese, *Tetrahedron* **1978**, *34*, 3143-3179.

101) A. M. Michelson, A. R. Todd, *J. Chem. Soc.* **1955**, 2632-2638.

102) R. H. Hall, A. Todd, R. F. Webb, *J. Chem. Soc.* **1957**, 3291-3296.

103) R. L. Letsinger, J. L. Finnan, G. A. Heavner, W. B. Carensford, *J. Am. Chem. Soc.* **1975**, *97*, 3278–3279.

104) S. L. Beaucage, M. H. Caruthers, *Tetrahedron Lett.* **1981**, *22*, 1859-1862.

105) M. D. Matteucci, M. H. Caruthers, *J. Am. Chem. Soc.* **1981**, *103*, 3185-3191.

106) M. H. Caruthers, *J. Biol. Chem.* **2013**, *288*, 1420-1427.

107) J. P. Genet, M. Savignac, *J. Organomet. Chem.* **1999**, *576*, 305-317.

108) K. H. Shaughnessy, *Chem. Rev.* **2009**, *109*, 643-710.

109) M. Yoshikawa, T. Kato, T. Takenishi, *Tetrahedron Lett.* **1967**, *8*, 5065-5068.

110) M.Yoshikawa, T. Kato, T. Takenishi, *Bull. Chem. Soc. Jpn.* **1969**, *42*, 3505-3508

111) J. Ludwig, F. Eckstein, *J. Org. Chem.* **1989**, *54*, 631–635.

112) K. B. Mullis, *Sci. Am*. **1990**, *262*, 56-61.

113) R. J. Roberts, K. Murray, *Crit. Rev. Biochem. Mol. Biol.* **1976***, 4,* 123–164.

114) R. J. Roberts, M. Belfort, T. Bestor, A. S. Bhagwat, T. A. Bickle, J. Bitinaite, R. M. Blumenthal, S. K. Degtyarev, D. T. Dryden, K. Dybvig, K. Firman, *Nucleic Acids Res.* **2003**, *31*, 1805–1812.

115) A. Pingoud, G. G. Wilson, W. Wende, *Nucleic Acids Res.* **2014**, *42*, 7489- 7527.

116) J. Jiricny, S. G. Wood, D. Martin, A. Ubasawa, *Nucleic Acids Res.* **1986**, *14*, 6579–6590.

117) H. Komatsu, S. G. Kim, I. Sakabe, T. Ichikawa, M. Nakai, H. Takaku, *Bioorg. Med. Chem. Lett.* **1992**, *2*, 565–570.

118) M. A. Marchionni, D. J. Roufa, *J. Biol. Chem.* **1978**, *253*, 9075–9081.

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

120) H. Macíčková-Cahová, M. Hocek, *Nucleic Acids Res.* **2009**, *37*, 7612–7622.

121) H. Macíčková-Cahová, R. Pohl, M. Hocek, *ChemBioChem* **2011**, *12*, 431–438.

122) M. Mačková, R. Pohl, M. Hocek, *ChemBioChem* **2014**, *15*, 2306–2312.

123) M. Mačková, S. Boháčová, P. Perlíková, L. Poštová Slavětínská, M. Hocek, *ChemBioChem* **2015**, *16*, 2225–2236.

124) P. Kielkowski, H. Macíčková-Cahová, R. Pohl, M. Hocek, *Angew. Chem. Int. Ed.* **2011**, *50*, 8727–8730.

125) Z. Vaníková, M. Hocek, *Angew. Chem. Int. Ed.* **2014**, *53*, 6734-6737.

126) S. Boháčová, Z. Vaníková, L. P. Slavětínská, M. Hocek, *Org. Biomol. Chem*. **2018**, *16*, 5427-5432.

127) S. Boháčová, L. Ludvíková, L. P. Slavětínská, Z. Vaníková, P. Klán, M. Hocek, *Org. Biomol. Chem.* **2018**, *16*, 1527-1535

128) M. Slavičková, M. Janoušková, A. Simonová, H. Cahová, M. Kambová, H. Šanderová, L. Krásný, M. Hocek, *Chem. Eur. J.* **2018**, *24*, 8311-8314

129) J. Jakubovska, D. Tauraitė, R. Meškys, *ChemBioChem* **2019**, *20*, 2504–2512.

130) D. Voet, J. G. Voet, *Biochemistry*, John Wiley and Sons, Inc., **2004.**

131) V. Raindlová, M. Janoušková, M. Slavíčková, P. Perlíková, S. Boháčová, N. Milisavljevič, H. Šanderová, M. Benda, I. Barvik, L. Krásný, M. Hocek, *Nucleic Acids Res.* **2016**, *44*, 3000-3012.

132) Z. Vaníková, M. Janoušková, M. Kambová, L. Krásný, M. Hocek, *Chem. Sci.* **2019**, *10*, 3937-3942.

133) J. Turner, N. Meeuwenoord, A. Rood, P. Borst, G. van der Marel, J. van Boom, *Eur. J. Org. Chem.* **2003**, *25*, 3832–3839.

134) R. K. Grover, S. J. K. Pond, Q. Cui, P. Subramaniam, D. A. Case, D. P. Millar, P. Wentworth, *Angew. Chem. Int. Ed.* **2007**, *46*, 2839–2843.

135) M. de Kort, P. de Visser, J. Kurzeck, N. Meeuwenoord, G. van der Marel, W. Rüger, J. van Boom, *Eur. J. Org. Chem.* **2001**, 2075–2082.

136) B. Steigenberger, S. Schiesser, B. Hackner, C. Brandmayr, S. K. Laube, J. Steinbacher, T. Pfaffeneder, T. Carell, *Org. Lett.* **2013**, *15*, 366– 369.

137) Q. Sun, J. Sun, S. S. Gong, C. J. Wang, S. Z. Pu, F. D. Feng, *RSC Adv.* **2014**, *4*, 36036– 36039

138) L. C. Sowers, G. P. Beardsley, *J. Org. Chem.* **1993**, *58*, 1664– 1665.

139) M. Münzel, D. Globisch, C. Trindler, T. Carell, *Org. Lett.* **2010**, *12*, 5671– 5673

140) Q. Dai, C. X. Song, T. Pan, C. He, *J. Org. Chem.* **2011**, *76*, 4182– 4188

141) M. Münzel, U. Lischke, D. Stathis, T. Pfaffeneder, F. A. Gnerlich, C. A. Deiml, S. C. Koch, K. Karaghiosoff, T. Carell, *Chem. Eur. J.* **2011**, *17*, 13782– 13788

142) A. S. Hansen, A. Thalhammer, A. H. El-Sagheer, T. Brown, C. J. Schofield, *Bioorg. Med. Chem. Lett.* **2011**, *21*, 1181– 1184

143) R. Bertolini, J. Hunziker, *Helv. Chim. Acta.* **2000**, *83*, 1962– 1976

144) V. A. Litosh, W. Wu, B. P. Stupi, J. Wang, S. E. Morris, M. N. Hersh, M. L. Metzker, *Nucleic Acids Res*. **2011**, *39*, e39.

145) B. P. Stupi, H. Li, J. Wang, W. Wu, S. E. Morris, V. A. Litosh, J. Muniz, M. N. Hersh , M. L. Metzker, *Angew. Chem. Int. Ed.* **2012**, *51*, 1724-1727.

146) B. C. Froehler, M. D. Matteucci, *Nucleic Acids Res.* **1983**, *11*, 8031– 8036.

147) Y. Zuo, T. A. Steitz, *Mol. Cell*. **2015**, *58*, 534– 540.

148) L. Sojka, T. Kouba, I. Barvík, H. Šanderová, Z. Maderová, J. Jonák, L. Krásný, *Nucleic Acids Res.* **2011**, *39*, 4598–4611

149) https://www.thermofisher.com/order/catalog/product/EO0831

150) https://international.neb.com/protocols/2018/01/16/protocol-for-m0649