https://doi.org/10.15388/vu.thesis.403
https://orcid.org/0000-0003-0474-5184

VILNIUS UNIVERSITY

Justina Medžiūnė

The synthesis and applications of oligonucleotide-modified nucleotides

DOCTORAL DISSERTATION

Natural Sciences,
Chemistry (N 003)

VILNIUS 2022

This dissertation was written between 2017 and 2021 at the Faculty of Chemistry and Geosciences, Vilnius University and the Research and Development department at UAB Thermo Fisher Scientific Baltics. The research was supported by UAB Thermo Fisher Scientific Baltics and Vilnius University.

Academic supervisor:
Prof. Dr. Inga Čikotienė (Vilnius University, Natural Sciences, Chemistry, N 003).

This doctoral dissertation will be defended in a public meeting of the Dissertation Defence Panel:
Chairman – **Prof. Dr. Edvinas Orentas** (Vilnius University, Natural Sciences, Chemistry, N 003).
Members:
Dr. Jonas Bucevičius (Max Planck Institute for Multidisciplinary Sciences (Germany), Natural Sciences, Chemistry, N 003).
Doc. Dr. Jelena Dodonova-Vaitkūnienė (Vilnius University, Natural Sciences, Chemistry, N 003).
Prof. Dr. Daumantas Matulis (Vilnius University, Natural Sciences, Biochemistry, N 004).
Prof. Dr. Rolandas Meškys (Vilnius University, Natural Sciences, Biochemistry, N 004).

The dissertation shall be defended at a public meeting of the Dissertation Defence Panel at 2.00 pm on 2nd December 2022 in meeting room A101 of the Center for Physical Sciences and Technology.
Address: Saulėtekio av. 3, LT-10257 Vilnius, Lithuania
Tel. +370 5 264 9211; e-mail office@ftmc.lt

The text of this dissertation can be accessed at the libraries of Center for Physical Sciences and Technology of Vilnius University, as well as on the website of Vilnius University: www.vu.lt/lt/naujienos/ivykiu-kalendorius

https://doi.org/10.15388/vu.thesis.403
https://orcid.org/0000-0003-0474-5184

VILNIAUS UNIVERSITETAS

Justina Medžiūnė

Oligonukleotidais modifikuotų nukleotidų sintezė ir panaudojimas

DAKTARO DISERTACIJA

Gamtos mokslai,
Chemija (N 003)

VILNIUS 2022

Disertacija rengta 2017– 2021 metais Vilniaus universiteto chemijos ir geomokslų fakultete ir UAB „Thermo Fisher Scientific Baltics“ mokslinių tyrimų ir eksperimentinės plėtros centre.
Mokslinius tyrimus rėmė UAB „Thermo Fisher Scientific Baltics“ ir Vilniaus universitetas.

Mokslinis vadovas:
prof. dr. Inga Čikotienė (Vilniaus universitetas, gamtos mokslai, chemija, N 003).

Gynimo taryba:
Pirmininkas – **prof. dr. Edvinas Orentas** (Vilniaus universitetas, gamtos mokslai, chemija, N 003).
Nariai:
dr. Jonas Bucevičius (Max Planck Institute for Multidisciplinary Sciences (Vokietija), gamtos mokslai, chemija, N 003).
doc. dr. Jelena Dodonova-Vaitkūnienė (Vilniaus universitetas, gamtos mokslai, chemija, N 003).
prof. dr. Daumantas Matulis (Vilniaus universitetas, gamtos mokslai, biochemija, N 004).
prof. dr. Rolandas Meškys (Vilniaus universitetas, gamtos mokslai, biochemija, N 004).

Disertacija ginama viešame Gynimo tarybos posėdyje 2022 m. gruodžio mėn. 2 d. 14.00 val. Fizinių ir technologijos mokslų centro A101 auditorijoje. Adresas: Saulėtekio al. 3, LT-10257 Vilnius, Lietuva
Tel. +370 5 264 9211, el. paštas office@ftmc.lt

Disertaciją galima peržiūrėti Vilniaus Universiteto, Fizinių ir technologijos centro bibliotekoje ir VU interneto svetainėje adresu:
https://www.vu.lt/naujienos/ivykiu-kalendorius

TABLE OF CONTENT

ABBREVIATIONS

ACN	acetonitrile
AMP	adenosine 5’-monophosphate
ATP	adenosine 5’-triphosphate
B	nucleobase
Cap	7-methylguanosine (m7G) connected by a triphosphate bridge to the first nucleotide of mRNA
CDI	1,1'-carbonyldiimidazole
cDNA	complementary DNA
CuAAC	copper-catalyzed alkyne-azide 1,3-dipolar cycloaddition
Cy3	cyanine dye Cy3
Cy5	cyanine dye Cy5
Cys	cysteine
dAMP	2’-deoxyadenosine 5’-triphosphate
DCC	*N*,*N*'-dicyclohexylcarbodiimide
DCI	4,5-dicyanoimidazole
ddN	2’,3’-dideoxynucleoside
dd^{I}N	5-iodo- or 7-deaza-7-iodo-2’,3’-dideoxynucleoside
dd^{I}NTP	5-iodo- or 7-deaza-7-iodo-2’,3’-dideoxynucleoside 5’-triphosphate
dd^{PA}NTP	5- or 7-deaza-7-(3-aminoprop-1-ynyl)- 2’,3’-dideoxynucleoside 5’-triphosphate
dd^{N3}NTP	5- or 7-deaza-7-(3-(azido-substituted)prop-1-ynyl)-2’,3’-dideoxynucleoside 5’-triphosphate
ddNTP	2’,3’-dideoxynucleoside 5’-triphosphate
dN	2’-deoxynucleoside
d^{I}N	5-iodo- or 7-deaza-7-iodo-2’-deoxynucleoside
d^{I}NTP	5-iodo- or 7-deaza-7-iodo-2’-deoxynucleoside 5’-triphosphate
d^{PA}NTP	5- or 7-deaza-7-(3-aminoprop-1-ynyl)- 2’-deoxynucleoside 5’-triphosphate
d^{N3}NTP	5- or 7-deaza-7-(3-(azido-substituted)prop-1-ynyl)-2’-deoxynucleoside 5’-triphosphate
DIPEA	*N*,*N*-diisopropylethylamine
DMF	*N*,*N*-dimethylformamide
DMSO	dimethyl sulfoxide
dNDP	2’-deoxynucleoside 5’-diphosphate
dNMP	2’-deoxynucleoside 5’-monophosphate
dNTP	2’-deoxynucleoside 5’-triphosphate
DNA	deoxyribonucleic acid
EtOH	ethanol
Et_3N	triethylamine

ETT	5-ethylthio-1*H*-tetrazole
FDA	United States Food and Drug Administration
FISH	fluorescence *in-situ* hybridization
His	histidine
HPLC	high performance liquid chromatography
IDP	inosine 5'-diphosphate
LC-MS	liquid chromatography – mass spectrometry
MeOH	methanol
mRNA	messenger ribonucleic acid
MstCl	mesitylenesulfonyl chloride
NDP	nucleoside 5'-diphosphate
NGS	next generation sequencing
NHS	*N*-hydroxysuccinimide
NMP	nucleoside 5'-monophosphate
NTP	nucleoside 5'-triphosphate
nt	nucleotides
ON	oligonucleotide
OTDN or $d^{ON}NTP$	oligonucleotide-tethered 2'-deoxynucleoside 5'-triphosphate
OTDDN or $dd^{ON}NTP$	oligonucleotide-tethered 2',3'-dideoxynucleoside 5'-triphosphate
PAGE	polyacrylamide gel electrophoresis
PEG	polyethylene glycol
PCR	polymerase chain reaction
PEX	primer extension
Py	pyridine
RNA	ribonucleic acid
RT	reverse transcriptase or reverse transcription
SAM	S-adenosyl-L-methionine
SELEX	systematic evolution of ligands by exponential enrichment
siRNA	small interfering RNA
SPAAC	strain-promoted alkyne-azide cycloaddition
TAMRA	carboxytetramethylrhodamine
TBA	tributylamine
TBAPP	tri-*n*-butylammonium pyrophosphate
TBP	tri-*n*-butylammonium phosphate
TBTA	tris((1-benzyl-4-triazolyl)methyl)amine
t-BuOH	*tert*-butanol
TEMED	tetramethylethylenediamine
TEP	triethyl phosphate
TdT	terminal deoxynucleotidyl transferase
TEAAc	triethylammonium acetate
THF	tetrahydrofuran
THPTA	tris(3-hydroxypropyltriazolylmethyl)amine

TMP	trimethyl phosphate
TPPMS	sodium 3-(diphenylphosphino)benzenesulfonate
TPPTS	trisodium 3,3′,3″-phosphanetriyltri(benzene-1-sulfonate
TXPTS	tris(2,4-dimethyl-5-sulfophenyl)phosphine trisodium salt
UMI	unique molecular identifier

INTRODUCTION

Natural nucleotides are the building blocks of nucleic acids that are vital for all living organisms. The exploration of living organisms has revealed a wide variety of naturally occurring nucleotide modifications and their important role in bioprocesses regulation, relation to human diseases, etc. [1] The evolution of diagnostics, bioprocess monitoring, and nucleic acid therapeutics induced an increasing demand for versatile nucleic acid modification and labelling. Modified nucleotides play an essential role in today's cutting-edge molecular biology applications such as fluorescence in-situ hybridization (FISH) [2], systematic evolution of ligands by exponential enrichment (SELEX) [3], next-generation sequencing (NGS) [4], and many other techniques as well as in the development of pharmaceuticals [5] and RNA vaccines [6].

While 2'-deoxynucleoside 5'-triphosphates (dNTPs) are the essence in many diagnostics and bioprocesses monitoring, less attention is paid to 2',3'-dideoxynucleoside 5'-triphosphates (ddNTPs), which are poorly explored. Initially, ddNTP analogues were investigated as promising antiviral [7] or anticancer drugs [8, 5, 9], and also further developed to enzymatically label nucleic acids for various array-based assays [10]. The revolutionary breakthrough in their applicability was the emergence of the Sanger sequencing methodology, where dye-labelled ddNTPs were used as DNA replication terminators. [11] This technology enabled the initial sequencing of the human genome. The ever-growing demand for inexpensive, rapid, and accurate exploration of genomes called for the refinement of existing sequencing techniques. The development of next-generation sequencing was a revolutionary milestone in genome analysis and became an everyday analysis technique in modern biology and biomedical sciences.

Regardless of the sequencing platform used, the nucleic acids have to be prepared by means of constructing the sequenceable libraries. Such pre-processing of nucleic acids is a multistep procedure that gained attempts to improve by scientific communities in recent years. While modified nucleotides already were inherent tools in sequencing and imaging, further modification of nucleotides enabled the expansion into even more diverse applications. The utility of 3'-azido labelled nucleotide during enzymatic incorporation followed by CuAAC click reaction with alkyne-oligonucleotide (resulted in triazole formation) developed into the chemical ligation concept and approach that was applied to simplify NGS library preparation. [12, 13] Generally, modified nucleic acids are designed to mimic natural ones with enhanced specific functionalities (e.g., resistance to nucleases, melting point

increase etc.). A triazole moiety was considered as the phosphate mimicking group. [14] The developed chemical ligation alternative proved to be feasible and advantageous for simplification of NGS library preparation. However, the replication through the unnatural backbone was not in high-fidelity since deletions were observed as well as copper-mediated degradation of DNA, low click reaction and read-through efficiencies. [15, 16] Nonetheless, the technological feasibility was demonstrated opening the perspectives for further development to improve conjugation nature and provide accuracy for read-through.

With the growing diversity of application empowered by modified nucleotides and nucleic acids, the necessity for efficient, simple, and universal synthesis methodologies increases alongside. To enable easy addition of the desired labels, functional groups such as amino and azido are usually used in the synthesis of modified dNTPs and ddNTPs. To retain the substrate properties of modified nucleotides, the design of the linker, which connects a nucleobase and a label is of high importance. [17] The commonly used synthetic strategy for functionalized dNTPs and ddNTPs consists of at least five chemical steps, resulting in low overall yields. [11] While modified dNTPs have gained more attention in attempts to propose alternative synthesis approaches [18], ddNTPs lack the scientific input regarding efficient and scalable preparation methodologies.

Aim and tasks

The **aim** of this work was to design and synthesize oligonucleotide-tethered 2'-deoxy- and 2',3'-dideoxynucleoside 5'-triphosphates (OTDNs and OTDDNs, respectively), that would serve as substrates for polymerases and exhibit unnatural backbone read-through capability.

The following **tasks** were defined to achieve the aim:

1. To develop simplified, alternative synthesis approach for azido-group bearing 2'-deoxynucleotides, 2',3'-dideoxynucleotides and to compare it with conventional method.
2. To adjust the reaction conditions for scale-up synthesis of azido-group bearing 2',3'-dideoxynucleotides.
3. To evaluate azido-group bearing 2'-deoxynucleotides and 2',3'-dideoxynucleotides substrate properties during enzymatic incorporation.
4. To study the efficiency of CuAAC and SPAAC reactions in relation to the nature of substrates.

5. To synthesize oligonucleotide-tethered 2’-deoxynucleotides, 2’,3’-dideoxynucleotides and investigate their properties in both enzymatic incorporation and read-through unnatural linkage assays.

6. To employ oligonucleotide-tethered 2’,3’-dideoxynucleotides for diverse applications in attempts to simplify NGS library preparation.

7. To determine the efficiency of read-through the unnatural linkages.

In terms of the significance of the work, it is stated that:

- The azido-group bearing 2’-deoxy- and 2’,3’-dideoxynucleotides are substrates for diverse DNA polymerases.
- The empirical design of oligonucleotide-tethered 2’-deoxynucleotides met both criteria: substrates properties during enzymatic incorporation into growing DNA strand and template properties for the following synthesis of a complementary strand.
- The efficient and less time-consuming synthesis strategy for modified 2’,3’-dideoxynucleotides has been developed, providing several functionalized 2’,3’-dideoxynucleotides that can be either prior or post-enzymatically labelled.
- The oligonucleotide-tethered 2’,3’-dideoxynucleotides are substrates for numerous DNA polymerases during enzymatic incorporation.
- Several empirically designed distinctive linkers of oligonucleotide-tethered 2’,3’-dideoxynucleotide were proven to be biocompatible, and some polymerases are capable to read through it during synthesis of a complementary strand with single-cycle extension efficiency of 75%.
- Developed oligonucleotide-tethered 2’,3’-dideoxynucleotide terminators provide nucleic acids labelling technique that enables to surpass fragmentation and specific adapters ligation in one enzymatic step, consequently simplifying next generation sequencing (NGS) library preparation workflow.
- The technology applicability in NGS was demonstrated, by performing a semi-targeted sequencing experiment of M13mp18 viral genome. The resulting sequencing data indicated capture of intended genomic loci with a characteristic insert structure.
- The practical perspective of developed technology was proven to be useful for both DNA and cDNA sequencing by analysis of

microbial communities, determination of unknown gene fusion events in cancer, and quantification of gene expression.

Statements to be defended

1. The azido-labelled 2'-deoxynucleotides can be obtained via a traditional synthesis strategy and an alternative simplified one.
2. The three steps, straightforward synthetic strategy can be applied to the synthesis of 2',3'-dideoxynucleotides bearing azido-group with good efficiency and yields.
3. The azido-group bearing 2'-deoxy- and 2',3'-dideoxynucleotides are substrates for diverse DNA polymerases during primer extension and/or polymerase chain reactions.
4. The CuAAC and SPAAC reactions rates depend on the nature of azide-group bearing substrates.
5. Oligonucleotide-tethered 2'-deoxy- and 2',3'-dideoxynucleotides are substrates for DNA polymerases.
6. The empirically designed linkers connecting nucleobase-to-oligonucleotide are biocompatible and can serve as templates for enzymatic synthesis of a complementary strand.
7. Nucleic acid termination by oligonucleotide-tethered 2',3'-dideoxynucleotides simplifies NGS library preparation workflow by integrating fragmentation and adapter addition in a single enzymatic step and is applicable for DNA and cDNA sequencing.

LITERATURE OVERVIEW

1. NUCLEOSIDES PHOSPHORYLATION REACTIONS

Nucleosides and their 5'-phosphates are essential molecules in biological systems, that play both functional and structural roles. Nucleoside 5'-triphosphates (NTPs) and dNTPs are important biological building blocks for polymerase-mediated enzymatic synthesis of nucleic acids (DNA, RNA) and genetic material of the living organisms. Apart from this significant role, nucleotides and their different conjugates are core molecules in a variety of enzymatic processes. For example, the first nucleoside 5'-triphosphate was isolated in 1929 from the muscle and liver extracts. [19, 20] The defined structure was attributed to adenosine 5'-triphosphate (ATP) which is now known to have an allosteric effect during the regulation of cell metabolism, as well as is a primary energy source for a variety of cellular functions (from metabolic transformations to signaling events). Such nucleotide conjugates as 3',5'-cyclic mononucleotides, cyclic dinucleotides, dinucleoside 5',5'-polyphosphates, nicotinamide mononucleotides, flavin mononucleotides, sugar nucleotides, nucleolipids and many other analogues were investigated for their functional importance in humans. [21] Even though this is just a part of functional roles of nucleotides, it is obvious that exploration of these bioprocesses inevitably includes the use of (d)NTPs and their conjugates.

For a deeper exploration of biochemical systems and processes, a variety of structural analogues of either nucleosides or nucleotides have been synthesized. Nucleoside analogues and their phosphorylated prodrugs are widely investigated for their antiviral [7] and antitumor [22] properties. [5, 23] Few more emerging fields exploring therapeutic properties of modified nucleic acids are RNA vaccines, antisense oligonucleotides (ASO) and small interfering RNA (siRNA). ASO field has created and explored a huge variety of nucleobase, ribose ring, phosphate linkage modified analogues. Moreover, the investigated mechanisms of action were manifold targeting even genetic diseases. [24, 25] Several therapeutics have been approved by FDA and many more are in clinical trials. Meanwhile, siRNA technology explored over 20 years, in 2018 resulted in the first approved drug – patisiran for hereditary transthyretin amyloidosis treatment, followed by two more (givosiran, lumasiran) recently authorized. [26] Mentioned therapeutics structure contains several modifications like fluorine, -OMe at the 2'-position of ribose ring and thiophosphate linkages.

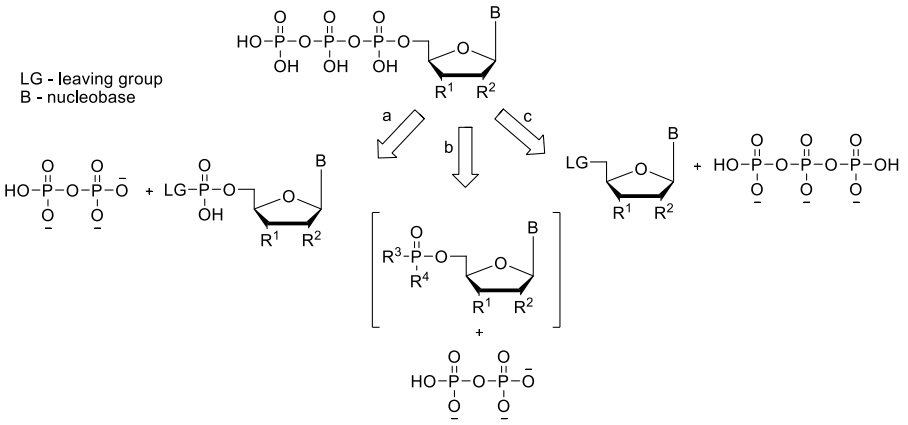
Just as the therapeutic properties of nucleotides are of great interest, diagnostics opened up even more diverse application. FFISH technology

utilizes fluorophore-labelled nucleotides for the synthesis of DNA hybridization probes. Fluorescent microscope detection enables to determine the presence or absence of a complementary sequence. Nowadays this technique enables prenatal diagnosis of chromosomal abnormalities, chromosomal aberrations in adults, detection of cancer and infectious diseases. [27] The technological advances in DNA microarrays enabled the miniaturization of DNA probe-based detections for gene expression profiling and genotyping. [28]

Since 1977 when Frederic Sanger established the first sequencing technique, the revolution of modified chain terminators has begun. Although Sanger sequencing was outrivaled by more efficient next-generation sequencing techniques, the chain terminators reached manifold structural diversity. To date, the vast majority of sequencing techniques require modified nucleotides. [4, 29] Represented technologies are just a small segment of areas where nucleotides and their conjugates play an essential role. Moreover, a great part of applications, based on enzymatic incorporation of modified nucleotides, require their 5'-triphosphate analogues. Consequently, efficient, rapid, and simple synthetic methodologies for triphosphorylation are of crucial importance. Different applications require functionally and chemically diverse modifications, labels, and tags (e.g., functional groups, fluorescent, spin, redox labels, lipids, enzymes, sugars [30]). However, to date there is no universal and at the same time efficient protocol for synthesis of various labeled nucleotides. Some methods that are suitable for natural nucleotides may not be sufficient for modified ones. Differences are even observed between phosphorylation of dNTPs and NTPs, since the impurity amount can increase significantly depending on the phosphorylation method used. This is in agreement with the literature data representing the large scope of developed phosphorylation agents, strategies, and methods. [31, 32, 21] Even though the introduction of orthogonal protecting groups to control regioselectivity solves part of the problems, currently the main objective is a less time-consuming, minimal steps requiring and more efficient methodology.

There are three main strategies for nucleoside 5'-triphosphates synthesis:

- synthesis of activated monophosphate and its reaction with nucleophilic pyrophosphate (Scheme 1a); [33, 34]
- direct single step "one-pot" synthesis via formation of active monophosphate intermediate followed by reaction with pyrophosphate (Scheme 1b); [35]
- formation of a 5'-O-leaving group for direct displacement by triphosphate nucleophiles (Scheme 1c). [36]



Scheme 1. Nucleoside 5‘-triphosphate synthesis strategies.

During the years other synthetic strategies have been developed as well. Examples are the nucleophilic attack of nucleoside diphosphate on the activated monophosphate or vice versa, and the activated cyclic triphosphate or trimetaphosphate reaction with nucleoside. However, these synthesis strategies are less employed and investigated to a lower extent. Several reviews covering different synthesis methodologies have been reported. [21, 31, 32] In most cases scientists note several limitations of nucleosides phosphorylation:

- synthesis is usually a multi-step procedure;
- no universal methodology;
- lack of regioselectivity regarding ribose ring OH groups;
- demanding reaction conditions;
- moisture-sensitive reagents and reactions;
- some active phosphorylation agents require protection of both nucleobases functional groups and ribose OH groups;
- only low to moderate yields;
- complex purification: formation of both inorganic and organic byproducts;
- formation of byproducts of similar nature (interfering with products use);
- limited diversity of reaction solvents;
- poor solubility of modified nucleosides in reaction media.

Generally, at least one of the above-mentioned limitations becomes the factor for the optimization of phosphorylation reaction conditions for nucleosides of choice.

1.1 Nucleotides synthesis: one-pot synthesis strategy

1.1.1 Nucleotides synthesis via formation of phosphorodichloridates

In most synthetic strategies, the objective and limiting step is monophosphorylation. Many different phosphorylation agents have been developed to increase reactivity and regioselectivity applying both P(III) and P(V) derivatives (Figure 1). Usually using P(V) agents the regioselectivity between ribose ring OH groups is higher in comparison with P(III). However, in some cases, the yields may be very low. The reaction rate can be enhanced by applying more reactive P(III) derivatives, however, both ribose ring OH groups and nucleobases reactive groups must be protected. [21]

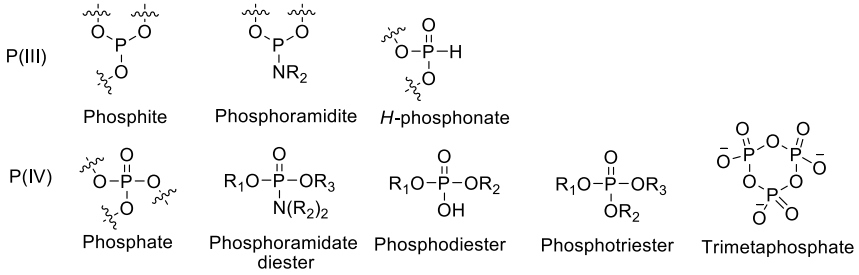
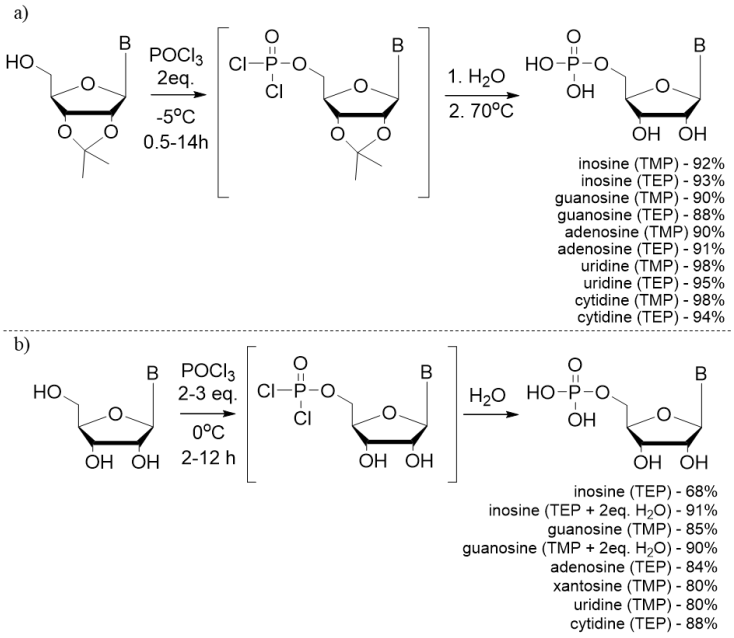


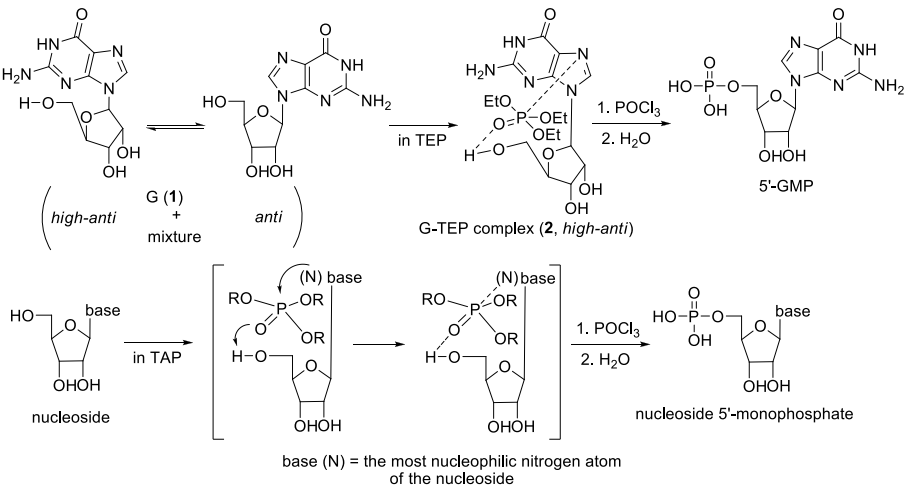
Figure 1. Nomenclature of several different P(III) and P(V) groups, derivatives and general structures of intermediates formed during phosphorylation reaction.

To date, the most widely used phosphorylation methodology relies on the formation of a phosphorodichloridate intermediate by using $POCl_3$ as a phosphorylation agent. In 1967 Yoshikawa and co-workers used $POCl_3$ for monophosphorylation of 2', 3'-*O*-isopropylidene nucleosides. The reactions were performed at -5 ^{o}C with the excess of phosphorus oxychloride in trialkyl phosphates media. After aqueous hydrolysis the products were obtained in 88 – 98% yield (Scheme 2a). Likewise, unprotected ribonucleosides were also phosphorylated in good yields (68 – 91%), though, formation of 2'- (or 3'-), 5'-diphosphates was observed. [37] However, with the help of advancing analytical techniques it was shown that excellent regioselectivity and yield was overestimated and impurity profile was more complex. Moreover, subsequent work towards optimization of reaction conditions showed that the addition of base, usually 1,8-bis(dimethylamino)naphtalene known as proton sponge, increases regioselectivity and reduces glycosidic bond decomposition. This destruction was found to be caused by highly acidic media. [38, 39]



Scheme 2. Synthesis of nucleoside 5‘-monophosphates according to Yoshikawa methodology.

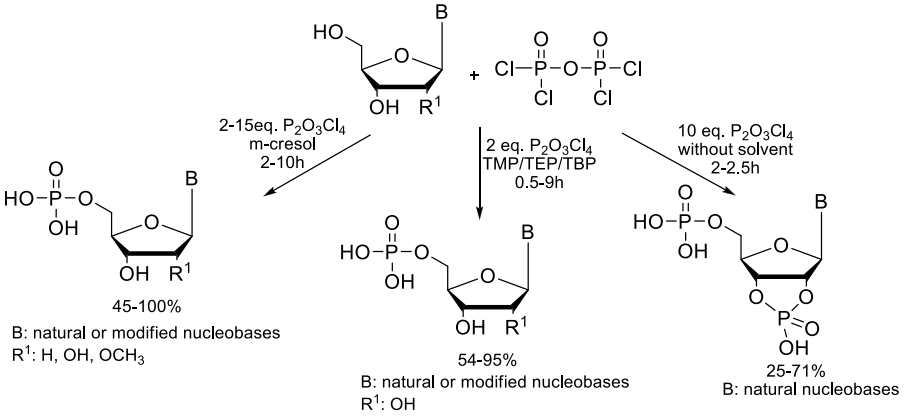
The other important finding Yoshikawa made was the utility of trialkyl phosphates (TMP and TEP) as reaction solvents. In 1995 Ikemoto and colleagues investigated the mechanism of phosphorylation reaction in triethyl phosphate (TEP) (Scheme 3). After testing different reaction conditions and



Scheme 3. Phosphorylation of nucleosides in trialkylphosphate.

employing NMR and X-ray analysis techniques, they proposed a reaction mechanism where an intermediate of TEP and nucleoside is formed (Scheme 3, G-TEP). [40] The utility of trialkyl phosphates resulted in increased regioselectivity during the reaction, which was further explained by probable oxygen atom activation of the 5'-hydroxy group upon the formation of G-TEP intermediate (Scheme 3). The effect of alkyl group nature on the phosphorylation efficiency was investigated. Trimethyl phosphate (TMP) and TEP performed efficiently (90% yield), however, tri-*n*-butyl phosphates reduced the intermediate complex formation, due to steric hindrance induced by a bulky alkyl group and resulted in poor yield (17%). [40, 41]

In the late 1960s and 1970s the pyrophosphoryl chloride ($P_2O_3Cl_4$) was used for ribo-, 2'-deoxy-, 2'-OMe-, and nucleobase-modified nucleosides phosphorylation. Reactions were performed in *m*-cresol and 2 – 15 eq. of $P_2O_3Cl_4$ were necessary to obtain natural and modified nucleoside 5'-monophosphates. The resulted yields ranged from 45% to quantitative, although long reaction times were required (Scheme 4a). For ribonucleotides synthesis, different solvents such as acetonitrile, benzonitrile, ethyl acetate, methyl acetate, ethyl benzoate, nitrobenzene were tested. In all cases products were obtained in good yields (71 – 88%) without any distinctive side products formation. [42]

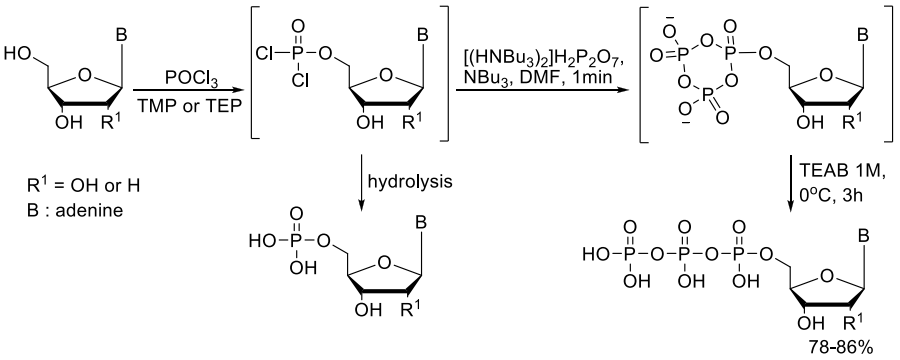


Scheme 4. Nucleosides phosphorylation applying pyrophosphoryl chloride.

Similar results were obtained for the synthesis of ribonucleotides using 2 eq. of pyrophosphoryl chloride in trialkyl phosphates (Scheme 4b). [41] $P_2O_3Cl_4$ was also adopted for the synthesis of 2',3'-cyclic phosphate 5'-phosphates. The synthesis of mentioned compounds required 10 eq. of the phosphorylation agent and 2 – 2.5 hours for reaction to proceed (Scheme 4c). The yield was moderate, while the only side-product observed was 2'(3')-

phosphate 5'-polyphosphate. [43] Later on, pyrophosphoryl chloride was employed for the preparation of modified diphosphates as 2'-diflouromethyluridine 3',5'-bisphosphate [44] and 2-azido-2'-deoxyadenosine 3',5'-bisphosphate [45] in either *m*-cresol or without solvent. In both cases products were obtained in moderate yields (50%). Even though pyrophosphoryl chloride proved to be efficient phosphorylation agent, it did not receive much interest during the time.

In 1981 Liudwig improved Yoshikawa's phosphorylation methodology to obtain corresponding 5'-triphosphates. Liudwig applied Yoshikawa's method for AMP and dAMP formation. However, instead of performing hydrolysis of formed phosphorodichloridate intermediate, he used bis(tri-*n*-butylammonium) pyrophosphate (TBAPP) as nucleophile in order to obtain the cyclic trimetaphosphate intermediate that would be further hydrolyzed to linear triphosphates (ATP and dATP). The synthesis was successful and resulted in 78 and 86% yield of the desired products (Scheme 5). It was the first straightforward, three-step, one-pot synthesis strategy for the preparation of nucleoside 5'-triphosphates. Due to its simplicity, to this date, this is still the most preferable methodology for the preparation of 5'-triphosphates in both academia and industry.

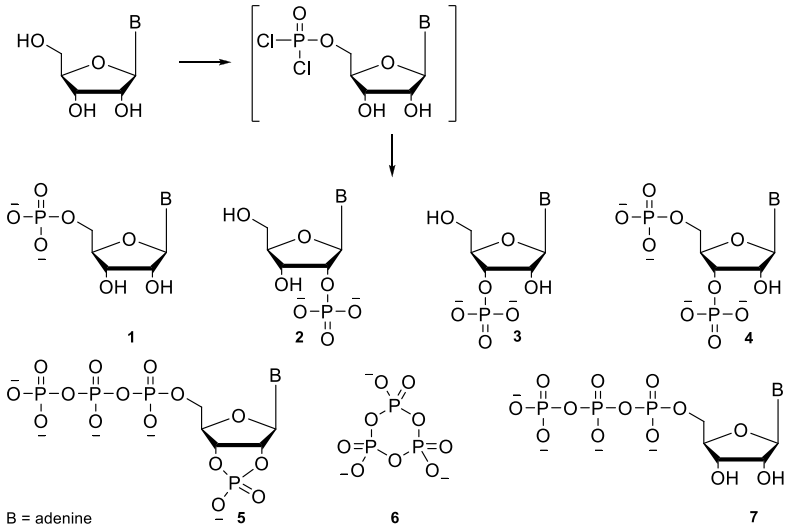


Scheme 5. One-pot synthesis methodology for preparation of nucleoside 5'-triphosphates.

The designed one-pot strategy was extensively used by many scientists for the preparation of diverse nucleotides. For the synthesis of 5-vinyl- and 5-ethynyl-2'-deoxyuridine-5'-triphosphates a slight modification of reaction conditions was necessary, since unsaturated side-chain modification as the vinyl and the ethynyl groups were not stable under acidic conditions. Therefore, several different bases were tested to neutralize the liberated HCl, whilst the most effective one was proton sponge (PS). [46] Likewise, in subsequent research proton sponge was applied to prevent the cleavage of

glycosidic bond, which was noticed during triphosphorylation of modified 7-deaza-2',3'-dideoxynucleosides. [47]

Despite the widespread use of this one-pot methodology for the synthesis of nucleosides 5'-triphosphates, it has some drawbacks and does not apply to all nucleoside derivatives. Reaction times and yields are highly dependent on the starting nucleoside, especially in cases of modified nucleosides phosphorylation. Deeper analysis of Ludwig's methodology revealed formation of more complicated reaction mixture than it was originally assumed. Gillerman and Fischer proceeded ATP synthesis according to direct Ludwig's methodology, they identified several occurring byproducts (Scheme 6 compounds 1 – 6). Compounds 2 – 4 are the result of regioselectivity deficiency during the phosphorodichloridate intermediate formation step. Although compounds 2 and 3 can be separated from the desired triphosphate (7), the branched diphosphate 4 complicates the purification. Byproducts 5 and 6 are hard to separate form triphosphate (7), due to their similar charge and elution via ion exchange chromatography (Scheme 6 compounds 5 – 6). [48]



Scheme 6. Structure of byproducts forming during ATP synthesis.

The formation of impurities 5 and 6 takes place during the second step, when a tri-*n*-butylammonium pyrophosphate solution in DMF is added. The residual amount of $POCl_3$ reacts with DMF producing the Vilsmeier reagent. [49] Whilst the formed intermediate has higher reactivity than $POCl_3$ and, consequently, reacts with 2'- and 3'-positions of nucleotide ribose ring forming a cyclic derivative which gives compound 5 after hydrolysis (Scheme

6). DMF-$POCl_3$ intermediate reacts similarly with pyrophosphate resulting in a chlorinated trimetaphosphate intermediate which gives impurity 6 after hydrolysis (Scheme 6). [48]

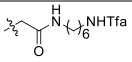
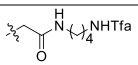
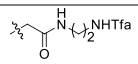
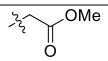
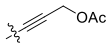
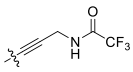
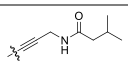
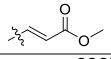
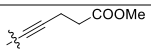
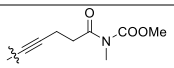
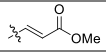
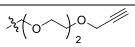
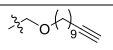
Attempts to improve reaction regioselectivity failed when different reaction solvents (pyridine, dioxane, triethylamine) and pyrophosphate salts (such as tetra-*n*-butylammonium or potassium) were applied. However, the breakthrough was achieved by manipulating the reaction temperature and time of the second step. It was concluded that lower reaction temperature improved regioselectivity towards the 5' position and, consequently, the ATP yield, even though the increased amount of $POCl_3$ was necessary. Both reaction steps were performed at -15 ^oC while treatment with tri-*n*-butylammonium pyrophosphate was expanded to 2 hours instead of 1 minute. In this case, no formation of byproducts 5 and 6 was observed resulting in a nearly quantitative ATP formation, with unreacted monophosphate (AMP) being the main impurity. Optimized conditions were applied to other ribonucleosides, obtained product yields ranged from 51 to 74%. Interestingly, unlike other ribonucleosides, cytidine resulted in the formation of another type of impurity. It was declared that the exocyclic amino group reacted with $POCl_3$ giving additional byproducts such as N^4-monophosphates. Consequently, protection of the amino group was necessary for efficient phosphorylation. [48] These findings are in agreement with previously reported. [50, 51]

The yields of one-pot synthesis in many cases are far from good, especially when modified nucleosides are phosphorylated. Table 1 summarizes just a small part of reported modified nucleoside 5'-triphosphates synthesis applying one-pot strategy. To obtain represented examples the standard methodology was applied using $POCl_3$ as monophosphorylating agent, maintaining reactions in trialkyl phosphate media, followed by the addition of TBAPP solution. The main differences in synthesis methodologies usually involve the change of reaction time and temperature, as well as usage of proton sponge. These summarized data represent the diversity in obtained products yields, that depends not just on the substituent itself but on the nucleobase as well. In the majority of cases, product yields are low and range from 1.1 to 72% (Table 1). This suggests that a convenient one-pot methodology may be acceptable for the small-scale preparation of nucleoside 5'-triphosphates, however, larger-scale synthesis requires either careful optimization or the development of other synthetic strategies.

Since the nucleoside 5'-monophosphates synthesis is usually the critical step, it is important to mention a technological improvement achieved regarding this matter. In 2014 Zhu *et al.* applied *meso*-flow reactor for rapid, scalable, and high yielding nucleoside 5'-monophosphates synthesis. This

technological advantage enabled to obtain various natural and modified NMPs along with dNMPs in 53 – 89% yields. Moreover, in comparison to batch synthesis, the reactions time was reduced to at least half, and so was the $POCl_3$ amount, while yields were increased nearly in all cases. [52]

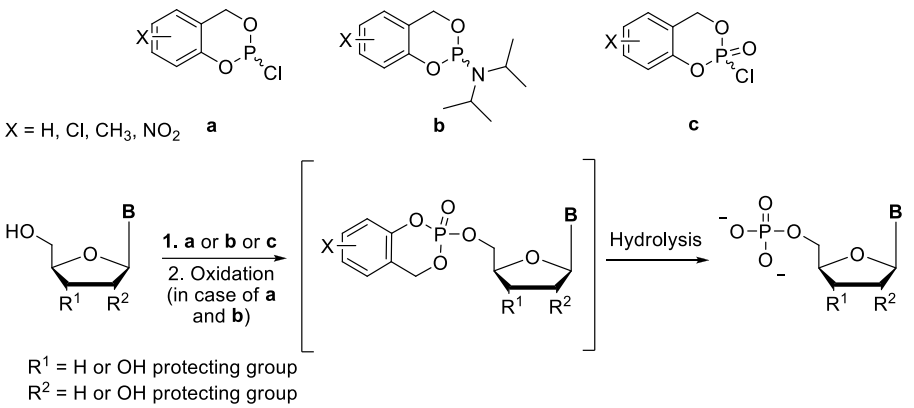
Table 1. The triphosphorylation yields of differently modified nucleotides prepared using the one-pot synthesis methodology.

Entry	Ribose	Substituent	Product	Yield	Ref.
1	2‘-deoxy-		5-substituted-dCTP	9.6%	[53]
2			5-substituted-dCTP 5-substituted-dUTP	5.6% 6.3%	
3			5-substituted-dCTP	6.9%	
4			5-substituted-dCTP 5-substituted-dUTP	65% 41%	
5			5-substituted-dCTP	4.9%	
6			5-substituted-dCTP 5-substituted-dUTP	1.1% 26%	
7	ribo	Br-	8-substituted-GTP	72%	[54]
8	2‘-deoxy-		5-substituted-dCTP 5-substituted-dUTP 7-deaza-7-substituted-dATP	10% 48% 16%	[55]
9			5-substituted-dCTP 5-substituted-dUTP	36% 30%	
10			5-substituted-dUTP	23%	
11			7-deaza-7-substituted-dATP	20%	
12			7-deaza-7-substituted-dATP 8-substituted-dGTP	19% 33%	
13			7-deaza-7-substituted-dGTP	19%	
14	2‘-deoxy-		5-substituted-dUTP	13%	[56]
15		I-	5-substituted-dUTP	38%	
16	2‘-deoxy-	-	7-deaza-dGTP	52%	[57]
17	2‘-deoxy-	I-	5-substituted-dCTP	39%	[58]
18	ribo-	I-	5-substituted-CTP	39%	
19	ribo-		5-substituted-UTP	26%	[59]
20	2‘-deoxy-		5-substituted-dUTP	22%	[60]
21			5-substituted-dUTP	48%	
22	ribo-	Me-	7-deaza-7-substituted-ATP 7-deaza-7-substituted-GTP	30% 62%	[61]
23			5-substituted-CTP	9%	

		5-substituted-UTP	18%
		7-deaza-7-substituted-ATP	8%
		7-deaza-7-substituted-GTP	8%
24		7-deaza-7-substituted-GTP	30%
25		5-substituted-UTP	30%
		7-deaza-7-substituted-GTP	26%
26		7-deaza-7-substituted-GTP	10%

1.2.2 Nucleotides synthesis via formation of P(III) – P(V) intermediates

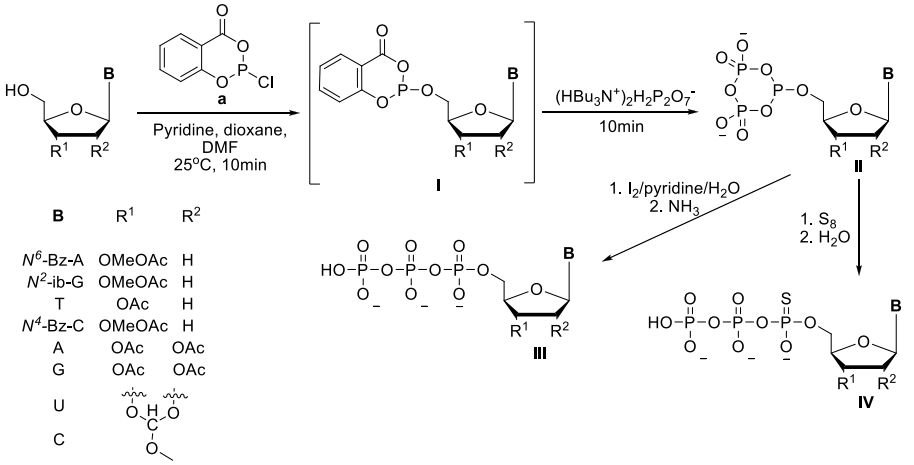
In 1990 Chris Meier used a phosphodiester called *cyclo*Sal (Scheme 7 c) for the preparation of nucleotide prodrugs. [9] Such activated compounds are easily hydrolyzed under basic pH (>7) resulting in corresponding nucleoside monophosphates. At a same time, *cyclo*Sal derivatives (Scheme 7, a, b, and c) were applied for phosphorylation of nucleosides. Depending on the substituent on the aromatic ring the phosphorus atom electrophilicity can be enhanced as well as the reaction rate can be increased. The electron withdrawing substituent as NO_2- or Cl- enhanced reactivity of *cyclo*Sal derivatives, while electron donating groups as CH_3- reduced (Scheme 7). [62] The reactivity of these phosphorylation agents is as well highly dependent on P valence, where P(III) derivatives are more reactive in comparison to P(V). [21]



Scheme 7. *Cyclo*Sal phosphorylation agents.

Ludwig and Eckstein applied 2-chloro-4*H*-1,3,2-benzodioxaphosphorin-4-one (Scheme 8, agent a) as a monophosphorylation agent in the one-pot synthesis of dTTP, α-thio- NTPs and other. Due to the trivalent phosphorus reagent, the protection of ribose ring OH groups was necessary, while

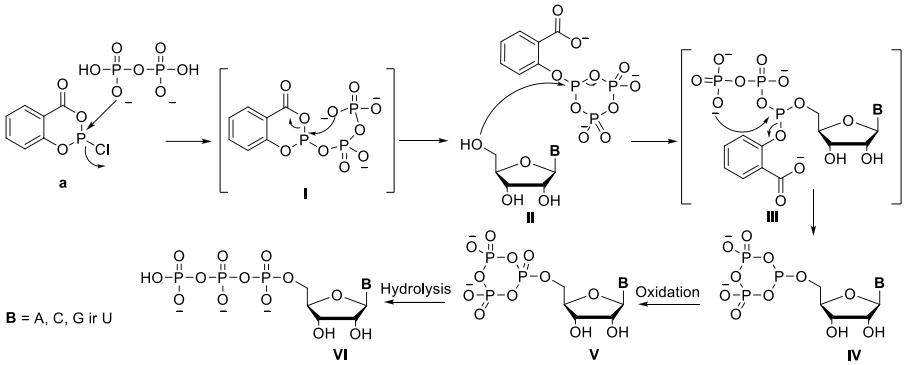
nucleobase protection was not required. The anhydrous reaction conditions were necessary to prevent the formation of H-phosphonates. For the monophosphorylation step, the phosphorylation agent was used either in pyridine, dioxane, DMF or their mixture. Reactions were performed at room temperature for 10 min. The phosphotriester intermediate was formed (I), and it further underwent two consecutive nucleophilic substitutions when pyrophosphate was added. During the first substitution, the carboxylate group acts as the leaving group, while the phenolate residue is substituted during the second intramolecular reaction. The formed cyclic derivative (II) can be further oxidized by iodine in aqueous media to provide the desired triphosphate (III), or directly treated with S_8 to obtain the corresponding α-thio-triphosphates (IV) (Scheme 8). Yields ranged from 60 to 75%. [63] This more reactive phosphorylation agent has been widely applied for the phosphorylation of numerous nucleosides. However, this strategy requires some level of nucleosides preprocessing, such as functional groups protection prior to phosphorylation.



Scheme 8. One-pot synthesis using *cyclo*Sal derivative.

In 2011, the Huang research group proposed the hypothesis that bulkier phosphorylation reagent would react with higher regioselectivity towards ribose ring -OH groups, thus avoiding protection and deprotection of competing functional groups. Their suggested method relies on the *in-situ* formation of a bulky phosphorylation (I) agent by reacting 2-chloro-4-*H*-1,3,2-benzodioxaphosphorin-4-one (a) with pyrophosphate prior to employment of nucleoside (Scheme 9). After phosphorus oxidation by iodine and hydrolysis, desired NTPs (VI) were synthesized in 10 – 50% yields. The formation of 3'- or 2'-triphosphate impurities was observed yielding

approximately 15%. This was partially attributed to residual amount of phosphorylation agent (a), that could compete in reaction with nucleoside resulting in reduced regioselectivity. [64] Likewise, the dNTPs were synthesized in yields ranging from 19 to 46%. [65]



Scheme 9. The proposed mechanism for nucleoside 5‘-triphosphates applying *in-situ* generated *cyclo*Sal-triphosphate derivative in one-pot synthesis.

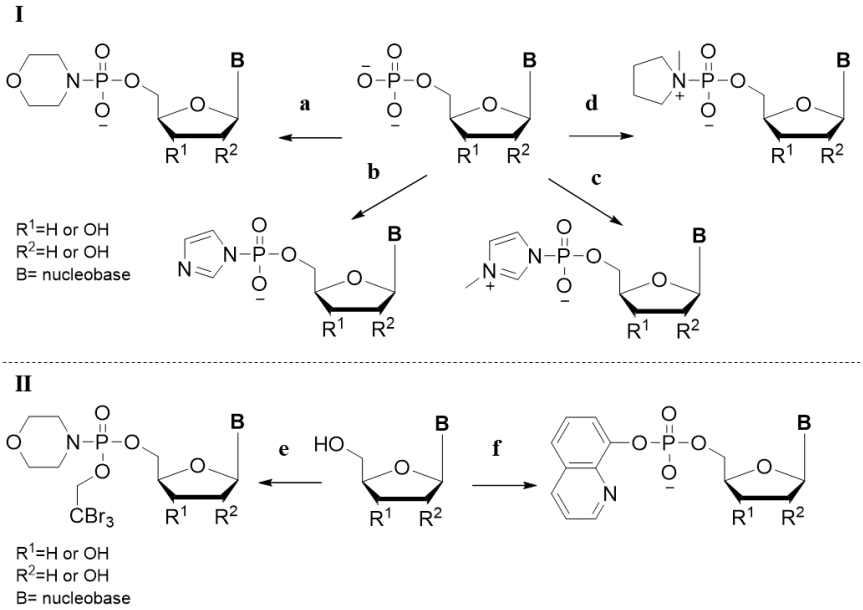
*Cyclo*Sal and its derivatives were widely used for numerous natural and modified nucleotides synthesis. [66, 67, 68, 69] In parallel, different oxidation conditions and agents were tested, showing that in some cases, the oxidation is a limiting step. Therefore, several oxidizing agents such as *t*-BuOOH, oxone, mCPBA, I_2/pyridine/water were tested. [62] Nevertheless, the I_2/pyridine/water system is still the most commonly used.

1.2 Nucleotides synthesis: monophosphate activation

The other synthetic strategy for the preparation of nucleoside 5’-triphosphates relies on a multistep synthesis. This approach is based on the synthesis of corresponding nucleoside 5’-monophosphates, followed by their activation during a nucleophilic substitution. The activating groups subsequently act as leaving groups in the following reactions with pyrophosphates. Although this strategy is a multi-step procedure with purification after each step, it has some advantages in comparison to one-pot synthesis. One of the most important aspects is the impurity profile that can be simplified by applying this synthetic strategy. Since the monophosphorylation step is not regioselective, various monophosphate intermediates are formed which, subsequently, react with pyrophosphate resulting in complex impurity profile. Thus, a properly purified nucleoside 5’-monophosphate reduces the probability of impurities formation in the following reaction with pyrophosphate. Moreover, it also reduces the number

of substances able to react with pyrophosphate, which is not consumed by unwanted side-reactions.

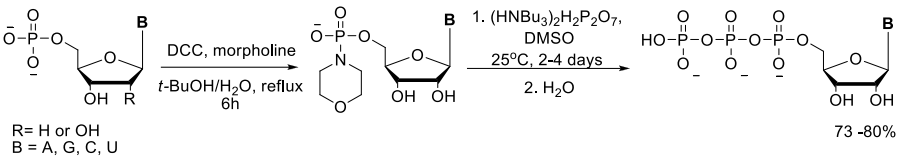
Some prior art regarding nucleoside monophosphate activation have been done during the late 1960s and 1970s. In the majority of reported articles DCC in pyridine was applied for activation of either orthophosphoric acid derivatives (e.g., phosphoric acid [70], β-cyanoethyl phosphate [71]) or nucleoside monophosphates. [72]. During the time many activated monophosphates preparation strategies were developed, but two main pathways can be distinguished. The first one follows the synthetic procedure and purification of nucleoside 5'-monophosphate, which is subsequently activated, usually by forming phosphoramidates of diverse nature (Scheme 10 I). The other pathway relies on the direct preparation of activated monophosphate using a phosphorylation agent activated prior to monophosphorylation reaction (Scheme 10 II). The most common technique is I and it is the main procedure to obtain nucleoside 5'-diphosphates. Despite of the existing variety of activating groups, to date, the most prevailing ones rely on the formation of either phosphoromorpholidates (a) or phosphoroimidazolidates (b) (Scheme 10 I).



Scheme 10. Synthetic pathways for activated monophosphates formation. a, b, c, c, e and f – indicates different synthesis strategies.

The activated phosphoromorpholidate was prepared using DCC and morpholine in *t*-BuOH/ H_2O mixture under reflux conditions. After 6 hours

the product was obtained in 91% yield (Scheme 11). To obtain dATP the activated monophosphate was reacted with tri-*n*-butylammonium pyrophosphate in DMSO. It is important to mention, that reaction is highly sensitive to moisture. Performing reaction under inert atmosphere for several days dATP was obtained in 80% yield. Under optimized conditions, several dNTPs and NTPs were successfully obtained in 73 – 80% yield (Scheme 11). [33]

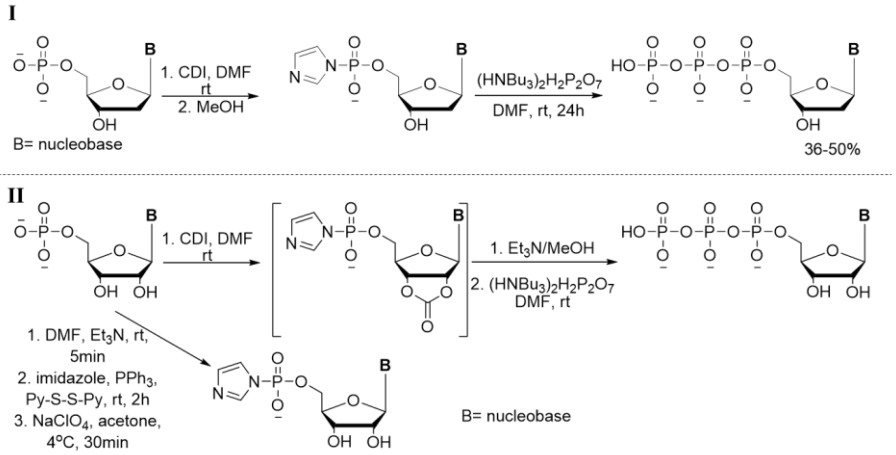


Scheme 11. Nucleoside 5'-triohosphate synthesis via formation of morpholidate intermediate.

Another common activation method relies on the formation of phosphoroimidazolidates. It is important to mention that phosphoroimidazolidates were reported to be more reactive than phosphoromorpholidates. [73] Hoard and Ott treated nucleoside monophosphates with CDI in DMF at room temperature. Obtained corresponding phosphoroimidazolidates were, subsequently, reacted with tri-*n*-butylammonium pyrophosphate to produce dNTPs in 36 – 50% yields (Scheme 12 I). [74] The attempts to apply the same methodology for the preparation of NTPs resulted in the formation of 2',3'-cyclic carbonates that were carried through as impurities in the subsequent triphosphorylation reaction (Scheme 12 IIa). Further investigation led to an efficient removal of carbonate functionality by applying basic conditions, usually by treatment with amines (e.g., ammonia, triethylamine). According to the described methodology, the AMP was successfully converted to ATP in 80% yield (Scheme 12 IIa). [75] Likewise, NDPs and dNDPs were prepared, when corresponding phosphoroimidazolidates were treated with tri-*n*-butylammonium orthophosphate. During the second round of activation with CDI followed by subsequent reaction with tri-*n*-butylammonium orthophosphate NTPs and dNPTs were obtained in high yields (84 – 93%). [76] This technique is widely used, not just for the synthesis of natural nucleotides, but for modified nucleotides as well, e.g., a series of carbocyclic nucleoside-5'-triphosphates analogues were obtained throughout the activation methodology in 6 – 61% yields. [77, 78]

To obtain nucleoside 5'-diphosphates the alternative methodology for the preparation of nucleoside phosphoroimidazolidate was reported. The

synthesis includes nucleoside-5'-monophosphate treatment with imidazole, triphenylphosphine and 2,2'-dithiopyridine in anhydrous DMF and Et_3N media. After completion of the reaction, product was precipitated by applying sodium perchlorate solution in acetone (Scheme 12 IIb). [79]



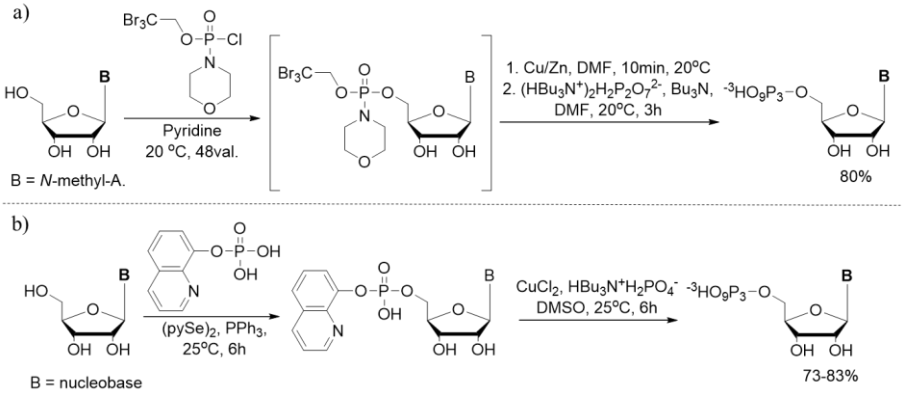
Scheme 12. Nucleoside 5'-triphosphates synthesis via formation of phosphoroimidazolidates.

Further development of activating agents led to several original ideas, reported by Borch and Taylor groups independently. Throughout *in-situ* formation of pyrrolidinium phosphoramidate zwitterion at the 5'-position of diverse NTPs and dNTPs, the fast and efficient subsequent coupling with tri(tetra-*n*-butylammonium) pyrophosphate was observed, producing the corresponding nucleoside 5'-triphosphates (Scheme 10d). [80] Taylor group used sulfonylimidazolium salt which is prepared by reacting phenylsulfonylimidazolide with methyl triflate in ether. The reaction results in mixed activated anhydrides where both intermediates were reported to be reactive towards TBAPP in the following triphosphorylation step (Scheme 10 c). [81] Both activation agents and methodologies afforded triphosphates in high yields (55 – 100%).

As mentioned before, the other way to obtain activated nucleoside 5'-monophosphates is by direct nucleoside treatment with beforehand activated phosphorylation agent. Several examples of such a strategy were reported. 2,2,2-Tribromoethyl phosphoromorpholinochloridate reacted with *N*-methyladenosine to obtain 2,2,2-tribromoethyl phosphoromorpholidate intermediate with high regioselectivity. Subsequently, the 2,2,2-tribromoethyl group was removed by treatment with Cu/Zn in DMF. The resulted

phosphoromorpholidate reacted with TBAPP to produce *N*-methyl-ATP in 80% yield (Scheme 13a). [82]

Hata *et al.* applied 8-quinolyl for nucleotides activation. By performing a reaction between ribonucleosides and 8-quinolyl phosphate in the presence of 2,2'-dipyridyl diselenide and triphenylphosphine, the activated monophosphates were obtained in good yields (70 – 87%). In the following reaction with TBAPP copper dichloride was used as an activating agent for the 8-quinolyl group. Natural NTPs were obtained in 73 – 83% yields (Scheme 13b). [83]



Scheme 13. Synthesis of nucleoside 5'-triphosphates applying beforehand activated phosphorylation agents.

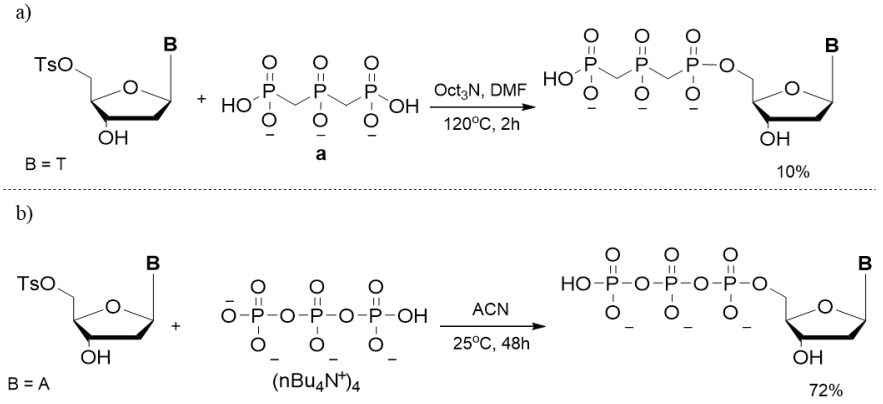
Preparation of activated nucleoside 5'-monophosphates is the essential strategy for subsequent synthesis of either natural or modified nucleoside 5'-diphosphates [21], dinucleotides [84], nucleotides bearing labels at the phosphate group [85] and nucleoside 5'-triphosphates. [73]

1.3 Nucleotides synthesis via direct reaction of triphosphate moieties with nucleoside

Synthetic strategies involving direct displacement of 5'-*O*-leaving groups of nucleosides by triphosphate nucleophiles or reaction of nucleoside -OH group with cyclic trimetaphosphate species were reported. The first mentioned example has been rarely applied in practice and only few reports exist. The other method was investigated more closely and was more widely applied for the synthesis of either natural or phosphate-modified nucleotides.

The first reported example standing for 5'-tosylated thymidine reaction with bis(dihydroxyphosphinomethyl)phosphinic acid was reported in 1979 by Stock. The reaction was performed at 120 °C in a tri-*n*-butylamine and DMF

mixture. The product was obtained in only 10% yield (Scheme 14a). During the reaction, it was noticed that esterification undergoes at the central P atom rather than the terminal, resulting in branched structure impurity. It is important to mention that Stock tested the cyclic analogue of bis(dihydroxyphosphinomethyl)phosphinic acid (a) in reaction with 5'-tosylated thymidine, however no satisfactory results were obtained. [86]

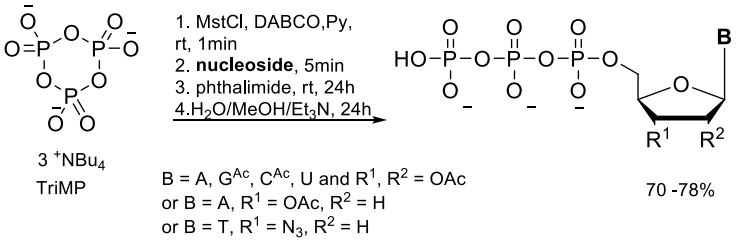


Scheme 14. Synthesis of nucleoside 5'-triphosphates via reaction of triphosphate nucleophiles and 5'-tosylated nucleosides.

After 10 years, a similar experiment was performed with 5'-tosylated adenosine, but in this case the monohydrogen tetra-*n*-buthylammonium triphosphate was used. When the reaction was performed in dry acetonitrile at 25 °C for 48 h the desired dATP was obtained in 72% yield (Scheme 14b). Likewise, the ADP was synthesized when monohydrogen tetra-*n*-buthylammonium diphosphate was used (74% yield). It was observed that, during the synthesis of ADP the protected 2',3'-*O*-isopropylideneadenosine 5'-tosylate was more reactive than unprotected analogue yielding 93% of ADP. [87] Correspondingly, applying either bridging-atom-modified or natural tetra-*n*-butylammonium hydrogen phosphate analogues (bridging atom: CH_2, CF_2, O), series of NDPs and dNDPs were synthesized with yields ranging from 50 to 83%. [36]

The utility of trimetaphosphate analogue for triphosphorylation of nucleosides was mentioned previously (by Stock), however with no success. The breakthrough was accomplished by Taylor group where a series of articles were reported from 2013 to 2015. The synthesis strategy included the use of activated trimetaphosphate derivatives. However, they were successful only by reacting them with diverse monophosphates producing corresponding tetraphosphates. [88, 89] Finally, the nucleoside 5'-triphosphates were synthesized in a similar manner. After optimization of reaction conditions, the

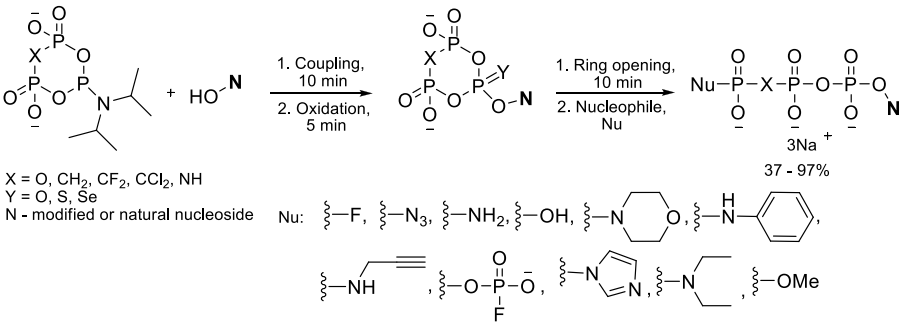
protected nucleosides were triphosphorylated as follows: MstCl, TriMP and DABCO were mixed in dry pyridine for 1 min, followed by the addition of the nucleoside, after 5 min the phthalimide was applied and the reaction mixture was stirred at room temperature for 24 h. Finally, the reaction mixture was cooled on ice and quenched with 100 mM triethylammonium acetate buffer. The desired triphosphates were obtained in 70 – 78% yield (Scheme 15). However, in this case, the authors could not determine the reaction mechanism since reaction monitoring with ^{31}P NMR was complicated by precipitate formation during the reaction. Nevertheless, they proposed several insights. They assumed that phthalimide may form a sufficiently reactive intermediate for phosphorylation of OH groups, moreover, polymeric form of TriMP could act as a phosphorylating agent as well. [89]



Scheme 15. Synthesis of nucleoside 5‘-triphosphates applying trimetaphosphate.

An innovative solution for direct triphosphorylation of nucleosides was proposed in 2019 by Jessen group. They described an easy synthesis of the whole family of phosphorylation reagents (Scheme 16) derived from natural or modified pyrophosphates and (diisopropyl-amido)dichlorophosphite ((*i*Pr)$_2$N-PCl$_2$). These phosphoramidites were directly coupled with nucleosides either using DCI or ETT in DMF and ACN mixture. The presence of P(III) in the cyclic agent enhanced its reactivity and enabled this direct triphosphorylation that was unsuccessful in case of P(V) analogues under same reaction conditions. The great structural diversity of these cyclic phosphorylation agents highly simplifies the preparation of some nonhydrolyzable analogues (e.g., where bridging O is replaced by CH_2, CCl_2, CF_2). Moreover, after P(III) oxidation with *m*CPBA or KSeCN, the triphosphate ring opening can be performed by numerous nucleophiles. Those nucleophiles can further serve as activating groups for dinucleoside polyphosphates synthesis or functional groups for subsequent labelling. Although the cyclic phosphorylation agent is bulky, the P(III) high reactivity reduces triphosphorylation regioselectivity. In cases of unprotected nucleosides, the 10 – 20% of their 3’-triphosphate analogues were obtained. In all cases the selective attack and ring opening by the nucleophile to produce

a linear structure with the modified termini was observed. During the mechanistic study, the nucleophile attack was shown to occur at the least sterically hindered site, where elimination takes place with preferential cleavage of anhydride bond resulting in the least charge accumulation on the leaving group. Series of natural, triphosphate bridge-modified, and/or terminal phosphate-modified nucleotides were synthesized in yields ranging from 37% to quantitative. [90]



Scheme 16. Synthesis of nucleoside 5‘-triphosphates applying different P(III)-P(V) cyclic triphosphate derivatives.

2. MODIFICATION OF HETEROCYCLIC BASES: PALLADIUM-CATALYZED SONOGASHIRA CROSS-COUPLING REACTIONS FOR C-C BOND FORMATION IN AQUEOUS MEDIA

In general, modifications can be introduced at any essential subunit of nucleotide structure: ribose ring, nucleobase, phosphate. Nucleobases can be modified to enhance the interaction with a complementary strand of nucleic acids by providing additional groups for hydrogen bonding. Such base modifications can be made to improve their natural functionalities. [91] However, the strategy to provide non-canonical functional abilities is of great interest, especially for bioprocesses monitoring and diagnostics. In the latter case, the modification bears its own functional ability and is usually considered as a tag or label (e.g., fluorescent label, biotin, enzyme, spin or redox labels, etc.). [30] Tags of this type are commonly introduced at the nucleobases since polymerases are less tolerant towards ribose ring modifications. Moreover, bulky tags can complicate the incorporation of subsequent nucleotides during enzymatic synthesis of nucleic acids. There are several positions as prospective modification sites (Figure 2), however, in most cases, modifications are introduced at the C5-position of pyrimidines and C7-position of 7-deaza-purines. This positioning of tags has the least effect on disturbance of natural double-stranded nucleic acids conformation since modifications point out to the major groove. The vast majority of modifications introduced at the nucleobases are based on C-C bond formation

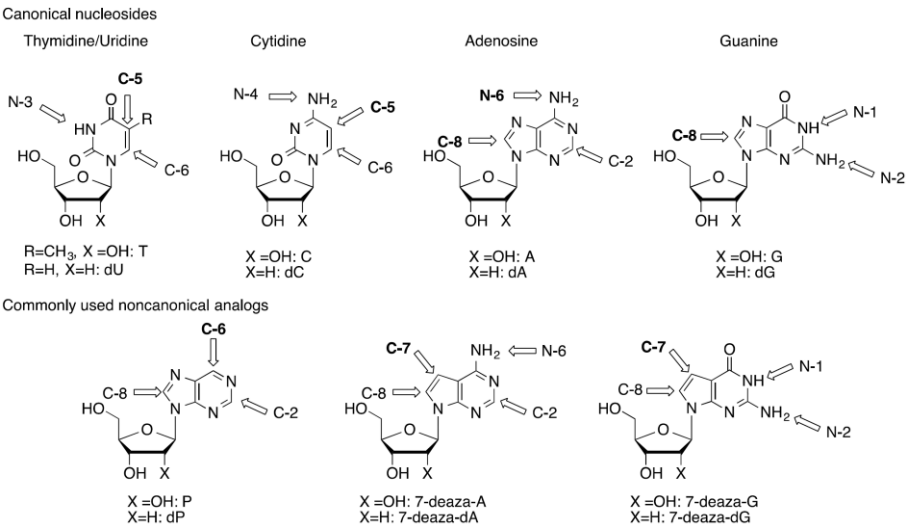


Figure 2. Most common modification sites of natural and unnatural nucleobases.

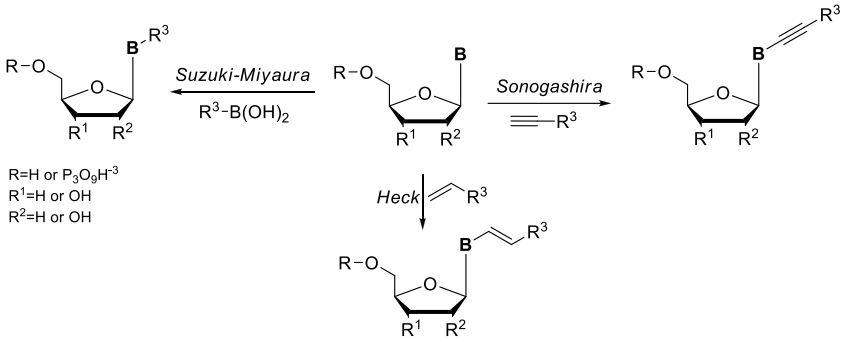
reactions. For this matter, the use of palladium-catalyzed cross-coupling reactions is the most common strategy. [92] Mostly, these reactions do not require harsh conditions and are highly tolerant towards various functional groups.

Although palladium-catalyzed coupling reactions are of great interest for the introduction of diverse modifications into nucleosides and nucleotides, these reactions can be very challenging on such substrates. Nucleosides possess the ability to coordinate metal catalysts during cross-coupling reactions subsequently deactivating them. [54] A deeper investigation of the coordination positions and abilities was performed during the Sonogashira cross-coupling. It was concluded that, guanosine is capable to coordinate with Pd and Cu in diverse ways. [93] The other important aspect is electronical difference of nucleosides in comparison to standard aromatic substrates. These differences are highly observed between heterocyclic bases and the positions of halogenation. For example, the 8-bromopurines possess lower reactivity than 6-halopurines or 5-halopyrimidines during the Pd-catalyzed cross-coupling reactions. This was attributed to the position of halogen where in the case of 8-halopurines it is present on the electron-rich azole ring. [94] Besides that, the solubility of highly polar nucleosides and nucleotides in typical organic solvents is one of the major issues. In attempts to overcome solubility matter, hydroxy groups of nucleosides were protected prior to cross-coupling reaction enhancing their lipophilicity. [95, 96] Such pre-processing allowed to apply standard cross-coupling conditions for various nucleoside substrates in good yields. From the practical perspective, the protection of hydroxy groups could be of use if the phosphorylation as the subsequent reaction proceeds. However, if the target compound is a modified nucleoside, the protection and deprotection additional steps result in lower overall yields and more time-consuming synthesis. Consequently, the most attractive strategy relies on direct labelling via cross-coupling reactions of unprotected nucleosides or nucleotides. This was accomplished for nucleoside labelling using polar organic solvents as DMF. Depending on nucleoside and label structure the necessity to select a solvent system is required, where both starting materials and highly hygroscopic ligands as triphenylphosphine would dissolve. Some cases might demand aqueous-organic solvents mixture.

2.1 The Sonogashira coupling with halogenated nucleosides

Many well-known C-C bond formations via Pd-catalyzed cross-coupling reactions were reported for modification of nucleosides or nucleotides. The Suzuki-Miyaura reaction is widely applied for single C-C bond formation with

introduction of either aromatic or aliphatic substituents using suitable boronic acids (Scheme 17). [95] Heck cross-coupling is used for the synthesis of vinyl derivatives [97], while the Sonogashira reaction for the formation of acetylenyl products (Scheme 17). [98] Regarding the following application of labelled nucleosides or nucleotides, the appropriate method for their modification is chosen. Alkynylation during the Sonogashira cross-coupling reaction has an advantage. It was identified that the triple C-C bond attached directly to the nucleobases has the least disturbance or significant structural changes to the nucleic acids' conformation. For example, the alkynylated

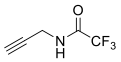
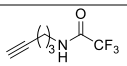
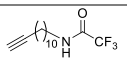
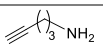
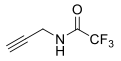
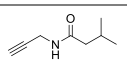
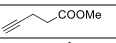
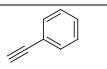
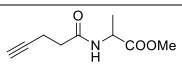
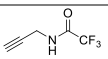
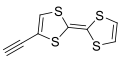
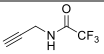
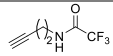
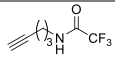
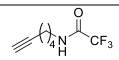
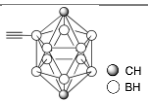
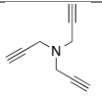


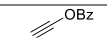
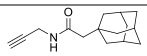
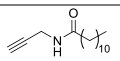
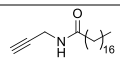
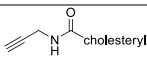
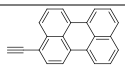
Scheme 17. Palladium-catalyzed cross-coupling reactions used for nucleobase-modified nucleosides and nucleotides synthesis.

uridine (at C5 position) can successfully replace thymidine with nearly no effect on DNA structure. [92] Consequently, triple bond formation in some cases might be of high preference towards double or single C-C bond.

With the emergence of DNA sequencing, the synthesis of chain terminators began. 2',3'-Dideoxynucleotides used in Sanger sequencing had fluorescent tags, that were attached through the linker connecting nucleobase and label. Although several biocompatible linkers were already known, the standard procedure for the preparation of all four nucleotides was lacking. In 1989 Hobbs selected several trifluoroacetamido-protected acetylenes and coupled them with halogenated 2',3'-dideoxynucleosides. His first attempts to apply standard procedure of the Sonogashira cross-coupling failed, since reactions were performed in triethylamine which was a poor solvent for nucleosides. When the reaction was performed in DMF in the presence of tetrakis(triphenylphosphine)palladium (0) and copper (I) iodide, differently modified nucleosides were obtained in good yields (Table 2, entry 1). In many cases a 2:1 mol ratio of CuI and $Pd(PPh_3)_4$ was necessary, yet long reaction times were observed (4 – 24 h). [99]

Table 2. Summary of different modifications introduced by alkynylation of halogenated nucleosides via the Sonogashira cross-coupling reaction.

Entry	Catalyst	Substituent	Halogenated nucleoside	Yield	Ref.
1	$Pd(PPh_3)_4$		5-I-ddU	77%	[99]
			5-I-dU	81%	
			5-I-U	80%	
			5-I-ddC	90%	
			5-I-ddC	74%	
			7-deaza-7-I-ddA	92%	
			7-deaza-7-I-ddG	78%	
			5-I-ddU	80%	
			5-I-ddU	73%	
			5-I-ddU	64%	
2	$PdCl_2(PPh_3)_2$	$HC{\equiv}CCH_3$	8-Br-dA	69%	[100]
		$HC{\equiv}C(CH_2)_2CH_3$		58%	
		$HC{\equiv}C(CH_2)_4CH_3$		86%	
3	$Pd(PPh_3)_4$		5-I-dC	76%	[55]
			5-I-dU	84%	
			5-I-dC	93%	
			5-I-dU	95%	
			7-deaza-7-I-dA	20%	
			7-deaza-7-I-dG	68%	
	$Pd(PPh_3)_4$ $PdCl_2(PPh_3)_2$		7-deaza-7-I-dA	72%	
			8-Br-dG	66%	
4	$Pd(PPh_3)_4$		5-I-dC	60%	[101]
5	$PdCl_2(PPh_3)_2$		5-I-dC	65%	[102]
			5-I-dU	61%	
			7-deaza-7-I-dA	69%	
			7-deaza-7-I-dG	35%	
6	$Pd(PPh_3)_4$		8-Br-dG	86%	[103]
			8-Br-dG	99%	
			8-Br-dG	67%	
			8-Br-G	89%	
7	$Pd(PPh_3)_4$		7-deaza-7-I-dA	80%	[104]
8	$Pd(PPh_3)_4$		7-deaza-7-I-dG	55%	[105]

9	$Pd(PPh_3)_4$		5'-DMTr-LNA-U	70%	[106]
			5'-DMTr-LNA-U	76%	
			5'-DMTr-LNA-U	87%	
			5'-DMTr-LNA-U	68%	
			5'-DMTr-LNA-U	53%	
			5'-DMTr-LNA-U	80%	
			5'-DMTr-LNA-U	80%	

Exploiting Sonogashira cross-coupling reaction a wide diversity of modifications was introduced at both purines and pyrimidines nucleobases (Table 2). The obtained yields slightly varied depending on either nucleoside nature or alkynyl substituent itself. Simple aliphatic substituents, as well as sterically small aromatic, were coupled with nucleosides. Reactions with sterically bulky alkynes as 2-ethynyl-*para*-carborane (Table 2, entry 7) or in Table 2, entry 9 represented condensed aromatic derivatives, cholesteryl-, adamantyl- substituent bearing alkynes resulted in the formation of a desired alkynylated nucleoside as well in good yields. Commonly $Pd(PPh_3)_4$ and $PdCl_2(PPh_3)_2$ in the presence of CuI in DMF were used for efficient Sonogashira coupling. Usually, the reactions were maintained at room temperature, however, for some substrates higher temperature (such as 70°C) was required (Table 2, entry 5). [102] Table 2 represents just a part of different modifications that have been introduced via the Sonogashira coupling reaction with halogenated nucleosides, indicating this modification approach as a versatile and productive method for nucleosides labelling.

2.2 The Sonogashira coupling with halogenated nucleosides in aqueous media

From the perspective of catalyst development, the major objective is usually driven by attempts to increase catalytic activity towards an extensive spectrum of substrates. Moreover, reducing the necessary amount of catalyst used for reaction is a very attractive goal. The challenges of metal-catalyzed reactions arise with the substrates of high hydrophilicity, due to their poor solubility in commonly used reaction media. The ability to adopt the aqueous or partly aqueous conditions would open up cross-coupling reactions

employment for a wider diversity of substrates. [107, 108] Consequently, water-soluble phosphine ligands were developed. [107]

The development of homogeneous aqueous-phase catalysis began with Rhone-Poulenc/Ruhrchemie, when industrial hydroformylation of propene was performed in biphasic (water/organic) media applying tri(3-sulfonatophenyl)phosphine (TPPTS) in 1985. [109] Much following research led to the development of various aqueous-phase catalysts, their synthesis and application techniques. [110, 111] Palladium-catalyzed Heck, Suzuki and Sonogashira reactions with aryl iodides which used $Pd(OAc)_2$/TPPTS as the catalytic system were reported. Moreover, it was shown that *in-situ* prepared palladium catalyst in water/acetonitrile media mixing $Pd(OAc)_2$ and TPPTS, could be directly applied in coupling reactions without the necessity of beforehand isolation. [112] In 1995 Savignac group applied ^{31}P NMR and voltammetry to prove the *in-situ* formation of active Pd(0) intermediate after mixing $Pd(OAc)_2$/TPPTS in water-acetonitrile media. [113, 114]

In 1990 report by Casalnuovo and Calabrese howed the diverse applicability of Pd-catalyzed cross-coupling reactions in water or water/organic solvents mixture and opened wide possibilities of palladium catalysis in aqueous media. [115] Since then, more than a decade was dedicated to the development of various water-soluble phosphine ligands for coupling reactions in aqueous media (Figure 3). Various ionic or neutral hydrophilic groups were introduced to increase the ligand's solubility in water. To date, the most commonly used are the TPPTS and the TXPTS ligands (Figure 3). [107, 108]

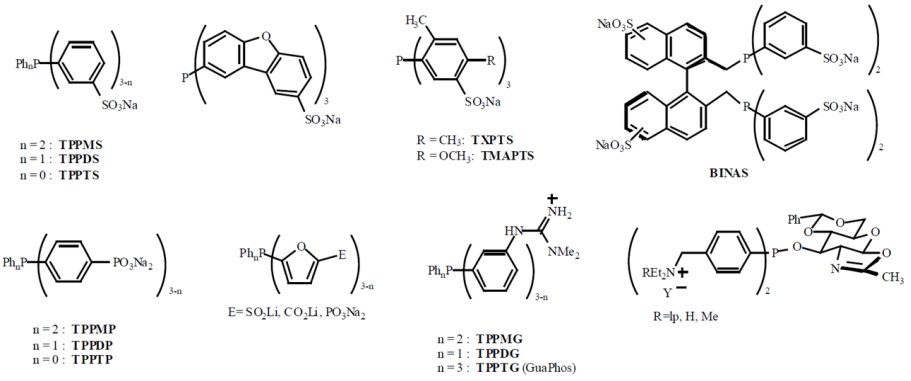


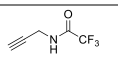
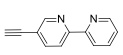
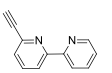
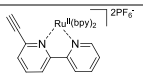
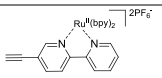
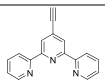
Figure 3. Hydrophilic phosphine ligands examples for aqueous-phase Pd-catalyzed cross-coupling reactions.

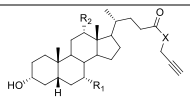
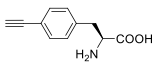
The same authors not just represent the wide applicability of aqueous-phase Pd-catalyzed reactions, but also the variety of substrates suitable for this

type of couplings. More importantly, they were the first scientists to apply aqueous-phase Sonogashira coupling for biomolecules such as unprotected nucleosides, nucleotides and amino acids. Usually, 1:1 water/ACN mixture was used, yet other ratios or organic solvents (EtOH, MeOH) could also be employed. For efficient Sonogashira coupling, Casalnuovo employed the previously mentioned TPPMS ligand (Figure 3) and Na_2PdCl_4 to obtain Pd(0) valent complex. 5-I-dU was coupled with propargyltrifluoroacetamide in the presence of 10 mol% of catalyst, 20 mol% of CuI and 2 eq. of Et_3N in water/ACN (1:1) media. After 4 hours the desired product was obtained in 95% yield performing the reaction at room temperature (Table 3, entry 1). [115] Despite the pioneering work of Casalnuovo on cross-coupling reactions with nucleosides in aqueous-media, it took nearly a decade for this methodology to be continuously adopted for nucleosides and nucleotides modification. However, nowadays it is gaining increasing attention for biomolecules modification and labelling.

To synthesize nucleotides with probes that possess electrochemical and photophysical properties, small functionalities, bile and amino acids, and many others were produced via the Sonogashira cross-coupling reaction. A brief summary of a small part of modification diversity is shown in Table 3. In the majority of cases, a higher reaction temperature was required, while yields broadly varied depending on both modification nature and nucleobase.

Table 3. Summary of different modifications introduced by alkynylation of halogenated nucleosides via aqueous-phase Sonogashira cross-coupling reaction.

Entry	Conditions	Substituent	Halogenated nucleoside	Yield	Ref.
1	Na_2PdCl_4/TPPMS, CuI, rt, Et_3N, 4h, H_2O/ACN (1:1)		5-I-dU	95%	[115]
2	[1]$Pd(OAc)_2$/TPPTS, CuI, Et_3N, 75°C, H_2O/ACN (2:1) or DMF [2]$Pd(OAc)_2$/TPPTS, CuI, Et_3N, rt, H_2O/ACN (2:1)		8-Br-dA[1] 7-deaza-7-I-dA[2] 5-I-dC[1] 5-I-dU[1]	96% 91% 67% 70%	[116] [117] [118]
			8-Br-dA[1] 7-deaza-7-I-dA[2] 5-I-dC[1] 5-I-dU[1]	90% 82% 70% 75%	
			8-Br-dA[1] 7-deaza-7-I-dA[2]	16% 47%	
			8-Br-dA[1] 7-deaza-7-I-dA[2]	0% 59%	
			5-I-dC[1] 5-I-dU[1]	63% 67%	

3	$Pd(OAc)_2$/TXPTS, CuI, Et_3N, 65ºC (for pyrimidines) or 80ºC (for purines), H_2O/ACN (1:1)		5-I-dU 8-Br-dA 8-Br-A 8-Br-dG	71% 88% 53% 86%	[94]
			5-I-dU 8-Br-dA 8-Br-A 8-Br-dG	42% 89% 74% 85%	
			5-I-dU 8-Br-dA 8-Br-dG	55% 98% 85%	
			5-I-dU 8-Br-dA 8-Br-dG	84% 98% 84%	
4	[1]$Pd(OAc)_2$/TPPTS, CuI, DIPEA, 75ºC, H_2O/ACN (2:1) [2]$Pd(OAc)_2$/TPPTS, CuI, DIPEA, 65 ºC, H_2O/ACN (1:1)	 **a**: R_1=R_2=OH; X=O	7-deaza-7-I-dA[2] 5-I-dC[2]	69% 39%	[119]
		b: R_1=H; R_2=OH; X=O	7-deaza-7-I-dA[2] 5-I-dC[2]	35% 88%	
		c: R_1=OH; R_2=H; X=NH	7-deaza-7-I-dA[2] 5-I-dC[2]	45% 90%	
		d: R_1=R_2=OH; X=NH	7-deaza-7-I-dA[1] 5-I-dC[1]	69% 66%	
		e: R_1=H; R_2=OH; X=NH	7-deaza-7-I-dA[1] 5-I-dC[1] 5-I-dU[2]	65% 61% 58%	
5	$Pd(OAc)_2$/TPPTS, CuI, Et_3N, 60 ºC, H_2O/ACN (2:1)		8-Br-dA 7-deaza-7-I-dA 5-I-dU	61% 94% 70%	[120]
6	$Pd(OAc)_2$/TPPTS, CuI, Et_3N, rt, H_2O/ACN (2:1)		5-I-dU	88%	[121]
			5-I-dC	84%	
			7-deaza-7-I-dA	98%	
			7-deaza-7-I-dG	69%	

1 and 2 indicates different reaction conditions. Same references present at the end of halogenated nucleoside name indicates which methodology was used for the following reaction.

2.3 The Sonogashira coupling with halogenated nucleotides in aqueous media

Even if cross-coupling reactions are performed with nucleosides with good efficiency, the following triphosphorylation may lead to poor yields. Due to the high hydrophilicity of nucleotides, their labelling in non-aqueous media is challenging. The development of water-soluble ligands for Pd-catalyzed

reactions had a significant impact on the synthesis of nucleobase-modified nucleotides. [122] The diverse labels were introduced applying aqueous-phase Sonogashira cross-coupling, several examples are given in Table 4.

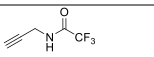
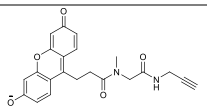
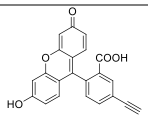
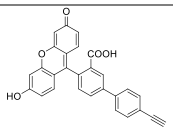
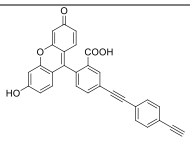
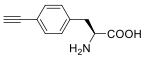
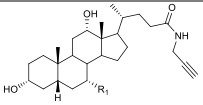
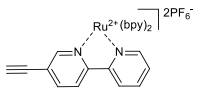
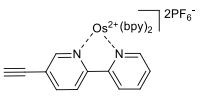
In 1990, Casalnuovo was the first to report the applicability of aqueous-phase Sonogashira cross-coupling reactions on halogenated nucleotides. Performing reaction at room temperature for 3 hours in presence of Casalnuovo obtained propargylaminated dCMP in 73% yield (Table 4, entry 1). Attempts to increase the amount of water to 96:4 ratio of H_2O/ACN respectively resulted in 52% yield after 6 hours. Likewise, Casalnuovo performed a reaction between 5-I-dUTP and T-505 dye, where dye-labelled dUTP was obtained in 50% yield (Table 4, entry 1). The versatility of their methodology opened a new, rapid, and straightforward synthetic route for T-505-labelled dUTP, which was used as chain-terminator in DNA sequencing. [115]

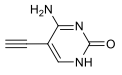
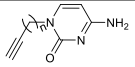
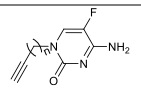
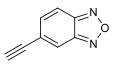
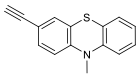
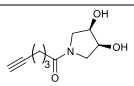
Burgess group employed aqueous-phase Sonogashira for coupling different fluorescein moieties with 5-I-d(d)UTP in completely aqueous media. The $Pd(OAc)_2$/TPPTS catalytic system along with CuI and Et_3N at room temperature provided the desired fluorescein-labelled d(d)UTPs with yields ranging from 17 to 42% (Table 4, entry 2). It is important to mention that the authors' attempts to obtain these products via phosphorylation of fluorescein-labelled uridine nucleosides completely failed. Poor solubility of starting materials and their possible aggregation were indicated as the main issues. [17]

The ferrocene label bearing purines and pyrimidines were successfully obtained. Coupling ethynylferrocene with 5-I-dUTP and 7-deaza-7-I-dATP the desired products were obtained in the single-step aqueous-phase reaction (Table 4, entry 3). [123] Likewise, using the same catalytic system the series of aminoacid-labelled dNTPs were obtained in 61 – 67% yields (Table 4, entry 4). The reactions reached full completion in 0.5-1 hour. [120] The same tendencies and similar results were obtained when bile acids were coupled with halogenated dNTPs (Table 4, entry 5). Maintaining reactions at 75 ^oC after 45 min. labelled dNTPs were isolated giving 32 – 57% yields. [119] Some smaller acetylenes like propargyl alcohol and propargyl phosphate were coupled with 5-I-dUTP (Table 4, entry 6). After one hour products were obtained in 43 – 48% yields. [56] All four dNTPs were successfully labelled with both Ru and Os bipyridine complexes applying aqueous-phase Sonogashira coupling (Table 4, entry 7). Although the obtained products yields were low (12 – 35%) the previously described attempts to synthesize their nucleoside analogues proceeded just slightly better (Table 3, entry 2).

However, obtained nucleosides were not further phosphorylated, probably due to hardly compatible phosphorylation conditions and label nature. [124]

Table 4. Summary of different modifications introduced by alkynylation of halogenated nucleotides via aqueous-phase Sonogashira cross-coupling reaction.

Entry	Conditions	Substituent	Halogenated nucleotide	Yield	Ref.
1	Na_2PdCl_4/TPPMS, CuI, rt, Et_3N, H_2O/ACN (1:1)		5-I-dCMP	73%	[115]
		 T-505	5-I-dUTP	50%	
2	$Pd(OAc)_2$/TPPTS, CuI, rt, Et_3N, phosphate buffer (pH=7.2)		5-I-dUTP 5-I-ddUTP	17% 41%	[17]
			5-I-dUTP 5-I-ddUTP	17% 42%	
			5-I-dUTP 5-I-ddUTP	42% 26%	
3	$Pd(OAc)_2$/TPPTS, CuI, 65ºC, Et_3N, H_2O/ACN (2:1)		5-I-dUTP 7-deaza-7-I-dATP	42% 48%	[123]
4	$Pd(OAc)_2$/TPPTS, CuI, Et_3N, 60 ºC, H_2O/ACN (2:1)		5-I-dUTP 7-deaza-7-I-dATP 8-Br-dATP	66% 67% 61%	[120]
5	$Pd(OAc)_2$/TPPTS, CuI, 75ºC, DIPEA, H_2O/ACN (2:1)	 **a**: R_1=OH	5-I-dCTP 7-deaza-7-I-dATP	32% 44%	[119]
		b: R_1=H	5-I-dCTP 7-deaza-7-I-dATP	53% 57%	
6	$Pd(OAc)_2$/TPPTS, CuI, 60ºC, Et_3N, H_2O/ACN (2:1)		5-I-dUTP	48%	[56]
			5-I-dUTP	43%	
7	$Pd(OAc)_2$/TPPTS, CuI, 70ºC, Et_3N, H_2O/ACN (2:1)		5-I-dUTP 5-I-dCTP 7-deaza-7-I-dATP 7-deaza-7-I-dGTP	12% 35% 21% 20%	[124]
			5-I-dUTP 5-I-dCTP 7-deaza-7-I-dATP 7-deaza-7-I-dGTP	17% 17% 26% 28%	

8	$Pd(OAc)_2$/TPPTS, CuI, 80°C, DIPEA, H_2O/ACN (2:1)		5-I-dUTP 5-I-dCTP 7-deaza-7-I-dATP	31% 35% 17%	[129]
9	$Pd(OAc)_2$/TPPTS, CuI, 80°C, DIPEA, H_2O/ACN (2:1)	 n=1	5-I-dCTP	30%	[130]
		n=2	5-I-dCTP	38%	
		n=3	5-I-dCTP	35%	
		 n=1	5-I-dCTP	80%	
		n=2	5-I-dCTP	39%	
		n=3	5-I-dCTP	37%	
10	$Pd(OAc)_2$/TPPTS, CuI, 75°C, DIPEA, H_2O/ACN (2:1)		5-I-dCTP 7-deaza-7-I-dATP	52% 54%	[125]
11	$Pd(OAc)_2$/TPPTS, CuI, 75°C, Et_3N, H_2O/ACN (1:1)		5-I-dCTP 7-deaza-7-I-dATP	48% 49%	[126]
12	$Pd(OAc)_2$/TPPTS, CuI, 80°C, Et_3N, H_2O/ACN (2:1)		5-I-dUTP 5-I-dCTP 7-deaza-7-I-dATP	67% 57% 83%	[127]
13	$Pd(OAc)_2$/TPPTS, CuI, 80°C, Et_3N, H_2O/ACN (2:1)		5-I-dCMP 5-I-dCTP	74% 59%	[128]

In attempts to introduce a benzofurazane-label (Table 4, entry 10) the aqueous-phase Sonogashira was performed with both iodinated nucleosides and nucleotides. When reactions were performed with iodinated nucleosides the benzofurazane-dCTP and -dATP were obtained in 45% and 28% yields, respectively (using $Pd(OAc)_2$/TPPTS). While Sonogashira coupling of iodinated dCTP and dATP resulted in 52% and 54% yields. The most efficient results were obtained when $Pd(PPh_3)_2Cl_2$ in DMF was applied for the coupling of iodinated nucleosides. Here, after one hour at 75°C benzofurazane-labelled dNs were obtained in good yields (60 – 70%). However, the triphosphorylation of benzofurazane-cytosine nucleoside resulted only in 24% yield. [125] Similar trends were observed during the synthesis of phenothiazine-labelled nucleotides (Table 4, entry 11). [126] This indeed represents that even though higher yields can be obtained from halogenated nucleosides and a higher reactivity catalytic system, the following triphosphorylation results in low to moderate yields. Consequently, the direct labelling of halogenated nucleotides holds the advantage of a less time-consuming, rapid, and simple methodology. In such cases, when functional groups of labels are incompatible with the triphosphorylation conditions, the

direct labelling of nucleotides is a preferred methodology. [127, 128] The representatives in Table 4, entries 12 and 13 are the best examples of that. The diol functional groups bearing labels can be successfully introduced at either iodinated nucleosides or nucleotides. However, in the case of nucleoside labelling the diol functional groups must be protected prior to the following triphosphorylation reaction.

The aqueous-phase Sonogashira cross-coupling provides direct labelling of halogenated nucleotides that allows to overcome the triphosphorylation of labeled nucleosides. Although in some cases the yields might be low or moderate, the ability to introduce labels in a single step is attractive from the perspective of time and resources. [92, 97]

3. NUCLEIC ACIDS LABELLING TECHNIQUES

The synthesis of functionalized or labelled nucleic acids can be achieved in two ways. The first one is based on solid-phase synthesis, where up to 100-200 nt long unmodified or modified oligonucleotides (ONs) are routinely prepared. [131] However, the mentioned strategy has several drawbacks and the main relates to the incompatibility of some functional groups with either phosphoramidite or solid-phase synthesis protocols. This is especially attributed to strongly nucleophilic or electrophilic groups, oxidizable functional species. Some functional groups can be protected either before phosphoramidite or solid-phase ON synthesis, yet in some cases, it can be hard to find a suitable protecting group for the whole cascade of reactions. Moreover, a higher degree of preprocessing leads to longer protocols and lower yields. [132] The other strategy to prepare modified nucleic acids is based on the enzymatic incorporation of modified 5'-triphosphates by various polymerases. Here, limitations of applicable functionalities are reduced to bare minimum. After all, several aspects must be taken into account when designing the labelled nucleotides. Usually, the major one is the steric hindrance induced by bulky labels that can suppress the enzymatic incorporation of such nucleotides. In 1981 Ward groups reported the synthesis of biotin labelled NTPs and dNTPs as affinity probes and their application in enzymatic assay, including incorporation by several polymerases and interaction with avidin and anti-biotin antibodies. [133] This pioneering work was extensively developed during all these years where various functionalized dNTPs and NTPs were tested as substrates for polymerases. [30, 134, 135]

When designing modified nucleotides there is another important aspect to be considered – modification positioning. As it was already mentioned, nucleobase modifications are usually introduced at the C5 position of pyrimidines and 7-deaza-C7 position of purines. [30] This positioning allows labels to point out to the major groove of double-stranded helix with minimal disturbance and destabilization of its native structure (Figure 4). The C8 position of purines is the second most often modified. Even though substituents at the C8 position are oriented at the major groove, bulky groups can initiate or stabilize the conformational change of DNA to Z form. This effect was attributed to the steric hindrance occurring between sugar hydrogens and substituent at the C8 position, due to the *anti*-conformation of modified nucleotide. [132] This is also in agreement with the poor acceptability of C8-substituted purines by various DNA polymerases. [120, 136] Substituents at other positions, for example, C2 of pyrimidines or purines appear in the minor groove and can completely destabilize the double-stranded

helix depending on the nature of modification. Moreover, nucleotides bearing substituents positioned at the minor groove can be discriminated by polymerases resulting in poor or no incorporation into the growing nucleic acid strand. [137] This is attributed to both the importance of minor groove for Watson-Crick base pairing and polymerase recognition of nucleotides during the chain extension. [138] The hindrance of substituents in various positions can be clearly seen from the three-dimensional structure of B-form DNA (Figure 4). [56, 139] Nevertheless, some relatively small modifications can be successfully introduced to the minor groove as well, without high destabilizing effect. [140, 141, 142]

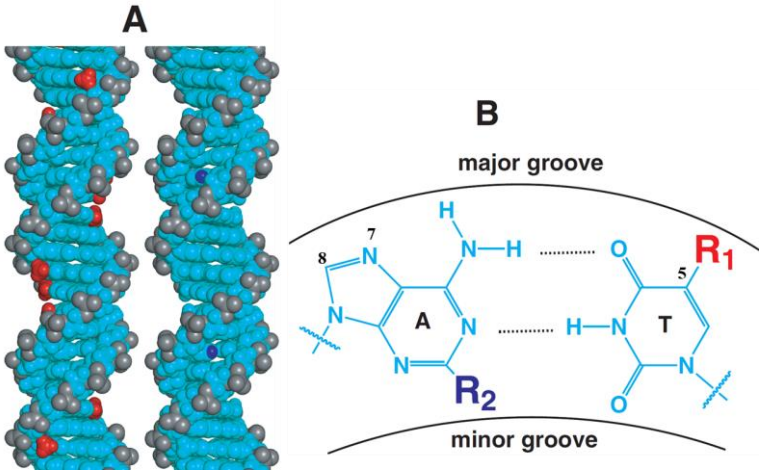


Figure 4. Positioning of substituents at the nucleobases in the double-stranded DNA. A) three-dimensional structure of B-DNA representing different substituent positions; B) Watson-Crick base pairing of A and T nucleobases. Substituents pointing out to the major groove are labelled in red, while the ones in the minor groove – blue.

When nucleic acids labelling is achieved by enzymatic incorporation of corresponding nucleotides, two types of strategies are applied. The first one is based on direct labelling when the final label structure is formed prior to enzymatic incorporation of nucleotide. Numerous examples were given in previous chapters on aqueous-phase cross-coupling. [119, 124, 143] The other strategy consists of small reactive groups bearing nucleotides synthesis, followed by their enzymatic incorporation. Finally, subsequent post-synthetic labeling is directly performed on the nucleic acid, during which the label is attached to the reactive groups. This labelling strategy is widely applicable for diverse modifications [144] employing various well-known nucleic acids amplification techniques such as primer extension reaction (PEX) or polymerase chain rection (PCR). [135]

3.1 Introduction of functional groups at the nucleobases for subsequent labelling

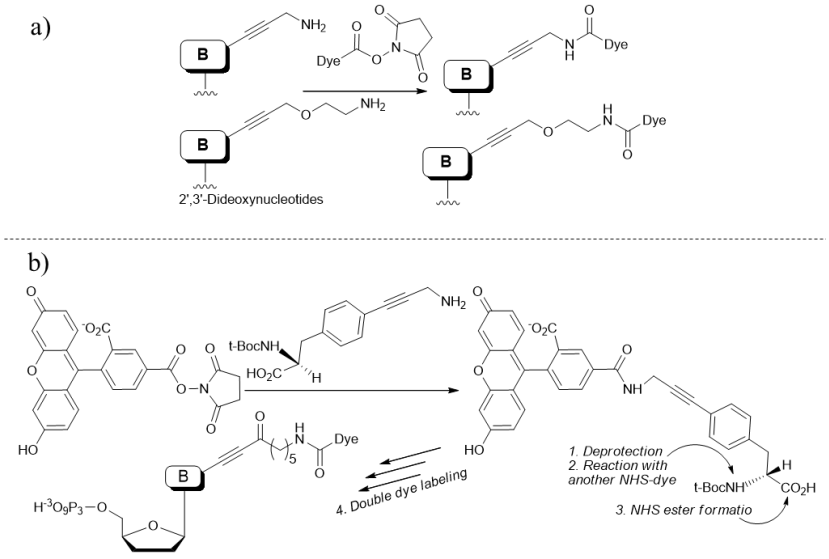
The aldehyde functional group is widely used and can be also found as a natural non-canonical modification of nucleobases. [145] A series of different reactions between the aldehyde group and other functionalities were performed to label nucleic acids. [144] Formation of oximes was reported during reaction of aldehyde bearing cytidine (at C5 position) and fluorescent-hydroxylamines. Several fluorescent probes (e.g., coumarin, BODIPY-B) were successfully attached for fluorescent labelling of nucleic acids. [146, 147] Along with hydrazones formation, nucleic acids bearing biotin-, nitrobenzofurazane-labels were obtained to detect the 5-formyl-U and C species in genomic DNA. [148, 147] The reductive amination reaction example of 5-formyl-cytidine was performed using $NaBH_3CN$ and lysine moiety of highly basic proteins – histones, for Schiff-base conjugate formation. [149] Various heterocyclic condensation [150], Wittig [151], Kneovenagel condensation [152] reactions with 5-formyl-U and C bearing nucleic acids and nucleotides were reported. The introduced labels vary from small functionalities to fluorophores, proteins, etc. [144, 153, 154]

Apart from numerous reported aldehyde reactions, a broad variety of other functional groups reactions were tested. In attempts to cross-link DNA with proteins the vinylsulfonamide- and acrylamide-nucleotides were synthesized and incorporated during PEX. The obtained Michael-acceptors bearing ONs were subsequently reacted with thiols of cysteine, cysteine-containing peptides or proteins. [155] The 2-vinyl or 2-ethynyl functionalized 2'-deoxyadenosine (dA) containing ONs were reacted with fluorophores bearing thiols. [156] The diols functionality was also applied for both nucleotides and nucleic acids modification. The labelling of various vicinal or cyclic *cis*-diols was performed with diverse osmium (VI) reagents. [128]

Amine is the other broadly used and very important functional group. The amine-based modifications were the first ones applied for the introduction of bulky substituents. The amine group can be reacted with various substrate functionalities for labels conjugation. However, one of the most common examples stand for their reaction with labelled-N-hydroxysuccinimide (NHS) esters resulting in amide bond formation. This technique was widely employed for the synthesis of the first dye-terminators used in Sanger sequencing. Later on, applied for numerous dye-labelled nucleotides. [157, 158] Amino functionalized nucleotides were obtained bearing different linkers connecting nucleobases (NB) to corresponding dye species as represented in Scheme 18a. Furthermore, the amine group is an attractive

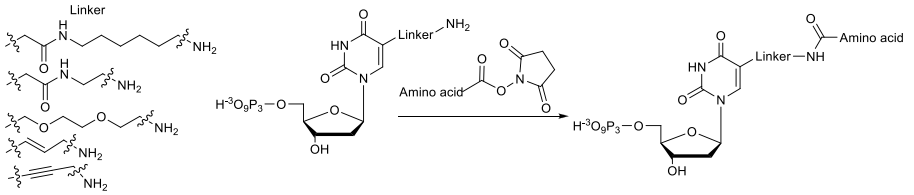
functionality not just for direct nucleotides labelling, but also for the formation of different linkers, attachment of necessary structural moieties, or multiple labelling as is represented in the Scheme 18b. Here, in attempts to have a rigid and at the same time linear linker the propargylaminated *t*-Boc-phenylalanine was reacted with the dye-NHS ester (Scheme 18b). The goal was to obtain the fluorescence resonance energy transfer (FRET) based dye-terminators. Consequently, each nucleotide had to have the same donor dye and four different acceptor dyes for identification of each nucleobase. For such compounds synthesis double dye labelling was applied, where the linker was first reacted with one of the dyes (Scheme 18b). After removal of *t*-Boc-protecting group the second labelling was applied with four different dye-NHS esters. Obtained products were subsequently reacted with amino-ddNTPs throughout the carboxylic acid group (Scheme 18b reaction 3). [159]

The numerous examples of nucleotides labelling throughout reaction of amine and NHS ester, or amine activation followed by reaction with carboxylic acids were reported. [160, 161]



Scheme 18. Nucleotides dye-labelling throughout amine and NHS-ester reaction.

In attempts to functionalize nucleic acids with amino acids, series of amino acid-NHS esters were synthesized. In this manner labelled nucleic acids could be applied for aptamer selection or for the analysis of substrate specificity of polymerases. Several different linkers bearing amino-dUTPs were synthesized and subsequently labelled with 15 differed amino acids (Scheme 19). Afterwards, all synthesized substrates were used in PEX or PCR with several polymerases. [162]



Scheme 19. Nucleotides amino acid-labelling throughout amine and NHS-ester reaction.

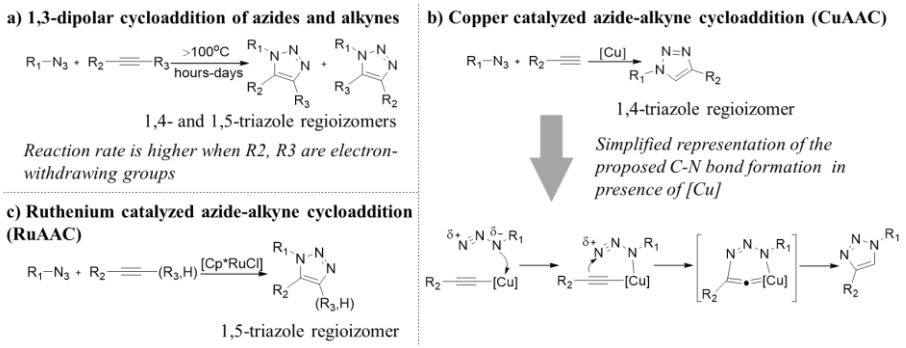
The wide applicability of amino group functionality for subsequent labeling probably relies on several main aspects. First, the synthesis of amino-nucleotides can be easily accomplished by cross-coupling reactions. A broad spectrum of different NHS esters are readily commercially available, while in some reports the preparation of NHS esters with various labels was efficient and without technical difficulties. Reaction conditions for amine interaction with NHS ester are simple, mild, and compatible with nucleotides and nucleic acids functionalities. [163]

3.2 Employment of alkyne-azide cycloaddition for nucleotides and nucleic acids labelling

The azide-alkyne 1,3-dipolar cycloaddition reaction did not gain much attention until the comprehensive mechanistic and kinetic studies were performed alongside with wide substrates applicability by Huisgens. [164] Although the reaction was highly versatile, several disadvantages remained. Reactions usually required high temperature and long times to reach full conversion. While lack of regioselectivity resulted in the formation of 1,4- and 1,5-linked isomers, consequently complicating the purification procedure (Scheme 20a). [165] In 2002 two independent research groups led by Meldal [166] and Sharpless [167] made significant progress in azide-alkyne 1,3-dipolar cycloaddition reaction applying Cu (I) salts. The utility of copper (I) catalysts allowed to perform reaction at room temperature in decent time range and afforded explicitly 1,4-disubstituted triazole regioisomer (Scheme 20b). Sharpless and colleagues examined several copper salts and found that $CuSO_4$ in the presence of reductants such as ascorbic acid or sodium ascorbate is highly efficient. Moreover, they tested azide and alkyne substrates of diverse chemical nature. Performing reactions at room temperature for 12 – 24 hours, desired products were obtained with regioselectivity and in high yields (82-94%). [167] The presence of mild reducing agents as ascorbate or sodium ascorbate are very important for reaction yields and impurity profile. The active Cu (I) species can easily undergo, resulting in formation of inactive

Cu (II). Moreover, in the presence of oxygen the copper-mediated oxidative side reaction occurs resulting in the formation of oxidative coupling byproducts (e.g., additional substitution at the 5-position of triazole, coupling of two triazole moieties via 5-positions). The presence of sodium ascorbate along with Cu (II) salts results in *in-situ* formation of active copper(I) acetylides while any dissolved dioxygen is rapidly reduced. The presence of ascorbate nearly diminishes the formation of oxidative coupling byproducts, while desired triazole moieties can be obtained in quantitative yields. [168]

In 2005 Jia and Sharpless focused their research to exclusively obtain the 1,5-disubstituted isomers. They investigated several Ru complexes among which pentamethylcyclopentadienylbis(triphenylphosphine)ruthenium (II) chloride ($[(C_{10}H_{15})Ru(P((C_6H_5)_3))_2Cl]$ indicated as Cp*RuCl) was identified as the most efficient with various aliphatic and aromatic azide and alkyne substrates, giving products in excellent yields (80 – 94%) and regioselectivity (Scheme 20c). [169]



Scheme 20. Azide-alkyne 1,3-dipolar cycloaddition.

Copper catalyzed azide-alkyne cycloaddition (CuAAC) reaction conditions are highly flexible, especially in case of solvent choice, where organic and aqueous media or mixed composition can be used. This makes the CuAAC of great interest for bioconjugation. [170] Moreover, the azide and alkyne functionalities are nearly inert towards the conditions of living cells as well as biomolecules due to their weak acid-base properties. These groups are orthogonal, meaning that they react just between themselves, and this makes them ideal functional groups for biomolecules labelling. Moreover, the resulting triazole ring is also very stable and non-toxic. Although CuAAC reactions can be carried out with high efficiency without any ligand for the metal catalyst, their presence can enhance the reaction rate. Many different ligands were investigated (e.g., nitriles, imines, halides, guanidines) while some labile ligands as amines could either interfere with the Cu redox stability

or were just too labile. On the other hand, strongly binding ligands decreased the catalytic activity. [171] Mildly binding ligands were found to be the most efficient for CuAAC and excellent performance was observed with ligands containing heterocyclic donor atoms. Folkins and colleagues came up with the idea to utilize triazole derivatives as possible Cu ligands to accelerate the CuAAC reaction. Out of 15 representatives, the best results were obtained when the TBTA ligand (Figure 5) was applied in the model reaction between phenylacetylene and benzyl azide. [171] To demonstrate bioconjugation feasibility virus coated with alkyne or azide-groups was fluorescently labelled via CuAAC. [172] It was also observed that tertiary amine center containing ligands displayed higher reactivity, where amine can serve as both base and coordination donor. [168] However, TBTA solubility in water was quite poor, consequently, the development of more hydrophilic ligands was prompted. Here the THPTA (Figure 5) was synthesized and applied in bioconjugation of bovine serum albumin (BSA) with either fluorescent dye or virus-like particles. [173] The sulfonated bathophenanthroline ligand (Figure 5) was also utilized for bioconjugation. However, due to the formation of a strongly electron-rich copper-ligand complex, high sensitivity to oxygen was noticed, leading to easy oxidation to its inactive form. [174] Other polydentate ligands based on trimethylamine core were synthesized bearing tri(benzimidazole) species (Figure 5). The significant catalytic activity was observed with those containing ester and carboxylic acid groups. [175] Examples of ligands often used for bioconjugation or biomolecules labelling are shown in Figure 5, where the TBTA and THPTA are the most commonly employed.

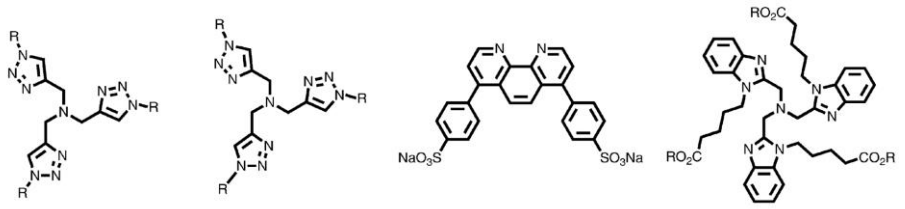


Figure 5. The most efficient ligands used for CuAAc reaction acceleration. [168]

A second azide-alkyne 1,3-dipolar cycloaddition reaction is based on efficient catalyst-free coupling driven by the activation of the alkyne. The scaffold of cyclooctyne was declared to be the smallest stable alkyne-containing ring. The conformation of a triple bond contained within a cyclooctyne is nearly linear, with a slight angle of 17° apart from the preferred

perfectly linear alkyne geometry. Such conformation of cyclooctyne ring results in ring strain which is sufficiently enough to activate the alkyne and enable efficient [3+2] cycloaddition with azide moieties. Consequently, this reaction was called strain-promoted azide-alkyne cycloaddition (SPAAC). [176] However, despite all the scientific input carried out during the years, just in 2004 the strained alkynes were used for bioconjugation purposes. [177, 178] Nevertheless, SPAAC application in biological systems was still complicated by several limitations. Slow reaction rate, poor solubility of cyclooctyne derivatives in aqueous media and lack of stability of starting materials were the main concerns. Consequently, a wide series of differently substituted cyclooctyne derivatives were designed, in order to reduce the previously mentioned drawbacks. [179, 180, 181, 182] Generally, SPAAC reaction has more advantages over the CuAAC and it is widely applied for various bioconjugation in live systems (Figure 6a). Even though the CuAAC reaction rate is higher the biggest disadvantage is the utility of copper, since Cu-ions were found to be toxic due to their ability to produce reactive oxygen species. [183] Later on, it was proved that the employment of certain ligands for CuAAC, can increase reaction rate and reduce the copper-induced toxicity in live systems (Figure 6b). [173, 183]

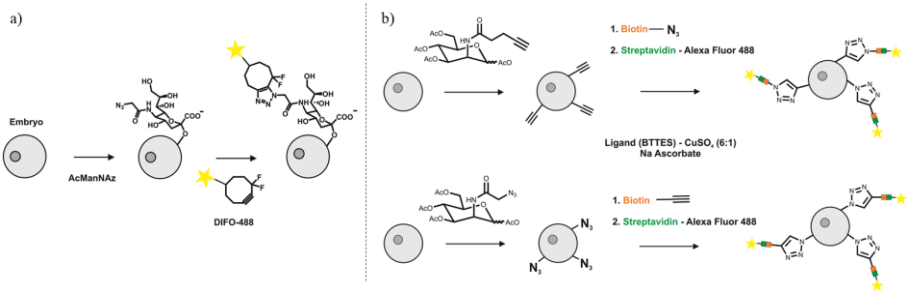
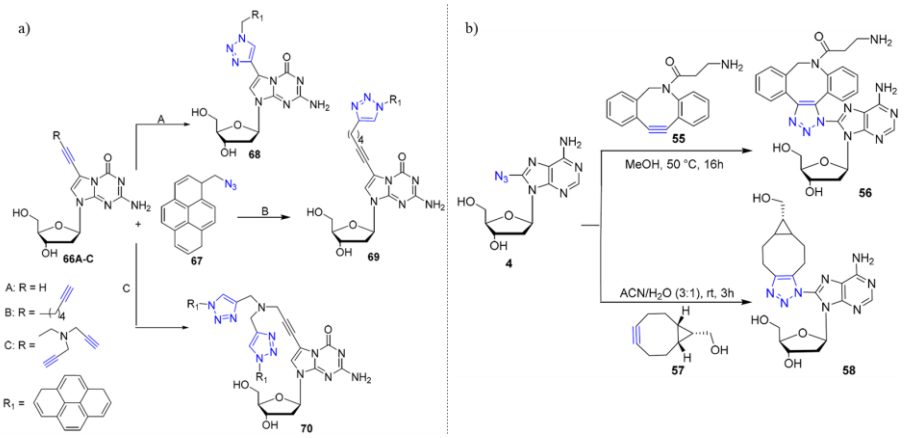


Figure 6. Cell-surface labelling via CuAAC and SPAAC reactions. [176]

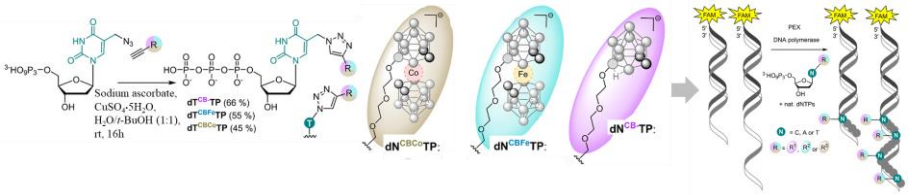
Throughout many click and cycloaddition reactions the most popular ones are CuAAC and SPAAC. The techniques are highly adaptable for nucleosides, nucleotides, nucleic acids and many other biomolecules labelling. Since the substrates for these click reactions require functional groups that are sterically small and unreactive towards other functionalities of biomolecules, the labelling of alkyne- or azide-modified nucleic acids, cell surfaces, proteins are more common than nucleosides or nucleotides. Several fluorescently labeled nucleosides were synthesized using different alkyne species bearing 5-aza-7-deaza-7-substituted-purines (Scheme 21a). Reactions were performed in *t*-BuOH/THF/H_2O (3:1:1) mixture in the presence of $CuSO_4 \cdot 5H_2O$ and sodium ascorbate. Depending on the substrate, the reaction took from 3 to 16 hours at

room temperature to result in 49 – 68% yields. [184] In attempts to design and synthesize substrates for live cancer cells (MCF-7) imaging by direct fluorescence light-up, 8- and 2-azido adenosines, 5-azido uridines and their corresponding 5'-triphosphates were synthesized. Followed by the subsequent SPAAC reaction with several different cyclooctyne derivatives bearing functional groups for further labeling (Scheme 21b). Regardless of whether nucleosides or nucleotides were applied in reaction, products were obtained in excellent yields (77 – 100%). However, in the cases of unsymmetrical cyclooctynes, regioisomers were formed in 1:1 ratio. [185]



Scheme 21. Functionalized nucleosides synthesis applying CuAAC and SPAAC.

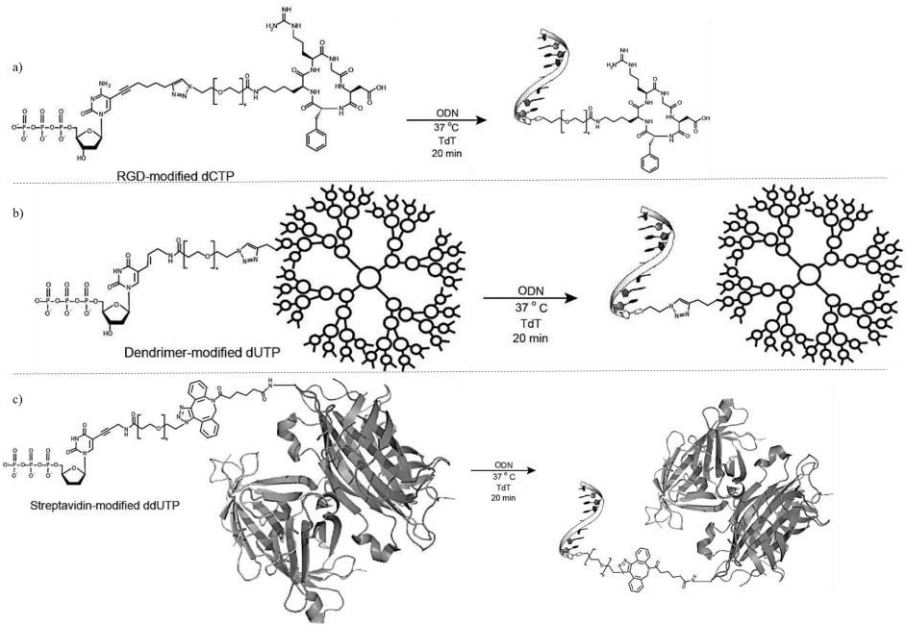
As an alternative route, nucleoside 5'-triphosphates can be labelled and subsequently used for enzymatic incorporation in growing nucleic acid strand. For this matter, several dicarba-*nido*-undecaborate groups bearing terminal alkyne moiety were reacted with azido group bearing dTTP via CuAAC. The products were obtained in 45 – 66% yields and were successfully incorporated by DNA polymerases (Scheme 22). [186]



Scheme 22. Synthesis of dicarba-*nido*-undecaborate bearing dTTPs.

Apart from relatively small labels bearing nucleotides, the bulky substituents are also introduced throughout CuAAC or SPAAC, followed by successful incorporation of such nucleotides by polymerases. Several

examples of this kind were prepared bearing integrin targeting cyclic peptide (Scheme 23a), dendrimer (Scheme 23b) and streptavidin (Scheme 23c) (45 – 80% yields). Due to the long poly(ethylene glycol) (PEG) linker, they all were substrates for TdT polymerase. [187]



Scheme 23. Nucleotides labelling throughout CuAAC and SPAAC with bulky substrates.

The growing applicability of both CuAAC and SPAAC is observed in post-enzymatic labelling of nucleic acids and other biomolecules. The introduced labels are diverse and depend on the desired function. ONs bearing terminal alkyne groups on cytosine and thymidine nucleobases were synthesized. The obtained alkynyl-modified nucleic acids were subsequently reacted with various azido groups bearing labels (dyes, biotin, carbohydrates moieties, various small labels) in the presence of CuBr, TBTA ligand and sodium ascorbate in DMSO/*t*-BuOH (3:1) media. Products were obtained in good yields (45-92%). [188]

CuAAc and SPAAC reactions are highly applied for connecting or introducing diverse labels for nanotechnology. For nucleic acids therapeutics delivery, self-assembling tails were attached at the end of ONs resulting in nanoparticles [189, 190]. Dimeric, trimeric and oligomeric DNA nanostructures were formed during multiple click reactions between modified ONs [191, 192]. Apart from nanotechnology, CuAAC and SPAAC reactions are also applied for bioprocesses imaging, aptamers selection. [193]

3.3 Influence of linker structure on nucleotides and nucleic acids labelling

Nucleotides can be modified by introducing labels into all main structural parts (ribose ring, nucleobase, phosphate moiety). Nevertheless, the nucleobase-modified dNTPs are predominantly employed. This is attributed to the acceptance of nucleobase-modified dNTPs by many DNA polymerases. [194] The incorporation efficiency can be affected by several aspects like the chemical nature of the modification, its ability to interact with the polymerase active site, its size and position at the nucleobase, chemical nature, length and flexibility of the linker connecting label with nucleobase. All these factors play a role during enzymatic incorporation and depending on modification can have a significant impact on the recognition of nucleotide as a substrate. Since the diversity of possible nucleobase labels is broad, understanding the principles of incorporation of natural and modified nucleotide analogues is of utmost importance. Even though some early empirical experiments briefly highlighted the structural aspects of efficient nucleobase-modified nucleotides incorporation, the first crystallographic structure analysis of DNA polymerase with base-modified dNTPs was reported only in 2010. [195] Marx *et al.*, have investigated several complexes of KlenTaq DNA polymerase and C5-modified dTTPs. The analyzed dTTPs were labelled by nitroxide- ($dT^{spin}TP$) and dendron-modifications ($dT^{dend}TP$) (Figure 7a). The bulkier $dT^{dend}TP$ was incorporated with 20 times higher efficiency in comparison to $dT^{spin}TP$ (Figure 7b). The orientational change of Arg660 amino acid residue induced

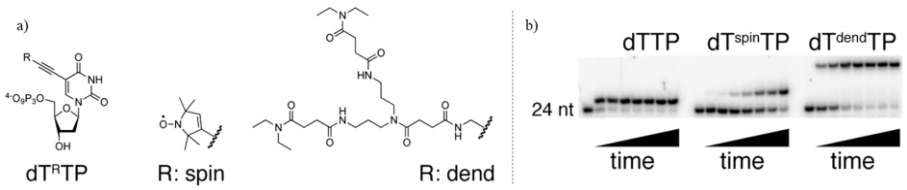


Figure 7. Incorporation efficiency measurement of $dT^{spin}TP$ and $dT^{dend}TP$ by KlenTaq DNA polymerase. a) Chemical structure of $dT^{spin}TP$ and $dT^{dend}TP$, b) Single nucleotide incorporation in a different period of time.

by bulky nitroxide, most likely prevented the stabilization of closed polymerase conformation, which resulted in the 2500-fold decrease in $dT^{spin}TP$ incorporation efficiency. Examining the KlenTaq complex with $dT^{dend}TP$, a similar Arg660 orientation change was observed. However, the branched parts of dendron modification were located outside the active site and additional hydrogen bonding between propargylamide linkage and

Arg660 was observed. This in turn, resulted in 20 times higher incorporation efficiency of $dT^{dend}TP$ over $dT^{spin}TP$. [195]

The research with more diverse linkers and KlenTaq was continued including both purines and pyrimidines. Nucleotides bearing aminopentynyl, aryl, (hydroxydecanoyl)-aminopentynyl modifications were tested. All dN*TPs (Figure 8) exhibited natural base pairing with complementary nucleotides. Moreover, the overall conformation of KlenTaq with modified nucleotides highly resembled the native one.

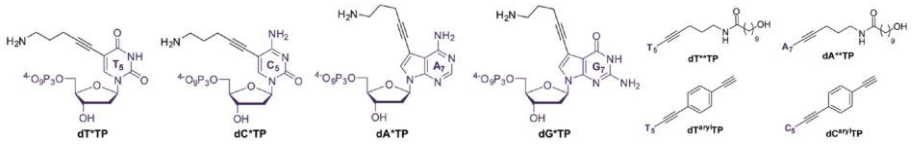


Figure 8. Chemical structure of modified nucleotides examined for KlenTag DNA polymerase incorporation efficiency.

Further examination of longer linkers, such as (hydroxydecanoyl)-aminopentynyl (Figure 8), revealed some enzyme-nucleotide complex changes between purines and pyrimidines. In the case of dT**TP, the C5-modification extended outside the protein throughout the cavity above O-helix (Figure 9), while the crystal structure with dA**TP revealed that C7-modification is orientated above the 5'-triphosphate moiety protruding through the cleft of the polymerase palm domain. This represents the ability of nucleotides bearing bulky substituents to be incorporated into DNA given that the linker is long enough to extend the modification outside the protein. [194]

Examining rigid nonpolar modifications like aryl moieties bearing nucleotides (Figure 8) similar Arg660 displacement tendencies were observed, where interaction with primer is diminished to make room for modification. However, the Arg587 and Lys663 excelled cation-π interaction with aromatic ring by additionally stabilizing the formed complex. Consequently, for 1:1 incorporation efficiency between natural dTTP and dT**TP just a 7-fold higher concentration of modified analogue was necessary. [195, 196] This research showed the polymerase ability to adopt the necessary active conformation for modified nucleotides and stabilize it. Additional interactions (like hydrogen bonding or cation-π interaction) occurring between amino acids in the active site and modification, enable the additional stabilization of protein and nonnatural substrates complex, without significant interference of the overall structure. [197, 198]

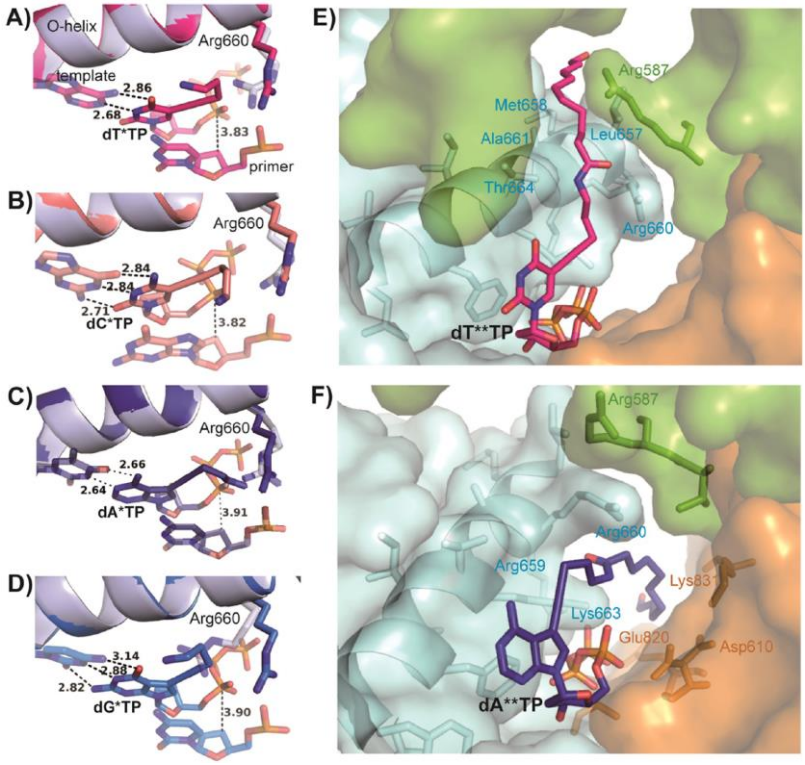
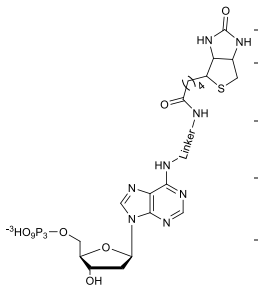
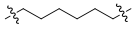
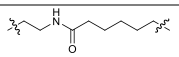
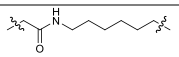
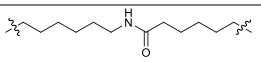
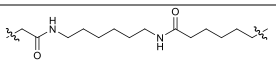


Figure 9. Structures of KlenTaq DNA polymerase bound to dN*TPs. (A-D) – representation of interactions in closed-up form of KlenTaq DNA polymerase with (A) dT*TP, (B) dC*TP, (C) dA*TP, (D) dG*TP. (E-F) – closed-up view of dT**TP (E) and dA**TP (F) with modifications pointing outside the protein. [194]

During many years of research in modified-nucleotides enzymatic incorporation, several aspects regarding linker were observed. In 1987, relative incorporation efficiencies were measured for six different linkers bearing biotinylated dATPs (Table 5). This research indicated that incorporation efficiencies of N6-substituted-dATP analogues depended on both the linker length and chemical nature. The best linker contained 7 atoms displacement from the biotin label (Table 5 Biotin-7-dATP). Likewise, the analogous N4-substituted-dCTPs were incorporated with same efficiency tendencies. [199, 200] A further research on biotinylated substrates incorporation investigated analogues bearing even longer linkers. The 22 and 36 atom arm connecting nucleobase to biotin was constructed bearing several PEG moieties as well as amide connections. However, neither of the substrates were accepted by Taq DNA polymerase, while the higher fidelity Tli DNA polymerase (also called Vent DNA polymerase) was capable to perform incorporation of such substrate. [201] All in all, the longer linker can serve for bulky modification in the way of reducing steric hindrance induced by label

excluding it out of polymerase active center. On the other hand, further attempts to prolong linkers resulted in the decrease of incorporation efficiency. Here several aspects could be accounted for: possibly the higher flexibility and length can provide the ability to form more complex structure involving label-linker, label-polymerase or label-label (in case of multiple incorporated modified nucleotides) interactions, resulting in decreased incorporation efficiencies.

Table 5. Incorporation of Biotin-substituted-dATP analogues into growing DNA strand (the incorporation reaction was maintained for 90 min).

Substrate	Linker	Name	Incorporation
		Biotin-3-dATP	26%
		Biotin-7-dATP	56%
		Biotin-10-dATP	16%
		Biotin-10'-dATP	7%
		Biotin-14-dATP	36%
		Biotin-17-dATP	7%

Besides the important role of linker length, geometry and flexibility can exert profound effects as well. Several articles were published analyzing and comparing alkyl-, alkenyl- and alkynyl- conjugation to the nucleobase to determine their impact on incorporation efficiency. [200] The poor substrate abilities were observed in the cases of *cis*-alkenyl and flexible alkyl conjugates. [202] In some cases, the alkyl-linkers bearing nucleotides were not substrates for some polymerases. [203] This discrimination of flexible alkyl- and *cis*-alkenyl-linked substrates was attributed to their unfavorable interaction with the polymerase-substrate complex interfering with its stabilization. More rigid, linear propynyl- and *trans*-alkenyl-linkers bearing nucleotides are better substrates and do not affect the substrate recognition and incorporation. [202, 101, 204] Figure 10 represents several linkers tested for C5-substituted-dUTP analogues incorporation efficiency, where alkynyl- and *trans*-alkenyl-conjugates exhibited higher efficiency with Taq DNA polymerase. [200]

Another important aspect of the triple bond attached to the nucleobase is its ability to additionally stabilize the double helix. Several ONs bearing C5-propyne-dU and C5-propyne-dC moieties were prepared for the measurement of melting temperature (Tm) of relatively short ONs. A sharp increase in Tm was observed in the presence of propyne-modification. [205] The similar

finding was observed with propynyl-purines as well. [206, 207] This was attributed to both the increased hydrophobicity and polarizability of the nucleobase. [200]

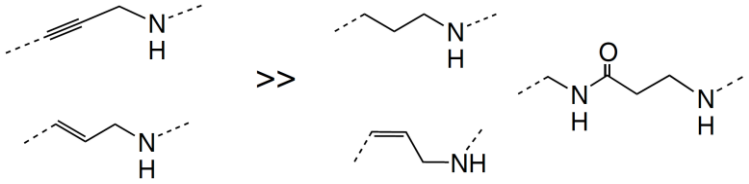


Figure 10. The preference for linker nature of C5-substituted-dUTP analogues for efficient incorporation by Taq polymerase.

4. APPLICATION OF MODIFIED NUCLEOTIDES FOR NUCLEIC ACIDS SEQUENCING

4.1 Sanger sequencing

The substantial progress in DNA sequencing was achieved with the establishment of Sanger sequencing in 1977. The 2',3'-dideoxynucleotides (ddNTPs) applied in this technique were called – chain-terminators, as the nucleic acid synthesis would stop after their incorporation due to the lack of the 3'-OH group. First chain-terminators applied by Sanger were radiolabelled ddNTPs containing ^{32}P isotope at the α-position of the 5'-triphosphate moieties. [208] This technique required four separate primer extension reactions where each contained one type of labelled ddNTP. Statistically, each reaction will result in a variety of DNA fragments terminated at each position corresponding to the specific ddNTP. Upon size-separation of DNA fragments and the analysis of the autoradiograph, the original DNA sequence can be defined. Throughout the time a number of improvements regarding Sanger sequencing were made where the most advantageous was the replacement of radiolabeling with fluorometric detection, that enabled the automatization of Sanger sequencing workflow. [209] Automated sequencing was first performed by applying dye-labelled primers containing the fluorescent label at the 5'-terminus. [210, 211] Although the technique was relatively simple it still required the four independent primer extension reactions to be performed until dye-labelled ddNTPs were developed. Here, the nucleobase-modified terminators, containing four different fluorescent labels, were applied and the revolution of structural diversity of dye-ddNTPs has begun. [212] This technique was more convenient since a single primer extension reaction per template was sufficient for the analysis. The aim in the development of dye labels was to obtain such a set that would possess high absorbance and exhibit both well-separated and strong fluorescence emission alongside high quantum yields. The most common dyes were rhodamine-based (ROX, REG, R110, TAMRA). Even though they met a majority of requirements, their quantum yields differed substantially (Figure 11a). [158, 213] To overcome this problem the energy transfer dye terminators were designed, where the transfer occurs from the excited chromophore (donor) to another chromophore (acceptor). The energy transfer efficiency can be manipulated by the small changes in the distance between two chromophores (Figure 11b). [213] Nampalli *et al.* constructed bifunctional nucleobase-modified nucleotides. To one of the functionalities the fluoresceine (donor) dye was attached while to the other – one of the rhodamine (acceptor) dyes (Figure 11b). These bulky

ddNTPs were substrates for Thermo Sequenase DNA polymerase. Moreover, labelled terminators produced an 18-fold brighter signal in comparison to single rhodamine moiety. [159, 213]

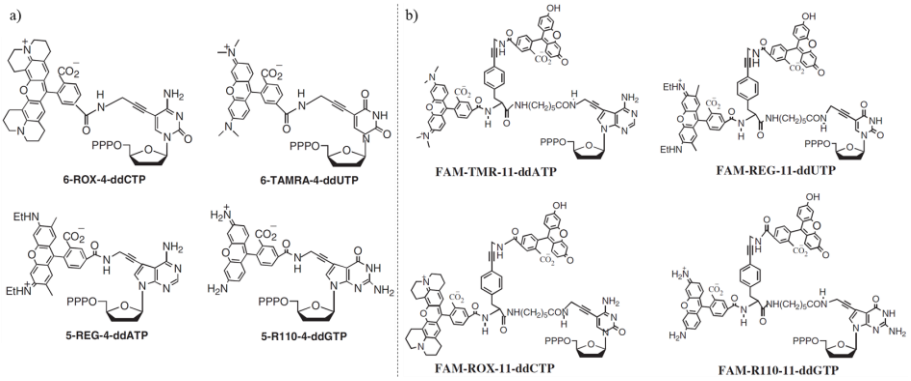


Figure 11. A four-color set of dye-labelled 2‘,3‘-dideoxynucleoside-5‘-triphosphate terminators.

4.2 Next-generation sequencing

The development of the first sequencing technique was not just a momentous breakthrough but also a significant milestone for the evolution of diverse modified nucleotides. ddNTPs have a relatively narrow application field in comparison to dNTPs or NTPs, and Sanger sequencing is one of the major fields. The research focused on improved functional abilities of ddNTPs to produce various dye labelled analogues. And to this day dye-labelled nucleotides, either ddNTPs or dNTPs, are the core for sequencing techniques. [213]

The establishment of NGS techniques has revolutionized research abilities in modern biology and biomedical sciences. The ability to perform analysis cheaply and sequence large numbers of genes or entire genomes at once made these techniques routine tools in life science research. Alongside, several dye-labelled nucleotides were developed, and various positioning of dye modifications were explored. The concept of continuous single fragment molecule sequencing without irreversible termination was achieved by various dye-dNTPs. Depending on the sequencing platform the dyes could be attached to the nucleobase, or to the phosphate group. To avoid probable interference induced by ribose ring modifications during incorporation of dye-nucleotide by DNA polymerase, several nucleotides bearing dye moiety at the terminal phosphate were developed (Figure 12). Such sequencing technique is called single-molecule real-time sequencing (SMRT, commercialized by

Pacific Biosciences), since DNA synthesis utilizing dye-nucleotides is not halted by any means. [214, 215, 216, 4]

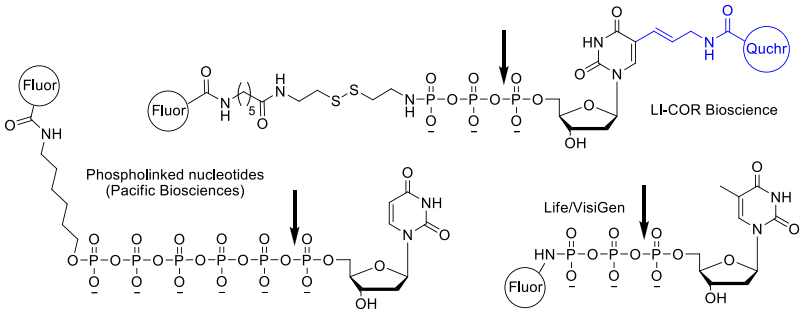


Figure 12. Modified nucleotides bearing fluorescent labels at the terminal phosphate moiety used in NGS sequencing platforms. Arrows indicates the cleavage site after incorporation of nucleotide.

Another class of dye-labelled nucleotides for sequencing techniques consists of nucleoside analogues fluorescently labelled at the nucleobase, and so called – reversible terminators. The sequencing accuracy is directly related to the detection and detector abilities. Usually, after incorporation of nucleotide a short period of time is necessary to obtain the signal that specifically indicates the nature of incorporated nucleotide. Otherwise, some data might be corrupted, consequently, decreasing the precision of sequencing. Reversible terminators are designed to produce the time pause for the detection of incorporated nucleotide. There are two main strategies of their action. The first is based on steric hindrance induced by the label attached to the nucleobase. The other one is based on the nucleotides bearing the reversible 3'-protecting group (Figure 13). LaserGen was the first to represent the reversible terminator bearing a bulky dye group that is attached closely to the nucleobase in turn inhibiting the incorporation of further nucleotides till the fluorophore is attached. After the cleavage at a specific site (Figure 13 shown with the arrow), the subsequent nucleotide can be incorporated. [217] Helicos BioSciences took the ability to temporarily halt the incorporation of subsequent nucleotide from a different angle. They hypothesized that alongside with attachment of dye moiety the additional nucleoside could be present to inhibit polymerase after incorporation until cleavage of both fluorophore and nucleoside is proceeded (Figure 13). Hereafter, so-called virtual terminators were developed and applied for the Helicos BioSciences sequencing platform. [218]

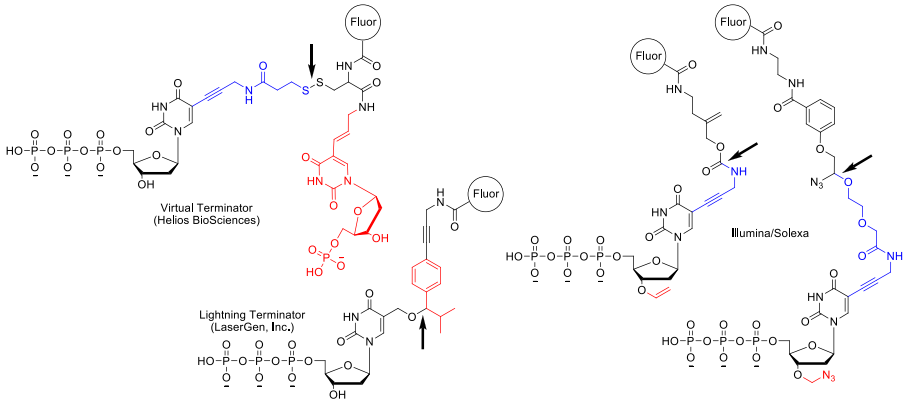


Figure 13. Reversible terminators bearing fluorophore at the nucleobase. Arrows indicates the cleavage site after fluorescent detection of incorporated nucleotide is completed.

The other sequencing platform using reversible terminators is Illumina (Figure 13). Here, dye-labelled nucleotides have the 3'-blocked -OH group and fluorophore-conjugated to the nucleobase. [219] Commonly, the blocking of 3'-position is performed through the formation of 3'-*O*-allyl- [220] or 3'-azidomethyl-groups [221]. The sequencing proceeds by applying all four nucleotides at once. After incorporation, dye-nucleotides are washed and four-color imagining is performed, followed by cleavage of dye and 3'-terminating group. The cleaved substances are washed away, and the cycle is repeated till the full complementary strand is synthesized. [4] Currently, the Illumina is the most widely used sequencing technique across the world, which dominates nearly 80 – 90% of NGS market. Illumina platform is specified to be able to sequence reads of a maximum 600 base pairs (bp) length and exhibits one of the highest base calling accuracies. One of the fundamental reasons for short read length is the structural vestiges ("scars") that remain at the nucleobase after dye cleavage. This accumulation of linker's residuals along the DNA major groove, eventually impairs the stability of the double-helix structure and can hinder the substrate. [222]

4.3 Fragment library preparation for NGS

Prior to sequencing, by any sequencing platform, some pre-processing of target nucleic acids is necessary. The fundamental aspect is to prepare the target RNA or DNA into the form that is suitable for the sequencing system of choice. Generally, the major steps in preparation of either DNA or RNA for NGS analysis consist of fragmentation of target nucleic acids to the fragments ranging in a particular size, enzymatic ligation with platform-specific adapters

of defined sequences, and amplification/quantitation of prepared libraries. [223]

The first core step in library preparation is fragmentation. The size of target nucleic acid in the final library is a critical parameter for NGS, since the sequencing platforms are defined by the capability to sequence fragments of certain length. Insert size (sequence between adapters) determines desired library size. There are three approaches used for nucleic acid fragmentation: physical - acoustic shearing and sonication, enzymatic – application of non-specific endonuclease mixture, chemical - complexes that generate hydroxyl radicals (such as iron-EDTA) [224]. Sonication and enzymatic fragmentation are the most commonly used methods. The enzymatic fragment preparation strategy is quite efficient, although, many nucleases exhibit certain levels of undesirable sequence bias. Consequently, sonication is preferred due to the robust and controllable generation of randomly sheared fragments of the desired size range. [225]

After fragmentation, the obtained DNA fragments must be prepared for ligation. This typically includes end-repair, phosphorylation of the 5' prime ends and A-tailing of the 3' ends. End-repair step is necessary to blunt any protruding DNA ends that might have been formed during fragmentation. Subsequently, 5'-end has to be phosphorylated applying kinase enzyme, and in some strategies A-tailing can be performed by either Taq polymerase or Klenow Fragment exo- to minimize undesirable intermolecular ligation and formation of adapter dimers. The ligation of specific adapters is of great importance since they are necessary for the sequencing platform. Enzymatic ligation of adapters is notorious for low efficiency [226] leading to decreased complexity of the original library and impoverishment of sequencing results. During adapter ligation, the ratio between fragments and adapters should be considered wisely, since too high excess of adapters can result in the formation of adapter dimers. Consequently, dimers would interfere with the amplification of target fragments. Finally, several cycles of PCR amplification are required for library enrichment alongside quantification. During the PCR different barcodes can be introduced to enable multiplexing.

The alternative method for the library preparation is available utilizing transposases. Here, a so-called tagmentation reaction is performed, where fragmentation and the simultaneous addition of adapters is carried out. [227] Such library preparation procedure, indeed, simplifies the whole workflow. However, the technique is highly sensitive to the amount of target DNA (in comparison to standard fragmentation), the ratio between the DNA sample and the transposase enzyme is crucial, and so is the fine tuning of all other

reaction parameters (temperature, time, etc.), since fragment size is dependent on reaction efficiency. [223]

The library preparation from RNA fragments is of even higher complexity and more time-consuming procedure than that for DNA. The vast majority of sequencing techniques are exceptionally customized for sequencing of DNA fragments. Consequently, RNA transcripts must be converted to complementary DNA (cDNA) via the reverse transcription (RT) before the fragment library preparation. [223, 228]

4.4 Alternative ligation techniques: chemical ligation

The first attempts to adopt chemical ligation for nucleic acids were not directed to simplify the library preparation for sequencing, but rather to develop techniques for the synthesis of long ONs. The solid-phase synthesis of ON has several challenges: imperfect coupling efficiency, side reaction resulting in chemical modifications, shorter oligonucleotides formation, limitations of the final length of the product. The longer the ON chain has to be synthesized the higher risk of mistakes in sequence and side reactions arises, alongside increased complexity of their detection and oligonucleotides purification. However, the nucleotide mismatch in the ON sequence is critical for the following enzymatic applications. Consequently, the chemical ligation of shorter ONs to produce long analogues seemed very appealing. Brown group came up with the idea of mimicking the natural phosphodiester linkage of oligonucleotides with alternative triazole moiety. Since azide-alkyne click chemistry was compatible with aqueous media, while the reaction itself is efficient, high yielding, rapid, suitable for biomolecules labelling, such chemical ligation approach met all the necessary requirements. The fundamental of this conjugation was to provide a biocompatible linkage for the subsequent enzymatic manipulations (e.g., PCR) and produce the natural nucleic acids during amplification. However, it was already known that triazole itself is not recognized by polymerases and no read-through was observed. For this matter, the 3’-azidothymidine (AZT) was attached to the 3’-terminus of the ON during solid-phase synthesis (Figure 14a). It is important to mention that the azido-group is not compatible with the P(III) chemistry which is applied during solid-phase ON synthesis, consequently, the conditions of reaction had to be adopted. The other ON at the 5’-terminus contained the 4’-propargylamidothymidine (Figure 14a). After the CuAAC reaction the triazole-linked (ribose-to-ribose connected) ONs were obtained and tested in PCR. Several DNA polymerases like GoTaq, Taq and Pfu were tested and showed ability to produce the desired amplification (Figure 14a)

through non-natural linkage. This, in turn, proved that the designed linker was biocompatible. To ascertain the fidelity of read-through, the sequencing of PCR products was performed. Regardless of the polymerase used in PCR, all the products had to produce the complementary strands for CT-triazole-TTC (giving GAAAG). However, all of them resulted in read CTTC (giving GAAT). In attempts to explain this phenomenon, many hypotheses were formulated, mainly relying on the amide bond. The authors explained this as if the polymerase incorporates the first thymidine ahead of the linkage while skipping the second one. The other theory was that the rigidness of the amide bond may cause a rotation of the thymidine, directly attached to the triazole, away from the template (Figure 14b). Although this proved that the linker was biocompatible and complementary strand can be enzymatically synthesized, the nucleotide skipping event was not satisfactory. [229]

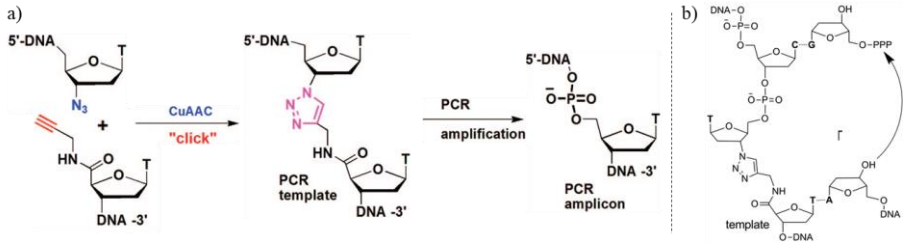


Figure 14. Chemical ligation of the oligonucleotides reacting AZT-ON and 4'-propargylamido-ON via CuAAC.

One more drawback of the present linker (Figure 14) was addressed to the lack of 3'-oxygen and 5'-methylene groups as recognition sites that could have a negative effect on substrate suitability for the polymerase. Consequently, by the same research group, the other linkage satisfying the mentioned requirements was designed. Here ON bearing 5'-azido group was coupled with the ON having 3'-O-propargyl group. The obtained ONs were PCR amplified, to produce the desired products, where the sequencing resulted in the correct sequence. However, it was noted that amplification during PCR might be efficient regardless of the fact of possibly rare read-through event. To prove the triazole linkage applicability *in vivo* the plasmids containing triazole linkage in both strands of its antibiotic marker gene were constructed and transformed into *E. coli.* After cell growth on LB-agar plates containing ampicillin the colonies from triazole-modified plasmids constituted 96.5% of the colonies transformed with native DNA. In general, the growth and survival of colonies that contained antibiotic marker gene linked with non-natural triazole linker proved that the gene is amplified by *E. coli* polymerase. [14]

In attempts to simplify the RNA library preparation the alternative backbone for ribose-to-ribose connection was applied as well. During the reverse transcription of viral RNAs or mRNA with random primers containing Illumina adapter, 3'-azido-ddNTPs were utilized for chain termination to produce specific length distributed cDNA fragments. Subsequently, 5'-hexynyl-oligonucleotides containing another Illumina adapter were applied in CuAAC reaction to produce cDNA products labelled by both Illumina adapters and containing artificial ribose-to-ribose connection (Figure 15). In this manner, the limiting and challenging enzymatic ligation of adapters is bypassed via chemical ligation. However, the reaction did not result in full conversion of starting ONs since a residual amount of 3'-azido-cDNA was observed and only 10% of 3'-azido-cDNA were successfully ligated. They also demonstrated that by varying the amount of 3'-azido-ddNTP the libraries of different fragment lengths can be generated. The obtained click-linked cDNA fragments were subsequently applied in PCR to obtain cDNA libraries, that were sequenced on an Illumina instrument. Afterward, results were

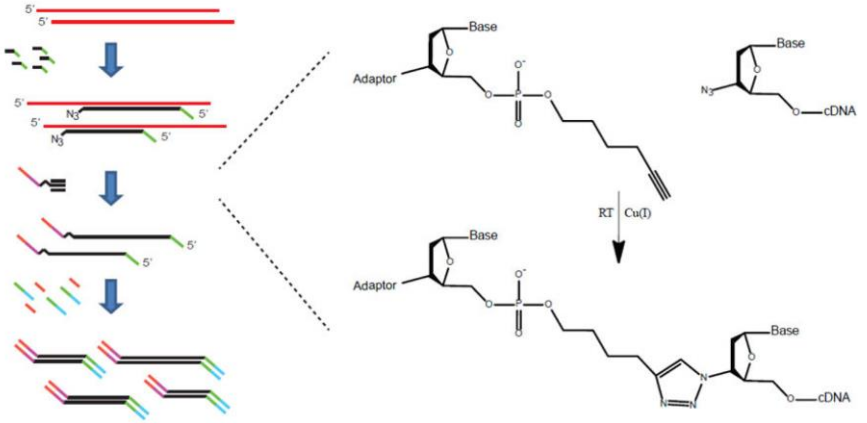


Figure 15. RNA library preparation workflow of Click-Seq approach for NGS Illumina platform.

compared with the sequencing data obtained from the same RNA prepared with another library preparation kit. Although the observed mistake rate was slightly higher in the case of triazole linkage, the obtained data quality was in the range for base-calling errors for Illumina platform. This so-called Click-Seq library preparation technique was claimed to exhibit an important advantage over traditional techniques in terms of the reduced formation of artificial recombination events, during which the sequence chimeras are formed. This was directly attributed to the chemical ligation itself since the alternative fragmentation is obtained by sequence termination with 3'-azido-

ddNTP, subsequently allowing fragments and adapters to be ligated only via click reaction, while terminated sequences cannot serve as priming substrates for template switching. However, the technology has two main and highly important drawbacks. The chemical ligation is far from being efficient, since only 10% of triazole-linked cDNA is formed. Read-through the unnatural linkage is also a rare event where its efficiency could not be measured since it was lower than the fluorescence detection limit of the method used and was attributed to being less than 4%. [16]

The Click-Seq technology was also applied for the examination of alternative polyadenylation (APA). APA is gaining more attention after discoveries of multiple potential cleavage sites of mRNA and polyadenylation is subsequently increasing the different mRNA isoforms of length. There are several known sequence elements inducing such cleavage, however, the uncovered cleavage events were also observed. This gained much attention and driven the desire to understand the APA mechanism, poly(A) tail position coherence to mRNA stability, degradation, and translation. Here oligo-dT primers containing p7 Illumina adapters were used for cDNA synthesis in the presence of low amount of 3'-azido-ddGTP, 3'-azido-ddCTP and 3'-azido-ddATP (Figure 16a). After fragments generation the chemical click ligation was applied to attach the p5 Illumina adapter. Finally, obtained cDNA library was amplified during PCR and sequenced (Figure 16b). The technique was robust and enabled accurate capture of authentic polyadenylation sites via highly simplified workflow. [230]

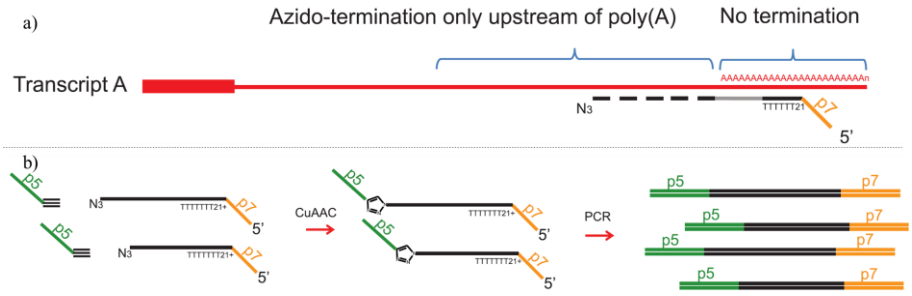


Figure 16. Schematic representation of PAC-seq technique: a) cDNA fragments preparation, b) chemical click ligation and library amplification.

Other applications of chemical ligation include library preparation for targeted sequencing, epigenomic profiling and even niche applications, such as sequencing of single-stranded DNA. For example, the investigation of deletion, insertion or single-nucleotide variations (SNVs), which is RNA recombination effects mostly attributed to the evolution of viruses and adaptations, was performed with the clinical samples of SARS-CoV-2. The

cDNA amplicons were formed in the similar matter as described previously. This technique proved to be potent, since full-length SARS-CoV-2 genomes were reconstructed, alongside with determination of either unique of already known SNV and recombination events. [231]

For epigenetic analysis of nucleic acids Mikutis and colleagues suggested interesting approach performing epitranscriptomes sequencing involving click chemistry. Methylation of RNAs is performed by diverse methyltransferases (MTase), moreover it is known that RNAs located in cancers cells characterize themselves with the greater degree of methylation, that play a pivotal role in maintenance of cancers. Mukitis and colleagues proposed the simplified way to determine the methylation region of specific MTase (METTL3 and METTL16) *in vivo* analyzing several cancer cells RNAs (Figure 17). Alkyne moieties were introduced to RNA by MTases of interest utilized SeAdoYn analogue followed by subsequent click reaction with azido-degrader. The method is based on the click-degrader – the PEG chain bearing azide with terminal imidazole moiety, that after click reaction of propargylated RNA result in the cleavage of RNA in the way of ribonuclease activity (Figure 17). Sequencing data obtained from transcriptome allowed to identify specific target sites of MTase. Moreover, the targeted cleavage of RNA was implemented via CuAAC in live cells, while lysed RNAs do not require any further preprocessing of RNAs *in vitro* prior to library preparation. [232]

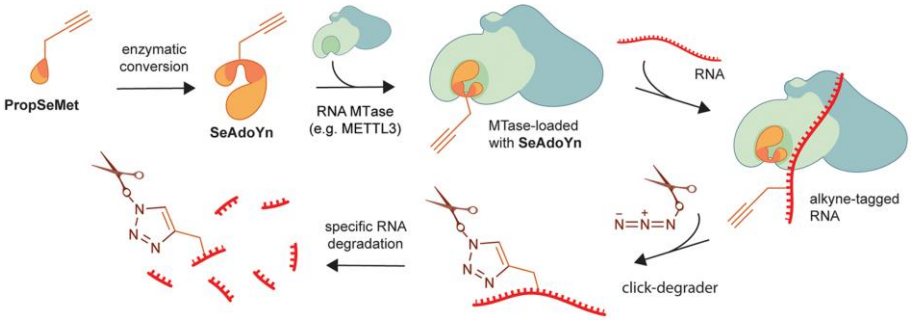


Figure 17. Proposed mechanism of small molecule-based methylated RNA editing platform of meCLICK-Seq.

Another example of epigenetic profiling utilizing chemical ligation approach was developed by Stasevskij and colleagues. Here, as in the previous example, methyltransferase (SssI) was also applied, only in this case for transfer of azido-moieties (by modified SAM) in order to determine the unmodified CG sites in DNA (Figure 18). Afterwards, the internal priming was achieved by tethering ON duplexes to the azido-CG sites. The following reading-through mechanism was not investigated. Nevertheless, the strand

invasion and template switching, requiring read-through event, produced the successful synthesis of complementary DNA. The obtained result analyzing bacterial and human genome with so called TOP-seq (tethered oligonucleotide-primed sequencing) technique showed better agreement with published bisulfite sequencing database in comparison to other commonly used techniques (MBD-seq, MRE-seq). [233]

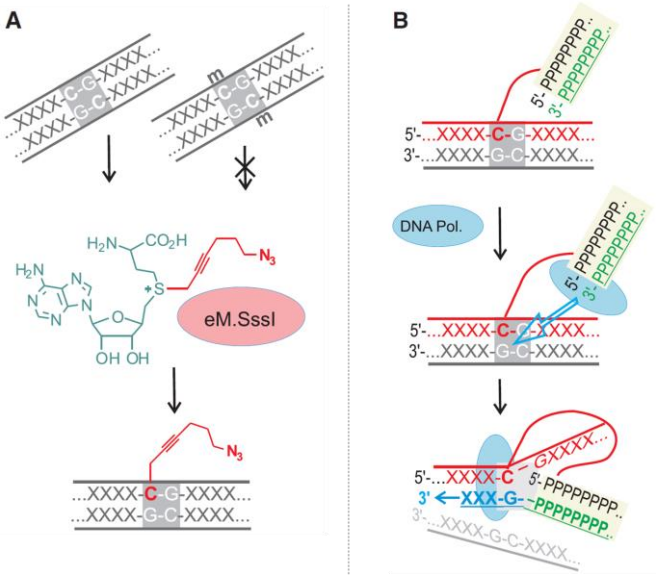


Figure 18. Schematic representation of tethered oligonucleotide-primed sequencing (TOP-seq), a) selective tagging of unmodified CG sites, b) DNA polymerase activity at the tethered oligonucleotide-primed CG site.

For DNA library preparation the chemical click-based ligation of platform specific adapters was also applied. In 2018 reported article by Ito group the TdT (terminal deoxynucleotidyl transferase) was used to add single 3'-azido-ddNTPs at the ssDNA termini. The scientist declared observations regarding the importance of distance between 3'- and 5'-positions in ribose-to-ribose chemical ligation. They stated that the natural nucleotides connection in nucleic acids contains five bonds and adjacent throughout the triazole should maintain this criterion. Consequently, they used 5'-ethynyl-ONs bearing sequencing platform specific adapter in the CuAAC reaction, subsequently generating chemically ligated ssDNA. However, several difficulties were declared. Firs, the chemical ligation devastate in low efficiency (20-30% conversion). The efficient ligation of two ONs could be proceeded only by performing templated ligation, where complementary strand is used for 3'azido-ON and 5'-ethynyl-ON hybridization to bring closely the reactive

groups. However, such approach is inapplicable for unknown 3'-terminal sequence ONs ligation. The slight degradation of ONs was observed, and it is crucial since reaction is performed with DNA fragments of interest. Here, degradation can compromise subsequent library amplification, generate artifactual fragments and impoverish sequencing data. The only Klenow fragment was capable of synthesizing identifiable amount of complementary strand.

Library preparation workflow still remains quite complicated (12 steps) (Figure 19 A). Finally, the fragment library was produced after PCR amplification. The overall efficiency of this protocol was valued by authors to be 0.6% (assuming that all PCR cycles performed in 100% efficiency). Analyzing read length authors observed that more than a half (57.8%) of reads were shorter than expected (102 nt) (Figure 19 B), the possible cause was attributed to be DNA degradation during click ligation. Analyzing sequencing

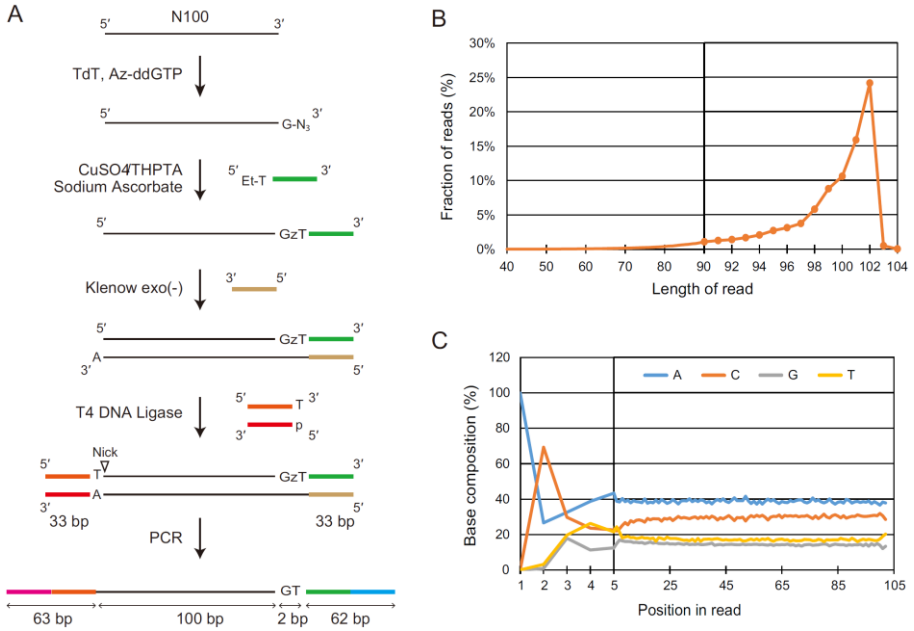


Figure 19. Library preparation workflow applying TCS ligation from ssDNA. A – schematic representation of library preparation workflow. B – read length distribution after trimming of adaptor sequences. C – mean base composition of reads at each position.

data, the correct incorporation of first nucleotide at the site of nonnatural linkage was 99.9%, which was expected to be. Interestingly, the second nucleotide was just 69.4% C, representing of incorrect read-through. Since in the model system the deletion event at the 3'-adjacent site of triazole linking

was a rare event, this was attributed to the nucleotide substitution where A was the most often case (26.6%) (Figure 19). The method was termed terminal deoxynucleotidyl transferase (TdT)-assisted, cycloaddition CuAAC-mediated ssDNA ligation (or TCS ligation). [15]

Although published techniques exhibit interesting advantages, they all face the same limitation related to the inefficient replication of click-ligated DNA. Therefore, the exploration of better molecular designs is important for the further development of this field. Nevertheless, and more importantly, techniques have extensive potential and opened doors for exploration of more diverse possibilities to simplify nucleic acids sequencing. Due to the ever-growing demand for rapid, inexpensive and straightforward nucleic acids sequencing and library preparation techniques, a multidisciplinary coalition of chemistry and molecular biology, not only empowers new possibilities, but expands the exploration of previously unattainable fields and biological insights.

RESULTS AND DISCUSSION

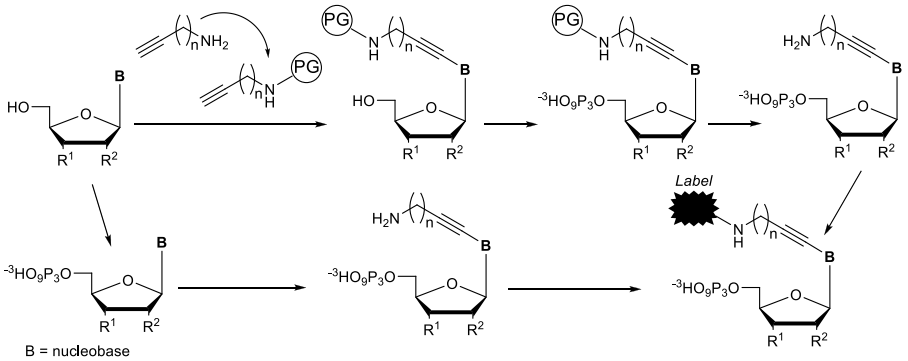
There are two main challenges in regard to this work: the empirical design of oligonucleotide-tethered nucleotides (OTDNs and OTDDNs) that would provide both efficient enzymatic incorporation of bulky oligonucleotide-modification bearing nucleotides and to enable read-through event during the polymerase induced synthesis of complementary strand. The fulfillment of these requirements relies on the conjugation approach and the design of linker connecting nucleotides with oligonucleotides. The ribose ring modification by introduction of bulky groups is usually hard, as numerous polymerases do not accept nucleotides bearing such modifications on the sugar moiety as substrates. However, the introduction of various modifications, regardless their size, have been successfully accomplished throughout nucleobase modification, without a loss of substrates ability for many polymerases.

In case of large labels attachment and such nucleotides incorporation, usually, the main challenge to overcome is to avoid steric hindrance caused by the label during the enzymatic nucleotide incorporation. Here, the linker connecting the nucleobase and label plays a pivotal role. As it already was described in the introduction and literature overview (chapter 3), several important criteria of linker design have to be considered. First, the position for introduction of modification has to be chosen (C5 for pyrimidines and C7 for purines), followed by the formation of C-C bond, where attachment of triple bond was found to be most effective (best accepted by polymerases). Secondly, introduction of reactive/functional group, that could be easily reacted applying chemical methods not intervening with other functionalities, has to be chosen. Finally, linker have to be long enough to avoid label-induced steric hindrance during enzymatic incorporation, but also, have to be length-compatible for polymerase to read through it. Considering all this, we designed functionalized azido-dNTPs and ddNTPs with desired linker properties, as substrates for Cu (I)-catalyzed alkyne-azide cycloaddition (CuAAC) with alkyne-oligonucleotides. Propargylamine was chosen for linker formation since it can be easily used for C-C bond formation via the Pd-catalyzed coupling. More importantly, the reactive amine group may serve for further modification, while the triple bond next to the heterocyclic base unit is not reactive in the following click chemistry. Finally, several diverse azido-group bearing linkers were selected to examine the influence of the linker length on the efficiency of both modified nucleotides incorporation and read-though unnatural linkage.

5. OLIGONUCLEOTIDE-TETHERED 2'-DEOXYNUCLEOTIDES SYNTHESIS AND ENZYMATIC EVALUATION

5.1 Azido-group bearing 2'-deoxynucleotides synthesis via conventional synthesis strategy

Nucleobase-labeled nucleotides are usually synthesized through linear synthetic pathway (Scheme 24a). This synthetic strategy was developed many years ago and became a fundamental path for nucleobase-labeled NTPs, dNTP and ddNTPS synthesis. The strategy was patented alongside with the dye terminators (dye-ddNTPs) for Sanger sequencing. Bulky labels like dyes, macromolecules, enzymes, nucleic acids, etc., usually, require some linker conjugation in between nucleobase and label. Meanwhile, formation of the linker requires several synthesis steps depending on the desired length and chemical nature. For example, dye terminator synthesis starts with the formation of the first part of the linker, containing a functional group for subsequent modification. The halogenated nucleoside is reacted via Pd-catalyzed cross-coupling reactions (Sonogashira, Suzuki, Heck, etc.). Here, bifunctional substrates are used such as compounds bearing both alkyne and amine functional groups at the opposite termini (e.g., propargylamine). However, the nucleophilic functionalities (like amines) selected for the following attachment of the label must be protected, since those will be reactive during the subsequent triphosphorylation reaction.



Scheme 24. Synthetic strategies of nucleobase-labeled nucleotides, a) conventional synthesis path, b) alternative three steps synthesis strategy.

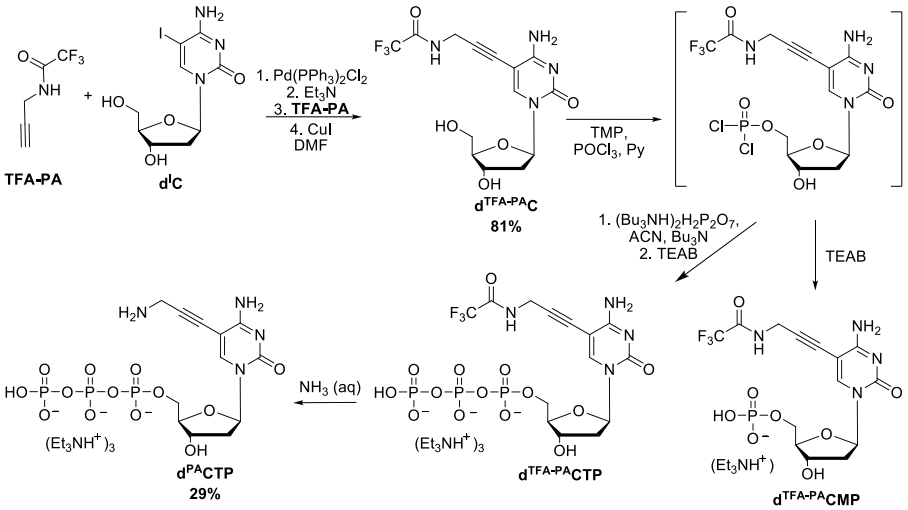
The other challenge regarding the synthesis of nucleobase-modified nucleotides mainly relies on efficient triphosphorylation of partly labeled nucleoside. This step is low yielding and highly dependable on the modification nature. When the desired triphosphate is obtained the

deprotection of functional group is proceeded. Finally, depending on anticipated linker length and structure the following step may be either the formation of longer linker or direct attachment of the label (Scheme 24a). All in all, this synthetic strategy consists of at least five steps and commonly results in less than 10% overall yield.

After development of water-soluble ligands for Pd-catalyzed reactions an alternative synthetic strategy was proposed, consisting of triphosphorylation of halogenated nucleoside, followed by an aqueous-phase cross-coupling reaction for initial linker introduction and, finally, by the direct attachment of the desired label (Scheme 24b). This straightforward, less time-consuming alternative became widely investigated for various modifications. The advantage of this pathway is the ability to overcome the triphosphorylation of reactive functional groups bearing nucleosides, that may need additional optimization depending on the modification nature.

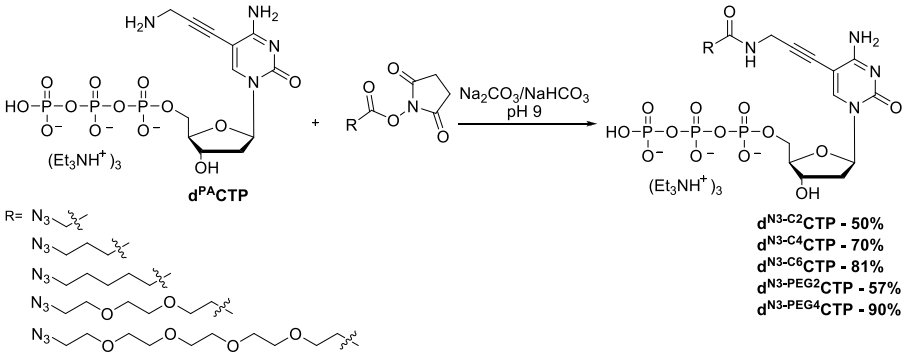
For the synthesis of propargylamine-functionalized dCTP both strategies were tested, beginning with conventional method. The synthesis started from the classical Sonogashira cross-coupling reaction in DMF media, using $Pd(Ph_3)_2Cl_2$ and reacting iodinated-dC with 2,2,2-trifluoro-N-(prop-2-yn-1-yl)acetamide (TFA-PA) (Scheme 25). The reaction required 4 equivalents of protected propargylamine and after 16 hours the full completion of the starting material, was achieved. After purification on silica gel (by column chromatography) the product $d^{TFA\text{-}PA}C$ was obtained in 81% yield, however, several disadvantages remained. Firstly, the purification from residual propargylamine required either a large silica gel column or several consecutive purifications (in both cases resulting in losses of target material). Secondly, the slow decomposition of product was observed. Nevertheless, the biggest challenge was the efficient monophosphorylation of $d^{TFA\text{-}PA}C$. The product was slightly yellow, while after addition of $POCl_3$ the color change was noticed without any differences in HPLC analysis. This proposed an assumption of the occurring reaction between decomposition products or residual propargyl amine and phosphorylation agent. This could partly explain the necessity of higher $POCl_3$ amount for $d^{TFA\text{-}PA}CMP$ formation. Several monophosphorylation reaction conditions varying in reaction temperature, amount of $POCl_3$ and base additive was tested. Unfortunately, the compound was poorly reactive when 2 equivalents of $POCl_3$ were used. If the reaction proceeded at room temperature, rapid impurities formation was observed. When reactions were performed without the base, even with additive amount of $POCl_3$ (5 eq. in total), only traces of $d^{TFA\text{-}PA}CMP$ were detected. The best results were obtained utilizing 2 eq. of pyridine (Py) as the base and 2 eq. of $POCl_3$. After maintaining reaction for 16 hours at (-20) – (-15) °C, 2 additional

equivalents of $POCl_3$ were applied leading to the 64% conversion to the $d^{TFA\text{-}PA}CMP$. However, several distinctive impurities alongside with diverse side products formation was noticed. After addition of TBAPP and TBA mixture in acetonitrile the desired triphosphate was difficult to identify from other impurities due to similar retention time on HPLC.



Scheme 25. $d^{PA}CTP$ synthesis according to the conventional synthetic strategy.

Aqueous ammonia was subsequently used for amino group deprotection without any further purification needed. The desired triphosphate $d^{PA}CTP$ was obtained only in 29% yield. The process lacked repeatability; it was clearly noticed between different batches of $d^{TFA\text{-}PA}CTP$ synthesis. This confirmed the assumption that the residual propargyl amine after Sonogashira cross-coupling and/or decomposition derivatives of $d^{TFA\text{-}PA}C$ interacts with $POCl_3$ and, consequently, the monophosphorylation reaction conditions may require seldom customization with each synthesis. Despite that, the reaction scale-up was successfully performed (0.45 mmol) producing the necessary amount of $d^{PA}CTP$ for the following labeling procedure (Scheme 26). Next, azido moieties were introduced utilizing different aliphatic and polyethylene glycol chains bearing azido-NHS esters in aqueous sodium carbonate/sodium bicarbonate solution (pH 9). In several hours the desired azido-labeled dCTPs were obtained yielding 50 – 90%. The diverse yields could be partly attributed to the utility of $d^{PA}CTP$ from different batches and its decomposition during the time.



Scheme 26. Synthesis of diverse azido-labeled dCTPs.

In general, for the preparation of nucleotides bearing reactive functional group, such synthetic strategy results into a multistep procedure, taking a long time and resulting in only 13% overall yield of $d^{PA}CTP$ (including the preparation of TFA-PA obtained in 55% yield). Such outcome was not satisfying, especially keeping in mind that additional synthetic steps for either further formation of the linker and/or direct labels attachment are necessary.

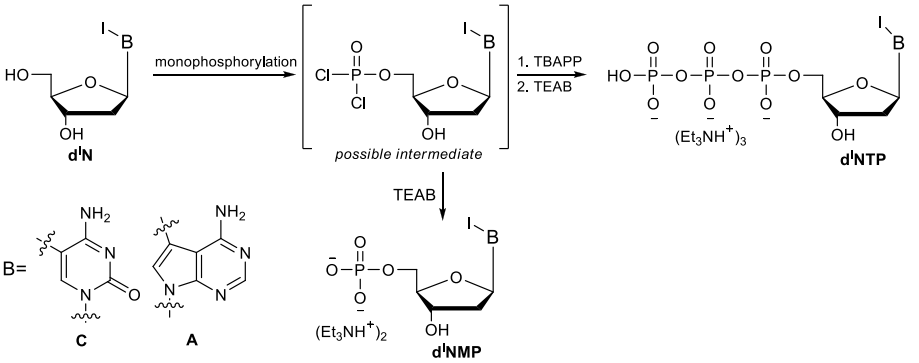
5.2 Azido-group bearing 2'-deoxynucleotides synthesis via alternative and efficient synthetic pathway

5.2.1 Optimization of phosphorylation reaction conditions for iodinated 2'-deoxynucleosides

The alternative three steps approach starts with iodinated dNTPs ($d^{I}NTPs$) synthesis. According to the literature data, usually, the standard conditions are applied resulting in moderate yields. Consequently, the triphosphorylation step was identified as the limiting and optimization of $d^{I}CTP$ and $d^{I}ATP$ synthesis was carried out.

Regarding all possible phosphorylation strategies discussed in the literature overview (chapter 1), in general, preparation of triphosphates can be divided into two main parts. First, formation of monophosphate active intermediate – the limiting and yield decisive step. Second, triphosphate synthesis, commonly, applying pyrophosphate salts. Our goal was to propose a simple, one step or so called one-pot synthetic methodology (Scheme 27).

We started our research with the optimization of monophosphorylation reaction for iodinated 2'-deoxynucleosides ($d^{I}Ns$), namely 5-I-dC ($d^{I}C$) and 7-deaza-7-I-dA ($d^{I}A$). The most common and probably first to test phosphorylation agent is $POCl_3$. Performing reaction at room temperature using 1.2 ekv $POCl_3$ only 19% conversion to the desired $d^{I}CMP$ was observed. Longer reaction times did not improve the result. On the contrary, diverse



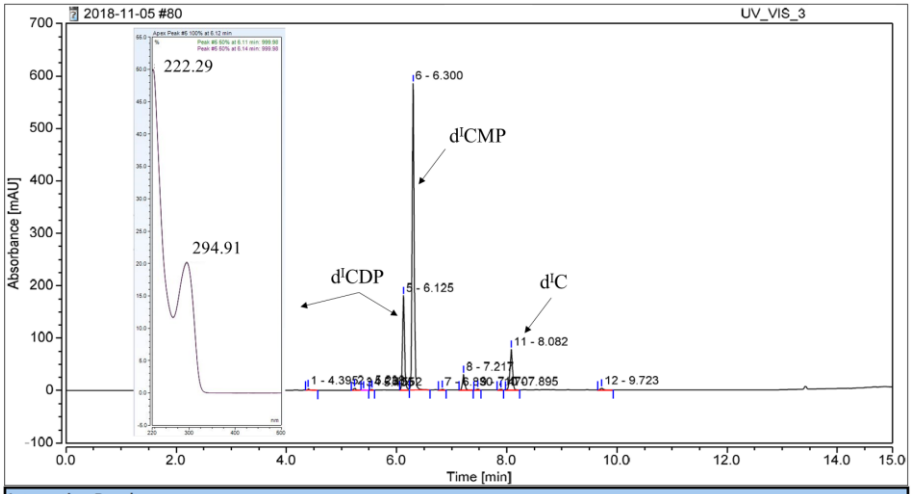
Scheme 27. Synthesis of d^{I}NMPs and d^{I}NTPs.

impurities formation was observed (Table 6, entry 1). To reduce amount of side products further reactions were carried out at 0 °C. Since some research articles attributed the addition of water as the reaction accelerating factor, 0.1 eq. of water was used, however, without any increasement in product formation (Table 6, entry 2). Slightly better results were obtained when reaction was performed maintaining the temperature at -10 to -5 °C. Even though less impurities were observed, the reaction proceeded sluggishly. This can be attributed to low solubility of the starting material under certain reaction temperature (Table 6, entry 3). Increasing $POCl_3$ quantity to 2 or 3 eq. had no significant effect on conversion to d^{I}CMP (Table 6, entry 4-5).

The other factor usually attributed to the reaction rate acceleration is the base additive. The most common bases used in phosphorylation reactions are either tributylamine or proton sponge, in some cases the utility of pyridine is also found to be beneficial. Moreover, the addition of base is also known to reduce impurities formation in some cases and prevent glycosidic bond decomposition. The tributylamine (TBA) was added equivalently to the phosphorus oxychloride amount. An increase of the reaction rate was observed. However, a formation of diverse impurities was also noticed, and only 29% conversion to the product was achieved (Table 6, entry 6). Interestingly, 3 equivalents of both base and $POCl_3$ proceeded in reaction accelerating manner. After half an hour a 56% conversion to the product was already observed alongside with distinctive impurity (Table 6, entry 7 and Figure 20). Analyzing selected reaction mixtures by mass spectrometry (MS) we identified molecular mass corresponding to the branched 3',5'-diphosphate (3',5'-d^{I}CDP). The enhanced formation of mentioned compound was observed alongside with higher amount of phosphorylation agent and base used (Figure 20). Different ratios of $POCl_3$ and TBA were also tested, but none of the experiments produced any significant improvement regarding

the conversion to d^{I}CMP. Meanwhile, the increase in reaction concentration resulted in such poor d^{I}C solubility, that no product formation was observed (Table 6, entries 8 and 9).

Since phosphorus oxychloride produced just moderate conversion to the target monophosphate along with diverse impurities, it was necessary to use other phosphorylation agent. However, the utility of commonly used P(III) derivatives would inevitably bring the need to protect the reactive groups of both ribose ring and nucleobases. These additional synthetic steps were undesirable. During the investigation of possible phosphorylation agent alternative, we found a pyrophosphoryl chloride ($P_2O_3Cl_4$). To our knowledge, $P_2O_3Cl_4$ was particularly unexplored. Only a few articles described synthesis of nucleoside polyphosphates or monophosphates using $P_2O_3Cl_4$. Nevertheless, this phosphorylation agent did not find its place in everyday nucleotides synthesis.



Integration Results

No.	Peak Name	Retention Time min	Area mAU*min	Height mAU	Relative Area %	Relative Height %	Amount n.a.
5		6.125	7.420	180.249	19.26	20.29	n.a.
6		6.300	25.089	584.627	65.11	65.82	n.a.
11		8.082	4.020	77.060	10.43	8.68	n.a.

Figure 20. HPLC analysis of reaction mixture during phosphorylation of d^{I}C. Sample taken after an hour where 3 eq. of both TBA and $POCl_3$ were used (Table 6 entry 7).

Performing reactions in TMP media using either 2 or 3 equivalents of $P_2O_3Cl_4$ caused the immediate gelation of the reaction mixture with no product formation (Table 6, entries 10 and 11). The additive of base had no influence over reaction course as well. Considering the presumably occurring side reaction of TMP and pyrophosphoryl chloride we further proceeded

phosphorylations in acetonitrile. Applying 2 eq. of pyrophosphoryl chloride 70% conversion to d^{I}CMP was obtained in one-hour time (Table 6, entry 12).

Table 6. Optimization of d^{I}C monophosphorylation reaction conditions.

Entry	Phosphorylation agent	Quantity of phosphorylation agent, eq.	Solvent	Base, eq.	Concentration	Time	Conversion to d^{I}CMP, %	d^{I}C, %
1[a]	$POCl_3$	1.2	TMP	-	0.1	1h	19	63
2[a,b]		1.2	TMP	-	0.1	1h	19	68
3[c]		1.2	TMP	-	0.1	2h	33	63
						4h	35	59
4		2	TMP	-	0.2	1h	22	74
						3h	38	56
5		3	TMP	-	0.2	1h	31	62
						3h	44	40
6		2	TMP	TBA, 2	0.2	1h	27	26
						3h	29	22
7		3	TMP	TBA, 3	0.2	30min	56	28
						1h	65	11
						3h	58	2
8		3	TMP	TBA, 3	0.3	3h	≤ 3	97
9		3	TMP	TBA, 3	0.4	3h	≤ 3	98
10	$P_2O_3Cl_4$	2	TMP	-	0.1	3h	gelled	99
11		3	TMP	TBA, 2	0.1	3h	gelled	98
12		2	ACN	-	0.1	30min	56	31
						1h	70	9
13[d]		3	ACN	-	0.1	1h	53	5
						3h	48	2
14		2	ACN	TBA, 1.3	0.1	10min	59	26
						1h	60	25
15		3	ACN	TBA, 2	0.1	10min	67	16
						1h	70	3
16		3	ACN	TBA, 3	0.1	10min	*	-
17		3	ACN	Py, 2	0.1	10min	57	24
						3h	60	4
18		3	ACN	PS, 2	0.1	3h	66	5
19[e]		**2**	**ACN**	**-**	0.1	**15min**	**68**	**7**

All the reactions were performed at 0 ℃ unless specified differently. a – reactions were caried out at room temperature; b – 0.1 eq, of H_2O was added; c – reaction was maintained at -10 – (-5) ℃, d – formation of distinctive impurity, e – scale-up reaction conditions and results (1g starting material), * - diverse impurities formation, hard to unambiguously identify and quantify the percentage of product.

However, increasing $P_2O_3Cl_4$ amount to 3 eq. 53% conversion to the d^{I}CMP was obtained along with 42% conversion to the previously discussed impurity (Table 6, entry 13). Further we investigated the impact of different bases (TBA, pyridine (Py) and proton sponge (PS)) for the course of the reaction. Generally, base additive resulted in faster formation of desired product, with no distinctive differences between the bases (Table 6, entries 14 – 18). Increase in base amount produced more diverse impurities formation (Table 6, entry 16). The best results were obtained using 3 eq. of $P_2O_3Cl_4$ and 2 eq. of TBA. After 10 min 67% conversion to the d^{I}CMP was obtained (Table 6, entry 15).

For one-pot d^{I}CTP synthesis the monophosphorylation conditions shown in Table 6, entry 12 were selected. After 15 min 68% conversion to the d^{I}CMP was observed (Table 6, entry 19), followed by immediate addition of mixture containing TBAPP (3 eq.) and TBA (6 eq.) in acetonitrile. Proceeding the triphosphorylation reaction at room temperature for 20 min the target d^{I}CTP was formed in 63%. After ion-exchange purification the product was obtained in 54% yield.

After successful optimization of d^{I}CMP reaction conditions and d^{I}CTP synthesis, we further examined the reaction conditions for d^{I}AMP. First, when $POCl_3$ was applied (2.2 eq.), only the 18% conversion to the d^{I}AMP was observed (Table 7, entry 1). The use of TBA increased the conversion to the product, nevertheless, only 38% of target material was formed (Table 7 entry 2). The other tested reaction conditions using $POCl_3$ did not produce any better results (Table 7, entries 3 – 4). Moreover, poor solubility of d^{I}A was observed, in all cases, the full dissolution of starting material was not attained, even after multiple additions of $POCl_3$. Further, the $P_2O_3Cl_4$ was tested in TMP media. Herein, neither the presence of different bases nor the addition of $P_2O_3Cl_4$ resulted in the formation of product. In all cases, as it was observed with d^{I}C, the reaction mixtures gelled (Table 7, entries 5 – 9). Using 3 equivalents of $P_2O_3Cl_4$, after 5 minutes 48% of d^{I}AMP was already formed. After additional equivalent of phosphorylation agent 66% conversion to the product was reached (Table 7 entry 10). The use of base did not produce any better results. On the contrary, the phosphorylation was slower and less efficient. Higher conversions to the product were obtained only after additional amount of $P_2O_3Cl_4$ was applied (Table 7, entries 11 – 15).

For scale-up d^{I}ATP synthesis monophosphorylation reaction conditions listen in Table 7, entry 10 were selected. After an hour 76% conversion to the target d^{I}AMP was observed, without necessity of additional amount of $P_2O_3Cl_4$ (Table 7, entry 16). TBAPP cocktail was immediately employed in the same reaction. However, when 3 eq. of TBAPP were utilized, d^{I}ATP was

obtained in only 56% conversion, while in the case of 6 eq. of TBAPP, a 71% conversion was observed. The following reaction mixture was purified by ion-exchange chromatography yielding 65% $d^{I}ATP$.

Table 7. Optimization of $d^{I}A$ monophosphorylation reaction conditions.

Entry	Phosphorylation agent	Quantity of phosphorylation agent, eq.	Solvent	Base, eq.	Reaction time	Conversion to $d^{I}AMP$, %	$d^{I}A$, %
1	$POCl_3$	1.2 + 1[a]	TMP	-	2h	18	77
2		1.2 + 1[a]	TMP	TBA, 2	1h	29	65
					2h	38	43
3[b]		1.2 + 2[c]	TMP	TBA, 2	3h	26	71
4		1.2 + 1[a]	TMP	PS, 1.5	2h	≤3	89
5	$P_2O_3Cl_4$	1.2	TMP	TBA, 2	2h	gelled	95
6		1.2	TMP	Py, 2	2h	gelled	89
7		1.2	TMP	PS, 1	2h	gelled	99
8[d]		1.2 + 1[a]	TMP	TBA, 2	2h	gelled	97
9		1.2 + 1[a]	TMP	-	2h	gelled	99
10		3 + 1[a]	ACN	-	5min	48	50
					1h	56	40
					2h	66	29
11		3	ACN	TBA, 2	5min	14	86
12		3 + 1[e]	ACN	Py, 2	5min	26	59
					3.5h	62	11
13		3	ACN	Py, 4	5min	20	78
14		3 + 3[e]	ACN	PS, 1	5min	42	57
					3.5h	53	42
15		3 + 3[f]	ACN	PS, 1 + 1[f]	30min	40	58
					1.5h	47	46
16[g]		**3**	**ACN**	**-**	**1h**	**76**	**17**

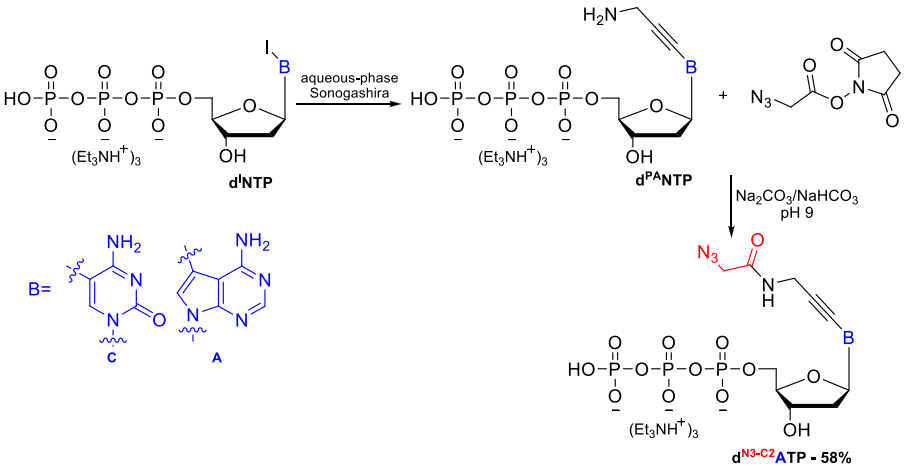
All the reactions concentration was 0.1M. All the reactions were performed at 0 °C unless specified differently. a – the additional equivalent was applied after an hour, b – the reaction was maintained at -11 – (-15) °C, c – the additional equivalents were added after each hour, d – the reaction was performed at room temperature, e – the additional equivalents were applied after 3 hours, f – the additional equivalents were added after half an hour., g – scale-up reaction conditions and results (0.5g starting material).

During the optimization of $d^{I}NMPs$ reaction conditions some general tendencies were observed. Phosphorus oxychloride was poor phosphorylation agent for both substrates. In case of $d^{I}CMP$ synthesis, base addition alongside with $POCl_3$ slightly accelerated reactions, as well as led to a greater amount

of impurities formation. Higher concentrations as well as low temperature resulted in lower reaction rate. In case of $P_2O_3Cl_4$ the TMP was insufficient solvent in all cases, leading to the gelation of reaction mixtures. For both d^INs $P_2O_3Cl_4$ was a suitable phosphorylation agent. Best results were obtained when monophosphorylation was performed in the absence of base. For d^ICMP formation 2 eq. of $P_2O_3Cl_4$ were effective, while for d^IAMP – 3 eq. were required.

5.2.2 Synthesis of amino- and azido-groups bearing 2'-deoxynucleotides

After successful synthesis of both d^ICTP and d^IATP the subsequent labelling with amino moieties was accomplished throughout the application of aqueous phase Sonogashira cross-coupling reaction (Scheme 28). Many literature data discuss the necessity of high temperature (about 70 °C) required for efficient synthesis of $d^{PA}NTPs$, however, at the same time moderate yields are obtained due to the decomposition of triphosphate moiety. Meanwhile, the forming impurities like corresponding diphosphates or even monophosphates can be difficult to purify, by reverse phase or ion-exchange chromatography. Considering all that, reaction mixtures were gently heated (30 °C), expecting lower conversion to the product but avoiding complex impurity profile. Literature overview indicated the possibility to implement reaction utilizing $[(C_6H_5)_3P]_2PdCl_2$ catalyst in DMF that results in higher yield of the product, though the common drawback relies on inferior solubility in comparison to aqueous-phase cross-coupling system.



Scheme 28. Synthesis of amino- and azido-groups bearing dNTPs.

For the synthesis of $d^{PA}ATP$ $[(C_6H_5)_3P]_2PdCl_2$ was applied resulting in 70% conversion to the product (Table 8, entry 1). The aqueous-phase

Sonogashira reaction was tested using 2 or 10 eq. of TEA. The slightly increased reaction rate was noticed in the case of greater amount of base alongside with faster decomposition of the product (Table 8, entries 2 and 3). Applying the reaction conditions shown in entry 2 (Table 8) the desired product was obtained in 77% yield. Synthesized $d^{PA}ATP$ was subsequently reacted with azido acetic acid-NHS ester to produce the desired $d^{N3\text{-}C2}ATP$ in a good 58% yield.

Interestingly, the $d^{I}CTP$ was far less reactive during Sonogashira reaction and after 16 hours 51% conversion to the product was reached (Table 8, entry 4). The reaction did not reach the full completion and 18% of starting material remained, alongside with two distinctive impurities. After purification by ion-exchange chromatography and desalting with C18 column, $d^{PA}CTP$ was obtained in 45% yield. It is important to mention that in all cases the decomposition of product was noticed.

Table 8. Aqueous-phase Sonogashira reaction conditions tested for the synthesis of $d^{PA}NTP$.

Entry	Nucleotide	Propargylamine, eq.	CuI, eq.	Catalyst	Et_3N, eq.	Time	Conversion to $d^{PA}NTP$, %
1[a]	$d^{I}ATP$	2	0.1	$[(C_6H_5)_3P]_2PdCl_2$	3	1h	70
2	**$d^{I}ATP$**	**2**	**0.1**	**$Pd(OAc)_2$ (5 mol%) /TPPTS (25 mol%)**	**2**	**1h** **3h** **16h**	**86** **84** **57**
3	$d^{I}ATP$	2	0.1	$Pd(OAc)_2$ (5 mol%) /TPPTS (25 mol%)	10	1h 3h 16h	90 73 34
4	$d^{I}CTP$	2	0.1	$Pd(OAc)_2$ (5 mol%) /TPPTS (25 mol%)	10	6h 16h	40 51

All the reactions concentration was 0.06 M. All the reactions were performed at 30 ℃ in ACN:H_2O (1:2) media unless specified differently. a – reaction was performed in DMF.

According to the previously described conventional synthesis method, $d^{PA}CTP$ was obtained in 13% overall yield (four steps). Meanwhile, applying the alternative approach after two steps $d^{PA}CTP$ was isolated in 24% overall yield.

5.3 Azido-group bearing 2'-deoxynucleotides testing in enzymatic assays and their reactivity during click reactions determination

The functional groups, like amino, azido, alkyne, etc., bearing nucleotides of different chemical nature have a growing demand for subsequent labelling either prior or after their enzymatic incorporation. Although the objective of this research relied on design and synthesis of oligonucleotide-tethered nucleotides, substrate properties of azido-nucleotides (and amino) are of great importance. Such functional groups have a high potential for diverse labelling abilities. Azide-alkyne cycloaddition reaction alongside with labelled-NHS esters utilization in reaction with amine moieties are considered as simple, rapid, and biorthogonal reactions that can be performed even in the living cells. Azido moieties bearing nucleotides are also attractive from the perspective of the ability to subsequently perform either strain-promoted alkyne-azide cycloaddition (SPAAC) or copper-catalyzed one (CuAAC). This would not be possible if the alkyne moiety would be introduced at the nucleotide, since SPAAC and CuAAC reactions require alkynes of different nature to proceed. Considering all that, it was important to prove d^{N3}CTPs substrates for polymerases properties. This was also of great relevance for the following labeling, since if the d^{N3}CTPs would turn out to be poor substrates, the same results could be expected from their highly labeled counterparts.

5.3.1 Substrate properties of diverse azido-2'-deoxycytidine 5'-triphosphates during enzymatic incorporation

One of the tested polymerases was KOD XL derived from *Thermococcus kodakarensis* KOD1 DNA polymerase. KOD XL polymerase is a mixture of wild type KOD polymerase and mutant form that is deficient in 3'→5' exonuclease activity. KOD XL DNA polymerase is widely applicable for modified nucleotides incorporation either during PCR or PEX and is well known for high tolerance towards nucleobase-modified 2'-deoxynucleotides. The d^{N3}CTPs ($d^{N3\text{-}C2}$CTP, $d^{N3\text{-}C4}$CTP, $d^{N3\text{-}C6}$CTP, $d^{N3\text{-}PEG2}$CTP, $d^{N3\text{-}PEG4}$CTP) were investigated for both single and multiple modified nucleotide incorporation. Templates and primers were selected to produce either single G in template or four in different distances (Figure 21).

The primer extension reactions were first performed for single nucleotide incorporation testing all d^{N3}CTP representatives (Figure 21). The polyacrylamide gel electrophoresis (PAGE) was afterwards ran indicating the efficient incorporation of all d^{N3}CTPs without any distinctive differences between substrates (Figure 22a). The residual amount of primer was observed

in all experiments, regardless of the attempts to increase the amount of template, utility of a new batch of primer or template. However, since the residual primer is detected in the positive control as well, this effect cannot be attributed to the less efficient incorporation of modified $d^{N3}CTP$ substrates, on the contrary, the incorporation capabilities are highly analogous (Figure 22a).

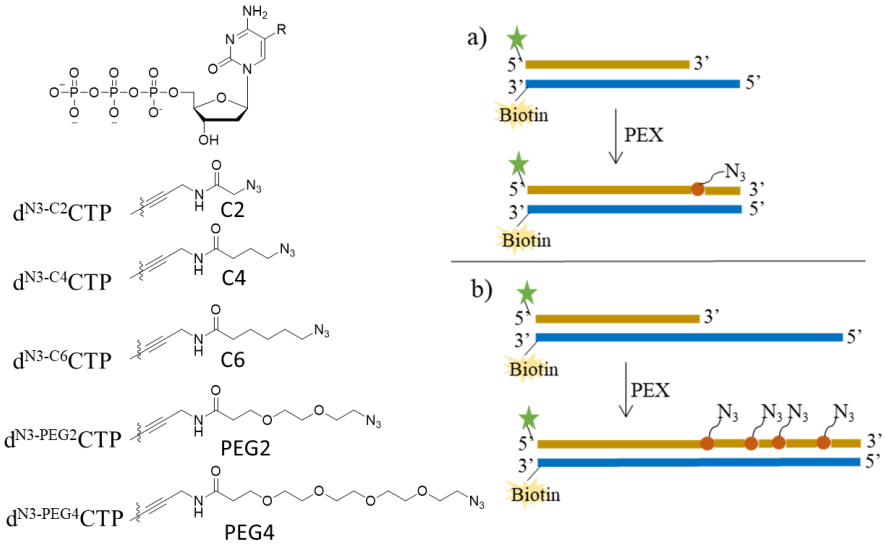


Figure 21. Schematic representation of different systems selected for $d^{N3}CTPs$ incorporation; a) single modified dCTP incorporation, b) multiple nucleotide incorporation.

After the successful incorporation of a single modified $d^{N3}CTP$ was accomplished, the other system, containing several incorporation sites for $d^{N3}CTP$, was tested. Here, the difficulties regarding efficient DNA synthesis can occur due to the protruding modification part. Experiments were carried out under likewise PEX reaction conditions and utilizing the template with four sites for incorporation of $d^{N3}CTP$. Herein all nucleotides were incorporated with the same efficiency, producing the full 31 nt length product (Figure 22b).

Substrates suitability testing was also performed with one of the commonly used polymerases Phusion exo-. The selected system for $d^{N3}CTPs$ testing is represented in Figure 23a where the single nucleotide incorporation was expected. The PAGE analysis represented efficient single nucleotide incorporation. Since the reaction did not contain any natural dNTPs, no full-length product could be synthesized. The formation of an additional nucleotide prolongation product was also observed, representing the ability of polymerase to add one more $d^{N3}CTP$ regardless of template sequence.

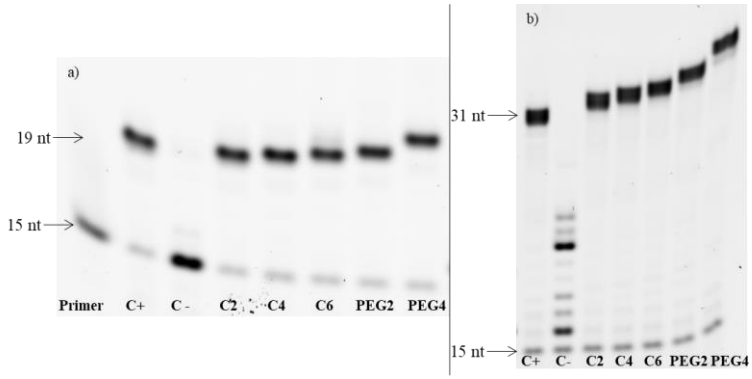


Figure 22. Enzymatic incorporation of d^{N3}CTPs via PEX with KOD XL; a) PEX experiments were performed using the system of template 1C and primer 1, C+ – stands for positive control in the presence of natural dCTP and dGTP, C- – stands for negative control and in the absence of natural dCTP, b) PEX experiments were performed using the system of template 4C and primer 1, C+ – stands for positive control in the presence of all four natural dNTPs, C- – stands for negative control and in the absence of natural dCTP. C2, C4, C6, PEG2, PEG4 – indicates the linker nature in between nucleobase and azido-group. All experiments were incubated for 30min at 60 °C.

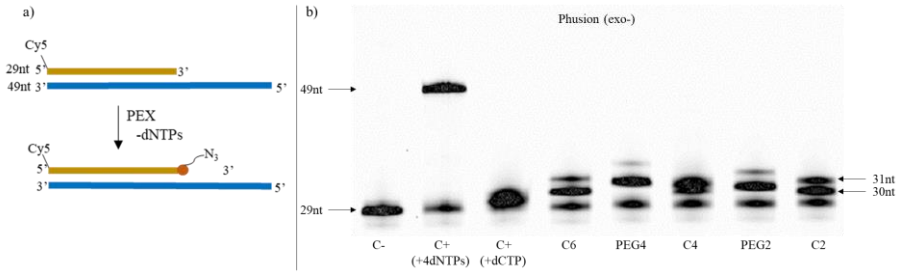


Figure 23. Enzymatic incorporation of d^{N3}CTPs via PEX with Phusion (exo-); a) schematic representation of the system used for incorporation assay, b) PAGE analysis of incorporation products where, C- - stands for negative control in the absence of all four dNTPs, C+ (+4dNTPs) – stands for positive control in the presence of four dNTPs, C+ (dCTP) – stands for positive control in the presence of only dCTP.[1]

Since both single and multiple incorporations of azido-dCTPs were efficient, the following experiments focused on the synthesis of highly modified nucleic acids obtained during PCR in the absence of natural dCTP. Such experiments would represent polymerase tolerance towards the multiple incorporations of d^{N3}CTPs, while in some cases several in the row modified

[1] These experiments were performed by Odeta Taujanskaitė and Artūras Berezniakovas.

nucleotides have to be inserted. Moreover, it would show the ability of polymerase to read through the already modified template during amplification. The selected template contained more than 50 dCTP incorporation sites. First experiments were performed according to the reaction conditions A (Table 9), where 40 PCR cycles were performed followed by the analysis of products on an agarose gel. Band intensity differences indicated the formation of lower amount of products containing nucleotides with C2 and C4 linkers, while other substrates worked efficiently. This could be attributed to both suboptimal PCR conditions for C2 and C4 substrates, or the modification itself, since longer linkers can produce more flexibility and ability to adopt the necessary conformation.

Table 9. PCR reaction conditions tested for d^{N3}CTPs.

Entry	Reaction conditions	Reaction steps	Temperature	Time	Cycles
1	A, B, C, D	Preheating	94 °C	3min	1
		Denaturation	94 °C	1min	A - 39 B - 9 C- 19 D - 29
		Annealing	55 °C	1min	
		Extension	72 °C	1.5min	
		Extension	72 °C	5min	1
2	E	Preheating	94 °C	3min	1
		Denaturation	94 °C	1min	39
		Annealing	55 °C	1min	
		Extension	72 °C	**2min**	
		Extension	72 °C	5min	1
3	F	Preheating	94 °C	3min	1
		Denaturation	94 °C	1min	39
		Annealing	**57 °C**	1min	
		Extension	72 °C	**2min**	
		Extension	72 °C	5min	1

To determine how many cycles are necessary for efficient PCR amplification, one of the best and poorest d^{N3}CTP candidates were selected. Several PCR experiments differing in the number of cycles were performed (Table 9, entry 1 conditions B, C and D). According to the electrophoreses results for incorporation of $d^{N3\text{-}PEG4}$CTP even 10 PCR cycles were sufficient enough to produce the desired product. Nevertheless, it was clearly seen that reactions with $d^{N3\text{-}C2}$CTP analogue produced less product, whereas in case of 10 cycles it was hardly noticeable (Figure 24a). Further optimization of PCR conditions focused on least efficient substrates ($d^{N3\text{-}C2}$CTP and $d^{N3\text{-}C4}$CTP). Prolongation of the extension step (Table 9, entry 2 conditions E) and

elevation of the annealing temperature (Table 9, entry 3 conditions F) resulted in formation of products with satisfying efficiency (Figure 24b).

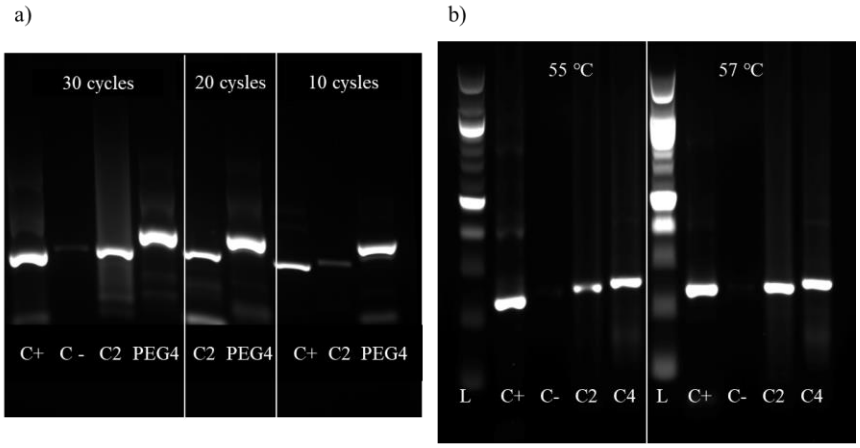
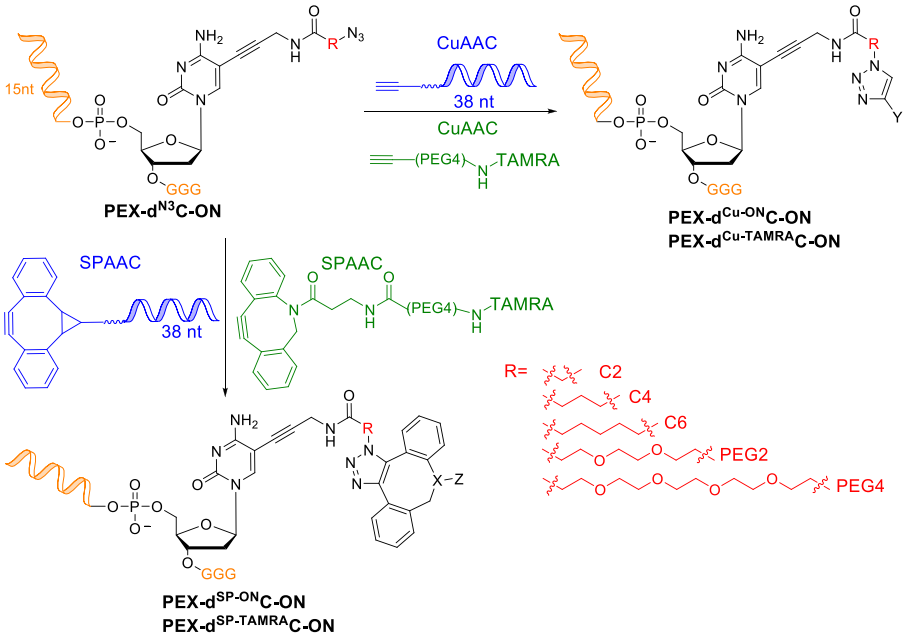


Figure 24. Enzymatic incorporation of d^{N3}CTPs via PCR; a) PCR results according to the reaction conditions B, C and D (Table 9), b) PCR results according to the reaction conditions E and F (Table 9). C+ – stands for positive control in the presence of all natural dNTPs, C- – stands for negative control in absence of all natural dNTPs. L – ladder.

5.3.2 Examination of azide-alkyne cycloaddition reactions rate dependency on substrates nature

The objective of this research was to design and synthesize oligonucleotide-tethered nucleotides. For this matter the CuAAC, SPAAC reactions were selected and different azido-dNTPs were synthesized. However, since the oligonucleotide is a large label that could have an influence on incorporation efficiency of such nucleotides, the ability to incorporate the azido-dNTP and subsequently label it with alkyne-oligonucleotide would be an alternative solution. The feasibility of this approach was tested alongside with evaluation of reaction rate dependency on the linkers structure, as well as alkyne-label. In the selected model system click reactions were performed between various linkers bearing PEX-d^{N3}C-ONs and differently labeled alkynes (Scheme 29). Oligonucleotides and TAMRA dyes bearing terminal alkynes for CuAAC reactions or cyclooctyne moieties for SPAAC reactions were selected (Scheme 29). The PEX-d^{N3}C-ONs (Scheme 29) were obtained by scale-up procedure of previously described PEX of single d^{N3}CTP incorporation with KOD XL (Figure 21a).



Scheme 29. Model system for CuAAC and SPAAC reactions rate dependency on substrates nature measurements.

The first experiments were performed with different linkers bearing PEX-d^{N3}C-ON and alkyne-oligonucleotides for both CuAAC and SPAAC (Scheme 29). In all reactions 5 equivalents of alkyne-oligonucleotide were used, reactions were carried out at 37 °C for four hours. The subsequently performed PAGE analysis of reaction mixtures revealed several important aspects. First, all PEX-d^{N3}C-ONs reacted sluggishly during SPAAC. Although the lower reaction rates were expected, only a trace amount of products was formed (Figure 25a, SPAAC experiments). On the other hand, CuAAC reactions were rapid and, in all cases, nearly full conversion of starting material (PEX-d^{N3}C-ON) was observed. Nevertheless, bands intensities of products bearing different linkers were greatly diverse especially for C6 (Figure 25a). This can be attributed to presumable copper-mediated decomposition of the oligonucleotides. Since C6 linker bearing PEX-d^{N3}C-ON performed poorest and was prone to degradation, it was selected as a model substrate for further optimizations.

To compare reaction rates with different substrates, conditions should be optimized in the manner for substrates to react not too rapid at the same time reaction should come close to the completion. To slow down CuAAC reaction the amount of the alkyne-oligonucleotide was reduced to 1 and 2.5 eq., whereas for SPAAC on the contrary – increased (25 eq.). In the case of CuAAC experiments very controversial results were obtained. A nearly full

starting material conversion was observed in both tested reactions after 2 hours. However, differences between formed product quantities were not consistent and indicated a decomposition process (Figure 25b). The further attempts to reduce both reaction and decomposition rates by decreasing Cu amount (2, 10 and 50 times) were not successful. The reaction reached a completion in 10 to 30 min, yet, in majority of cases decomposition was observed. When Cu amount was reduced even more (50 times), the reaction did not take place. Consequently, determination of CuAAC reaction rate dependency on substrates would be irrational using selected system and substrates. Therefore, the following measurements of reaction rete dependency on substrates structure were performed for SPAAC. The drastic increase in alkyne-oligonucleotide (25 eq.) was still not enough and full reaction completion was not achieved even after 16 hours (Figure 25b). However, such large amount of alkyne seemed irrational to use.

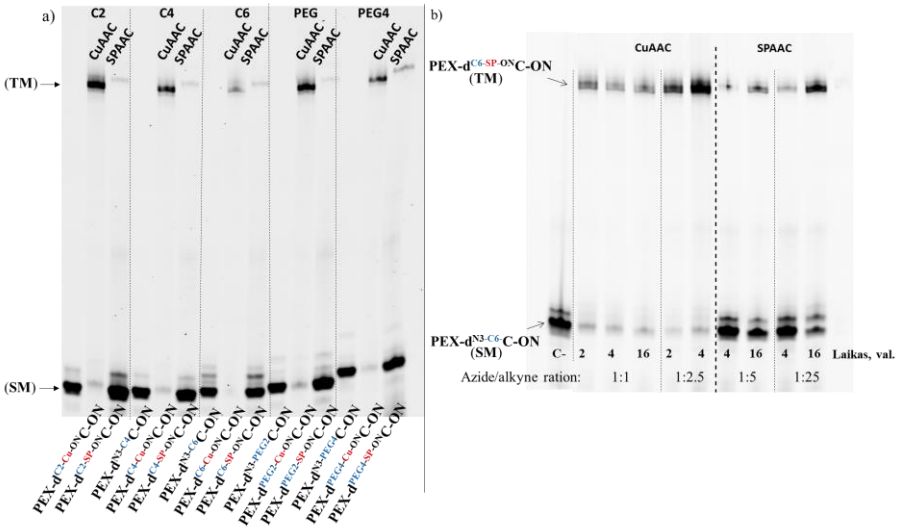


Figure 25. CuAAC and SPAAC reactions with PEX-d^{N3}C-ONs; a) CuAAC experiments were performed at 37 °C, four hours, in the presence of 5eq. alkyne-oligonucleotide, CuBr/TBTA (1:4) as a catalytic system in DMSO:t-BuOH (3:1); SPAAC experiments were carried out in water at 37 °C for four hours in the presence of 5eq. alkyne-oligonucleotide. b) Experiments were performed specifically with PEX-d$^{N3\text{-}C6}$C-ON, different amounts of alkyne-oligonucleotide were used other reaction specifics are the same as in experiments a. C- – negative control: PEX-d$^{N3\text{-}C6}$C-ON. TM – target material, SM – starting material.

We hypothesized that such low reactivity, presumably, could be attributed to the existing repulsion between two reacting oligonucleotides, caused by their phosphate linkages. Moreover, it is well known that click reaction rate could

be increased by applying higher temperature. Consequently, SPAAC was performed in the presence of either spermine tetrahydrochloride or $MgCl_2$ salt excess at 55 °C. Herein after 2 hours the reaction was at its full completion when spermine tetrahydrochloride was used, with even lower alkyne-oligonucleotide amount used (10 eq.). On the other hand, $MgCl_2$ did not had such a breakthrough (Figure 26a). After further optimizations the alkyne-oligonucleotide amount was reduced to its minimum (1.1eq.), and in the presence of spermine tetrahydrochloride at either 55 or 37 °C the reaction was completed in a couple of hours (Figure 26b).

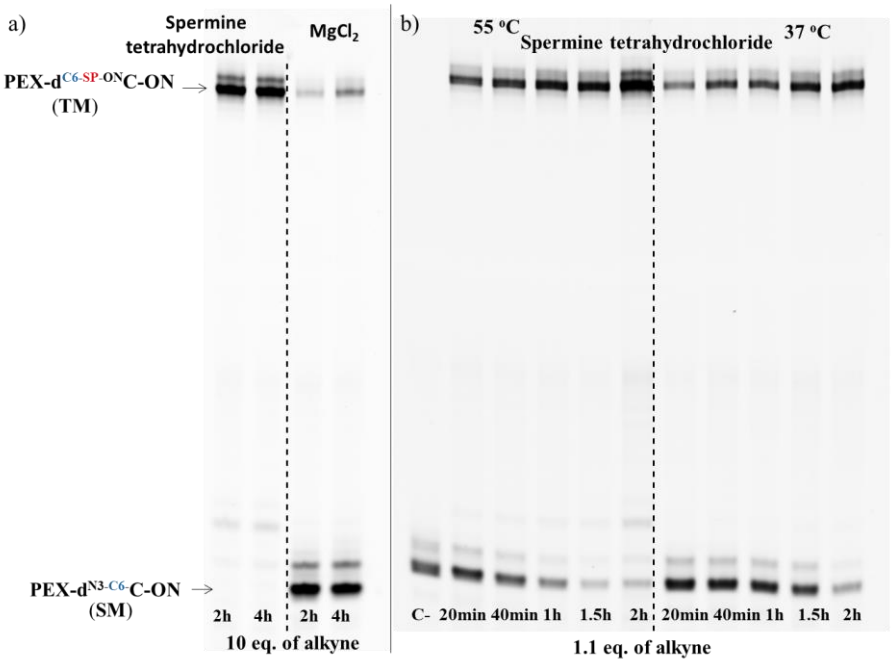


Figure 26. Optimization of SPAAC reaction conditions; a) SPAAC experiments were performed between PEX-$d^{N3\text{-}C6}$C and alkyne-oligonucleotide (10 eq.) at 55 °C in the presence of either spermine tetrahydrochloride or $MgCl_2$ salt (50 eq.), for four hours in water media. b) SPAAC experiments were performed as described in part a, the reactions were carried out at either 55 or 37 °C. TM – target material, SM – starting material.

The SPAAC reaction rate dependency on the structure of PEX-d^{N3}C-ON was analyzed performing experiments in water at 37 °C, applying 1.1 eq. of alkyne-oligonucleotide; samples were taken every 10 to 20 minutes. With each PEX-d^{N3}C-ON representative at least three independent experiments were performed. According to obtained PAGE analysis densitometry data, average values were calculated, and data represented graphically (Figure 27) (For PAGE analysis examples see Supporting Information Figure S1).

Obtained data represented that the linker has a huge influence on reaction rate and substrates lined up in such order: PEX-d$^{N3\text{-}PEG2}$C-ON>PEX-d$^{N3\text{-}C6}$C-ON>PEX-d$^{N3\text{-}C4}$C-ON>PEX-d$^{N3\text{-}C2}$C-ON>PEX-d$^{N3\text{-}PEG4}$C-ON. In general, the longer the linker (the more distant azido-moiety from its oligonucleotide), the higher reaction rate. This can be attributed to steric hindrance of oligonucleotides that complicate reactants interaction. However, the PEX-d$^{N3\text{-}PEG4}$C-ON bearing the longest linker produced poorest results. There are some literature data declaring drawbacks of too long linkers that are attributed to the back-folding on oligonucleotide effect, specific conformation accommodation and occurring interactions with itself or oligonucleotides. PEG4 linker is the longest of tested ones and most probably has a structural flexibility, functionalities for several hydrogen bonds formation, and high solvation degree. Consequently, this phenomenon of low reactivity might be attributed to the previously mentioned drawback of long linkers.

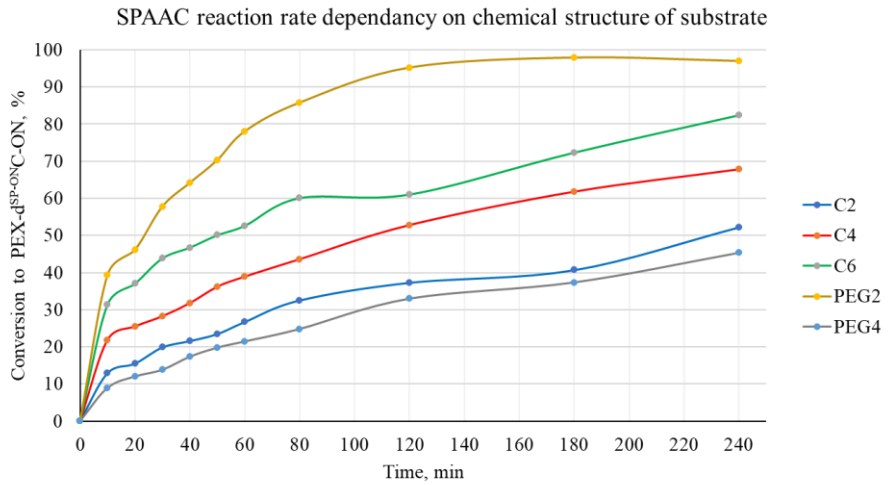


Figure 27. SPAAC reaction rate dependency on PEX-d^{N3}C-ON linker nature (standard deviation varies in the 10% range).

Likewise, CuAAC and SPAAC experiments were performed reacting PEX-d^{N3}C-ON substrates with alkynes bearing fluorescent labels (Scheme 29, TAMRA dye, for structures see Supporting Information Figure S2). In this case, no additive of salt was necessary for SPAAC reactions. Fluorescence of the samples was measured at the points of 15 and 60 min. Obtained values were normalized according to the highest value, consequently full completion of the reaction was not necessary. These experiments once again represent how reaction rate depends on PEX-d^{N3}C-ON substrate chemical structure. Even though click reactions were performed with smaller substrates as dye-

alkynes, the obtained results completely correspond to the previous ones. Longer linker produced the highest reaction rate except for the PEX-d$^{N3\text{-}PEG4}$C-ON. Same PEX-d^{N3}C-ON reactivity tendencies were noticed in both CuAAC and SPAAC reactions. In case of CuAAC after 60 min some substrates reached similar emission values, while in SPAAC case significant changes were not observed (Figure 28). This can be attributed to the higher CuAAC reaction rate due to the presence of catalyst.

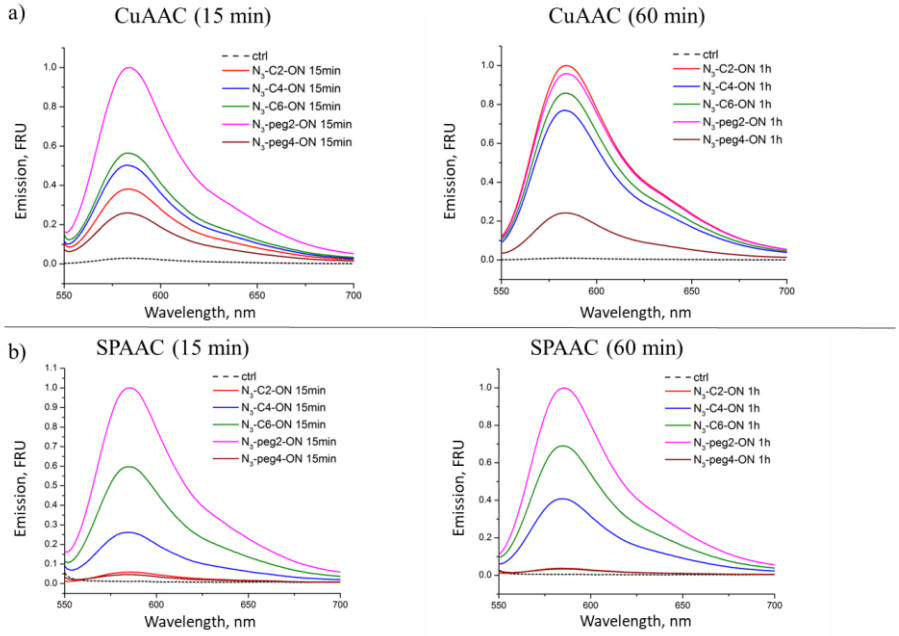


Figure 28. CuAAC and SPAAC reactions rate dependency on the structure of substrates reacting PEX-d^{N3}C-ONs with fluorescently labelled alkynes.[2]

Performed simplified click reaction kinetics measurements showed the same PEX-d^{N3}C-ON substrates reactivity tendencies, regardless, either dye-alkyne or alkyne-oligonucleotide were employed. Moreover, it also emphasizes the importance of linker design for an efficient subsequent labelling. However, the unsuccessful attempts to measure CuAAC reaction rate dependency on substrates nature between two oligonucleotides, reveals problems of higher complexity. Since the focus is to label unknown nucleic acids sequences with known ONs, that would be used as a universal priming site, such decomposition of ONs during CuAAC would make this approach

[2] These experiments were performed by Alessandro Panattoni.

problematic to apply. The products would be partly degraded at the different sites, consequently, it would complicate the following enzymatic applications and distort the obtained data. Accordingly, the utility of oligonucleotide-tethered nucleotides prepared by reacting d^{N3}NTPs with alkyne-oligonucleotides followed by their enzymatic incorporation strategy solves the above-mentioned problems and was further investigated.

5.4 Oligonucleotide-tethered 2'-deoxynucleotides synthesis and application in enzymatic assays

NGS library preparation is a long workflow procedure having several drawbacks (as it was described in literature overview, chapter 4). All current sequencing platforms require nucleic acid pre-processing to generate library suitable for sequencing. Generally, this includes DNA or RNA fragmentation to a platform-specific size range, followed by end polishing and specific adapter ligation to the 3' and 5' termini. Enzymatic ligation of adapters is notorious for low efficiency leading to decreased complexity of the original library and impoverishment of sequencing results. Consequently, novel library preparation techniques need to be developed to improve conversion efficiency and simplify the workflow. The ability to utilize nucleotides containing an oligonucleotide modification as a universal priming site would enable overcoming the ligation of platform-specific adapters. This, indeed, would highly simplify the NGS library preparation workflow alongside with elimination of the most critical step.

Two hypotheses regarding the idea of such workflow being applicable were raised. First, oligonucleotide-tethered 2'-deoxynucleotides (OTDNs) incorporation events could be controlled by the present amount of OTDN versus natural dNTP in the reaction. Since the oligonucleotide is a large modification, it was hypothesized that OTDN incorporation would be a rare event where several OTDNs would not be incorporated in a short distance due to steric hindrance induced by the label (the workflow is shown in Figure 29a). To prove raised hypothesis, first it was necessary to synthesize OTDNs and investigate whether they are the substrates for polymerases.

For technology feasibility studies and synthesis of OTDNs ($d^{B1\text{-}ON}$dCTPs and $d^{B2\text{-}ON}$dCTP) two alkyne-oligonucleotides were selected. B1 – contained hexynyl-modification at the first nucleobase of 5'-termini, while B2 – at the second nucleobase (Figure 30). B2 – oligonucleotide design was selected with the hypothesis that protrusion of one nucleotide may simplify the read-through step for the polymerase. CuAAC reactions were performed between all d^{N3}CTPs and both B1 and B2 alkyne-oligonucleotides. Experiments were

maintained on a very small scale (8.4 nmol of alkyne) with the excess of d^{N3}NTPs (2 eq.).

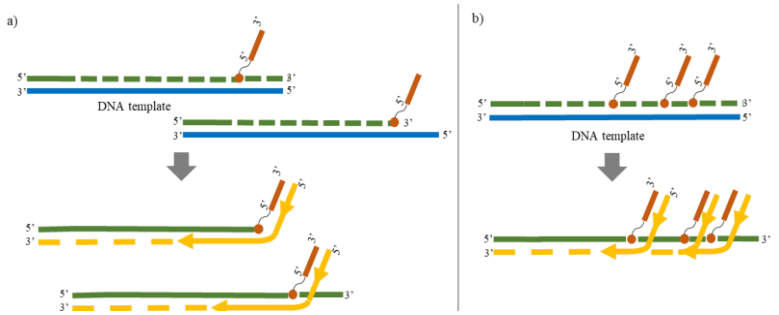


Figure 29. OTDNs utility in NGS library preparation workflow; a) expected workflow representation, b) presumably observed assay.

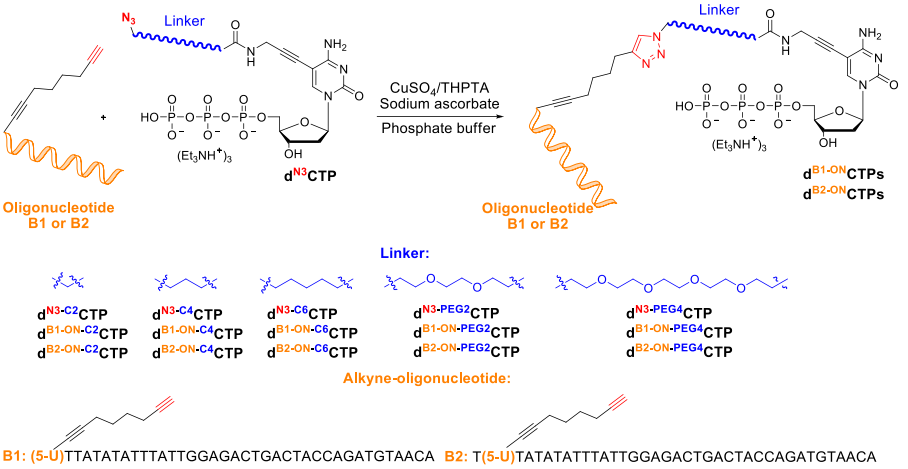


Figure 30. Synthesis of diverse OTDNs via CuAAC.

The course of the reactions was monitored by HPLC, and good conversions to the OTDNs were obtained. However, decomposition of oligonucleotides was observed and confirmed by liquid chromatography–mass spectrometry (LC/MS). The most challenging part remained the purification of OTDNs. The complex reaction mixture with reactants of similar nature and retention times on the C18 column indicated their possible complicated purification. Nevertheless, the biggest obstacle was the reaction scale, since none of the available equipment was suitable for the purification of such small amount. First, spin-columns with a specific pore size (3 kDa) were tested, since the critical impurity was the residual amount of d^{N3}NTP that would compete with OTDN during enzymatic incorporation. However, the full disposal of d^{N3}NTP

was not achieved, moreover, high amount of OTDN stuck inside pores of spin-column, while the alkyne-oligonucleotide could not be purified in this manner. Consequently, the analytical HPLC was used for the purification of OTDNs. The reaction mixture was purified in four portions, and the product was combined and concentrated. The moderate products yields (Table 10) can be attributed to the multiple purification steps performed.

Table 10. Synthesized OTDNs.

Entry	ON	$d^{N3}CTP$	Time	OTDN	Conversion to the OTDN, %	Yield, %
1	B1 (38 nt)	$d^{N3\text{-}C2}CTP$	1h	$d^{B1\text{-}ON\text{-}C2}CTP$	81	26
2		$d^{N3\text{-}C4}CTP$	1.5h	$d^{B1\text{-}ON\text{-}C4}CTP$	76	17
3		$d^{N3\text{-}C6}CTP$	1h	$d^{B1\text{-}ON\text{-}C6}CTP$	70	19
4		$d^{N3\text{-}PEG2}CTP$	1h	$d^{B1\text{-}ON\text{-}PEG2}CTP$	72	17
5		$d^{N3\text{-}PEG4}CTP$	1.5h	$d^{B1\text{-}ON\text{-}PEG4}CTP$	80	22
6	B2 (38 nt)	$d^{N3\text{-}C2}CTP$	1h	$d^{B2\text{-}ON\text{-}C2}CTP$	82	26
7		$d^{N3\text{-}C4}CTP$	3h	$d^{B2\text{-}ON\text{-}C4}CTP$	51	13
8		$d^{N3\text{-}C6}CTP$	1h	$d^{B2\text{-}ON\text{-}C6}CTP$	47	14
9		$d^{N3\text{-}PEG2}CTP$	2h	$d^{B2\text{-}ON\text{-}PEG2}CTP$	66	16
10		$d^{N3\text{-}PEG4}CTP$	1.5h	$d^{B2\text{-}ON\text{-}PEG4}CTP$	68	21

ON – oligonucleotide. The reactions were proceeded using 8.4 nmol of alkyne-oligonucleotide (B1 or B2) and 2 eq. of $d^{N3}NTP$ in phosphate buffer in the presence of $CuSO_4$/THPTA catalyst and sodium ascorbate (reaction volume 200 µL). Reactions were maintained at 37 °C. Products were purified by analytical HPLC system using C18 column.

After the synthesis of OTDNs, their as substrates properties during PEX were evaluated. For PEX reactions with duplex Dup1 (Table 11) Taq and Phusion exo- polymerases were tested. The obtained PAGE analysis results indicated that all OTDN were successfully incorporated into the growing DNA strand by both polymerases, with Phusion exo- showing a slightly higher efficiency. Beside bands of target material (at 68nt length), the formation of shorter side-products was also observed (indicated as 30nt length) (Figure 31).

Table 11. Oligonucleotide duplexes used for OTDNs incorporation.

Name	Oligonucleotide duplex	dNTPs
Dup1	5'–Cy5–TGCAGACATGGGTAGGCATCCTTGGCGTA–3' 3'–ACGTCTGTACCCATCCGTAGGAACCGCAT**G**ACATCGA CTCAACTCGCTG–5'	None

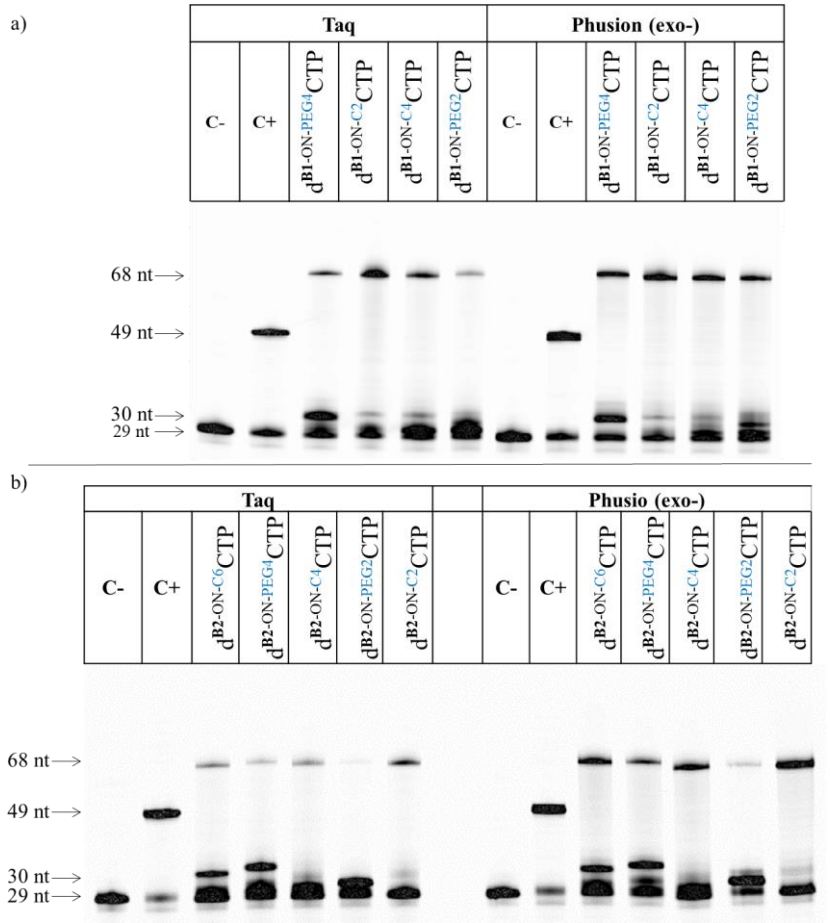


Figure 31. PAGE analysis of OTDNs incorporation by Taq and Phusion (exo-) polymerases in PEX assay; a) incorporation of OTDNs containing B1 oligonucleotide, b) incorporation of OTDNs containing B2 oligonucleotide. C- – negative control in absence of dNTPs, C+ – positive control with the presence of all four natural dNTPs. All OTDNs incorporation experiments were proceeded with absence of natural dNTPs to monitor a single OTDN incorporation event.[3]

This can be attributed to the unsuccessful removal of starting d^{N3}NTPs, although, their presence in the final product was not detected by HPLC (amount lower than detection limit). To confirm the presence of residual d^{N3}NTPs, LC/MS analysis was additionally performed. Only by extracting mass spectrum according to corresponding d^{N3}NTP molecular mass a minor amount of precursors was noticed. Azido-dNTPs were significantly better substrates and competed during incorporation with OTDNs. Since such a minor amount of d^{N3}NTP was undetectable with UV detector of HPLC, it was

[3] These experiments were performed by Artūras Berezniakovas and Odeta Taujanskaitė.

impossible to accurately define their quantity. Consequently, substrate capabilities of OTDNs could not be determined and compared. Despite that, it was obvious that OTDNs were substrates and could be incorporated during PEX (Figure 31). From the perspective of the anticipated application of OTDNs, the residual amount of d^{N3}NTPs was crucial, therefore, purification conditions were further optimized. To select suitable chromatographic conditions, different elution buffers were used, changing the ion pairing reagent concentration and composition of mobile phases. Different linear gradients as well as diverse gradient formation patterns were also tested. After numerous attempts, suitable conditions were found, and products were successfully obtained in the absence of unreacted d^{N3}NTPs.

To show technology feasibility and prove biocompatibility of linkers the system with the single incorporation site for OTDN was used with fluorescently labeled both template (Cy5 – red) and primer (Cy3 – green) for read-through assay. After incorporation of OTDN, the reaction mixture had to be purified to produce pure template for the subsequent read-through assay. Consequently, primer (labeled with Cy5 dye) for incorporation of OTDN also contained biotin modification for the subsequent ON-d^{ON}Cs purification with streptavidin-coated beads. Purified incorporation products were applied in the following reaction where ON-d^{ON}Cs served as templates and OTDN site as a priming site. The PEX reactions were performed with Phusion exo- polymerase in 10 linear cycles. Read-through results for diverse ON-d^{ON}Cs are given in Figure 32 lanes 5 – 13. Controls indicating read-through product length (Figure 32, lane 4) and extended primer till the unnatural linkage (Figure 32, lane 3) were prepared. PAGE analysis results showed that synthesis of complementary strand, throughout unnatural linker, was successfully proceeded with all substrates (Figure 32). The green bands in between the primer and read-through products indicate that polymerase stalls at the linker in attempts to read throughout the full-length template. Some bands may indicate stopped synthesis before the unnatural linker or just after the linker.

In general, obtained results prove that all linkers are biocompatible and Phusion exo- is capable of reading through them. However, this model system is not sufficient to determine which linker is the most suitable. Furthermore, the reading through unnatural linkage is not accomplished easily and such an obstacle for polymerase produces shorter side products. All in all, the feasibility of technology was proven to be possible. However, the use of oligonucleotide-tethered 2'-deoxynucleotides in real NGS library preparation applications had some drawbacks. When OTDN is incorporated several times, branched products are obtained, with very complex structure. This

complicates the subsequent complementary strand synthesis, obtained products are difficult to identify and polymerase have even more difficulties reading through such templates. Therefore, OTDNs could not be used for NGS library preparation application, since here the full coverage of precise-length reads is of greatest importance.

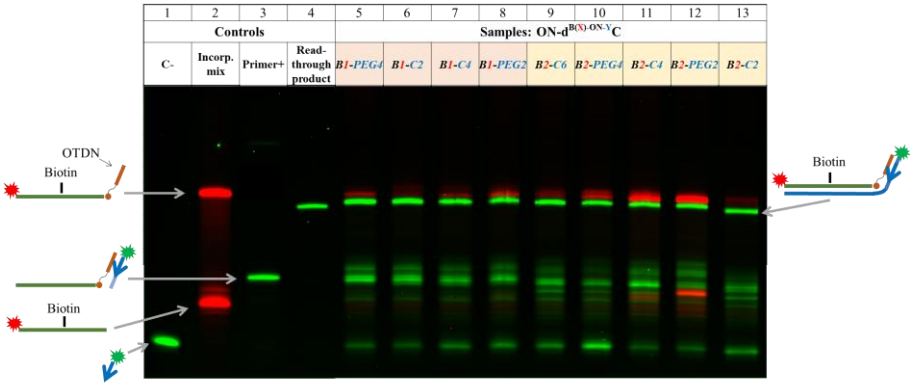


Figure 32. Diverse ON-d^{ON}Cs testing in read-through unnatural linkage assay with Phusion (exo-) polymerase. C- – negative control containing Cy3-labelled primer, Incorp. mix – OTDN incorporation product and primer mixture, Primer+ – length standard of primer extended till the linker, read-through product – length standard of expected product. Reactions were performed at the conditions: 95 ℃ – 1min → 61 ℃ – 1min → 72 ℃ – 1min (10 cycles). [4]

The most important part – read-through unnatural linker – was already proven. While challenges caused by multiple incorporations of OTDNs and branched structures formation could be solved by utilizing corresponding 2',3'-dideoxynucleotide analogues. Therefore, the following research focused on corresponding oligonucleotide-tethered 2',3'-dideoxynucleotides synthesis.

[4] These experiments were performed by Artūras Berezniakovas and Odeta Taujanskaitė.

6. SYNTHESIS OF OLIGONUCLEOTIDE-TETHERED 2',3'-DIDEOXYNUCLEOTIDES

The revolutionary breakthrough in applicability of 2',3'-dideoxynucleotides was the emergence of Sanger sequencing methodology where dye-labeled ddNTPs were used as DNA replication terminators. Initially, ddNMP and ddNTP analogues were investigated as promising antiviral, antitumor agents, while later on further developed to enzymatically label nucleic acids for various array-based assays. While dNTPs are the essence in many diagnostics and bioprocesses monitoring, ddNTPs are less explored. This obstacle is conspicuous regarding the existing literature data for nucleobase-modified ddNTPs synthesis. The main strategy here remains the long, multistep procedure described and tested for nucleobase-modified dNTPs synthesis (Scheme 24a).

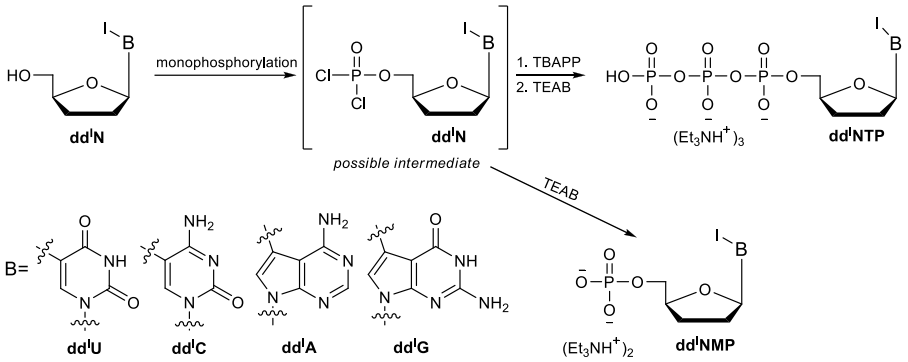
The proposed OTDN based technological approach was proven to be feasible, but critical challenges remained. The applicability of oligonucleotide-tethered 2',3'-dideoxynucleotides (OTDDNs) could significantly simplify and even improve the desired application. Therefore, the following chapter will focus on efficient synthesis of nucleobase-modified ddNTPs, OTDDNs and their properties in enzymatic assays.

6.1 Optimization of phosphorylation reaction conditions for iodinated 2',3'-dideoxynucleosides

Since the purpose of labeled nucleotides relies on their modification nature associated with the specific application, our objective was to optimize synthesis strategy that would be universal, scalable and provide functionalized nucleotides for diverse labels introduction. Three steps strategy for modified nucleotides synthesis was used for preparation of amino group bearing dNTPs in chapter 5.2. It was shown that this synthesis approach is more efficient in comparison to traditional five steps methodology, therefore, it was applied for modified-ddNTPs synthesis. The first goal was to optimize the monophosphorylation reaction conditions for iodinated 2',3'-dideoksinucleosides (dd^{I}Ns) and perform one-pot synthesis of corresponding iodinated 2',3'-dideoksinucleoside 5'-triphosphates (dd^{I}NTPs) (Scheme 30).

A standard phosphorylation conditions were applied for dd^{I}CMP synthesis, referring to this methodology used in many modified dNTPs or ddNTPs synthesis. However, after addition of 2 ekv. of phosphorus oxychloride to the reaction mixture no product formation was detected by HPLC analysis (Table 12, entry 1). Moreover, the increase of $POCl_3$ or the addition of base did not

result in the formation of the target material (Table 12, entries 2 – 4).



Scheme 30. Synthesis of ddINMPs and ddINTPs.

Interestingly, the similar experiments with dIC produced corresponding monophosphate in moderate conversion, while ddIC was unreactive towards $POCl_3$. It is important to mention that iodinated-ddNs exhibit poorer solubility in standard phosphorylation media than iodinated-dNs or natural dNs. The already proven to be an efficient phosphorylation agent – $P_2O_3Cl_4$ was further employed. The deficiency of hydroxy groups at 2' and 3' positions of ddNs eliminate regioselectivity problem, that was presumably observed in dIC phosphorylation. Consequently, the usage of pyrophosphoryl chloride seemed even more attractive. Utilizing 2 eq. of $P_2O_3Cl_4$ the low conversion to the corresponding monophosphate was observed, moreover, formation of specific impurity was occurring (Table 12, entries 5 – 7). The analysis of HPLC data revealed that impurity had a different absorption maximum shifted to the shorter wavelength (impurity – 288 nm, ddIC – 295 nm) in comparison to starting material (Figure 33). In addition, the reaction mixture turned yellow which led us to assume the possibility of occurring iodine exchange with chlorine. This hypothesis correlates with observed change in absorption maximum [234] and is further supported by earlier reports [52, 54].

Increase in $P_2O_3Cl_4$ amount (3 eq.) forced the reaction to the formation of the desired monophosphate rather than impurity (Table 12, entry 8). The addition of 0.25 equivalents of base triggered even faster formation of ddICMP (conversion 86%) and drastically diminished formation of impurity (Table 12, entry 9). However, the further increase in tributylamine amount did not result in any improvement. On the contrary, controversial results were obtained (Table 12, entries 10 – 11). While the equivalent amount of base to phosphorylation reagent resulted in almost no formation of the product (Table 12, entry 12).

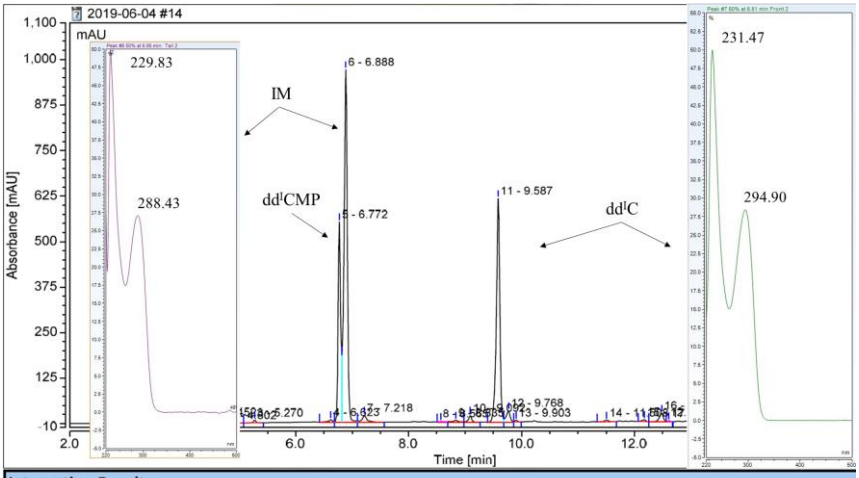
Table 12. Optimization of dd^{l}C monophosphorylation reaction conditions.

Entry	Phosphorylation agent	Quantity of phosphorylation agent, eq.	Solvent	TBA, eq.	Reaction time	Conversion to dd^{l}CMP, %	Conversion to impurity, %	dd^{l}C, %
1	$POCl_3$	2	TMP	-	2h	≤3	-	98
2		2	TMP	2	2h	≤3	-	97
3		3	TMP	-	2h	5	-	92
4		3	TMP	2	2h	≤3	-	98
5	$P_2O_3Cl_4$	2	ACN	-	0min	8	1	90
					30min	14	23	59
					1h	15	33	47
					2h	14	52	29
6		2	ACN	0.5	30min	19	24	50
					1h	19	45	29
					2h	18	66	10
7		2	ACN	2	2h	≤3	≤3	95
8		3	ACN	-	0min	20	1	78
					30min	63	13	16
					1h	68	18	7
					2h	67	25	4
9		3	ACN	0.25	30min	82	3	7
					1h	86	4	2
					2h	86	6	1
10		3	ACN	0.5	30min	51	4	37
					1h	59	4	26
					2h	57	29	13
11		3	ACN	1	30min	86	1	4
					1h	88	1	2
					2h	89	2	1
12		3	ACN	3	2h	10	1	84
13[a]		**3**	**ACN**	**-**	**20min**	**87**	**2**	**5**

Reactions were maintained at 0 – 5 °C, reactions concentration was 0.1 M. Small scale experiments were performed using 0.05 – 0.1 g of starting nucleoside. a – scale-up reaction conditions and results (1g starting material). The reaction was carried out at -15 – -10 °C.

For scale-up synthesis, reaction conditions listen in Table 12, entry 9 were employed. Although the reaction rate increased, a more complex impurity profile was observed in comparison to experiments in a smaller scale. Consequently, the reaction temperature was decreased to -15 °C and after 20 min 87% conversion to the intermediate material was obtained without any presence of the base (Table 12, entry 13). Since the objective was to synthesize 5'-triphosphates, after formation of active monophosphate derivative the

TBAPP and TBA solution in acetonitrile was immediately added. In 30 min the desired triphosphate was synthesized. During one-pot synthesis dd^{I}CTP was isolated in 74% yield.



Integration Results

No.	Peak Name	Retention Time min	Area mAU*min	Height mAU	Relative Area %	Relative Height %
5		6.772	27.814	550.502	18.74	24.20
6		6.888	66.225	968.311	44.63	42.57
11		9.587	43.276	614.481	29.16	27.02

Figure 33. HPLC analysis of the reaction mixture. The sample was taken after one hour from the reaction mixture where 2 eq. of $P_2O_3Cl_4$ and 0.5 eq. of TBA were used (Table 12 entry 6). IM – impurity.

For the synthesis of dd^{I}AMP, phosphoro oxychloride was also tested (Table 13, entries 1 – 6). First, the reaction was held at 0 – 5 °C using 2 eq. of $POCl_3$. With no product formation after an hour, additional equivalent of $POCl_3$ was added (Table 13, entry 1), however, only a trace amount of dd^{I}AMP was formed. Further attempts to increase $POCl_3$ amount, add base or perform reaction at higher temperature (5 – 10 °C) were ineffective (Table 13, entries 2 and 4 – 6). After unsuccessful results with $POCl_3$, the pyrophosphoryl chloride was tested. The poor solubility issue was also observed for dd^{I}A where too low reaction temperature (below 0 °C) would lead to low dissolution of starting nucleoside. When 2 eq. of $P_2O_3Cl_4$ and different tributylamine quantities were applied only a trace amount of product was detected (Table 13, entries 7 – 9). However, addition of $P_2O_3Cl_4$ (3 eq. in total) in the absence of base has moved reaction forward resulting in 37 – 48% conversion to the desired product (Table 13, entries 10 – 11). When 3 eq. of $P_2O_3Cl_4$ were applied at once, followed by one additional equivalent after an hour, nearly full conversion of the starting material was reached, where dd^{I}AMP accounted for 77 – 85% (Table 13, entries 12 – 13).

Table 13. Optimization of dd^{I}A monophosphorylation reaction conditions.

Entry	Phosphorylation agent	Quantity of phosphorylation agent, eq.	Solvent	TBA, eq.	Reaction time	Conversion to dd^{I}AMP, %	dd^{I}A, %
1[a]	$POCl_3$	2 + 1[b]	TMP	-	2h	4	90
2		2 + 1[b]	TMP	-	2h	trace	91
3		3	TMP	2	2h	3	93
4		3 + 1[b]	TMP	-	2h	4	90
5		3 + 1[b]	TMP	-	2h	3	90
6		3	TMP	3	2h	4	92
7	$P_2O_3Cl_4$	2	ACN	0.25	2h	5	92
8		2	ACN	0.5	2h	3	95
9		2	ACN	2	2h	4	93
10[a]		2 + 1[b]	ACN	-	0min	4	90
					30min	15	79
					1h	17	77
					2h	48	46
11		2 + 1[b]	ACN	-	0min	4	90
					30min	7	86
					1h	7	86
					2h	37	55
12[a]		3 + 1[b]	ACN	-	0min	7	86
					30min	50	44
					1h	51	42
					2h	85	6
13		3 + 1[b]	ACN	-	0min	11	83
					30min	34	59
					1h	35	57
					2h	77	14
14[c]		**3**	**ACN**	**-**	**30min**	**96**	**-**
					1h	**95**	**-**

Concentration of reactions was 0.1M. Small scale experiments were performed using 0.05 – 0.1 g of starting nucleoside. All reactions were performed at 5 – 10 ℃ unless specified differently. a – reaction was carried out at 0 – 5 ℃, b – the additional equivalent was applied after an hour, c – scale-up reaction conditions and results (1g starting material).

The reaction conditions listen in Table 13, entry 14 represents scale-up results of the monophosphorylation step. As it was noticed with dd^{I}CMP scale-up experiment, the reaction rate during dd^{I}AMP synthesis increased as well. After 30 min 96% conversion to the desired dd^{I}AMP was reached. Followed by the addition of TBAPP and TBA mixture in acetonitrile,

conversion to the desired ddA^{I}TP was 79%. After purification by ion exchange chromatography product was obtained in 73% yield.

Monophosphorylation of dd^{I}U was successful utilizing phosphorus oxychloride. A modest excess of $POCl_3$ (1.5 eq.) resulted into the formation of a trace amount of dd^{I}UMP (Table 14, entry 1). The use of tributylamine (1 eq.) alongside with 3 eq. of $POCl_3$ added in two portions resulted in product formation. However, the reaction rate was quite low, after nearly 5 hours the conversion to dd^{I}UMP was 78% (Table 14, entry 2). After finding the optimal phosphorylation agent amount, the reaction rate and higher conversion were further obtained by increasing the tributylamine quantity. Applying 2 eq. of base, 89% conversion was obtained as well as a slightly higher formation of

Table 14. Optimization of dd^{I}U monophosphorylation reaction conditions.

Entry	Phosphorylation agent	Quantity of phosphorylation agent, eq.	Solvent	TBA, eq.	Reaction time	Conversion to dd^{I}UMP, %	Conversion to impurity, %	dd^{I}U, %
1	$POCl_3$	1.5	TMP	-	30min	≤3	1	97
					2h	6	1	92
2		1.5 + 1.5[a]	TMP	1	30min	16	-	77
					3h	60	-	34
					4.5h	78	-	17
3		3	TMP	1	30min	55	-	41
					2.5h	84	-	11
					4h	89	-	4
4[b]		**3**	**TMP**	**1.5**	**30min**	**68**	**-**	**28**
					1h	**84**	**-**	**10**
					2h	**93**	**-**	**2**
5		3	TMP	2	30min	55	-	31
					1h	84	-	2
					1.5h	89	-	2
6	$P_2O_3Cl_4$	3	ACN	-	30min	22	64	6
					1h	20	75	1
					3h	13	82	-
					5h	8	88	-
7		3	ACN	3	30min	63	17	-
					1h	62	18	-

All the reactions concentration was 0.1M. Small scale experiments were performed using 0.05 – 0.1 g of starting nucleoside. All the reactions were performed at 0 – 5 °C. a – the additional equivalent was applied after two hours, b – scale-up reaction conditions and results (1g starting material).

impurities in comparison to reaction conditions, when 1.5 eq. of tributylamine were used (Table 14, entries 4 – 5).

From previous experience of our group on the synthesis of modified dNTPs, the pyrophosphoryl chloride was ineffective on uridine and guanosine derivatives. Nevertheless, for dd^{I}UMP synthesis we applied the best conditions from previous findings regarding the synthesis of dd^{I}CMP and dd^{I}AMP. The reaction was not taking place towards the formation of the desired product, on the contrary, immediate formation of specific impurity was observed (Table 14, entry 6). With no significant absorption maximum shift and without any retention on ion exchange resin during purification, we assumed that this could be attributed to the possible glycosidic bond decomposition. Such effect is widely described by many articles referring to nucleosides phosphorylation and acidic media catalyzed glycosidic bond destruction. [30, 54, 52] This also complies with the findings of base utilization to reduce the decomposition rate [54, 52], which we as well observed when 3 eq. of tributylamine were used (Table 14, entry 7).

The scale-up experiment was performed according to conditions 4 (Table 14), where in 2 hours conversion to the dd^{I}UMP reached 93%, and after triphosphorylation with TBAPP, dd^{I}UTP was obtained in 82% yield.

After all experiments on dd^{I}Ns monophosphorylation, the synthesis of dd^{I}GMP was successfully achieved by slightly modifying dd^{I}UMP synthetic protocol. Since all dd^{I}Ns phosphorylation required 3 equivalents of either $POCl_3$ or $P_2O_3Cl_4$ for dd^{I}G monophosphorylation we directly applied 3 eq. of $POCl_3$ (Table 15, entry 1). In the absence of tributylamine, the reaction rate was low and after 5 hours only 42% conversion to dd^{I}GMP was obtained (Table 15, entry 1). For efficient conversion to dd^{I}GMP the addition of tributylamine was necessary, yet the presence of 1 – 2 eq. of base was still not satisfying (Table 15, entries 2 – 3). When 2.5 eq. of base were utilized after 1 hour 92% conversion to the product was reached (Table 15, entry 4). The further increase in base amount did not produce any better results (Table 15, entries 5 – 6).

For reaction scale-up, conditions listed in Table 15, entry 4 were selected and final dd^{I}GTP was obtained in 57% yield. The relatively small yield was obtained due to the suboptimal purification conditions, since dd^{I}GTP is strongly detained on the Sepharose Fast Flow ion exchange resin and a large part of the product was locked on the column. The product was removed only during resin regeneration using 2 M NaCl solution.

Table 15. Optimization of dd^{I}G monophosphorylation reaction conditions.

Entry	Phosphorylation agent	Quantity of phosphorylation agent, eq.	Solvent	TBA, eq.	Reaction time	Conversion to dd^{I}GMP, %	dd^{I}G, %
1	$POCl_3$	3	TMP	-	5h	42	55
2		3	TMP	1	2.5h	62	36
3		3	TMP	2	3.5h	87	11
4[a]		**3**	**TMP**	**2.5**	**1h** **1.5h** **2h**	**92** **92** **90**	**3** **1** **-**
5		3	TMP	2.8	1h 2h	93 81	1 -
6		3	TMP	3	1.5h	92	2

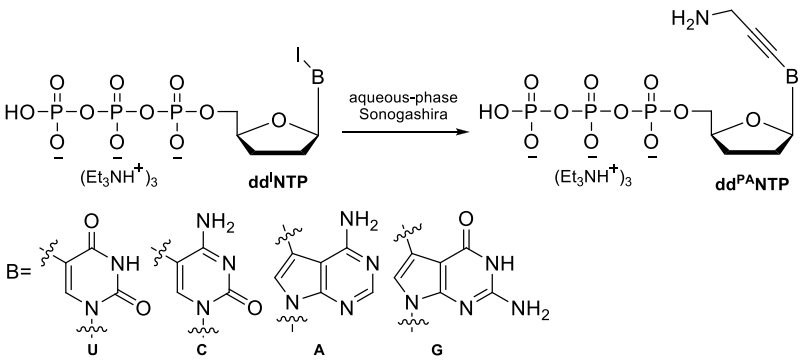
All reactions concentration was 0.1M. Small scale experiments were performed using 0.05 – 0.1 g of starting nucleoside. All reactions were performed at 0 – 5 °C. a – scale-up reaction conditions and results.

Monophosphorylation reaction conditions were optimized for all dd^{I}Ns independently, while for triphosphorylation step the standard methodology was effective in all cases with slight differences in TBAPP amount used. This enabled us to adopt an efficient one-pot triphosphorylation methodology to obtain dd^{I}NTPs. In addition, the synthesis of triphosphates was successfully scaled to 1 g (small scale experiments were performed using 0.05 – 0.1 g of nucleoside) resulting in good products yields. The study demonstrated that reaction conditions are not always readily transferable to even 10 to 20 times scaled-up synthesis procedures. Moreover, although it is usually declared that the same conditions can be efficiently utilized for all four nucleotides, in this case the same monophosphorylation conditions did not apply to all.

During the optimization of (d)d^{I}Ns monophosphorylation, we observed some nucleotides similarity patterns. The analogy in reactivity and suitable monophosphorylation reaction conditions was observed between (d)d^{I}C and (d)d^{I}A as well as dd^{I}U and dd^{I}G. For the synthesis of dd^{I}CMP and dd^{I}AMP phosphorus oxychloride was ineffective. However, as it was observed with 2'-deoxy-analogues, the $P_2O_3Cl_4$ was suitable for dd^{I}CMP and dd^{I}AMP synthesis. Moreover, during scale-up experiments reactions times reduced significantly and slight adjustment of conditions was necessary. For the synthesis of dd^{I}UMP and dd^{I}GMP phosphorus oxychloride was efficient phosphorylation agent. To obtain good results base addition was required. Here, TBA amount significantly increased reactions rate.

6.2 Optimization of aqueous-phase Sonogashira cross coupling reaction conditions for iodinated 2',3'-dideoxynucleosides

The second and one of the most important steps in the chosen synthetic strategy (Scheme 24b) is linker formation by introducing functional group for subsequent labeling. In the previous chapter 5.2.2 the successful utilization of aqueous-phase Sonogashira cross-coupling reaction for d^{PA}NTPs was shown. However, the obtained conversions to the products and yields were not very high. Consequently, comprehensive investigation on finding the efficient reaction conditions, for propargylamine (PA) attachment to dd^{I}NTPs via Sonogashira cross-coupling reaction, was carried out (Scheme 31).



Scheme 31. Synthesis of amino-functionalized ddNTPs (dd^{PA}NTPs).

To optimize coupling reaction conditions for the synthesis of dd^{PA}CTP, two different palladium catalysts were used. Due to poor solubility of dd^{I}CTP in DMF, bis(triphenylphosphine)palladium (II) chloride was moderately effective giving 45% conversion to the desired product. When DMF:H_2O (2:1) mixture was used the conversion decreased to 29% (Table 16, entries 1–2). Employing $Pd(OAc)_2$ alongside with the water-soluble ligand TPPTS ($P(C_6H_4SO_3Na)_3$) in water-acetonitrile (2:1) mixture and maintaining reaction at room temperature only 16% conversion was obtained (Table 16, entry 3). Neither the increase of propargylamine or Et_3N quantity nor reaction temperature to 70 °C gave any positive results. On the contrary, some decomposition of triphosphate was observed when the reaction was carried out at 70 °C (Table 16, entries 4–5), which is in agreement with previous reports. [18] Significant improvement was achieved when reaction concentration was increased to 0.14 M. More importantly, no addition of base or high reaction temperature was required (Table 16, entries 6–7).

Table 16. Optimization of the Sonogashira reaction conditions for dd^{PA}CTP.

Entry	Propargylamine, eq.	CuI, eq.	Catalyst	Et_3N, eq.	Concentration, M	Time	Conversion to dd^{PA}CTP, %
1[a]	2	0.2	$[(C_6H_5)_3P]_2PdCl_2$ (10 mol%)	3	0.05	16h	45
2[b]	2	0.2	$[(C_6H_5)_3P]_2PdCl_2$ (10 mol%)	3	0.05	16h	29
3	2	0.1	$Pd(OAc)_2$ (5 mol%) /TPPTS (25 mol%)	2	0.05	16h	16
4	3	0.1	$Pd(OAc)_2$ (5 mol%) /TPPTS (25 mol%)	10	0.05	16h	≤3
5 [c]	5	0.1	$Pd(OAc)_2$ (5 mol%) /TPPTS (25 mol%)	10	0.05	1h	9
6	2	0.1	$Pd(OAc)_2$ (5 mol%) /TPPTS (25 mol%)	2	0.14	1h 2h 3h 4h 16h	52 58 62 64 76
7	2	0.1	$Pd(OAc)_2$ (5 mol%) /TPPTS (25 mol%)	-	0.14	1h 4h 16h	69 72 79
8[d]	2	0.1	$Pd(OAc)_2$ (5 mol%) /TPPTS (25 mol%)	-	0.14	30min 1h 1.5h	45 74 75
9[e]	**2**	**0.1**	**$Pd(OAc)_2$ (5 mol%) /TPPTS (25 mol%)**	**-**	**0.14**	**1h**	**92**

All reactions were maintained at room temperature and in ACN:H_2O (1:2) media unless specified differently. Small scale experiments were performed utilizing 0.05 – 0.1 mmol of starting nucleotide. a – reaction was carried out in DMF, b – reaction was carried out in DMF:H_2O (2:1), c – reaction was carried out at 70 °C temperature, d – reaction scale-up results (1 – 2.5 mmol), e – scale-up reaction carried out at 40 °C temperature.

The dd^{PA}CTP scale-up synthesis was performed at room temperature, using 2 eq. of propargylamine and $Pd(OAc)_2$/TPPTS catalytic system after 1.5 hours giving 75% conversion to the desired product (Table 16, entry 8). To increase the conversion, an additional scale-up procedure was performed maintaining the exact conditions as indicated in entry 8 (Table 16) except for slight 40 °C heating. At such temperature, no decomposition of triphosphate was observed and in 1 hour 92% conversion to dd^{PA}CTP was reached. The purification by ion exchange chromatography afforded 88% yield of the final product.

Since reactivity similarities between dd^{I}C and dd^{I}A were observed during monophosphorylation, the best Sonogoshira reaction conditions of dd^{PA}CTP were applied for dd^{PA}ATP. Maintaining reaction at room temperature without base additive, after 1 hour the 72% conversion to dd^{PA}ATP was obtained (Table 17, entry 1). An even higher reaction rate was observed when 2 eq. of triethylamine were applied. Here, after an hour conversion to the desired product reached 85%, while prolonged reaction time led to slow degradation of the product (Table 17, entry 2). The successful dd^{PA}ATP synthesis scale-up was performed according to conditions 2 (Table 17) where after 30 min 86% conversion to the product was attained (Table 17, entry 3). The reaction mixture was purified yielding 69%.

Table 17. Optimization of the Sonogashira reaction conditions for dd^{PA}ATP.

Entry	Propargylamine, eq.	CuI, eq.	Catalyst	Et_3N, eq.	Time	Conversion to dd^{PA}ATP, %
1	2	0.1	$Pd(OAc)_2$ (5 mol%) /TPPTS (25 mol%)	-	1h	72
					2h	76
					3h	79
					4h	78
					16h	83
2	2	0.1	$Pd(OAc)_2$ (5 mol%) /TPPTS (25 mol%)	2	1h	85
					2h	84
					3h	80
					4h	79
					16h	61
3[a]	**2**	**0.1**	**$Pd(OAc)_2$ (5 mol%) /TPPTS (25 mol%)**	**2**	**30min**	**86**

All reactions were maintained at room temperature and in ACN:H_2O (1:2) media unless specified differently. Small scale experiments were performed utilizing 0.05 – 0.1 mmol of starting nucleotide. a – scale-up reaction results (1 – 2.5 mmol).

Based on previous results, a brief screening of reaction conditions for dd^{PA}UTP synthesis was carried out. It was noticed that slight heating (40 °C) was necessary for efficient synthesis. Moreover, if the amount of the additional base increases, the conversion to target material in time decreases, and a more complex impurity profile is generated (Table 18, entries 1–2). The use of DIPEA instead of TEA had no influence, while the increase in concentration shortened reaction time and enhanced conversion to desired

product to 92% (Table 18, entries 3–5). After purification dd^{PA}UTP was attained in 73% yield.

Table 18. Optimization of the Sonogashira reaction conditions for dd^{PA}UTP.

Entry	Propargylamine, eq.	CuI, eq.	Catalyst	Base, eq.	Concentration, M	Time	Conversion to dd^{PA}UTP, %
1	2	0.1	$Pd(OAc)_2$ (5 mol%) /TPPTS (25 mol%)	TEA, 2	0.07	1h	67
						2h	79
						3h	76
						4h	74
						16h	69
2	2	0.1	$Pd(OAc)_2$ (5 mol%) /TPPTS (25 mol%)	TEA, 10	0.07	1h	7
						2h	22
						3h	37
						4h	43
						16h	41
3	2	0.1	$Pd(OAc)_2$ (5 mol%) /TPPTS (25 mol%)	DIPEA, 2	0.07	1h	59
						2h	76
						3h	74
						4h	72
						16h	60
4	2	0.1	$Pd(OAc)_2$ (5 mol%) /TPPTS (25 mol%)	-	0.07	1h	67
						2h	67
						3h	68
						4h	68
5[a]	**2**	**0.1**	**$Pd(OAc)_2$ (5 mol%) /TPPTS (25 mol%)**	**-**	**0.1**	**1h**	**92**

Reactions were carried out at 40 °C in ACN:H_2O (1:2) mixture. Small scale experiments were performed utilizing 0.05 – 0.1 mmol of starting nucleotide. a – scale-up reaction results (1 – 2.5 mmol).

Tendencies similar to dd^{PA}UTP synthesis were observed in the case of dd^{PA}GTP. When the reaction was carried out at room temperature only 44% conversion in 4 hours was detected (Table 19, entry 1). Slight heating (40 °C) increased formation of the product as well as the absence of base did (Table 8, entries 2 – 3). Higher concentration resulted in 72% conversion to the target material in 3 hours. dd^{PA}GTP was purified to yield 41%, moderate yield is a result of the strong interaction of product with ion exchange resin during purification as it was observed in the case of dd^{I}GTP.

Table 19. Optimization of the Sonogashira reaction conditions for dd^{PA}GTP.

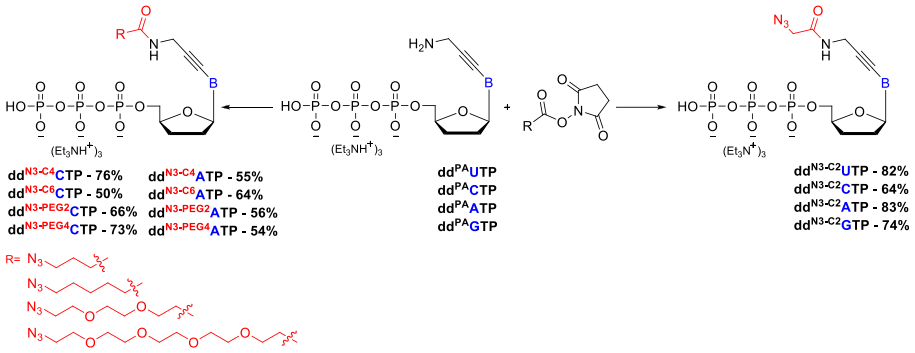
Entry	Propargylamine, eq.	CuI, eq.	Catalyst	Et_3N, eq.	Concentration, M	Time	Conversion to dd^{PA}GTP, %
1[a]	2	0.1	$Pd(OAc)_2$ (5 mol%) /TPPTS (25 mol%)	-	0.07	4h	44
2	2	0.1	$Pd(OAc)_2$ (5 mol%) /TPPTS (25 mol%)	2	0.07	4h	56
3	2	0.1	$Pd(OAc)_2$ (5 mol%) /TPPTS (25 mol%)	-	0.07	4h	61
4[b]	**2**	**0.1**	**$Pd(OAc)_2$ (5 mol%) /TPPTS (25 mol%)**	**-**	**0.1**	**3h**	**72**

Reactions were carried out at 40 °C in ACN:H_2O (1:2) mixture. Small scale experiments were performed utilizing 0.05 – 0.1 mmol of starting nucleotide. a – reaction was carried out at room temperature. b – scale-up reaction conditions (1 – 2.5 mmol).

The functionalized dd^{PA}NTPs were synthesized in a scalable straightforward two-step methodology in comparison to the conventional four-step synthetic strategy. The Sonogashira reactions were successfully carried out on 1 – 2.5 mmol scale without any loss on products yield. Such functionalized compounds could be used for direct polymerase-catalyzed incorporation into DNA strand for post-enzymatic modification of nucleic acids or could be precursors for further labeling (fluorescent label, dye, biotin, etc.).

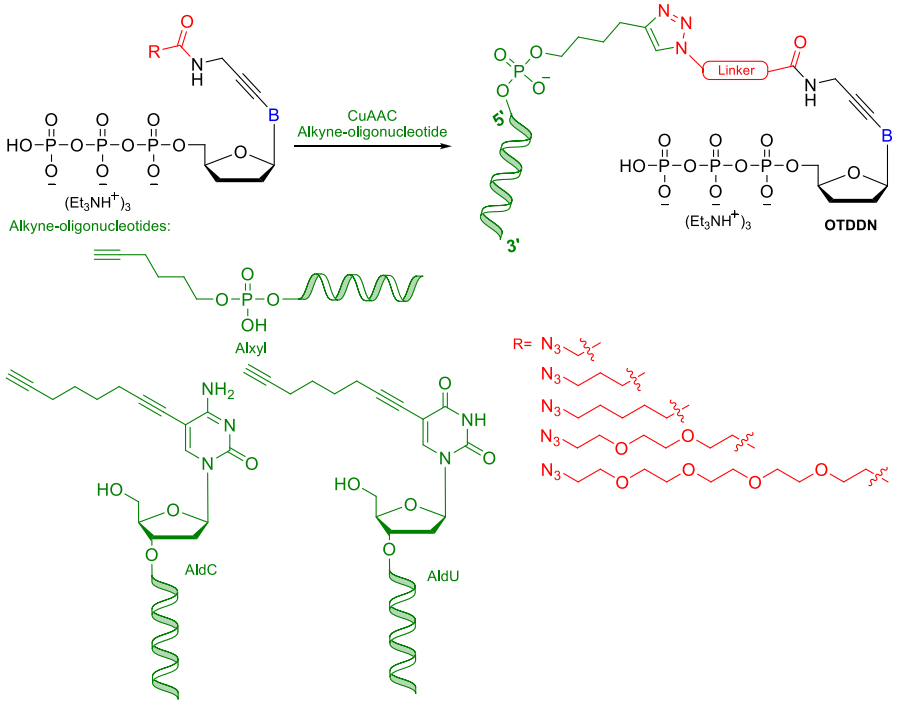
6.3 Synthesis of azido-group bearing and oligonucleotide-tethered 2',3'-dideoxynucleotides

After dd^{PA}NTPs were successfully synthesized and purified azido-moieties were introduced utilizing different azido-NHS esters (Scheme 32). Reactions were maintained in phosphate buffer (pH 9) and DMF mixture at room temperature and in the presence of different azido-NHS ester. All four dd$^{N3\text{-}C2}$NTPs were obtained in good yields (64 – 83%). Diverse linkers bearing azido-ddCTPs and ddATPs were also synthesized (50 – 76% yield) (Scheme 32).



Scheme 32. Synthesis of diverse azido-group bearing ddNTPs.

Different applications required the diversity in oligonucleotide-tethered 2',3'-dideoxynucleotide (OTDDNs) by means of changing oligonucleotide sequences and modifications. Terminal alkyne moieties were introduced by two types of modifications: Al – octa-1,7-diene attached to the pyrimidine nucleobases C5-posisition, Alxyl – hexynyl-modification attached directly to the oligonucleotide 5'-termini phosphate (Scheme 33). Al modification could be attached only at C and U nucleobases, while Alxyl – to all four. Consequently, Al modification was eventually replaced by Alxyl.



Scheme 33. Synthesis of oligonucleotide-tethered 2',3'-dideoxynucleotides (OTDDN).

For barcodes introduction randomized sequences were also necessary (indicated as N), several oligonucleotides were selected for such matter (Table 20, ON5 and ON8 – 11). Phosphorothioate modification at the 3'-termini was used to increase nuclease resistance (indicated as *). 3'-Termini, in some cases, was additionally labeled by biotin for the ability to purify nucleic acids using streptavidin coated magnetic beads, while 3'-phosphate – to increase exonuclease resistance (Table 20).

Table 20. Oligonucleotides applied in CuAAC for OTDDNs synthesis.

Entry	Name	Sequence	5'-termini	3'-termini
1	ON1	**C**TATAGTGAGTCGTATTA	AldC	-
2	ON2	**U**AGATCGGAAGAGCACACGTCTG	AldU	Biotin
3	ON3	**U**AGATCGGAAGAGCACACGTCTG	AldU	Phosphate
4	ON4	**U**AGATCGGAAGAGCACACGTCTGA ACTCCAGTCACATGCCTAAATCTCG TATGCCGTCTTCTGCTTG	AldU	Biotin
5	ON5	**U**NNNNNNNNAGATCGGAAGAGCGT CGTGTA	AldU	Biotin
6	ON6	**A**GATCGGAAGAGCACACGTCTG	Alxyl	Biotin
7	ON7	**A**GATCGGAAGAGCACACGT*C*T*G	Alxyl	Phosphate
8	ON8	**N**NNNNNNNAGATCGGAAGAGCACA CGT*C*T*G	Alxyl	Phosphate
9	ON9	**N**NNNNNNNAGATCGGAAGAGCGTC GTGTA	Alxyl	Biotin
10	ON10	**N**NNNNNNNAGATCGGAAGAGCACA CGTCTG	Alxyl	Biotin
11	ON11	**N**NNNNNNNAGATCGGAAGAGCGTC GTGTAGGGAAAGAG	Alxyl	Phosphate
12	ON12	**A**GATCGGAAGAGCACACGTCTG	Alxyl	Phosphate
13	ON13	**C**TGTCTCTTATACACATCTCCGAG	Alxyl	Biotin
14	ON14	**AT**CACCGACTGCCCATAGAGAGGA AAGCGGAGGCGTAGTGG	Alxyl	Phosphate

* – indicates phosphorothioate (PTO) junction, N – indicates randomized sequence, where N = any nucleotide, AldU – alkynyl-modification at the C5- position of uridine nucleobase (octa-1,7-diyne), AldC – alkynyl-modification at the C5-position of cytidine nucleobase (octa-1,7-diyne), Alxyl – hexynyl-modification attached to the oligonucleotide 5'-termini phosphate.

Small scale reactions (8.4 nmol) were performed according to the guidance provided by Jena Bioscience "CuAAC Biomolecule Reaction Buffer Kit" with some modifications. Due to the small reaction scale low concentration (42μM) and large excess of Cu/THPTA (48 eq.) catalyst were required. Monitoring reactions by HPLC the decomposition of oligonucleotide, mediated by Cu, was observed. The reactions reach their highest conversion to the product

values disregarding the residual amount of alkyne-oligonucleotide left. The intensity of OTDDN peak, eventually, starts to decrease, indicating of a possible oligonucleotide decomposition. The same results were monitored in the previous experiments with OTDNs synthesis and were confirmed by LC/MS analysis. Moreover, such decomposition is well-known and broadly discussed in many articles. [235, 236, 237] The CuAAC reactions scale-up experiments (to 84 – 400 nmol) were successfully proceeded. This enabled to increase reaction concentration (100 – 400μM) and, consequently, to reduce Cu amount twice, maintaining the same or increased reaction rate and conversion to the product. Series of OTDDNs were successfully synthesized and are present in the Table 21 (Scheme 33). The vast majority of OTDDNs synthesized were the cytidine (C) and uridine (U) derivatives. While in case of ON6, representatives of all four nucleotides were obtained (Table 21, entries 8, 10, 15 and 17). Hypothesizing that linker length would play a pivotal role in efficient read-through and not yet having the suitable method to accurately measure that, the shortest linker (C2) was selected as the priority for the further functional testing. After the development of proper method for read-through efficiency measurements, the OTDDNs with different linkers were synthesized bearing the ON6 oligonucleotide attached to the dd^{N3}CTPs (Table 21, entries 10 – 14).

The biggest challenge was the purification of the complex click reaction mixtures. Alkyne-oligonucleotide differs from dd^{ON}NTPs by a single nucleotide making them hard to separate. Modified azido-ddNTPs have long retention times which in the case of purines are very alike to both starting-oligonucleotide and product. A residual amount of each reactant in the final product formulation is unacceptable for enzymatic applications. Copper would inhibit polymerases, azido-ddNTPs would compete with dd^{ON}NTPs for incorporation (that was already observed with OTDNs), while the residual amount of alkyne-oligonucleotide can deplete primers in the further read-through assay. To obtain accurate and unambiguous results the precise HPLC purification methods had to be developed. The diversity of oligonucleotides length, modifications, sequences also required the different HPLC methods. Nearly 20 methods were developed for OTDDNs purification, according to which, products were obtained in ≥95% purity. More importantly, neither of OTDDNs contained any critical impurities – residual dd^{N3}NTPs, Cu or alkyne-oligonucleotide, as it was observed in OTDNs case.

Table 21. Synthesized OTDDNs for enzymatic applications.

Entry	ON	Azide	OTDDN	SM, nmol	TM, nmol	Yield, %
1	ON1	dd$^{N3\text{-}C2}$UTP	dd$^{ON1\text{-}C2}$UTP	8.4	1.4	17
2	ON2	dd$^{N3\text{-}C2}$UTP	dd$^{ON2\text{-}C2}$UTP	8.4	2.5	30
3		dd$^{N3\text{-}C2}$CTP	dd$^{ON2\text{-}C2}$CTP	8.4	3.4	41
4	ON3	dd$^{N3\text{-}C2}$UTP	dd$^{ON3\text{-}C2}$UTP	8.4	1.8	21
5	ON4	dd$^{N3\text{-}C2}$CTP	dd$^{ON4\text{-}C2}$CTP	8.4	2.6	31
6	ON5	dd$^{N3\text{-}C2}$UTP	dd$^{ON5\text{-}C2}$UTP	8.4	3.5	42
7	ON6	dd$^{N3\text{-}C2}$UTP	dd$^{ON6\text{-}C2}$UTP	8.4	2.1	25
8		dd$^{N3\text{-}C2}$UTP	dd$^{ON6\text{-}C2}$UTP	200	68	34
9		dd$^{N3\text{-}C2}$CTP	dd$^{ON6\text{-}C2}$CTP	84	24	28
10		dd$^{N3\text{-}C2}$CTP	dd$^{ON6\text{-}C2}$CTP	210	82	39
11		dd$^{N3\text{-}C4}$CTP	dd$^{ON6\text{-}C4}$CTP	210	51	24
12		dd$^{N3\text{-}C6}$CTP	dd$^{ON6\text{-}C6}$CTP	210	39	18
13		dd$^{N3\text{-}PEG2}$CTP	dd$^{ON6\text{-}PEG2}$CTP	210	30	14
14		dd$^{N3\text{-}PEG4}$CTP	dd$^{ON6\text{-}PEG4}$CTP	210	25	12
15		dd$^{N3\text{-}C2}$ATP	dd$^{ON6\text{-}C2}$ATP	100	29	29
16		dd$^{N3\text{-}C2}$GTP	dd$^{ON6\text{-}C2}$GTP	8.4	1.6	19
17		dd$^{N3\text{-}C2}$GTP	dd$^{ON6\text{-}C2}$GTP	210	90	43
18	ON7	dd$^{N3\text{-}C2}$UTP	dd$^{ON7\text{-}C2}$UTP	16.8	3.1	18
19		dd$^{N3\text{-}C2}$UTP	dd$^{ON7\text{-}C2}$UTP	200	24	12
20		dd$^{N3\text{-}C2}$CTP	dd$^{ON7\text{-}C2}$CTP	210	50	24
21	ON8	dd$^{N3\text{-}C2}$UTP	dd$^{ON8\text{-}C2}$UTP	16.8	2.7	16
22		dd$^{N3\text{-}C2}$CTP	dd$^{ON8\text{-}C2}$CTP	8.4	1.5	18
23	ON9	dd$^{N3\text{-}C2}$UTP	dd$^{ON9\text{-}C2}$UTP	8.4	2.2	27
24		dd$^{N3\text{-}C2}$CTP	dd$^{ON9\text{-}C2}$CTP	8.4	0.9	11
25	ON10	dd$^{N3\text{-}C2}$UTP	dd$^{ON10\text{-}C2}$UTP	100	32	32
26		dd$^{N3\text{-}C2}$CTP	dd$^{ON10\text{-}C2}$CTP	210	115	55
27	ON11	dd$^{N3\text{-}C2}$UTP	dd$^{ON11\text{-}C2}$UTP	200	40	20
28		dd$^{N3\text{-}C2}$CTP	dd$^{ON11\text{-}C2}$CTP	210	78	37
29	ON12	dd$^{N3\text{-}C2}$UTP	dd$^{ON12\text{-}C2}$UTP	400	116	29
30		dd$^{N3\text{-}C2}$CTP	dd$^{ON12\text{-}C2}$CTP	300	143	48
31	ON13	dd$^{N3\text{-}C2}$CTP	dd$^{ON13\text{-}C2}$CTP	210	77	37
32	ON14	dd$^{N3\text{-}C2}$UTP	dd$^{ON14\text{-}C2}$CTP	100	36	36
33		dd$^{N3\text{-}C2}$CTP	dd$^{ON14\text{-}C2}$CTP	100	32	32

In all CuAAC reactions 2 – 3 eq. of corresponding azide was used. Reactions were maintained at 37 ℃. SM – starting material (alkyne-oligonucleotide), TM – target material (corresponding OTDDN). All reaction components were used as solutions in water except alkyne-oligonucleotide (solution in sodium phosphate buffer). Where 8.4 nmol of SM were used, reactions volume was 200 uL (42μM), reaction volume for 84 – 400 nmol was 1ml (100 – 400μM). All OTDDNs were obtained in ≥95% purity.

Copper mediated decomposition of oligonucleotides is the crucial factor, moreover, any residual amount of Cu in the final product defines the quality of OTDDN. When the huge excess of copper is used the risk of incomplete

disposal during purification increases. In several OTDDN synthesis cases there have been observed inefficient removal of Cu. The future outlook for efficient OTDDNs synthesis should focus on minimization of copper amount per reaction, this should be achievable for larger scale experiments by further increasing reaction concentration and decreasing Cu. Performing click reactions in 84 – 400 nmol scale we were already able to decrease Cu quantity by half, as mentioned previously. To determine the least amount of Cu needed for CuAAC reactions in our used average scale (100 nmol), several reaction conditions were tested. The conditions C1 indicates the present experimental data used in OTDDN scale-up experiments, while C2, C3 and C4 represents the twice reduced Cu/THPTA catalytic system amount in comparison to every previous condition (Table 22). The reactions were performed with both $dd^{N3\text{-}C2}$UTP and $dd^{N3\text{-}C2}$CTP and alkyne oligonucleotide ON6. Reactions were monitored for 25 minutes by HPLC.

Table 22. CuAAC reaction conditions for Cu reduction.

Reagent	C1, µmol	C2, µmol	C3, µmol	C4, µmol
ON6	0.1	0.1	0.1	0.1
$dd^{N3\text{-}C2}$UTP/ $dd^{N3\text{-}C2}$CTP	0.21	0.21	0.21	0.21
$CuSO_4$	**2.5**	**1.25**	**0.625**	**0.3**
THPTA	12.5	6.25	3.125	1.5
Na-ascorbate	125	62.5	31.25	15

C – conditions. Reactions were performed at 37 °C temperature, for 25 minutes, reactions concentration was 125µM (reaction volume 0.8 ml). All reaction components were used as solutions in either water or sodium phosphate buffer (pH 7).

Graphical representation of conversion to the corresponding OTDDN dependency on Cu amount utilized in the reaction is given in Figure 34. The obtained results for both reactants ($dd^{N3\text{-}C2}$UTP and $dd^{N3\text{-}C2}$CTP) show that the catalyst amount can be reduced by four times more (Figure 34 C3) in comparison to conditions C1. Such decrease in catalyst did not have any significant impact on the amount of product formed. However, conditions C4 (Cu amount reduced by 8 times) resulted in less than half lower reaction rate for both $dd^{N3\text{-}C2}$NTPs. The slightly better reactivity was observed with $dd^{N3\text{-}C2}$UTP. Employing reaction conditions C1, C2 or C3 after 25 min conversion to the product was obtained between 80% and 90%. While $dd^{N3\text{-}C2}$CTP according to same reaction conditions resulted in 63% to 70% conversion to OTDDN.

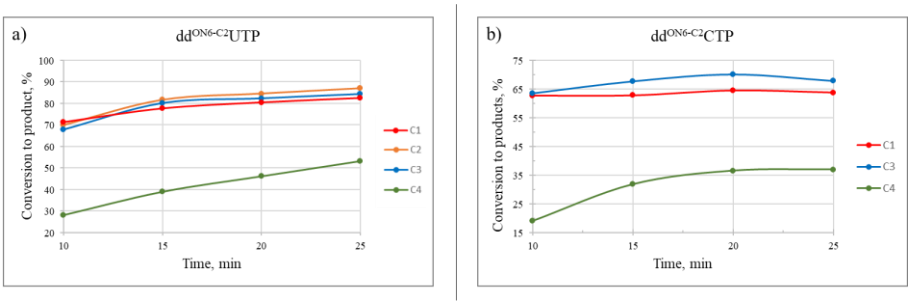


Figure 34. CuAAC reaction conversion to the OTDDN dependency on the Cu amount used.[5]

All in all, azido-groups bearing ddNTPs were successfully obtained in good yields according to optimized three-step synthetic strategy. Numerous OTDDNs were obtained applying copper catalyzed alkyne-azide cycloaddition reaction. Successful scaled-up experiments were performed with various alkyne-oligonucleotides and azido-ddNTPs. Diverse HPLC methods for OTDDN purification were developed enabling disposal of critical impurities. The brief experiments in attempts to reduce copper amount per CuAAC reaction were performed indicating perspective possibilities for reaction improvement and reduction of products degradation.

[5] These experiments were performed by Simona Žeimytė.
Simona Žeimytė also contributed to the preparation of oligonucleotide-tethered 2',3'-dideoxynucleotides.

7. OLIGONUCLEOTIDE-TETHERED 2',3'-DIDEOXYNUCLEOTIDES APPLICATION FOR NEXT-GENERATION SEQUENCING LIBRARY PREPARATION

As it was mentioned before the library preparation for sequencing is a multistep procedure with several drawbacks. In the context of NGS library preparation, the incorporation of dd^{ON}NTP would fulfill two requirements at once: the fragmentation step would be integrated into the workflow and the obtained fragments would readily be labelled by platform-specific adapters. In the literature overview (chapter 4), we discussed some techniques such as Click-Seq that are directed to the use of 3'-azido-2',3'-dideoxynucleotide terminators for NGS library preparation. Even though the notorious platform-specific adapter enzymatic introduction was simplified by performing CuAAC instead (Figure 35a and 34b), the several drawbacks remained: 1) after fragmentation, additional purification prior to chemical ligation was necessary to remove the residual free 3'-azido-ddNTP; 2) for efficient click reaction a complementary template to 3'-azido oligonucleotide was necessary, therefore it was impossible to implement for unknown sequences; 3) non-templated chemical ligation required high DNA amount which is far from practical for many NGS applications; 4) to avoid Cu-induced inhibition of polymerases, additional purification had to be performed after chemical ligation; 5) more than half of the obtained reads were shorter than expected, caused by Cu-mediated DNA degradation; [15] 6) the read-through efficiency

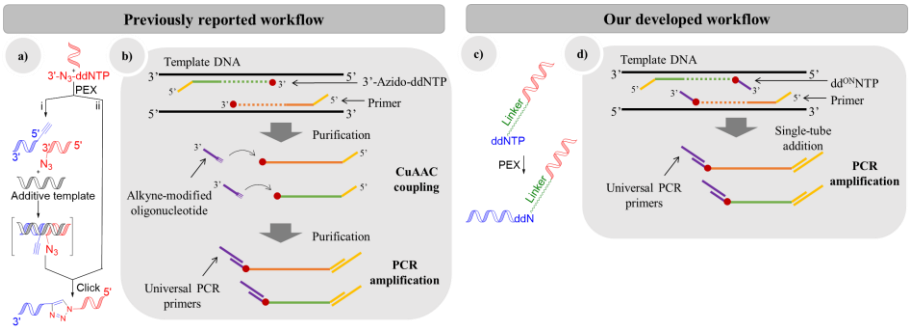


Figure 35. Comparison of NGS library preparation strategies: Previously reported workflow a) schematic representation of chemical ribose-to-ribose ligation, i) templated click reaction; ii) non-templated click reaction b) workflow for NGS library preparation using ribose-to-ribose ligation; Our reported workflow c) ligation, PEX/termination by dd^{ON}NTP and d) NGS library preparation workflow using dd^{ON}NTP.

was low, as single primer extension cycle generated undetectable amount of product (less than 4%) (Figure 35a and 34b). Despite mentioned drawbacks, these studies have shown the possibilities to design and successfully use nucleic acids with artificial backbones in molecular biology applications.

According to all drawbacks of the existing techniques the workflow proposed here overcomes all of them (Figure 35c and 34d). The utility of the OTDDN instead of azido-ddNTP in the PEX would result not only in termination (alternative to fragmentation) providing the specific range of DNA fragments but at the same time adapter would be also introduced. The significant benefit here underlies the evade of click reaction with samples to be tested, since this step causes most of the problems. Moreover, the proposed workflow is substantially simplified in comparison even to the Click-Seq technique (Figure 35b and 34d).

7.1 Azido-group bearing and oligonucleotide-tethered 2',3'-dideoxynucleotides enzymatic incorporation

After dd^{N3}NTPs and dd^{ON6}NTPs were synthesized and purified primer extension experiments were run using polymerases of families A, B, X and RT. In general, DNA polymerases are classified into seven families (A, B, C, D, X, Y and RT) according to their crystallographic structure and sequence homology, [238] which could lead to their different performance with our substrates.

To begin with, KOD XL polymerase was tested for dd^{N3}NTPs incorporation, similarly as in the case with d^{N3}NTPs. The used template contained four overhanging nucleotides; reactions were incubated for 20 min at 60°C. The obtained PAGE analysis showed that all dd^{N3}NTPs were successfully incorporated by KOD XL DNA polymerase with no significant differences (Figure 36).

Hereinafter, more polymerases were tested for $dd^{N3\text{-}C2}$NTPs and $dd^{ON6\text{-}C2}$NTPs incorporation. Oligonucleotide duplex that served as a template contained 10 nt overhang. Hence, during PEX in the presence of dNTPs up to 10 nucleotides could be added to the primer, while the incorporation of dd^{ON6}NTP would result in primer elongated by 23 nt (Table 23). All tested polymerases were 3'-5' exonuclease-deficient and accepted dd^{N3}NTPs and dd^{ON6}NTPs as substrates. The efficiency of dd^{ON6}NTP incorporation varied between polymerases (Figure 37). In addition, a slight preference for the nucleobase of terminator was observed. Klenow exo- exhibited better incorporation efficiency with oligonucleotide-modified 2',3'-dideoxypurines rather than pyrimidines, and similar trend was observed with dd^{N3}NTP

derivatives. Thermo Sequenase, another DNA polymerase from family A, was able to incorporate all types of dd^{N3}NTPs and dd^{ON6}NTPs with superior efficiency compared to other tested enzymes, this could be attributed to its ability not to discriminate between dNTPs and ddNTPs. [239]

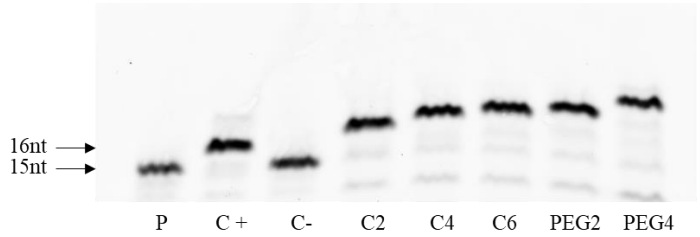


Figure 36. Electropherograms representing PEX results of d^{N3}NTPs incorporation by DNA polymerase KOD XL. P – primer (15nt), C+ – positive control, PEX reaction contained single ddCTP nucleotide, C- – negative control, PEX reaction in the absence of ddNTPs. C2, C4, C6, PEG2, PEG4 – indicate the linkers connecting nucleobase to the azido-group.

Table 23. Oligonucleotide duplexes used for dd$^{N3\text{-}C2}$NTP and dd$^{ON6\text{-}C2}$NTP incorporation testing.

Name	Oligonucleotide duplex
Dup^{A}	5’-AAAAAAAAAATACGCCAAGGATGCCTACCCATGTCTGCA-3’ 3’-ATGCGGTTCCTACGGATGGGTACAGACGT-Cy5-5’
Dup^{T}	5’-TTTTTTTTTTTACGCCAAGGATGCCTACCCATGTCTGCA-3’ 3’-ATGCGGTTCCTACGGATGGGTACAGACGT-Cy5-5’
Dup^{G}	5’-GGGGGGGGGGTACGCCAAGGATGCCTACCCATGTCTGCA-3’ 3’-ATGCGGTTCCTACGGATGGGTACAGACGT-Cy5-5’
Dup^{C}	5’-CCCCCCCCCCTACGCCAAGGATGCCTACCCATGTCTGCA-3’ 3’-ATGCGGTTCCTACGGATGGGTACAGACGT-Cy5-5’

Phusion exo- DNA polymerase from family B catalyzed the addition of dd^{ON6}NTPs with lower efficiency compared to Thermo Sequenase, especially in the case of dd^{ON6}UTP. Terminal deoxynucleotidyl transferase (TdT) performed with the lowest efficiency, although TdT is known to tolerate bulky nucleobase modifications. [187] This result could be attributed to suboptimal reaction conditions, such as the dd^{ON6}NTP-to-template ratio. The dd^{ON6}NTP addition by SuperScript™ IV reverse transcriptase (RT) had similar efficiency as the Thermo Sequenase (Figure 37). In practice an dd^{ON6}NTP-compatible RT enzyme could enable cDNA labelling, while a thermostable DNA polymerase could be used in nucleic acid analysis workflows with ssDNA or dsDNA templates. [240]

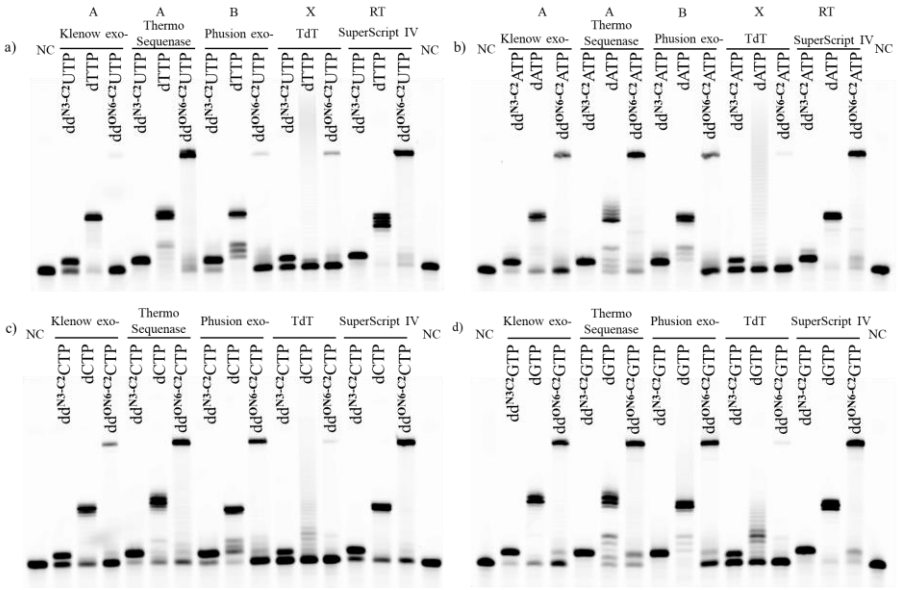


Figure 37. dd^{N3}NTP and dd^{ON}NTP incorporation by DNA polymerases. Electropherograms showing PEX results obtained using various DNA polymerases with either dd^{N3}NTPs, dNTPs or dd^{ON}NTPs in the presence of the following templates: a) Dup^{A}, b) Dup^{T}, c) Dup^{G}, d) Dup^{C}. NC – negative control. [6]

Diverse linkers bearing dd^{ON6}NTPs were also evaluated as substrates during PEX. Here the same system was used, and three polymerases were tested. Same tendencies between polymerases were observed, with no significant differences regarding the nature of linkers, where Thermo Sequenase exhibited the best incorporation efficiency, similarly to what was previously observed (Figure 38).

All azido-ddNTPs were good substrates for polymerases of diverse families and were successfully incorporated during PEX reaction. No significant differences during enzymatic incorporation were noticed due to the structure of linkers. On the contrary, the OTDDN incorporation was shown to be of higher difficult for some polymerases than others, furthermore, in some cases the discrimination of specific nucleobases was also observed. Thermo Sequenase and SuperScript IV reverse transcriptase exhibited the highest incorporation efficiency, while TdT – the lowest. The polymerases that incorporated azido-ddNTPs with the lowest efficiency, also gave low incorporation of the corresponding OTDDN, suggesting that OTDDN incorporation efficiency is not determined just by bulky oligonucleotide-label,

[6] These experiments were performed by Žana Kapustina and Rūta Sindikevičienė.

but also by the substrate properties of 2‘,3‘-dideoxy-ribose ring and/or nucleobase modification. This also is in agreement with the incorporation of diverse linkers bearing OTDDNs where no significant differences in-between were observed.

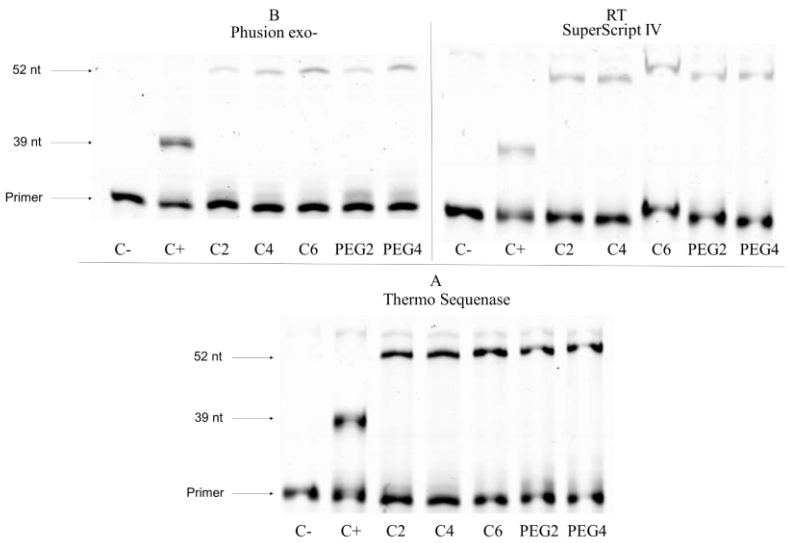


Figure 38. Incorporation of dd^{ON6}CTPs bearing different linkers by three DNA polymerases. C- - negative control, C+ - positive control (PEX in the presence of dCTPs), C2 – PEX in the presence of dd$^{ON6\text{-}C2}$CTP, C4 - PEX in the presence of dd$^{ON6\text{-}C4}$CTP, C6 - PEX in the presence of dd$^{ON6\text{-}C6}$CTP, PEG2 - PEX in the presence of dd$^{ON6\text{-}PEG2}$CTP, PEG4 - PEX in the presence of dd$^{ON6\text{-}PEG4}$CTP.

7.2 DNA library preparation enabled by oligonucleotide-tethered 2’,3’-dideoxynucleotides

7.2.1 Semi-targeted sequencing of single-stranded M13mp18 viral genome

To prove the technical feasibility, low complexity single-stranded genome of M13mp18 bacteriophage was selected. Target-specific primers targeting two loci within the M13mp18 genome were designed to contain a partial Illumina P5 adapter sequence (Figure 39a – Specific primer #1 and 2). Thermo Sequenase was applied for dd$^{ON6\text{-}C2}$UTP incorporation, where 15 linear extension cycles were performed. Upon incorporation, dd$^{ON6\text{-}C2}$UTP terminated DNA synthesis and simultaneously labelled the nascent DNA strand with a partial Illumina adapter sequence attached to its nucleobase. This step generated fragments of suitable length for short read sequencing. The obtained fragments were subsequently subjected for indexing PCR utilizing primers targeting P5 and P7 Illumina adapters (Figure 39a). Phusion exo-

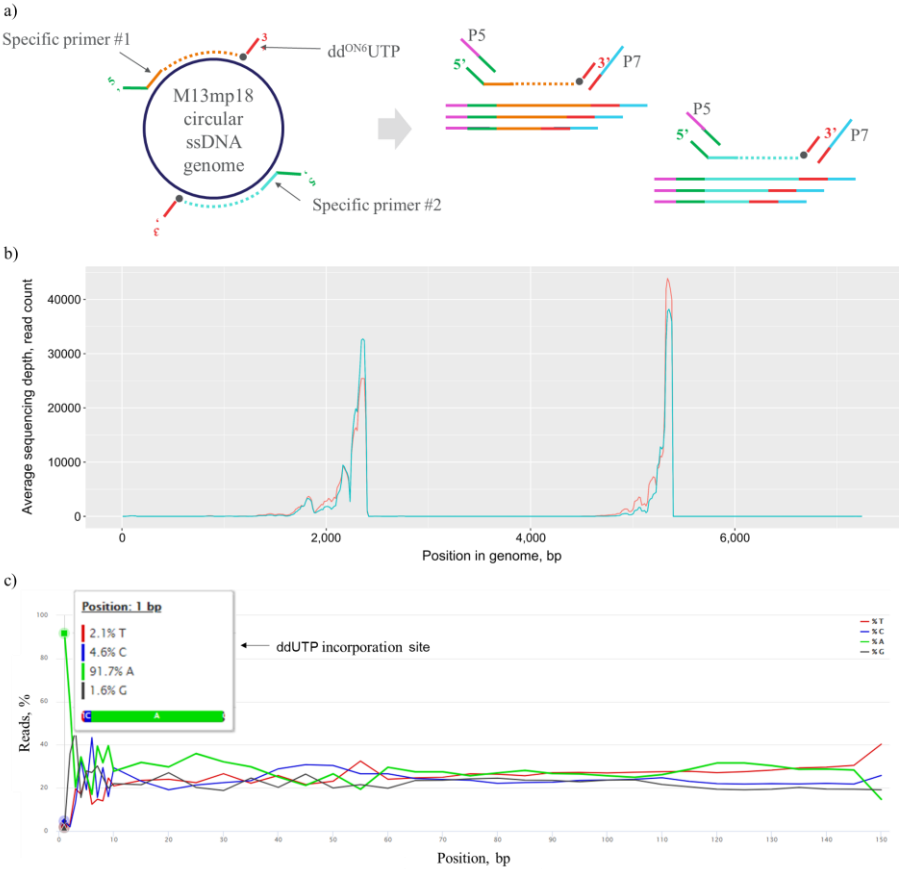


Figure 39. Semi-targeted sequencing of the M13mp18 viral genome. a) A schematic overview of library preparation with dd$^{ON6\text{-}C2}$UTPs. b) M13mp18 genome coverage. The reads concentrated at two loci with one terminus of sequenced inserts fixed at the specific priming sites. Another terminus corresponds to the stochastic positions of dd$^{ON6\text{-}C2}$UTP incorporation. The orange and blue lines represent technical replicates. c) Base composition of sequenced reverse reads. The dominance of A base at the first position indicates a seamless copying of the template around the linker position.[7]

polymerase was used in library amplification reaction as a helper enzyme to achieve reading through the unnatural linkage within OTDDN. After indexing PCR, the prepared fragment library was ready for subsequent sequencing. The obtained sequencing data indicated the capture of two genomic loci with a

[7] These experiments were performed by Žana Kapustina.

characteristic read structure: the fixed insert terminus corresponded to the specific priming site, while the randomly distributed terminus illustrated the stochastic nature of dd$^{ON6\text{-}C2}$UTP incorporation (Figure 39b). The inspectionof sequencing reads has not revealed any systematic errors or artefacts which could be associated with the presence of an unnatural backbone within the fragment library. The analysis of base distribution per sequenced position showed clear dominance (>90%) of A base at the first position of reverse reads (Figure 39c), which corresponds to ddUTP incorporation site. This indicates that the vast majority of sequenced molecules derived from a seamless read-through, i.e., the copying polymerase incorporated a complementary nucleotide immediately downstream of the dd$^{ON6\text{-}C2}$UTP linker. To compare, 69% of the reads indicated correct read-through in library preparation method based on chemical ligation. The authors reasoned that the presence of an incorrect base at the position downstream of the chemical linkage was a result of substitution events [15]. As the termini of reverse reads in our sequencing experiment aligned to the reference genome even in the case when the first base is not A, imperfect read-through likely originated from base skipping leading to deletions. Importantly, this rare effect does not impair the analysis of sequencing data and does not require additional trimming of sequencing reads. [240]

We termed the developed library preparation technique semi-targeted sequencing emphasizing that only a single primer is needed to capture a desired locus, while dd^{ON}NTPs can label primer extension products irrespective of the sequence context of the template downstream of the priming site.

7.2.2 Microbiome analysis by semi-targeted sequencing of 16S rRNA gene

A microbiome is a community of microorganisms or combined genetic material of microorganisms in a particular habitat (e.g., soil, water of a specific area, humans', animals' excreta). Microbiome profiling is of high importance and has a great potential to determine the microorganisms and uncover their population evolution in specific habitat, shed light on functional differences between communities, or investigate how microbiome population variety is influenced by external factors (e.g., fertilization of soil). One of the possibilities to determine microbial communities is based on sequencing the specific region of bacteria's genome – 16S rRNA gene. This gene consists of nine hypervariable regions (V1 – V9) flanked by conserved sequences and is present in all bacteria (Figure 40a), since it is a prerequisite for survival. However, the determination of closely related species can be hard to assign only by sequencing the 16S rRNA gene since differences might be as little as

single nucleotide polymorphisms (SNPs) and not necessarily present within the hypervariable regions. Moreover, several copies of 16S rRNA gene might be present in a single bacterial genome making it hard to profile the population quantitatively. The long-read sequencing technologies enabled the sequencing of the whole 16S rRNA gene, notwithstanding, sequencing errors and ambiguities regarding the variability of 16S rRNA gene copy numbers challenge the accurate determination of microbial diversity. We hypothesized that performing sequencing in combination of highly conserved 16S rRNA gene sequences with regions of lower conservation might provide the necessary data to determine the composition of microbial community on a species level. [241]

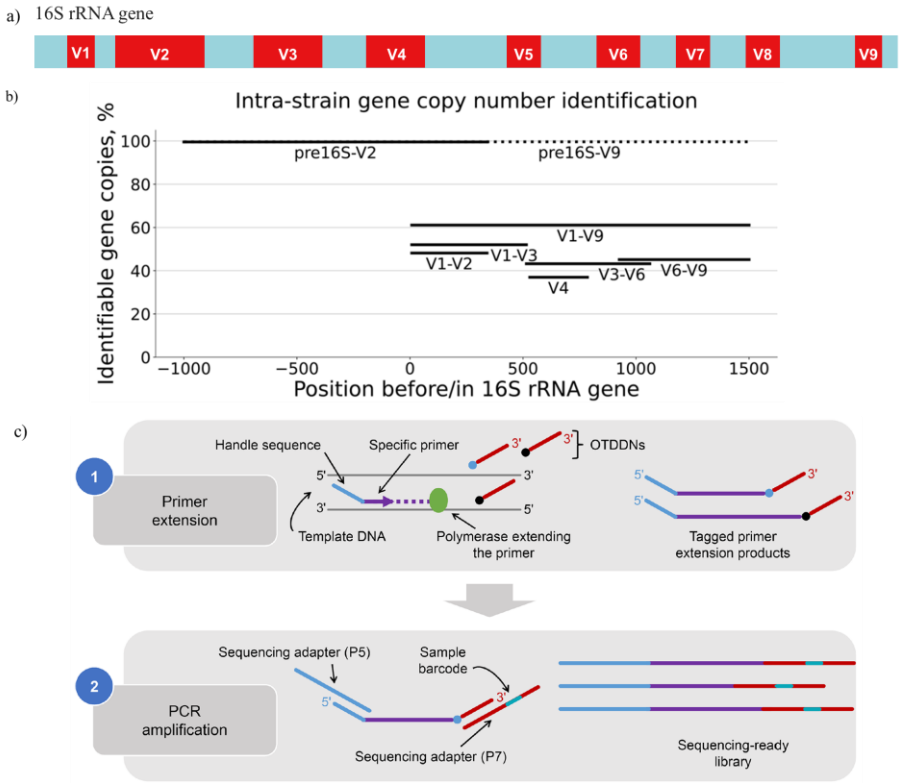


Figure 40. Determination of microbial communities in soil samples, a) representation of 16S rRNA gene: V1 – V9 indicates hypervariable regions, blue inclusions represent conserved regions, b) The percentage of identifiable 16S rRNA gene copy numbers as assessed by various regions of the 16S rRNA gene, c) library preparation of 16S rRNA gene workflow for semi-targeted sequencing.[8]

[8] These experiments were performed by Žana Kapustina.

First, the *in-silico* extraction of publicly available bacterial genome assemblies of 16S rRNA sequences was performed encompassing the 1 kb upstream regions. Obtained database indicated the percentage of identifiable microbiome communities dependent on gene regions selected for sequencing (Figure 40b). For example, sequencing of V1 – V9 region would result in the identification of 60% of the population, V1 – V2 in around 50%, while sequencing upstream the 16S rRNA gene would enable to identify nearly 100% of the population. Consequently, for the investigation of microbiome communities in selected soil samples, the library preparation included sequencing of V1 – V2 and part of near-16S region.

To combine the near-16S region with the sequences of 16S rRNA gene, the workflow represented in Figure 40c was developed. Specific primers, containing universal PCR handles for further library amplification, were designed to target region between V2 and V3 hypervariable sites and were oriented towards the 16S rRNA upstream sequence. During PEX reaction the primer is terminated by incorporation of OTDDN, hereby introducing the second PCR handle (Figure 40c). Manipulating the ratio of OTDDN and natural dNTPs the desired fragment size (read length) distribution is obtained. The obtained primer extension products are subsequently amplified via indexing PCR, where full P5 and P7 Illumina adapters are introduced (Figure 40c). After sequencing with Illumina platform, the obtained forward sequencing reads contain 16S rRNA V1 – V2 reads starting from the specific priming site, while the reverse reads contain part of the genomic region upstream the 16S rRNA gene starting from random positions.

As it was predicted from *in-silico* analysis, the library preparation including part of near-16S sequences improved the classification accuracy at the species level. Some bacterial species were distinguished by applying the developed method while it was impossible to achieve such resolution with alternative analysis strategies. The strategies based on sequencing of within-gene 16S rRNA sequences devastate from the inability to identify gene copy numbers, since the diversity of within-gene sequences is limited. The OTDDN-based approach solves this problem by establishing the direct connection between 16S rRNA sequences and adjacent genomic regions, which was demonstrated to be highly variable and of high diagnostic value. [241]

7.3 RNA library preparation enabled by oligonucleotide-tethered 2',3'-dideoxynucleotides

7.3.1 Prostate cancer RNA transcripts' fusion sequencing via terminator-assisted synthesis

Nowadays, cancer is the biggest challenge for medicine and the scientific community attempting to find the cure and proper treatment for this deadly illness. Since many different cancer forms exist, their triggering causes and mechanisms of action are not unambiguously known. Due to the high clinical importance of this field, continuous research is undergoing in attempts to prevent and cure it. The growing investigation of various cancers accelerated the boost of the assisting methods enabling the determination and monitoring of occurring changes in prospective cancer cells.

Genetic rearrangements that fuse unrelated, different coding material containing parts of protein-coding genes, are frequent events in prostatic cancer. TMPRSS2 is a protein reacting to male hormone changes. ERG is a transcription factor regulating cell division and metabolism by turning on the expression of certain genes. In case of the overexpression of ERG, cells tend to become malignant. The fusion events of *TMPRSS2* and *ERG* genes have been discovered in prostate cancer cells. Since fusion can produce not only conjugation of two full-length *TMPRSS2* and *ERG* genes, but their incomplete parts the diversity of *TMPRSS2-ERG* fusions is observed. Different fusion isoforms lead to variable outcomes, e.g., some fusion events produce mRNA that can be translated to corresponding enzymes. However, they can be inactive, partially active or of unknown activity. If fusion results in changes in the reading frame, the coding information will not be translated into functional protein. The identification of *TMPRSS2-ERG* fusion types and what is expressed in malignant cells is one of the ways enabling the selection of proper treatment. Sequencing of fusion transcripts is the method of choice to precisely identify the particular isoforms and predict the possible outcome. Here, the utility of novel semi-targeted RNA sequencing technique, that was called fusion sequencing via terminator-assisted synthesis (FTAS-seq) was proposed. FTAS-seq allows to enrich the gene of interest and simultaneously profile the whole spectrum of its 3'-terminal fusion partners (Figure 41).

For the NGS library preparation, total RNA was first reverse transcribed, then treated with RNase H for RNA strand hydrolysis. The specific primer targeting *TMPRSS2* exon 1 was used for the synthesis of second cDNA strand, which in turn was stochastically terminated by OTDDN. Next,

oligonucleotide-labeled fragments were amplified yielding sequencing-ready library (Figure 41).

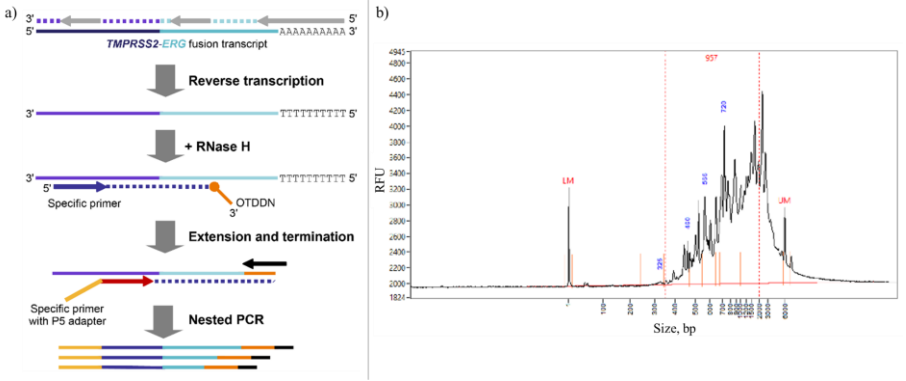


Figure 41. Library preparation for FTAS-seq, a) NGS library preparation workflow for fused transcripts, b) typical FTAS-seq library trace.[9]

After sequencing, the plethora of *TMPRSS2* 3'-terminal fusion isoforms was identified, moreover, 11 previously uncharacterized *TMPRSS2* fusion partners were discovered. This work further expands the applicability of OTDDNs and suggests the semi-targeted sequencing technique for transcriptomic analyses.

7.3.2 Analysis of gene expression by mRNA sequencing via terminator-assisted synthesis

Transcriptome is a full set of RNA transcripts in a specific type of cell. Each and every cell in a multicellular organism contains the same genetic information within the genome, but cell functions are determined by how this genetic information is expressed into mRNA and then into proteins. Transcriptome analysis provides comprehensive information about cell regulatory events and has a great potential to increase diagnostic capacity and precision. For example, identifying tumor cells, their functional properties, and expression patterns could enable to select the necessary treatment.

For transcriptome analysis, the simple and quick single-tube mRNA sequencing workflow was suggested utilizing OTDDNs (Figure 42). The method – mRNA sequencing via terminator-assisted synthesis (MTAS-seq) – leverages OTDDNs for the identification and quantitative profiling of 3'-terminal regions of eukaryotic mRNAs. Since all mRNAs must contain polyA

[9] These experiments were performed by Ugnė Drazdauskienė and Žana Kapustina.

tail to be expressed, reverse transcription primer was designed to target polyA tails and was extended by reverse transcriptase (Figure 42a). Here, as in previous workflow, the primer is terminated by stochastic incorporation of OTDDN, producing the specific size distributions cDNA fragments (Figure 42b). To make this workflow even more simplified, same experiments were performed not only with mRNAs purified out of cell lysates, but directly with cells. The reaction was performed in a single tube containing cell of interest, lysis buffer and all RT (reverse transcription) reaction components. Subsequently, the indexing PCR was performed for library amplification and Illumina adapters introduction. The approach described eliminated the need for intermediate purification of cDNA fragments before amplification thus making the library preparation a single-tube process with no material losses throughout the procedure.

Since gene expression profiling requires accurate estimation of transcript abundance the OTDDN applied here contained the region of randomized sequence that served as a unique molecular identifier (UMI) in order to distinguish between PCR duplicates and unique OTDDN incorporation events (Figure 42c).

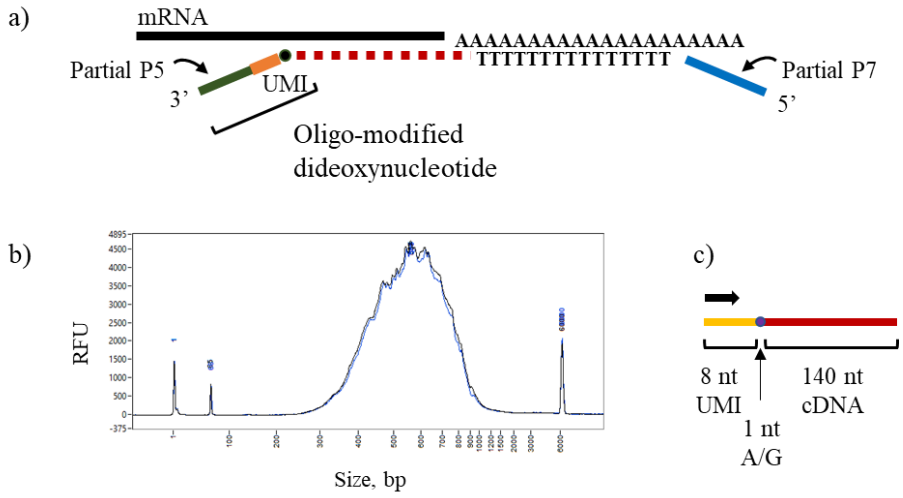


Figure 42. MTAS-seq technique and workflow, a) NGS library preparation workflow, b) a typical MTAS-seq library trace, c) the structure of sequencing reads.[10]

The represented transcriptome analysis technique demonstrated that the use of OTDDNs enables the simple and quick generation of cDNA libraries

[10] These experiments were performed by Žana Kapustina.

enriched for mRNA 3′-terminal sequences and significantly reduces the number of workflow steps. Randomized sequences within oligonucleotide chain of OTDDN tag each fragment with a unique barcode providing highly accurate estimation of transcript abundance in sequencing data. This approach might greatly facilitate gene expression profiling studies to unravel molecular signatures of complex diseases. [242]

7.4 Efficiency measurements of reading through unnatural nucleobase-to-ribose linkage

The alternative technologies, utilizing 3'-azido-ddNTPs for termination of extension of nucleic acids, followed by CuAAC reaction with alkyne-oligonucleotides, mainly focused on phosphodiester linkage replacement, through the perspective to utilize mimicking groups (Figure 43). Such conjugation approach was also called as ribose-to-ribose connection. Even though, read-through events were monitored, they were determined to be highly inefficient, as a single primer extension cycle generated undetectable amount of product (less than 4%). [16] Our designed linkers do not mimic the phosphate group. On the contrary, the design presents the completely different conjugation approach for oligonucleotides, connecting nucleobase-to-phosphate. Consequently, it was of high importance to measure whether suggested nucleobase-to-phosphate conjugation (Figure 43) was of greater

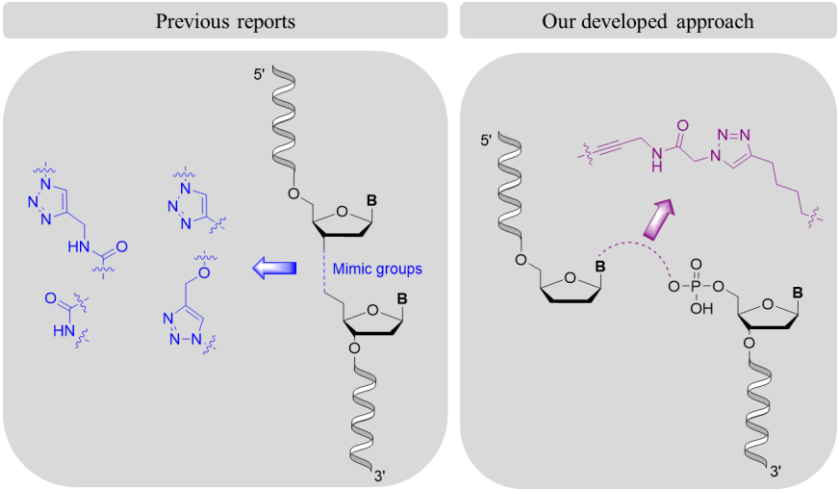


Figure 43. The general principles of chemical conjugation of nucleic acids, a) previously reported ribose-to-ribose conjugated analogues, b) proposed nucleobase-to-phosphate conjugation analogue (represented with C2 linker).

efficiency. Keeping in mind that good biocompatibility of the linker is the parameter of technology excellence.

To unambiguously measure read-through efficiency, a method must meet requirements such as: accuracy, no intervention of similar reactions, as little as possible intermediate steps in between PEX experiments and read-through measurements. To meet these requirements, the incorporation product must be purified from all the reaction components - this is hardly achievable via biochemical purification methods. Consequently, PEX reactions utilizing different linkers bearing dd^{ON6}CTPs were performed in large scale, followed by purification by HPLC. Since the purification scale was still very small, analytical HPLC was used, and a suitable method was optimized (Figure 44a). Next, a single-stranded DNA fragment with an incorporated dd^{ON6}C was used as a template for PEX to evaluate the efficiency of reading through the unnatural linker (Figure 44b). Phusion exo- enzyme exhibited the best single-cycle read-through efficiencies as compared to SuperScript IV, Thermo Sequenase and Klenow exo- (Figure 44c).

Interestingly, linkers longer than C2 that was used in previously shown library preparation examples performed better, with C4 and PEG2 demonstrating single-cycle read-through efficiencies of >70%. Moreover, in all cases the extension until the linker product bands can be seen indicating that polymerases stall at the unnatural conjugation site (Figure 44c)[11]. Since the amide bond has a conjugated electron system where nitrogen bonds are not pyramidal as in the case of amines but planar, this restricts rotation around the nitrogen linkage, consequently resulting in low amide bond flexibility. The triple bond at the nucleobase is also planar, making the whole (acetamido)prop-1-ynyl-part lacking conformational flexibility. The linker length present between the amide bond and constrained triazole ring has a substantial influence on the following read-through assay. It appears that this linker part primary feature is not to provide distance in between oligonucleotides but enable adjustable flexibility. We hypothesize that polymerase copies through the template downstream the linker in an event of sufficiently close alignment between the oligonucleotide modification of OTDDN and the rest of the DNA strand. A highly flexible linker might facilitate a higher probability of forming a favorable configuration for the polymerase to bypass the modification site. These results indicate that linker design can be tailored to improve the performance of modified ddNTPs in

[11] These experiments were performed by Justina Medžiūnė and Žana Kapustina.

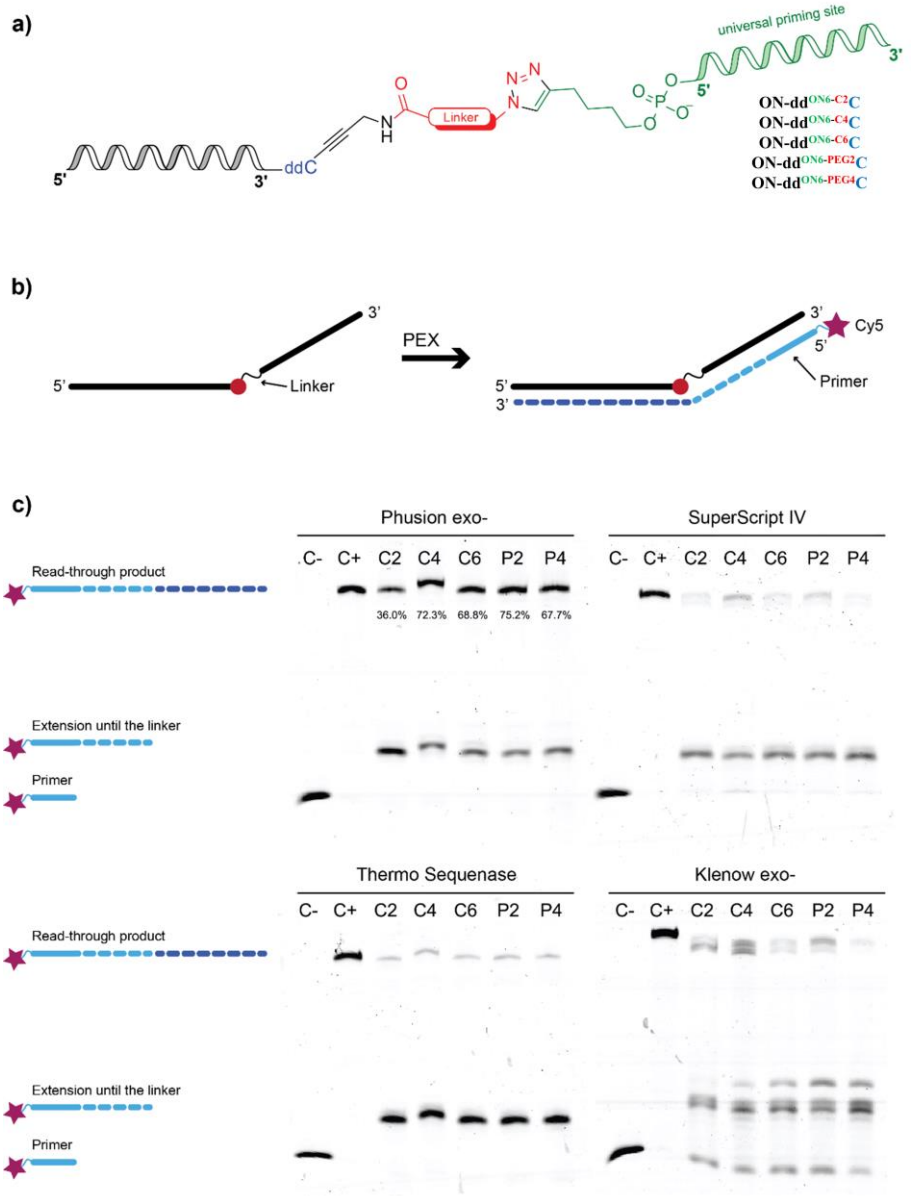


Figure 44. Evaluation of read-through efficiency of five different linkers bearing dd^{ON}CTPs, a) The structure of templates (ON-dd^{ON6}Cs) for PEX in read-through measurements, b) Schematic representation of PEX experiment, c) Electrophoregrams of polyacrylamide gels showing PEX reaction products labeled with Cy5 dye. Linkers and tested DNA polymerases are indicated above the lanes. In the case of Phusion exo-, numbers below the read-through product band stand for read-through efficiency for each respective linker. C- – negative control, i.e., unextended primer, C+ - positive control, i.e., an oligonucleotide of the same length as the anticipated read-through product. P2 and P4 – PEG2 and PEG4 linkers, respectively.

molecular biology assays, such as NGS library preparation. Improvement of read-through efficiency is especially important for the processing of low amounts of nucleic acids or for transcriptome analysis applications demanding maximal assay sensitivity to capture lowly expressed genes.

EXPERIMENTAL PART

All chemicals and solvents were obtained commercially and used without further purification. Iodinated nucleosides $dd^{I}U$, $dd^{I}C$, $dd^{I}A$, $dd^{I}G$ were obtained from AEchem Scientific; $d^{I}C$ and $d^{I}A$ were obtained from Biosynth® Carbosynth. Moisture sensitive reactions were performed under dry argon atmosphere using oven-dried glassware. Reactions, fractions and final purities were monitored by Vanquish™ UHPLC system using TSKgel™ ODS- 100V 150x4.6 mm, S – 3 μm column (mobile phases: A – 100 mM TEAAc, B – 40% 100 mM TEAAc + 60 % ACN). Modified nucleotides were purified by Biotage™ IsoleraTM One or Biotage™ Isolera™ Prime liquid chromatography systems with UV detectors using Q Sepharose™ Fast Flow, Biotage SNAP ULTRA C18 columns. Purification of oligonucleotide-tethered 2',3'-dideoxynucleotides was held with Waters 2555 Quaternary gradient system using YMC-Actus Triart C18 100x20 mm S – 5 μm column (in case of OTDDNs synthesis scale ≥84 nmol), while product for read-through efficiency measurements and oligonucleotide-tethered 2'-dideoxynucleotides and 2',3'-dideoxynucleotides synthesized in small scale (8.4 – 16.8 nmol) was purified by Vanquish UHPLC system using either TSKgel™ ODS- 100V 3 μm or YMC-Triart C18 150x4.6 mm S – 3 μm columns. Conductivity was measured with Mettler-Toledo™ FiveGo™ F3 conductometer (prior to ion exchange chromatography reaction mixtures were diluted to 3.5 – 4 mS/m2 conductivity). Absorbance of products were measured by Thermo Fisher Evolution™ 201 UV-Visible spectrophotometer. Oligonucleotide-tethered nucleotides amount was determined using NanoDrop™ 2000 spectrophotometer. Monoisotopic masses measurements were performed using triple quadrupole mass spectrometer TSQ Endura™ with ESI ion source, oligonucleotides molecular masses were determined using hybrid Quadrupole-Orbitrap™ mass spectrometer Q Exactive plus™ (Thermo Fisher Scientific). NMR spectra were recorded using Bruker Avance III 400 MHz (9.0 T). Chemical shifts are reported in parts per million (ppm) relative to solvent resonance signal as an internal standard. Signal patterns are indicated as br, broad; s, singlet; d, doublet; t, triplet; q, quartet; m, multiplet. The complete assignment of ^{1}H and ^{13}C signals was performed by an analysis of the correlated homonuclear H,H-COSY, and heteronuclear H,C-HSQC and H,C-HMBC spectra.

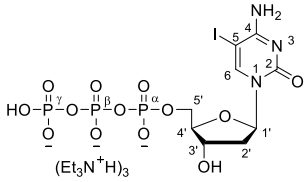
All enzymes and buffers were manufactured by Thermo Fisher Scientific unless specified otherwise. Oligonucleotides were synthesized by Metabion GmbH requesting HPLC purification.

8. CHEMICAL SYNTHESIS

8.1 Synthesis of reactive groups bearing nucleotides

5-Iodo-2'-deoxycytidine 5'-triphosphate (d^{I}CTP)

Reaction was carried out under argon atmosphere. 5-Iodo-2'-deoxycytidine (d^{I}C) (1 g, 2.83 mmol) was suspended in ACN (28.3 ml) under argon atmosphere. The suspension was stirred for 30 min at room temperature, then cooled to 0 °C, followed by the dropwise addition of pyrophosphoryl chloride (0.78 ml, 2 eq). The reaction was monitored by HPLC (Table 24 AM2) and after 15 – 25 min tributylammonium pyrophosphate (TBAPP) cocktail consisting of: 0.5 M TBAPP (17 ml, 3 eq.) solution in ACN, NBu_3 (4 ml, 6 eq.) and ACN (10 ml) was added to the reaction mixture and stirred for 20 min at room temperature. The reaction mixture was quenched with cooled 1 M triethylammonium bicarbonate (TEAB) buffer (100 ml, pH 7). The product was purified by ion exchange chromatography on Q Sepharose FF resin applying 50 mM TEAB/1 M TEAB (0-50%) gradient. The desired triphosphate was obtained in 63 % (1.78 mmol) yield. UV: λ_{max} = 294 nm, ε = 5700 l·mol^{-1}·cm^{-1}.

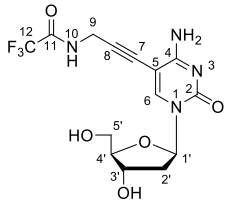


^{1}H NMR (400 MHz, D_2O) δ 8.24 (1H, s, H-6), 6.25 (1H, t, *J* = 6.7 Hz, H-1'),4.59 – 4.64 (1H, m, H-3'), 4.19 – 4.28 (3H, m, H-4' ir H-5'), 2.42 (1H, ddd, *J* = 14.0, 6.1, 3.5 Hz, H-2'), 2.23 – 2.36 (1H, m, H-2'). ^{31}P NMR (162 MHz, D_2O): δ -10.97 (d, *J* = 19.9 Hz, P_γ), -11.77 (d, *J* = 20.1 Hz, P_α), -23.42 (t, *J* = 20.1 Hz, P_β). HRMS (ESI-) m/z: $[M-H]^-$ calcd for $C_9H_{14}IN_3O_{13}P_3$ 591.8790; found 591.8797. Obtained product signals are in agreement with literature data [58].

5-(3-(2,2,2-trifluoroacetamido)prop-1-ynyl)-2'-deoxycytidine (d^{TFA-PA}C)

Reaction was carried out under argon atmosphere. 5-Iodo-2'-deoxycytidine (d^{I}C) (0.5 g, 1.4 mmol) was dissolved 10 ml DMF. Subsequently the $[(C_6H_5)_3P]_2PdCl_2$ (0.2 eq., 0.28 mmol, 0.164 g) and Et_3N (3 eq., 4.2 mmol, 0.524 ml) were poured, followed by dropwise addition of 2,2,2-trifluoro-N-(prop-2-ynyl)acetamide. After 10 min the CuI (0.2 eq., 0.28 mmol, 0.053 g) was applied and reaction was carried out for at room temperature overnight (around 16 h). The reaction was monitored by TLC using DCM:MeOH (9:1) mixture. The reaction was quenched with 1 M ammonium sulphate solution, concentrated under reduced pressure and purified on 50 g silica gel column using DCM:MeOH gradient elution (from

95:5 to 80:20). The product containing fractions were combined and concentrated. $d^{TFA\text{-}PA}C$ was obtained in 81% (0.41 g).



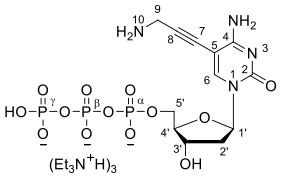
^{1}H NMR (400 MHz, DMSO-d6) δ 9.95 (1H, t, *J* = 4.4 Hz, **NH**COCF$_3$), 8.15 (1H, s, H-6), 7.83 (1H, s, NH$_2$), 6.85 (1H, s, NH$_2$), 6.10 (1H, t, *J* = 6.5 Hz, H-1'), 5.21 (1H, d, *J* = 4.2 Hz, 3'-OH), 5.06 (1H, t, *J* = 5.1 Hz, 5'-OH), 4.28 (2H, d, *J* = 4.9 Hz, **CH$_2$**NHCO), 4.17 – 4.21 (1H, m, H-3'), 3.78 – 3.80 (1H, m, H-4'), 3.57 – 3.63 (1H, m, H-5'), 3.51 – 3.56 (1H, m, H-5'), 2.15 (1H, ddd, *J* = 13.1, 5.9, 3.4 Hz, H-2'), 1.94 – 2.00 (1H, m, H-2'). HRMS (ESI+) m/z: $[M+H]^+$ calcd for $C_{14}H_{16}F_3N_4O_5$ 377.1067; found 377.1066. Obtained product signals are in agreement with literature data [55].

5-(3-aminoprop-1-ynyl)-2'-deoxycytidine 5'-triphosphate ($d^{PA}CTP$)

Reaction was carried out under argon atmosphere. 5-(3-(2,2,2-trifluoroacetamido)prop-1-ynyl)-2'-deoxycytidine ($d^{TFA\text{-}PA}C$) (0.45 mmol, 0.17 g) was suspended in TMP (6 ml) and mixed for 20 min at room temperature. Pyridine (2 eq., 0.91 mmol, 73 µl) was then applied. Reaction mixture was cooled to -20 – -15 ℃, followed by dropwise addition of $POCl_3$ (2 eq. 0.91 mmol, 85 µl). After maintaining reaction overnight 2 additional equivalents of $POCl_3$ were applied. Reaction progress was monitored by HPLC, were after 3 hours TBAPP cocktail consisting of: 0.5 M TBAPP 3.6 ml, 4 eq.) solution in ACN, NBu_3 (0.65 ml, 6 eq.) and ACN (2 ml) was added to the reaction mixture and stirred for 30 min at 0 ℃ temperature. The reaction mixture was quenched with cooled 1 M triethylammonium bicarbonate (TEAB) buffer (100 ml, pH 7). The product was purified by ion exchange chromatography on Q Sepharose FF resin applying 50 mM NaCl/1 M NaCl (0-20%) gradient. Product containing fraction were combined and concentrated. The product was desalted utilizing C18 column water/ACN (0-100%) gradient. Product was obtained in 33% (0.15 mmol) yield. HRMS (ESI-) m/z: [M-H]- calcd for $C_{14}H_{17}F_3N_4O_{14}P_3$ 614.9906; found 614.9910.

5-(3-(2,2,2-trifluoroacetamido)prop-1-ynyl)-2'-deoxycytidine 5'-triphosphate ($d^{TFA\text{-}PA}CTP$) was subsequently treated with aqueous 20% ammonium solution (36 ml) and left stirring overnight at room temperature. The product was purified by ion exchange chromatography on Q Sepharose FF resin applying 50 mM TEAB/1 M TEAB (0-47%) gradient. The desired triphosphate was obtained in 29 % (0.13 mmol) yield. It is important to mention that after phosphorylation of $d^{TFA\text{-}PA}C$, product do not need to be purified. Ammonia solution can be applied directly to phosphorylation reaction mixture to produce the unprotected $d^{PA}CTP$ (in majority of $d^{PA}CTP$ synthesis experiment such technique was used). UV: λ_{max} = 294 nm, ε =

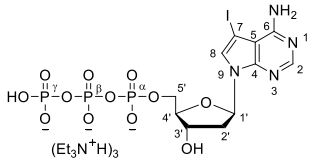
9300 l·mol^{-1}·cm^{-1}. HRMS (ESI-) m/z: [M-H]- calcd for $C_{12}H_{18}N_4O_{13}P_3$ 519.0083; found 519.0089. Obtained product signals are in agreement with literature data [55].



^{1}H NMR (400 MHz, D_2O) δ 8.45 (1H, s, H-6), 6.17 (1H, t, *J* = 5.4 Hz, H-1'), 4.31 (2H, s, **CH$_2$**NH), 4.25 (1H, m, H-3'), 4.10 – 4.20 (3H, m, H-4' ir H-5'), 2.39 (1H, m, H-2'), 2.23 – 2.32 (1H, m, H-2'). ^{31}P NMR (162 MHz, D_2O) δ -6,66 (d, *J* = 20.4 Hz, P_γ), -11.47 (d, *J* = 19.3 Hz, P_α), -22.24 (t, *J* = 19.8 Hz, P_β).

7-Deaza-7-iodo-2'-deoxyadenosine 5'-triphosphate (d^{I}ATP)

7-Deaza-7-iodo-2'-deoxyadenosine (d^{I}A) (0.5 g, 1.33 mmol) was suspended in ACN (12.5 ml) under argon atmosphere. The suspension was stirred for 30 min at room temperature, then cooled to 0 °C followed by dropwise addition of pyrophosphoryl chloride (3 eq., 3.99 mmol, 0.552 ml). The reaction was monitored by HPLC and after an hour tributylammonium pyrophosphate (TBAPP) cocktail consisting of: 0.5 M TBAPP (16 ml, 6 eq.) solution in ACN, NBu_3 (3 ml, 9 eq.) and ACN (5 ml) was added to the reaction mixture and stirred for 30 min at room temperature. The reaction mixture was quenched with cooled 1 M triethylammonium bicarbonate (TEAB) buffer (100 ml, pH 7). The product was purified by ion exchange chromatography on Q Sepharose FF resin applying 50 mM TEAB/1 M TEAB (0-55%) gradient. The desired triphosphate was obtained in 65 % (0.86 mmol) yield. UV: λ_{max} = 283 nm, ε = 8500 l·mol^{-1}·cm^{-1}.



^{1}H NMR (400 MHz, D_2O) δ 7.80 (1H, s, H-2), 7.37 (1H, s, H-8), 6. 28 (1H, t, *J* = 6,8 Hz, H-1'), 4.52 – 4.55 (1H, m, H-3'), 3.94 – 4.03 (3H, m, H-4' ir H-5'), 2.42 – 2.49 (1H, m, H-2'), 2.30 – 2.37 (1H, m, H-2'). ^{31}P NMR (162 MHz, D_2O) δ -10.53 (d, *J* = 20,2 Hz, P_γ), -11.76 (d, *J* = 20.4 Hz, P_α), -23.57 (t, *J* = 20.2 Hz, P_β). HRMS (ESI-) m/z: [M-H]- calcd for $C_{11}H_{15}IN_4O_{12}P_3$ 614.8944; found 614.8956. Obtained product signals are in agreement with literature data [120].

7-Deaza-7-(3-aminoprop-1-ynyl)-2'-deoxyadenosine 5'-triphosphate (d^{PA}ATP)

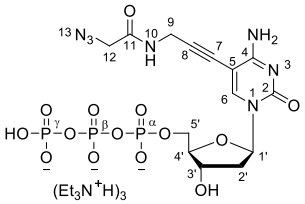
Water-acetonitrile mixture 2:1 (1 ml) was added through septum to an argon flushed flask containing 7-deaza-7-iodo-2'-deoxyadenosine 5'-triphosphate (d^{I}ATP) (0.1 mmol), followed by the addition of propargylamine (13 ul, 2 eq.) and CuI (1.9 mg, 0.1 mmol, 10 mol%) solution in ACN (0.1 M, 100 µl). In a separate flask under argon atmosphere $Pd(OAc)_2$ (1.2 mg, 0.05 mmol, 5 mol%) and TPPTS (14.2 mg, 0.25 mmol, 5 eq. to Pd) were

combined, water-acetonitrile mixture 2:1 (0.5 ml) was then added and stirred till full dissolution of solids. Prepared catalyst solution was applied to the reaction mixture and stirred for 1 h at 30 °C. After completion of the reaction the product was purified by ion exchange chromatography on Q Sepharose FF resin applying water/1M TEAB gradient (5-65 %). Desalting was carried out on C18 column with water/acetonitrile gradient (0-100 %). The product was obtained in 77 % (0.077 mmol) yield. UV: λ_{max} = 280 nm, ε = 12700 l·mol^{-1}·cm^{-1}. HRMS (ESI-) m/z: [M-H]- calcd for $C_{14}H_{19}N_5O_{12}P_3$ 542.0243; found 542.0249. The compound is not stable, consequently, it was immediately reacted with azidoacetic acid NHS ester. Finale product NMR is given below (**d$^{N3\text{-}C2}$ATP**, psl. 144).

General procedure for synthesis of azido-dNTPs

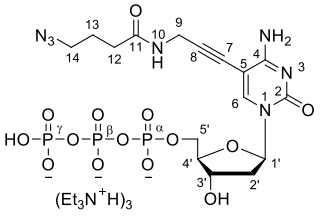
The corresponding d^{PA}NTP (d^{PA}CTP or d^{PA}ATP) (0.032 mmol) was dissolved in Na_2CO_3/$NaHCO_3$ 1:9 buffer (1 ml, 0.1 M, pH 9) and DMF (200 µl). Solution of corresponding azido-NHS ester (3 eq., 0.096 mmol) in DMF (0.4 ml) was then added. Reactions were monitored by HPLC and after stirring for 0.5 – 4h at room temperature, reactions mixture were purified by ion exchange chromatography on Q Sepharose FF resin applying water/1 M TEAB gradient (5-30 % - for d^{N3}CTPs and 5-40 % for d$^{N3\text{-}C2}$ATP). Desalting was carried out on C18 column with water/acetonitrile gradient (0-100 %).

5-(3-(4-azidoacetamido)prop-1-ynyl)-2'-deoxycytidine 5'-triphosphate (d$^{N3\text{-}C2}$CTP) was obtained in 50% (0.018 mmol) yield. UV: λ_{max} = 294 nm, ε = 9300 ml·mol^{-1}·cm^{-1}.



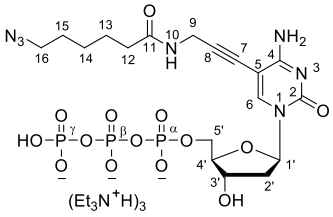
^{1}H NMR (400 MHz, D_2O) δ 8.19 (s, 1H, H – 6), 6.31 (t, *J* = 6.5 Hz, 1H, H – 1'), 4.64 (dd, *J* = 6.6, 3.4 Hz, 1H, H – 3'), 4.34 (s, 2H, CH_2 – 9), 4.31 – 4.24 (3H, H – 4', CH_2 – 5'), 4.16 (s, 2H, CH_2 – 12), 2.59 – 2.46 (m, 1H, H – 2'a), 2.45 – 2.29 (m, 1H, H – 2'b). ^{13}C NMR (101 MHz, D_2O) δ 170.52 (s, C=O – 11), 165.18 (s, C – 4), 156.13 (s, C=O – 2), 145.14 (s, CH – 6), 92.15 (s, C – 5 or 8), 91.56 (s, C – 5 or 8), 86.48 (s, CH – 1'), 85.48 (d, *J* = 8.7 Hz, CH – 4'), 73.49 (s, C – 7), 70.38 (s, CH – 3'), 65.39 (d, *J* = 5.9 Hz, CH_2 – 5'), 52.01 (s, CH_2 – 12), 39.31 (s, CH_2 – 2'), 29.90 (s, CH_2 – 9). ^{31}P NMR (162 MHz, D_2O) δ -7.55 (d, *J* = 18.5 Hz, P_γ), -10.76 (d, *J* = 19.0 Hz, P_α), -21.41 (t, *J* = 18.7 Hz, P_β). HRMS (ESI-) m/z: [M-H]- calcd for $C_{14}H_{19}N_7O_{14}P_3$ 602.0203; found 602.0207.

5-(3-(4-azidobutanamido)prop-1-ynyl)-2'-deoxycytidine 5'-triphosphate (d$^{N3\text{-}C4}$CTP) was obtained in 70% (0.023 mmol) yield. UV: λ_{max} = 294 nm, ε = 9300 ml·mol^{-1}·cm^{-1}.



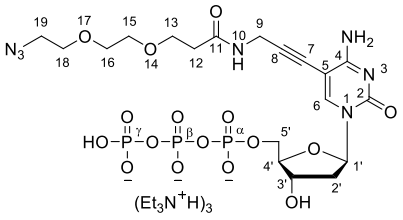
^{1}H NMR (400 MHz, D^{2}O) δ 8.38 (s, 1H, H – 6), 6.63 (t, *J* = 6.5 Hz, 1H, H – 1’), 4.67 – 4.54 (m, 5H, H – 3’, H – 4’, H – 5’a, CH_2 – 9), 4.54 – 4.44 (m, 1H, H – 5’b), 3.79 (t, *J* = 6.6 Hz, 2H, CH_2 – 14), 2.87 (ddd, *J* = 11.5, 6.4, 5.6 Hz, 1H, H – 2’a), 2.82 – 2.69 (m, 3H, H – 2’b, CH_2 – 12), 2.29 (p, *J* = 7.0 Hz, 2H, CH_2 – 13). ^{13}C NMR (101 MHz, D_2O) δ 176.34 (s, C=O – 11), 165.45 (s, C – 4), 156.61 (s, C=O – 2), 145.40 (s, CH – 6), 92.76 (s, C – 5 or 8), 92.68 (s, C – 5 or 8), 86.87 (d, *J* = 6.9 Hz, CH – 1’), 85.26 (d, *J* = 7.3 Hz, CH – 4’), 73.60 (s, C – 7), 70.74 (s, CH – 3’), 66.20 (d, *J* = 3.5 Hz, CH_2 – 5’), 51.18 (s, CH_2 – 14), 39.08 (s, CH_2 – 2’), 33.69 (s, CH_2 – 12), 30.48 (s, CH_2 – 9), 25.02 (s, CH_2 – 13). ^{31}P NMR (162 MHz, D_2O) δ -4.14 (d, *J* = 16.3 Hz, P_γ), -9.53 (d, *J* = 17.4 Hz, P_α), -19.01 (d, *J* = 18.5 Hz, P_β). HRMS (ESI-) m/z: [M-H]- calcd for $C_{16}H_{23}N_7O_{14}P_3$ 630.0516; found 630.0523.

5-(3-(6-azidohexanamido)prop-1-ynyl)-2’-deoxycytidine 5’-triphosphate (d$^{N3\text{-}C6}$CTP) was obtained in 81% (0.023 mmol) yield. UV: λ_{max} = 294 nm, ε = 9300 ml·mol^{-1}·cm^{-1}.



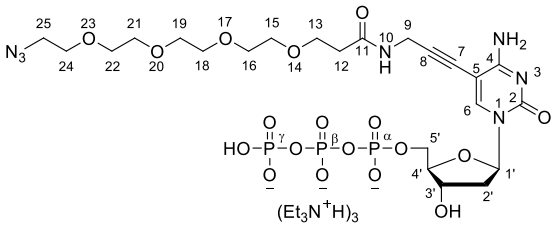
^{1}H NMR (400 MHz, D_2O) δ 8.08 (s, 1H, H – 6), 6.15 (t, *J* = 6.4 Hz, 1H, H – 1’), 4.52 (dt, *J* = 6.6, 4.4 Hz, 1H, H – 3’), 4.20 – 4.06 (m, 5H, H – 4’, CH_2 – 5’, CH_2 - 9), 3.20 (t, *J* = 6.8 Hz, 2H, CH_2 - 16), 2.35 (ddd, *J* = 14.0, 6.3, 4.8 Hz, 1H, H – 2’a), 2.26 – 2.15 (m, 3H, H – 2’b, CH_2 – 12), 1.57 (dd, *J* = 15.0, 7.5 Hz, 2H, CH_2 – 13), 1.49 (dd, *J* = 14.9, 7.2 Hz, 2H, CH_2 – 15), 1.34 – 1.20 (m, 2H, CH_2 – 14). ^{13}C NMR (101 MHz, D_2O) δ 177.16 (s, C=O – 11), 165.18 (s, C – 4), 157.07 (s, C=O – 2), 144.89 (s, CH – 6), 92.12 (s, C – 5 or 8), 92.10 (s, C – 5 or 8), 86.19 (s, CH – 1’), 85.52 (d, *J* = 8.7 Hz, CH – 4’), 73.10 (s, C – 7), 69.86 (s, CH – 3’), 64.94 (d, *J* = 6.5 Hz, CH_2 – 5’), 51.15 (s, CH_2 – 16), 39.26 (s, CH_2 – 2’), 35.41 (s, CH_2 – 9), 27.74 (s, CH_2 – 15), 25.31 (s, CH_2 – 14), 24.86 (s, CH_2 – 13). ^{31}P NMR (162 MHz, D_2O) δ -5.78 (d, *J* = 20.2 Hz, P_γ), -11.01 (d, *J* = 19.5 Hz, P_α), -21.65 (t, *J* = 19.8 Hz, P_β). HRMS (ESI-) m/z: [M-H]- calcd for $C_{18}H_{27}N_7O_{14}P_3$ 658.0829; found 658.0837.

5-(3-(3-(2-(2-azidoethoxy)ethoxy)propanimido)prop-1-ynyl)-2’-deoxycytidine 5’-triphosphate (d$^{N3\text{-}PEG2}$CTP) was obtained in 81% (0.026 mmol) yield. UV: λ_{max} = 294 nm, ε = 9300 ml·mol^{-1}·cm^{-1}.



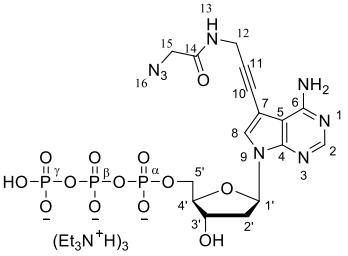
^{1}H NMR (400 MHz, D_2O) δ 8.17 (s, 1H, H – 6), 6.25 (t, *J* = 6.4 Hz, 1H, H – 1’), 4.61 (dt, *J* = 7.0, 4.2 Hz, 1H, H – 3’), 4.32 – 4.12 (m, 5H, H – 4’, CH_2 – 5’, CH_2 - 9), 3.82 (t, *J* = 5.9 Hz, 2H, CH_2 - 13), 3.73 – 3.59 (m, 6H, CH_2 – 15, 16, 18), 3.45 (t, *J* = 5.9, 2H, CH_2 – 19), 2.58 (t, *J* = 5.9 Hz, 2H, CH_2 – 12), 2.45 (ddd, *J* = 13.9, 6.1, 4.9 Hz, 1H, H – 2’a), 2.36 – 2.26 (m, 1H, H – 2’b). ^{13}C NMR (101 MHz, D_2O) δ 176.80 (s, C=O – 11), 167.69 (s, C – 4), 158.59 (s, C=O – 2), 147.50 (s, CH – 6), 94.67 (s, C – 5 or 8), 94.50 (s, C – 5 or 8), 88.76 (s, CH – 1’), 88.05 (d, *J* = 8.8 Hz, CH – 4’), 75.65 (s, C – 7), 72.49 (s, CH – 3’), 72.12 (s, 2xCH_2 – 15 or 16 or 18), 71.82 (s, CH_2 – 15 or 16 or 18), 69.28 (s, CH_2 – 13), 67.56 (d, *J* = 5.8 Hz, CH_2 – 5’), 52.71 (s, CH_2 – 19), 41.78 (s, CH_2 – 2’), 38.57 (s, CH_2 – 12), 32.26 (s, CH_2 – 9). ^{31}P NMR (162 MHz, D_2O) δ -3.25 (d, *J* = 19.9 Hz, P_γ), -8.41 (d, *J* = 19.3 Hz, P_α), -18.95 (t, *J* = 19.6 Hz, P_β). HRMS (ESI-) m/z: [M-H]- calcd for $C_{19}H_{29}N_7O_{16}P_3$ 704.0884; found 704.0898.

5-(1-azido-15-oxo-3,6,9,12-tetraoxa-16-azanonadec-18-yn-19-yl)-2’-deoxycytidine 5’-triphosphate (dd$^{N3\text{-}PEG4}$CTP) was obtained in 90% (0.029 mmol) yield. UV: λ_{max} = 294 nm, ε = 9300 ml·mol^{-1}·cm^{-1}.



^{1}H NMR (400 MHz, D_2O) δ 8.16 (s, 1H, H – 6), 6.25 (t, *J* = 6.4 Hz, 1H, H – 1’), 4.66 – 4.55 (m, 1H, H – 3’), 4.30 – 4.14 (m, 5H, H – 4’, CH_2 – 5’, CH_2 – 9), 3.82 (t, *J* = 6.0 Hz, 2H, CH_2 – 13), 3.77 – 3.56 (m, 14H, 7xCH_2 – (15 – 24)), 3.55 – 3.44 (m, 2H, CH_2 – 25), 2.64 – 2.52 (m, 2H, CH_2 – 12), 2.45 (ddd, *J* = 13.9, 6.2, 4.7 Hz, 1H, H – 2’a), 2.37 – 2.23 (m, 1H, , H – 2’b). ^{13}C NMR (101 MHz, D_2O) δ 174.22 (s, C=O – 11), 165.13 (s, C – 4), 156.02 (s, C=O – 2), 144.95 (s, CH – 6), 92.13 (s, C – 5 or 8), 91.95 (s, C – 5 or 8), 86.25 (s, CH – 1’), 85.49 (d, *J* = 8.7 Hz, CH – 4’), 73.11 (s, C – 7), 69.99 (s, CH – 3’), 69.64 (s, CH_2 – (15 – 24)), 69.59 (br. s, 4xCH_2 – (15 – 24)), 69.55 (s, CH_2 – (15 – 24)), 69.24 (s, CH_2 – (15 – 24)), 66.72 (s, CH_2 – 13), 65.06 (d, *J* = 5.6 Hz, CH_2 – 5’), 50.20 (s, CH_2 – 25), 39.22 (s, CH_2 – 2’), 36.00 (s, CH_2 – 12), 29.74 (s, CH_2 – 9). ^{31}P NMR (162 MHz, D_2O) δ -5.44 (d, *J* = 19.7 Hz, P_γ), -10.86 (d, *J* = 19.3 Hz, P_α), -21.25 (t, *J* = 19.5 Hz, P_β). HRMS (ESI-) m/z: [M-H]- calcd for $C_{23}H_{37}N_7O_{18}P_3$ 792.1408; found 792.1427.

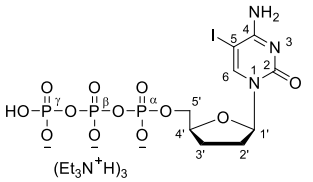
7-Deaza-7-(3-(2-azidoacetamido)prop-1-ynyl)-2'-deoxyadenosine 5'-triphosphate ($d^{N3\text{-}C2}ATP$) was obtained in 58% (0.019 mmol) yield. UV: λ_{max} = 280 nm, ε = 12700 l·mol^{-1}·cm^{-1}.



^{1}H NMR (400 MHz, D_2O) δ 8.24 (s, 1H, H – 2), 7.75 (s, 1H, H – 8), 6.56 (t, *J* = 6.8 Hz, 1H, H – 1'), 4.71 (s, 1H, H – 3'), 4.25 (br. s, 5H, H – 4', CH_2 – 5', CH_2 – 12), 4.13 (s, 2H, CH_2 – 15), 2.62 – 2.41 (m, 2H, CH_2 – 2'). ^{13}C NMR (101 MHz, D_2O) δ 170.50 (s, C=O – 14), 151.88 (s, C – 6), 146.44 (s, C – 4), 145.51 (s, CH – 2), 128.24 (s, CH – 8), 101.58 (s, C – 5), 97.34 (s, C – 7), 89.14 (s, C – 11), 85.57 (d, *J* = 8.6 Hz, CH – 4'), 83.72 (s, CH – 1'), 73.34 (s, C – 10), 71.01 (s, CH – 3'), 65.57 (s, *J* = 6.1 Hz, CH_2 – 5'), 51.91 (s, CH_2 - 15), 39.74 (s, CH_2 – 2'), 29.84 (s, CH_2 – 12). ^{31}P NMR (162 MHz, D_2O) δ -10.64 (d, *J* = 20.1 Hz, P_γ), -11.18 (d, *J* = 18.0 Hz, P_α), -23.04 (t, *J* = 18.3 Hz, P_β). HRMS (ESI-) m/z: [M-H]- calcd for $C_{16}H_{20}N_8O_{13}P_3$ 625.0363; found 625.0372.

5-Iodo-2',3'-dideoxycytidine 5'-triphosphate ($dd^{I}CTP$)

5-Iodo-2',3'-dideoxycytidine ($dd^{I}C$) (1 g, 2.97 mmol) was suspended in ACN (29.7 ml) under argon atmosphere, Bu_3N (0.734 ml, 1.04 eq.) was added. The suspension was stirred for 30 min at room temperature, then cooled to -15 – -10 °C, followed by the dropwise addition of diphosphoryl chloride (1.21 ml, 3 eq). After 20min tributylammonium pyrophosphate (TBAPP) cocktail consisting of: 0.5 M TBAPP (23.8 ml, 4 eq.) solution in ACN, NBu_3 (4.94 ml, 7 eq.) and ACN (10 ml) was added to the reaction mixture and stirred for 30min at room temperature. The reaction mixture was quenched with cooled 1 M triethylammonium bicarbonate (TEAB) buffer (100 ml, pH 7). The product was purified by ion exchange chromatography on Q Sepharose FF resin applying water/1 M TEAB (5-42 %) gradient. The desired triphosphate was obtained in 74 % (2.2 mmol) yield. UV: λ_{max} = 294 nm, ε = 5700 l·mol^{-1}·cm^{-1}.

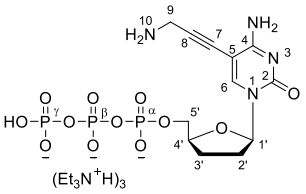


^{1}H NMR (400 MHz, D_2O) δ 7.99 (s, 1H, H-6), 5.81 (dd, *J* = 6.7, 3.5 Hz, 1H, H-1'), 4.22 – 4.13 (m, 1H, H-4'), 4.08 (ddd, *J* = 11.4, 5.9, 2.7 Hz, 1H, H-5'a), 3.93 (dt, *J* = 11.8, 6.0 Hz, 1H, H-5'b), 2.34 – 2.15 (m, 1H, H-2'a), 1.97 – 1.81 (m, 2H, H-2'b and H-3'a), 1.81 – 1.71 (m, 1H, H-3'b). ^{13}C NMR (101 MHz, D_2O) δ 164.37 (C-4), 156.33 (C=O-2), 147.56 (C-6), 87.18 (CH-1'), 80.60 (d, *J* = 8.5 Hz, CH-4'), 66.92 (d, J = 5.6 Hz, CH-5'), 57.37 (C-5), 31.84 (CH_2-2'), 24.76 (CH_2-3'). ^{31}P NMR (162 MHz, D_2O) δ -10.21 (d, *J* = 20.3 Hz,

P_γ), -11.60 (d, J = 20.0 Hz, P_α), -23.47 (t, J = 20.2 Hz, P_β). HRMS (ESI^-) m/z: $[M-H]^-$ calcd for $C_9H_{14}IN_3O_{12}P_3$ 575.8835; found 575.8828.

5-(3-aminoprop-1-ynyl)-2',3'-dideoxycytidine 5'-triphosphate ($dd^{PA}CTP$)

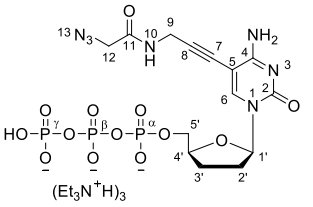
Water-acetonitrile mixture 2:1 (6.6 ml) was added through septum to an argon flushed flask containing 5-iodo-2',3'-dideoxycytidine 5'-triphosphate ($dd^{I}CTP$) (1.64 mmol), followed by the addition of propargylamine (209 µl, 2 eq.) and CuI (0.032 g, 0.1 mmol, 10 mol%). In a separate flask under argon atmosphere $Pd(OAc)_2$ (18.7 mg, 0.05 mmol, 5 mol%) and TPPTS (0.234 g, 0.25 mmol, 5 eq. to Pd) were combined, water-acetonitrile mixture 2:1 (5 ml) was then added and stirred till full dissolution of solids. Prepared catalyst solution was applied to the reaction mixture and stirred for 1 h at 40 °C. The product was purified by ion exchange chromatography on Q Sepharose FF resin applying water/1M TEAB gradient (5-52 %). Desalting was carried out on C18 column with water/acetonitrile gradient (0-100 %). The product was obtained in 88 % (1.45 mmol) yield. UV: λ_{max} = 294 nm, ε = 9300 $l \cdot mol^{-1} \cdot cm^{-1}$.



^1H NMR (400 MHz, D_2O) δ 8.31 (s, 1H, H-6), 5.88 (d, J = 5.9 Hz, 1H, H-1'), 4.32 – 4.21 (m, 2H, H-4' and H-5'a), 4.11 – 4.01 (m, 1H, H-5'b), 3.96 (s, 2H, H-9), 2.44 – 2.29 (m, 1H, H-2'a), 2.05 – 1.83 (m, 3H, H-2'b and H-3'a,b). ^{13}C NMR (101 MHz, D_2O) δ 164.17 (C-4), 155.25 (C=O-2), 146.27 (CH-6), 87.49 (C-8), 87.02 (CH-1'), 81.36 (d, J = 7.2 Hz, CH-4'), 77.57 (C-7), 65.95 (CH_2-5'), 32.72 (CH_2-2'), 29.82 (CH_2-9), 23.52 (CH_2-3'). ^{31}P NMR (162 MHz, D_2O) δ -10.59 (br s, P_γ), -11.52 (br s, P_α), -23.17 (br s, P_β). HRMS (ESI^-) m/z: $[M-H]^-$ calcd for $C_{12}H_{18}N_4O_{12}P_3$ 503.0134; found 503.0233.

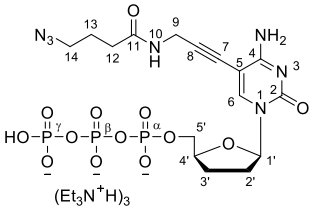
5-(3-(2-azidoacetamido)prop-1-ynyl)-2',3'-dideoxyuridine 5'-triphosphate ($dd^{N3\text{-}C2}CTP$)

5-(3-aminoprop-1-ynyl)-2',3'-dideoxycytidine 5'-triphosphate ($dd^{PA}CTP$) (0.031 mmol) was dissolved in $Na_2CO_3/NaHCO_3$ 1:9 buffer (1 ml, 0.1 M, pH 9) and DMF (0.2 ml). Solution of azidoacetic acid NHS ester (0.018 g, 0.092 mmol, 3 eq.) in DMF (0.4 ml) was then added. After stirring for 30 min at room temperature, reaction mixture was purified by ion exchange chromatography on Q Sepharose FF resin applying water/1 M TEAB gradient (5-52 %). Desalting was carried out on C18 column with water/acetonitrile gradient (0-100 %). Product was obtained 64 % (0.020 mmol) yield. UV: λ_{max} = 294 nm, ε = 9300 $ml \cdot mol^{-1} \cdot cm^{-1}$.



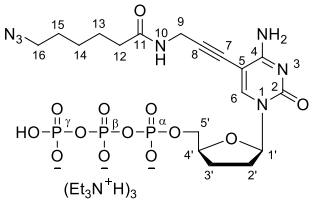
^{1}H NMR (400 MHz, D_2O) δ 8.11 (s, 1H, H-6), 5.93 (dd, *J* = 6.3, 2.2 Hz, 1H, H-1'), 4.33 – 4.26 (m, 1H, H-4'), 4.25 – 4.19 (m, 1H, H-5'a), 4.17 (s, 2H, H-9), 4.12 – 4.02 (m, 1H, H-5'b), 3.99 (s, 2H, H-12), 2.45 – 2.31 (m, 1H, H-2'a), 2.06 – 1.93 (m, 2H, H-2'b and H-3'a), 1.88 – 1.75 (m, 1H, H-3'b). ^{13}C NMR (101 MHz, D_2O) δ 170.14 (C=O-11), 164.02 (C-4), 154.53 (C=O-2), 145.24 (CH-6), 91.64 (2xC, C-5, C8), 87.42 (CH-1'), 80.98 (d, *J* = 7.6 Hz, CH-4'), 73.03 (C-7), 66.66 (d, *J* = 3.5 Hz, CH_2-5'), 51.68 (CH_2-12), 32.10 (CH_2-2'), 29.72 (CH_2-9), 24.46 (CH_2-3'). ^{31}P NMR (162 MHz, D_2O) δ -11.12 (d, *J* = 15.7 Hz, P_γ), -11.50 (d, *J* = 18.9 Hz, P_α), -23.57 (br s, P_β). HRMS (ESI^-) m/z: $[M-H]^-$ calcd for $C_{14}H_{19}N_7O_{13}P_3$ 586.0254; found 586.0251.

5-(3-(4-azidobutanamido)prop-1-ynyl)-2',3'-dideoxycytidine 5'-triphosphate (dd^{N3-C4}CTP) was prepared according to the procedure applied for preparation of dd^{N3-C2}CTP and obtained in 76% (0.028 mmol) yield. UV: λ_{max} = 294 nm, ε = 9300 ml·mol^{-1}·cm^{-1}.



^{1}H NMR (400 MHz, D_2O) δ 8.33 (s, 1H, H-6), 6.03 (d, *J* = 6.1 Hz, 1H, H-1'), 4.46 – 4.36 (m, 1H, H-4'), 4.37 – 4.29 (m, 1H, 5'a), 4.20 (s, 2H, H-9), 4.24 – 4.13 (m, 1H, H-5'b), 3.37 (t, *J* = 6.6 Hz, 2H, H-14), 2.55 – 2.40 (m, 1H, H-2'a), 2.38 (t, J = 7.3 Hz, 2H, H-12), 2.21 – 2.03 (m, 2H, H-2'b, H-3'a), 2.01 – 1.81 (m, 3H H-3'b, H-13). ^{13}C NMR (101 MHz, D_2O) δ 175.91 (C=O-11), 161.99 (C4), 151.58 (C=O-2), 146.35 (CH-6), 92.72 (2xC, C8, C5), 87.90 (CH-1') 81.50 (d, *J* = 8.5 Hz, CH-4'), 71.80 (C-7), 66.42 (d, *J* = 5.7 Hz, CH_2-5'), 50.54 (CH_2-14), 32.94 (CH_2-12), 32.23 (CH_2-2'), 29.74 (CH_2-9), 24.45 (CH_2-13), 24.06 (CH_2-3'). ^{31}P NMR (162 MHz, D_2O) δ -11.00 (d, *J* = 20.8 Hz, P_γ), -11.38 (d, *J* = 20.1 Hz, P_α), -23.48 (t, *J* = 17.7 Hz, P_β). HRMS (ESI^-) m/z: $[M-H]^-$ calcd for $C_{16}H_{23}N_7O_{13}P_3$ 614.0572; found 614.0553.

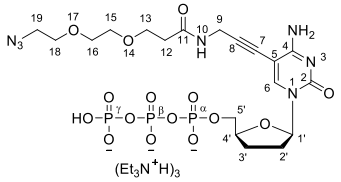
5-(3-(6-azidohexanamido)prop-1-ynyl)-2',3'-dideoxycytidine 5'-triphosphate (dd^{N3-C6}CTP) was prepared according to the procedure applied for preparation of dd^{N3-C2}CTP and obtained in 50% (0.016 mmol) yield. UV: λ_{max} = 294 nm, ε = 9300 ml·mol^{-1}·cm^{-1}.



^{1}H NMR (400 MHz, D_2O) δ 8.17 (s, 1H, H-6), 6.00 (dd, *J* = 7.0, 2.8 Hz, 1H, H-1'), 4.41 – 4.31 (m, 1H, H-4'), 4.31 – 4.24 (m, 1H, H-5'a), 4.16 (s, 2H, H-9), 4.19 – 4.10 (m, 1H, H-5'b), 3.25 (t, *J* = 6.7 Hz, 1H, H-16), 2.51 – 2.38 (m, 1H, H-2'a), 2.27 (t, *J* = 7.2 Hz, 2H, H-12), 2.12 – 2.00

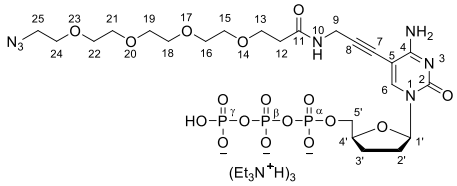
(m, 2H, H-2'b, H-3'a), 1.94 – 1.81 (m, 1H, H-3'b), 1.66 – 1.58 (m, 2H, H-13), 1.58 – 1.49 (m, 2H, H-15), 1.39 – 1.28 (m, 2H, H-14). ^{13}C NMR (101 MHz, D_2O) δ 176.91 (C=O-11), 164.78 (C-4), 155.39 (C=O-2), 145.01 (CH-6), 92.26 (2xC, C-8, C-5), 87.53 (CH-1'), 81.05 (d, *J* = 5.2 Hz, CH-4'), 73.09 (C-7), 66.79 (CH_2-5'), 51.12 (CH_2-16), 35.40 (CH_2-12), 32.17 (CH_2-2'), 29.64 (CH_2-9), 27.76 (CH_2-15), 25.36 (CH_2-14), 24.87 (CH_2-13), 24.55 (CH_2-3'). ^{31}P NMR (162 MHz, D_2O) δ -10.99 (d, *J* = 19.1 Hz, P_γ), -11.34 (d, *J* = 16.6 Hz, P_α), -23.48 (br s, P_β). HRMS (ESI^-) m/z: $[M-H]^-$ calcd for $C_{18}H_{27}N_7O_{13}P_3$ 642.0885; found 642.0879.

5-(3-(3-(2-(2-azidoethoxy)ethoxy)propanimido)prop-1-ynyl)-2',3'-dideoxycytidine 5'-triphosphate (dd$^{N3\text{-}PEG2}$CTP) was prepared according to the procedure applied for preparation of dd$^{N3\text{-}C2}$CTP and obtained in 66% (0.021 mmol) yield. UV: λ_{max} = 294 nm, ε = 9300 ml·mol^{-1}·cm^{-1}.



^{1}H NMR (400 MHz, D_2O) δ 8.17 (s, 1H, H-6), 6.02 (dd, *J* = 8.2, 2.0 Hz, 1H, H-1'), 4.41 – 4.33 (m, 1H, H-4'), 4.32 – 4.24 (m, 1H, H-5'a), 4.19 (s, 2H, H-9), 4.22 – 4.10 (m, 1H, H-5'b), 3.78 (t, *J* = 5.7 Hz, 2H, H-13), 3.66 – 3.62 (m, 6H, H-15 – H-18), 3.41 (t, *J* = 4.2 Hz, 2H, H-19), 2.55 (t, *J* = 5.9 Hz, 2H, H-12), 2.47 – 2.41 (m, 1H, H-2'a), 2.13 – 2.00 (m, 2H, H-2'b, H-3'a), 1.92 – 1.81 (m, 1H, H-3'b). ^{13}C NMR (101 MHz, D_2O) δ 174.08 (C=O-11), 165.06 (C-4), 155.87 (C=O-2), 145.00 (CH-6), 91.96 (2xC, C-8, C-5), 87.42 (CH-1'), 80.99 (d, *J* = 7.8 Hz, CH-4'), 73.21 (C-7), 69.55 (2xCH_2, C-15 – C-18), 69.27 (CH_2, C-15 – C-18), 66.83 (d, *J* = 4.7 Hz, CH_2-5'), 66.70 (CH_2-13), 50.16 (CH_2-19), 36.00 (CH_2-12), 32.12 (CH_2-2'), 29.72 (CH_2-9), 24.63 (CH_2-3'). ^{31}P NMR (162 MHz, D_2O) δ -10.64 (d, *J* = 19.4 Hz, P_γ), -11.39 (d, *J* = 20.0 Hz, P_α), -23.25 – -23.69 (m, P_β). HRMS (ESI^-) m/z: $[M-H]^-$ calcd for $C_{19}H_{29}N_7O_{15}P_3$ 688.0940; found 688.0939.

5-(1-azido-15-oxo-3,6,9,12-tetraoxa-16-azanonadec-18-yn-19-yl)-2',3'-dideoxycytidine 5'-triphosphate (dd$^{N3\text{-}PEG4}$CTP) was prepared according to the procedure applied for preparation of dd$^{N3\text{-}C2}$CTP and obtained in 73% (0.023 mmol) yield. UV: λ_{max} = 294 nm, ε = 9300 ml·mol^{-1}·cm^{-1}.

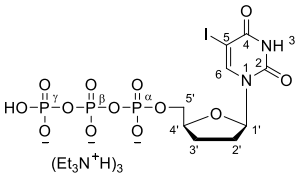


^{1}H NMR (400 MHz, D_2O) δ 8.13 (s, 1H, H-6), 6.00 (dd, *J* = 6.9, 2.9 Hz, 1H, H-1'), 4.40 – 4.29 (m, 1H, H-4'), 4.30 – 4.19 (m, 1H, H-5'a), 4.17 (s, 2H, H-9), 4.15 – 4.06 (m, 1H, H-5'b), 3.76 (t, *J* = 6.0 Hz, 2H, H-13), 3.67 – 3.58 (m, 14H, H-15 – H-24), 3.43 (t, *J* = 4.8 Hz, 2H, H-25), 2.53 (t, *J* = 6.0 Hz, 2H, H-12), 2.48 – 2.38 (m, 1H, H-2'a), 2.12 – 1.96 (m, 2H, H-2'b, H-3'a), 1.94 – 1.77 (m, 1H, H-3'b). ^{13}C

NMR (101 MHz, D_2O) δ 173.93 (C-11), 164.90 (C-4), 155.67 (C=O-2), 144.96 (CH-6), 92.03 (2xC, C-8, C-5), 87.37 (CH-1'), 80.92 (d, *J* = 8.4 Hz, CH-4'), 73.13 (C-7), 69.65 (CH_2, CH_2-15 – CH_2-22), 69.62 (br. s, 3xCH_2, 15 – 22), 69.58 (CH_2, CH_2-15 – CH_2-22), 69.56 (CH_2, CH_2-15 – CH_2-22), 69.26 (CH_2-24), 66.89 (d, *J* = 5.7 Hz, CH_2-5'), 66.69 (CH_2-13), 50.21 (CH_2-25), 35.97 (CH_2-12), 32.10 (CH_2-2'), 29.74 (CH_2-9), 24.74 (CH_2-3'). ^{31}P NMR (162 MHz, D_2O) δ -10.90 (d, *J* = 20.0 Hz, P_γ), -11.40 (d, *J* = 20.2 Hz, P_α), -23.42 (t, *J* = 20.2 Hz, P_β). HRMS (ESI^-) m/z: $[M-H]^-$ calcd for $C_{23}H_{36}N_7O_{17}P_3$ 798.1284; found 798.1300.

5-Iodo-2',3'-dideoxyuridine 5'-triphosphate (dd^{I}UTP)

NBu_3 (1.0 ml, 1.5 eq.) was added to the solution of 5-iodo-2',3'-dideoxyuridine (dd^{I}U) (1 g, 2.96 mmol) in trimethyl phosphate (TMP) (14 ml) under argon atmosphere. The suspension was stirred for 30 min at room temperature, then cooled in an ice bath (0 – 5 °C), followed by the dropwise addition of phosphorus oxychloride (0.79 ml, 3 eq.). After 2 h TBAPP cocktail consisting of: 0.5 M TBAPP (22.52 ml, 4 eq.) solution in ACN, Bu_3N (4.67 ml, 7 eq) and ACN (10 ml) was added to the reaction mixture and stirred for 30min at room temperature. The reaction mixture was then quenched with cooled 0.5 M TEAB buffer (100 ml, pH 7). The product was purified by ion exchange chromatography on Q Sepharose FF resin applying water/1 M TEAB gradient (5-56 %). The desired triphosphate was obtained in 82 % (2.44 mmol) yield. UV: λ_{max} = 287 nm, ε = 7700 $l \cdot mol^{-1} \cdot cm^{-1}$.

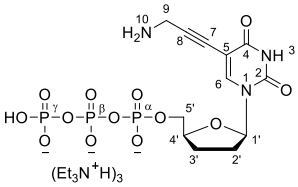


^{1}H NMR (400 MHz, D_2O) δ 8.12 (s, 1H, H-6), 5.94 (dd, *J* = 6.8, 3.7 Hz, 1H, H-1'), 4.31 – 4.23 (m, 1H, H-4'), 4.15 (ddd, *J* = 11.1, 5.8, 2.6 Hz, 1H, H-5'a), 4.01 (dt, *J* = 11.8, 6.0 Hz, 1H, H-5'b), 2.42 – 2.27 (m, 1H, H-2'a), 2.12 – 1.97 (m, 2H, H-2'b and H-3'a), 1.95 – 1.87 (m, 1H, H-3'b). ^{13}C NMR (101 MHz, D_2O) δ 162.82 (C=O-4), 151.19 (C=O-2), 146.24 (CH-6), 86.75 (CH-1'), 80.43 (d, *J* = 8.5 Hz, CH-4'), 67.83 (C-5), 67.07 (d, *J* = 5.6 Hz, CH_2-5'), 31.12 (CH_2-2'), 25.01 (CH_2-3'). ^{31}P NMR (162 MHz, D_2O) δ -11.18 (d, *J* = 20.1 Hz, P_γ), -11.64 (d, *J* = 20.3 Hz, P_α), -23.64 (t, *J* = 20.1 Hz, P_β). HRMS (ESI^-) m/z: $[M-H]^-$ calcd for $C_9H_{13}IN_2O_{13}P_3$ 576.8675; found 576.8684.

5-(3-aminoprop-1-ynyl)-2',3'-dideoxyuridine 5'-triphosphate (dd^{PA}UTP)

dd^{PA}UTP was synthesized from 5-iodo-2',3'-dideoxyuridine 5'-triphosphate (dd^{I}UTP) (2.2 mmol) according to the procedure applied for preparation of dd^{PA}CTP. The product was purified by ion exchange chromatography on Q Sepharose FF resin applying water/1M TEAB gradient (5-46 %). Desalting was carried out on C18 column with water/acetonitrile

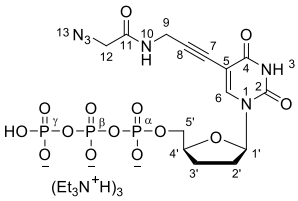
gradient (0-100 %). The product was obtained in 73 % (1.6 mmol) yield. UV: λ_{max} = 290 nm, ε = 13000 l·mol^{-1}·cm^{-1}.



^{1}H NMR (400 MHz, D_2O) δ 8.28 (s, 1H, H-6), 5.82 (d, *J* = 5.6 Hz, 1H, H-1'), 4.19 (br s, 2H, H-4' and H-5'a), 3.97 (br s, 1H, H-5'b), 3.83 (br s, 2H, H-9), 2.37 – 2.18 (m, 1H, H-2'a), 2.07 – 1.93 (m, 1H, H-2'b), 1.93 – 1.75 (m, 2H, H-3'). ^{13}C NMR (101 MHz, D_2O) δ 164.02 (C-4), 150.22 (C=O-2), 146.33 (CH-6), 97.15 (C-5), 87.14 (CH-1'), 84.99 (C-8), 81.49 (d, *J* = 2.2 Hz, CH-4'), 78.13 (C-7), 65.81 (CH_2-5'), 32.52 (CH_2-2'), 29.77 (CH_2-9), 23.50 (CH_2-3'). ^{31}P NMR (162 MHz, D_2O) δ -12.18 (br s, P_γ), -12.73 (d, *J* = 15.1 Hz, P_α), -24.84 (br s, P_β). HRMS (ESI$^-$) m/z: $[M-H]^-$ calcd for $C_{12}H_{17}N_3O_{13}P_3$ 503.9974; found 503.9983.

5-(3-(2-azidoacetamido)prop-1-ynyl)-2',3'-dideoxyuridine 5'-triphosphate (dd^{N3-C2}UTP)

dd^{N3-C2}UTP was synthesized from 5-(3-aminoprop-1-ynyl)-2',3'-dideoxyuridine 5'-triphosphate (dd^{PA}UTP) (1.37 mmol) according to the procedure applied for preparation of dd^{N3}CTP. The product was purified by ion exchange chromatography on Q Sepharose FF resin applying water/1M TEAB gradient (5-55 %). Desalting was carried out on C18 column with water/acetonitrile gradient (0-100 %). The product was obtained in 82 % (1.13 mmol) yield. UV: λ_{max} = 290 nm, ε = 13000 l·mol^{-1}·cm^{-1}.

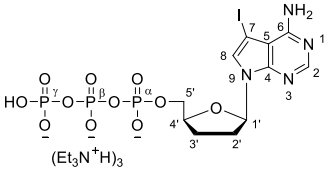


^{1}H NMR (400 MHz, D_2O) δ 8.10 (s, 1H, H-6), 5.97 (dd, *J* = 6.8, 3.2 Hz, 1H, H-1'), 4.30 (br s, 1H, H-4'), 4.25 – 4.18 (m, 1H, H-5'a), 4.15 (s, 2H, H-9), 4.11 – 4.02 (m, 1H, H-5'b), 3.99 (s, 2H, H-12), 2.47 – 2.30 (m, 1H, H-2'a), 2.13 – 1.97 (m, 2H, H-2'b and H-3'a), 1.97 – 1.81 (m, 1H, H-3'b). ^{13}C NMR (101 MHz, D_2O) δ 170.01 (C=O-11), 164.25 (C=O-4), 150.31 (C=O-2), 145.03 (CH-6), 98.33 (C-5), 89.29 (C-8), 86.89 (CH-1'), 80.83 (d, *J* = 7.9 Hz, CH-4'), 73.83 (C-7), 66.79 (d, *J* = 3.2 Hz, CH_2-5'), 51.70 (CH_2-12), 31.45 (CH_2-2'), 29.64 (CH_2-9), 24.72 (CH_2-3'). ^{31}P NMR (162 MHz, D_2O) δ -11.10 (d, *J* = 10.2 Hz, P_γ), -11.47 (d, J = 15.5 Hz, P_α), -23.53 (br s, P_β). HRMS (ESI$^-$) m/z: $[M-H]^-$ calcd for $C_{14}H_{18}N_6O_{14}P_3$ 587.0094; found 587.0104.

7-Deaza-7-iodo-2',3'-dideoxyadenosine 5'-triphosphate (dd^{I}ATP)

7-Deaza-7-iodo-2',3'-dideoxyadenosine (dd^{I}A) (1 g, 2.78 mmol) was suspended in ACN (27.8 ml) under argon atmosphere. The suspension was stirred for 30 min at room temperature, then cooled in an ice bath. Immediately before dropwise addition of diphosphoryl chloride (1.14 ml, 3 eq.) the flask

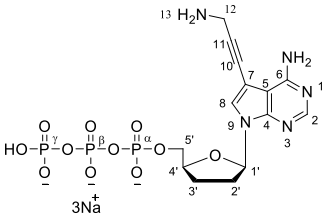
was lifted out of the ice bath. When the solid was completely dissolved (~7min) the flask was immersed in ice bath and stirred for 10 min. TBAPP cocktail consisting of: 0.5 M TBAPP solution in ACN (22.2 ml, 4 eq.), Bu_3N (4.62 ml, 7 ekv.) and ACN (10 ml) was then added to the reaction mixture and stirred for 30 min at room temperature. The reaction mixture was quenched with cooled 1 M TEAB buffer (100 ml, pH 7). The product was purified by ion exchange chromatography on Q Sepharose FF resin applying water/1 M TEAB gradient (5-52 %). The desired triphosphate was obtained in 73 % (2.03 mmol) yield. UV: λ_{max} = 283 nm, ε = 8500 l·mol^{-1}·cm^{-1}.



^{1}H NMR (400 MHz, D_2O) δ 7.98 (s, 1H, H-2), 7.54 – 7.39 (m, 1H, H-8), 6.23 – 6.13 (m, 1H, H-1'), 4.32 (br d, *J* = 2.7 Hz, 1H, H-4'), 4.23 – 4.09 (m, 1H, H-5'a), 4.06 – 3.90 (m, 1H, H-5'b), 2.56 – 2.39 (m, 1H, H-2'a), 2.22 – 2.05 (m, 2H, H-2'b and H-3'a), 2.04 – 1.90 (m, 1H, H-3'b). ^{13}C NMR (101 MHz, D_2O) δ 153.65 (C-6), 147.53 (C-4), 147.24 (CH-2), 128.01 (CH-8), 102.59 (C-5), 84.15 (CH-1'), 80.00 (d, *J* = 8.2 Hz, CH-4'), 67.45 (d, *J* = 5.7 Hz, CH_2-5'), 51.73 (C-7), 31.65 (CH_2-2'), 25.65 (CH_2-3'). ^{31}P NMR (162 MHz, D_2O) δ -10.78 (d, *J* = 19.6 Hz, P_γ), -11.34 (d, *J* = 19.9 Hz, P_α), -23.35 (t, *J* = 19.7 Hz, P_β). HRMS (ESI$^-$) m/z: [M-H]$^-$ calcd for $C_{11}H_{15}IN_4O_{11}P_3$ 598.8995; found 598.9009.

7-(3-Aminoprop-1-ynyl)-7-deaza-2',3'-dideoxyadenosine 5'-triphosphate (dd^{PA}ATP)

dd^{PA}ATP was synthesized from 7-deaza-7-iodo-2',3'-dideoxyadenosine (dd^{I}ATP) (2.03 mmol) according to the procedure applied for preparation of dd^{PA}CTP. In this reaction addition of Et_3N (0.565 ml, 4.06 mmol, 2 eq.) prior to Pd/TPPTS catalyst system was necessary. The product was purified by ion exchange chromatography on Q Sepharose FF resin applying water/1 M NaCl gradient (5-24 %). Desalting was carried out on C18 column using water/acetonitrile gradient (0-100 %). The product was obtained in 69 % (1.4 mmol) yield. UV: λ_{max} = 280 nm, ε = 12700 l·mol^{-1}·cm^{-1}.

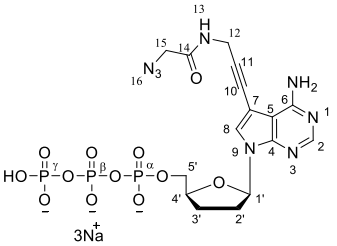


^{1}H NMR (400 MHz, D_2O) δ 7.43 (s, 1H, H-2), 7.18 (s, 1H, H-8), 5.83 (br s, 1H, H-1'), 4.20 (br s, 1H, H'4'), 4.06 (br s, 1H, H-5'a), 4.00 – 3.74 (m, 3H, H-12 and H-5'b), 2.23 (br d, *J* = 6.6 Hz, 1H, H-2'a), 1.97 (br s, 1H, H-3'a), 1.76 (br s, 1H, H-2'b), 1.64 (br s, 1H, H-3'b). ^{13}C NMR (101 MHz, D_2O) δ 155.01 (C-6), 150.10 (CH-2), 146.16 (C-4), 126.97 (CH-8), 101.48 (C-5), 93.97 (C-7), 84.28 (CH-1'), 83.35 (C-10), 79.96 (d, *J* = 7.4 Hz, CH-4'), 78.53 (C-9), 67.37 (d, *J* = 4.1 Hz, CH_2-5'), 31.89 (CH_2-2'),

29.81 (CH_2-11), 25.32 (CH_2-3'). ^{31}P NMR (162 MHz, D_2O) δ -6.43 (d, *J* = 18.6 Hz, P_γ), -10.57 (d, *J* = 18.0 Hz, P_α), -20.93 (t, *J* = 18.2 Hz, P_β).HRMS (ESI^-) m/z: $[M-H]^-$ calcd for $C_{14}H_{19}N_5O_{11}P_3$ 526.0294; found 526.0304.

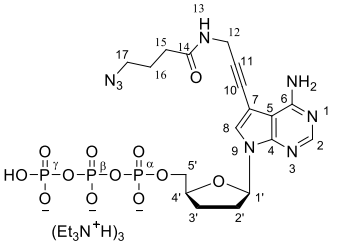
7-(3-(2-Azidoacetamido)prop-1-ynyl)-7-deaza-2',3'-dideoxyadenosine 5'-triphosphate (dd^{N3-C2}ATP)

dd^{N3-C2}ATP was synthesized from 7-deaza-7-PA-2',3'-dideoxyadenosine (dd^{PA}ATP) (0.73 mmol) according to the procedure applied for preparation of dd^{N3}CTP. The product was purified by ion exchange chromatography on Q Sepharose FF resin applying water/1 M TEAB gradient (5-60 %). Desalting was carried out on C18 column using water/acetonitrile gradient (0-100 %). The product was obtained in 83 % (0.61 mmol) yield. UV: λ_{max} = 280 nm, ε = 12700 l·mol^{-1}·cm^{-1}.



^{1}H NMR (400 MHz, D_2O) δ 7.99 (s, 1H, H-2), 7.52 (s, 1H, H-8), 6.31 (dd, *J* = 6.7, 3.9 Hz, 1H, H-1'), 4.46 – 4.36 (m, 1H, H-4'), 4.27 (s, 2H, H-12), 4.20 – 4.13 (m, 1H, H-5'a), 4.10 (s, 2H, H-15), 4.06 – 3.95 (m, 1H, H-5'b), 2.61 – 2.48 (m, 1H, H-2'a), 2.37 – 2.20 (m, 2H, H-2'b and H-3'a), 2.11 – 1.98 (m, 1H, H-3'b). ^{13}C NMR (101 MHz, D_2O) δ 170.50 (C=O-14), 156.97 (C-6), 151.87 (CH-2), 147.92 (C-4), 126.21 (CH-8), 102.77 (C-5), 95.32 (C-7), 87.60 (C-11), 84.04 (CH-1'), 79.95 (d, *J* = 8.2 Hz, CH-4'), 75.34 (C-10), 67.60 (d, *J* = 5.9 Hz, CH_2-5'), 51.90 (CH_2-15), 31.15 (CH_2-2'), 29.89 (CH_2-12), 26.01 (CH_2-3'). ^{31}P NMR (162 MHz, D_2O) δ -6.81 (d, *J* = 19.7 Hz, P_γ), -10.73 (d, *J* = 18.9 Hz, P_α), -21.74 (t, *J* = 19.2 Hz, P_β). HRMS (ESI^-) m/z: $[M-H]^-$ calcd for $C_{16}H_{20}N_8O_{12}P_3$ 609.0414; found 609.0425.

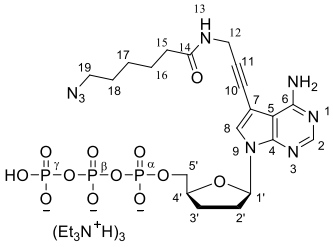
7-(3-(4-Azidobutanamido)prop-1-ynyl)-7-deaza-2',3'-dideoxyadenosine 5'-triphosphate (dd^{N3-C4}ATP) was prepared according to the procedure applied for preparation of dd^{N3-C2}ATP and obtained in 55% (0.11 mmol) yield. UV: λ_{max} = 280 nm, ε = 12700 l·mol^{-1}·cm^{-1}.



^{1}H NMR (400 MHz, D_2O) δ 7.95 (s, 1H, H – 2), 7.42 (s, 1H, H-8), 6.24 (dd, *J* = 6.8, 3.5 Hz, 1H, H-1'), 4.39 (tt, *J* = 9.2, 3.2 Hz, 1H, H-4'), 4.19 (s, 2H, CH_2-12), 4.18 – 4.09 (m, 1H, H-5'a), 4.08 – 3.93 (m, 1H, H-5'b), 3.35 (t, *J* = 6.6 Hz, 2H, CH_2-17), 2.60 – 2.44 (m, 1H, H-2'a), 2.38 (t, *J* = 7.3 Hz, 2H, CH_2-15), 2.31 – 2.16 (m, 2H, H-2'a, H-3'a), 2.09 – 1.93 (m, 1H, H-3'a), 1.88 (p, *J* = 6.9 Hz, 2H, CH_2-16). ^{13}C NMR (101 MHz, D_2O) δ 175.77 (s, C=O-14), 156.24 (s, C-6), 151.05 (s, CH-2), 147.52 (s, C-4), 126.05 (s, CH-8), 102.52 (s, C-5), 95.52

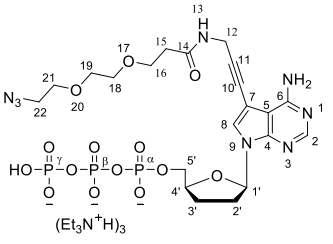
(s, C-7), 88.19 (s, C-11), 84.06 (s, CH-1'), 79.92 (d, *J* = 8.1 Hz, CH-4'), 74.85 (s, C-10), 67.68 (d, *J* = 5.7 Hz, CH_2-5'), 50.54 (s, CH_2-17), 32.92 (s, CH_2-15), 31.19 (s, CH_2-2'), 29.88 (s, CH_2-12), 26.00 (s, CH_2-3'), 24.57 (s, CH_2-16). ^{31}P NMR (162 MHz, D_2O) δ -8.05 (d, *J* = 19.2 Hz, P_γ), -10.75 (d, *J* = 18.8 Hz, P_α), -21.88 (t, *J* = 19.0 Hz, P_β). HRMS (ESI^-) m/z: $[M-H]^-$ calcd for $C_{18}H_{24}N_8O_{12}P_3$ 637.0725; found 637.0737.

7-(3-(6-Azidohexanamido)prop-1-ynyl)-7-deaza-2',3'-dideoxyadenosine 5'-triphosphate (dd^{N3-C6}ATP) was prepared according to the procedure applied for preparation of dd^{N3-C2}ATP and obtained in 64% (0.13 mmol) yield. UV: λ_{max} = 280 nm, ε = 12700 l·mol^{-1}·cm^{-1}.



^{1}H NMR (400 MHz, D_2O) δ 7.98 (s, 1H, H – 2), 7.45 (s, 1H, H-8), 6.25 (dd, *J* = 6.8, 3.6 Hz, 1H, H-1'), 4.47 – 4.29 (m, 1H, H-4'), 4.24 – 4.09 (m, 3H, CH_2-12, H-5'a), 4.08 – 3.94 (m, 1H, H-5'b), 3.13 (t, *J* = 6.8 Hz, 2H, CH_2-19), 2.58 – 2.41 (m, 1H, H-2'a), 2.32 – 2.16 (m, 4H, H-2'b, H – 3'a, CH_2-15), 2.07 – 1.88 (m, 1H, H-3'b), 1.59 (p, *J* = 7.6 Hz, 2H, CH_2-16), 1.48 (p, *J* = 7.6 Hz, 2H, CH_2-18), 1.35 – 1.24 (m, 2H, CH_2-17). ^{13}C NMR (101 MHz, D_2O) δ 176.89 (s, C=O-14), 156.29 (s, C-6), 151.10 (s, CH-2), 147.56 (s, C-4), 126.05 (s, CH-8), 102.61 (s, C-5), 95.65 (s, C-7), 88.61 (s, C-11), 84.13 (s, CH-1'), 79.94 (d, *J* = 8.2 Hz, CH-4'), 74.74 (s, C-10), 67.70 (d, *J* = 5.7 Hz, CH_2-5'), 51.06 (s, CH_2-19), 35.43 (s, CH_2-15), 31.32 (s, CH_2-2'), 29.80 (s, CH_2-12), 27.79 (s, CH_2-18), 26.00 (s, CH_2-3'), 25.44 (s, CH_2-17), 25.00 (s, CH_2-16). ^{31}P NMR (162 MHz, D_2O) δ -8.60 (d, *J* = 19.2 Hz, P_γ), -10.81 (d, *J* = 19.0 Hz, P_α), -22.06 (t, *J* = 19.0 Hz, P_β). HRMS (ESI^-) m/z: $[M-H]^-$ calcd for $C_{20}H_{28}N_8O_{12}P_3$ 665.1037; found 665.1044.

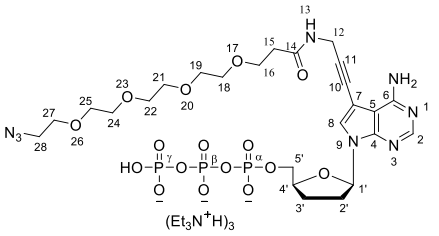
7-(3-(3-(2-(2-Azidoethoxy)ethoxy)propanimido)prop-1-ynyl)-7-deaza-2',3'-dideoxyadenosine 5'-triphosphate (dd$^{N3-PEG2}$ATP) was prepared according to the procedure applied for preparation of dd^{N3-C2}ATP and obtained in 56% (0.11 mmol) yield. UV: λ_{max} = 280 nm, ε = 12700 l·mol^{-1}·cm^{-1}.



^{1}H NMR (400 MHz, D_2O) δ 7.98 (s, 1H, H-2), 7.47 (s, 1H, H-8), 6.27 (dd, *J* = 6.7, 3.5 Hz, 1H, H-1'), 4.48 – 4.34 (m, 1H, H-4'), 4.20 (s, 2H, CH_2-12), 4.19 – 4.11 (m, 1H, H-5'a), 4.05 – 3.94 (m, 1H, H-5'b), 3.79 (t, *J* = 5.8 Hz, 2H, CH_2-16), 3.65 – 3.58 (m, 2H, CH_2-18), 3.57 – 3.52 (m, 2H, CH_2-19), 3.52 – 3.46 (m, 2H, CH_2-21), 3.28 (dd, *J* = 5.5, 4.2 Hz, 2H, CH_2-22), 2.60 – 2.48 (m, 3H, CH_2 – 15, H-2'a), 2.31 – 2.20 (m, 2H, H-2'b, H-3'a), 2.07 – 1.94 (m, 1H, H-3'b). ^{13}C

NMR (101 MHz, D_2O) δ 174.24 (s, C=O-14), 156.66 (s, C-6), 151.62 (s, CH-2), 147.72 (s, C-4), 126.01 (s, CH-8), 102.65 (s, C-5), 95.52 (s, C-7), 88.30 (s, C-11), 84.05 (s, CH-1'), 79.91 (d, *J* = 8.2 Hz, CH-4'), 74.92 (s, C-10), 69.58 (s, CH_2-18 or 19), 69.50 (s, CH_2-18 or 19), 69.18 (s, CH_2-21), 67.70 (d, *J* = 5.7 Hz, CH_2-5'), 66.79 (s, CH_2-16), 50.03 (s, CH_2-22), 36.10 (s, CH_2-15), 31.23 (s, CH_2-2'), 29.87 (s, CH_2-12), 26.06 (s, CH_2-3'). ^{31}P NMR (162 MHz, D_2O) δ -7.42 (d, *J* = 19.1 Hz, P_γ), -10.66 (d, *J* = 18.7 Hz, P_α), -21.61 (t, *J* = 18.9 Hz, P_β). HRMS (ESI^-) m/z: $[M-H]^-$ calcd for $C_{21}H_{30}N_8O_{14}P_3$ 711.1092; found 711.1101.

7-(1-Azido-15-oxo-3,6,9,12-tetraoxa-16-azanonadec-18-yn-19-yl)-7-deaza-2',3'-dideoxyadenosine 5'-triphosphate (dd$^{N3\text{-}PEG2}$ATP) was prepared according to the procedure applied for preparation of dd$^{N3\text{-}C2}$ATP and obtained in 54% (0.10 mmol) yield. UV: λ_{max} = 280 nm, ε = 12700 l·mol^{-1}·cm^{-1}.

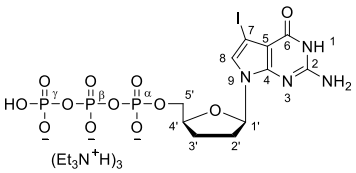


^{1}H NMR (400 MHz, D_2O) δ 8.01 (s, 1H, H-2), 7.52 (s, 1H, H-8), 6.29 (dd, *J* = 6.7, 3.5 Hz, 1H, H-1'), 4.45 – 4.33 (m, 1H, H-4'), 4.21 (s, 2H, CH_2-12), 4.20 – 4.11 (m, 1H, H-5'a), 4.09 – 3.95 (m, 1H, H-5'b), 3.80 (t, *J* = 5.8 Hz, 2H, CH_2-16), 3.68 – 3.59 (m, 4H, 2xCH_2 – from 18-27), 3.59 – 3.53 (m, 4H, 2xCH_2 – from 18-27), 3.53 – 3.44 (m, 6H, 3xCH_2 – from 18-27), 3.44 – 3.37 (m, 2H, CH_2-28), 2.56 (t, *J* = 5.8 Hz, 2H, CH_2-15), 2.54 – 2.47 (m, 1H, H-2'a), 2.31 – 2.20 (m, 2H, H – 2'b, H-3'a), 2.02 (qd, *J* = 11.9, 7.0 Hz, 1H, H-3'b). ^{13}C NMR (101 MHz, D_2O) δ 174.21 (s, C=O-14), 156.45 (s, C-6), 151.32 (s, CH-2), 147.63 (s), C-4, 126.21 (s, CH-8), 102.61 (s, C-5), 95.66 (s, C-7), 88.38 (s, C-11), 84.10 (s, CH-1'), 79.92 (d, *J* = 8.2 Hz, CH-4'), 74.82 (s, CH-10), 69.60 (s, CH_2, CH_2-18 – CH_2-24), 69.52 (s, CH_2, CH_2-18 – CH_2-24), 69.38 (s, 2xCH_2, CH_2-18 – CH_2-24), 69.17 (s, CH_2, CH_2-18 – CH_2-24), 67.67 (d, *J* = 5.9 Hz, CH_2-5'), 66.75 (s, CH_2-16), 50.13 (s, CH_2-25), 36.05 (s, CH_2-15), 31.36 (s, CH_2-2'), 29.86 (s, CH_2-12), 26.00 (s, CH_2-3'). ^{31}P NMR (162 MHz, D_2O) δ -8.70 (d, *J* = 19.2 Hz, P_γ), -10.80 (d, *J* = 19.0 Hz, P_α), -22.08 (t, *J* = 19.1 Hz, P_β). HRMS (ESI^-) m/z: $[M-H]^-$ calcd for $C_{25}H_{38}N_8O_{16}P_3$ 799.1616; found 799.1619.

7-Deaza-7-iodo-2',3'-dideoxyadenosine 5'-triphosphate (dd^{I}GTP)

7-Deaza-7-iodo-2',3'-dideoxyguanosine (1.13 g, 3.0 mmol) was suspended in TMP (16.8 ml) under argon atmosphere, NBu_3 (1.8 ml, 7.5 mmol, 2.5 eq.) was added. The suspension was stirred for 30 min at room temperature, then cooled in an ice bath followed by the dropwise addition of phosphorus oxychloride (0.84 ml, 3 eq.). After 1h TBAPP cocktail consisting

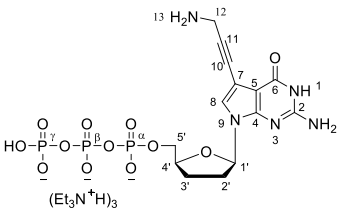
of: 0.5 M TBAPP solution in ACN (24 ml, 4 eq.), Bu_3N (4.67 ml, 7 eq.) and ACN (10 ml) was added to the reaction mixture and stirred for 30 min at room temperature. The reaction mixture was quenched with cooled 0.5 M TEAB buffer (100 ml, pH 7). The product was purified by ion exchange chromatography on Q Sepharose FF resin using water/1M TEAB gradient (5-75 %). The desired triphosphate was obtained in 57 % (1.71 mmol) yield. UV: λ_{max} = 267 nm, ε = 11000 l·mol^{-1}·cm^{-1}.



^{1}H NMR (400 MHz, D_2O) δ 6.97 (s, 1H, H-8), 5.96 (dd, *J* = 6.7, 3.9 Hz, 1H, H-1'), 4.27 – 4.17 (m, 1H, H-4'), 4.09 – 3.98 (m, 1H, H-5'a), 3.90 – 3.81 (m, 1H, H-5'b), 2.43 – 2.26 (m, 1H, H-2'a), 2.19 – 2.04 (m, 2H, H-2'b and H-3'a), 1.98 – 1.84 (m, 1H, H-3'b). ^{13}C NMR (101 MHz, D_2O) δ 159.98 (C-6), 152.40 (C-2), 150.35 (C-4), 123.27 (CH-8), 100.40 (C-5), 83.62 (CH-1'), 79.39 (d, *J* = 8.3 Hz, CH-4'), 67.91 (d, *J* = 5.8 Hz, CH_2-5'), 54.05 (C-7), 30.81 (CH_2-2'), 26.18 (CH_2-3'). ^{31}P NMR (162 MHz, D_2O) δ -11.13 (d, *J* = 19.8 Hz, P_γ), -11.47 (d, *J* = 20.1 Hz, P_α), -23.48 (t, *J* = 20.0 Hz, P_β). HRMS (ESI$^-$) m/z: [M-H]$^-$ calcd for $C_{11}H_{15}IN_4O_{12}P_3$ 614.8944; found 614.8952.

7-(3-Aminoprop-1-ynyl)-7-deaza-2',3'-dideoxyadenosine 5'-triphosphate (dd^{PA}GTP)

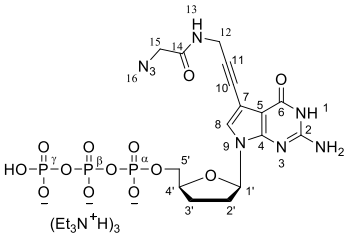
dd^{PA}GTP was synthesized from 7-deaza-7-iodo-2',3'-dideoxyguanosine (dd^{I}GTP) (0.2 mmol) according to the procedure applied for preparation of dd^{PA}CTP. The product was purified by ion exchange chromatography on Q Sepharose FF resin applying water/1M TEAB gradient (5-62 %). Desalting was carried out on C18 column using water/acetonitrile gradient (0-100%). The product was obtained in 41 % (0.08 mmol) yield. UV: λ_{max} = 272 nm, ε = 11900 l·mol^{-1}·cm^{-1}.



^{1}H NMR (400 MHz, D_2O) δ 7.24 (s, 1H, H-8), 5.96 (dd, J = 6.6, 3.6 Hz, 1H, H-1'), 4.32 – 4.25 (m, 1H, H-4'), 4.17 – 4.09 (m, 1H, H-5'a), 4.02 – 3.94 (m, 3H, H-12 and H-5'b), 2.43 – 2.31 (m, 1H, H-2'), 2.15 – 2.04 (m, 2H, H-2' and H-3'a), 2.00 – 1.89 (m, 1H, H-3'). ^{13}C NMR (101 MHz, D_2O) δ 159.75 (C=O-6), 152.77 (C-2), 149.33 (C-4), 124.50 (CH-8), 99.59 (C-5), 96.94 (C-7), 84.21 (CH-1'), 81.61 (C-11), 80.02 (C-10), 79.81 (d, *J* = 8.5 Hz, CH-4'), 67.43 (d, *J* = 5.6 Hz, CH_2-5'), 31.56 (CH_2-2'), 30.19 (CH_2-12), 25.61 (CH_2-3'). ^{31}P NMR (162 MHz, D_2O) δ -10.99 (d, *J* = 19.3 Hz, P_γ), -11.30 (d, *J* = 19.7 Hz, P_α), -23.34 (t, *J* = 18.7 Hz, P_β). HRMS (ESI$^-$) m/z: [M-H]$^-$ calcd for $C_{14}H_{19}N_5O_{12}P_3$ 542.0243; found 542.0251.

7-(3-(2-Azidoacetamido)prop-1-ynyl)-7-deaza-2',3'-dideoxyadenosine 5'-triphosphate (dd$^{N3\text{-}C2}$GTP)

dd^{N3}GTP was synthesized from 7-deaza-7-PA-2',3'-dideoxyguanosine (dd^{PA}GTP) (0.04 mmol) according to the procedure applied for preparation of dd^{N3}CTP. The product was purified by ion exchange chromatography on Q Sepharose FF resin applying water/1M TEAB gradient (5-70%). Desalting was carried out on C18 column using water/acetonitrile gradient (0-100%). The product was obtained in 74 % (0.03 mmol) yield. UV: λ_{max} = 272 nm, ε = 11900 l·mol^{-1}·cm^{-1}.



^{1}H NMR (400 MHz, D_2O) δ 7.21 (s, 1H, H-8), 6.09 (dd, *J* = 6.7, 4.1 Hz, 1H, H-1'), 4.37 – 4.29 (m, 1H, H-4'), 4.23 (s, 2H, H-15), 4.18 – 4.10 (m, 1H, H-5'a), 4.05 (s, 2H, H-12), 4.01 – 3.91 (m, 1H, H-5'b), 2.51 – 2.38 (m, 1H, H-2'a), 2.29 – 2.14 (m, 2H, H-2'b and H-3'a), 2.07 – 1.95 (m, 1H, H-3'b). ^{13}C NMR (101 MHz, D_2O) δ 170.07 (C=O-14), 160.15 (C=O-6), 152.92 (C-2), 149.87 (C-4), 123.71 (CH-8), 99.74 (C-5), 98.07 (C-7), 85.86 (CH-11), 83.88 (CH-1'), 79.65 (d, *J* = 8.3 Hz, CH-4'), 75.97 (C-10), 67.79 (d, *J* = 5.6 Hz, CH_2-5'), 51.84 (CH_2-15), 30.87 (CH_2-2'), 29.88 (CH_2-12), 26.04 (CH_2-3'). ^{31}P NMR (162 MHz, D^2O) δ -11.09 (br s, P_γ and P_α), -23.33 (P_β). HRMS (ESI$^-$) m/z: [M-H]$^-$ calcd for $C_{16}H_{20}N_8O_{13}P_3$ 625.0363; found 625.0376.

8.2 Synthesis of oligonucleotide-tethered 2'-deoxynucleotides and 2',3'-dideoxynucleotides

Analytical methods used for OTDNs and OTDDNs synthesis monitoring and purification are given in Table 24.

Table 24. Analytical methods specifications

Method name	Time, min	B, %	Curve*	Method name	Time, min	B, %	Curve*
AM1	0.0	15	5	**AM13**	0.0	12	5
	1.0	15	5		3.0	12	5
	12.0	27	7		12.0	19	5
	14.0	95	5		13.5	90	5
	16.0	95	5		15.0	9	5
	17.0	15	5		16.0	12	5
	22.0	15	5		22.0	12	5
AM2[a]	0.0	0	5	**AM14**	0.0	11	5
	10.0	40	5		3.0	11	5

	12.0	75	5		12.0	19	5
	14.5	75	5		13.5	90	5
	15.0	0	5		15.0	90	5
	22.0	0	5		16.0	11	5
					22.0	11	5
	0.0	10.0	5		0.0	11	5
	1.0	10.0	5		3.0	11	5
	12.0	30.0	5		12.0	18	5
AM3	14.0	95.0	5	**AM15**	13.5	90	5
	16.0	95.0	5		15.0	90	5
	17.0	10.0	5		16.0	11	5
	22.0	10.0	5		22.0	11	5
	0.0	10	5		0.0	11	5
	1.0	10	5		3.0	11	5
	12.0	30	7		12.0	19	7
AM4	14.0	95	5	**AM16**	13.5	90	5
	16.0	95	5		15.0	90	5
	17.0	10	5		16.0	11	5
	22.0	10	5		22.0	11	5
	0.0	10	5		0.0	10	5
	1.0	10	5		1.0	10	5
	12.0	25	5		14.0	15	5
AM5	14.0	95	5	**AM17**	16.0	90	5
	16.0	95	5		18.0	90	5
	17.0	10	5		19.0	10	5
	22.0	10	5		27.0	10	5
	0.0	10	5				
	1.0	10	5		0.0	11	5
	15.0	25	5		1.0	11	5
AM6	17.0	90	5	**PM1**	16.0	18	5
	19.0	90	5		23.0	100	5
	20.0	10	5		30.0	100	5
	28.0	10	5				
	0.0	12	5				
	1.0	12	5		0.0	11	5
	14.0	35	8		3.5	11	5
AM7	16.0	95	5	**PM2**	8.0	14	5
	18.0	95	5		20.0	14	5
	19.0	12	5		27.0	100	5
	27.0	12	5		34.0	100	5
	0.0	10	5				
	1.0	10	5		0.0	10	5
	15.0	25	7		2.0	10	5
AM8	17.0	95	5	**PM3**	20.0	20	5
	19.0	95	5		27.0	100	5
	20.0	10	5		34.0	100	5
	27.0	10	5				
AM9	0.0	14	5	**PM4**	0.0	11	5

	16.0	14	5		3.5	11	5
	18.0	95	5		15.0	19	5
	20.0	95	5		22.0	100	5
	21.0	14	5		30.0	100	5
	23.0	14	5				
	0.0	10	5				
	1.0	10	5		0.0	11	5
	15.0	25	8		3.5	11	5
AM10	17.0	95	5	**PM5**	17.0	17	5
	18.5	95	5		21.0	100	5
	19.5	10	5		30.0	100	5
	25.0	10	5				
	0.0	12	5				
	1.0	12	5		0.0	11	5
	12.0	40	9		3.0	11	5
AM11	14.0	95	5	**PM6**	20.0	19	5
	15.5	95	5		27.0	100	5
	16.5	12	5		39.0	100	5
	22.0	12	5				
	0.0	14	5				
	1.0	14	5		0.0	10	5
	17.5	49	9		3.5	10	5
AM12	18.5	100	5	**PM7**	18.0	15	5
	20.0	100	5		22.0	100	5
	21.0	14	5		33.0	100	5
	27.0	14	5				
					0.0	5	5
	0.0	5	5		8.0	5	5
DS1[b]	14.0	5	5	**DS2**[c]	15.0	90	5
	25.0	95	5		17.0	90	5
	40.0	95	5		18.0	5	5
					24.0	5	5

In all methods mobile phases, A – 100mM TEAAc and B – ACN, were used unless specified overwise. In all analytical chromatographic (AM) methods the column camber temperature was maintained at 40°C, flow rate 0.8 ml/min. All preparative chromatographic methods (PM) flow rate was 9 ml/min and column was not termostated. * – Gradient formation curve type specific to Chromeleon™ chromatography data system software. a – the mobile phases A – 100mM TEAAc and B – 40% 100 mM TEAAc + 60% CAN were used; b – chromatographic method used for desalting, mobile phases A – water, B – ACN were used (only for scale-up reactions); c - chromatographic method used for desalting, mobile phases A – water, B – ACN were used (only for small scale reactions).

General procedure for synthesis of OTDNs

All reaction components were applied to the reaction mixture as solutions in water unless specified differently. Corresponding $d^{N3}CTP$ (3 eq., 2-4 mM)

solution was added to alkyne-oligonucleotide B1 or B2 (8.4 nmol) (Table 25) solution in sodium phosphate buffer (150 µl, 100 mM, pH 7). In a separate vial $CuSO_4$ (100 mM, 4 µl) and THPTA (5 eq. to $CuSO_4$, 250 mM, 8 µl) were premixed and added to the reaction mixture, followed by the addition of sodium ascorbate (1 M, 20 µl). Reaction mixture was stirred for 1 – 3h at 37 °C, quenched with 0.5 M EDTA-Na_2 solution (40 µl, pH 8). The products were purified by analytical HPLC applying reverse-phase chromatography (TSKgel™ ODS- 100V 3 µm column). Two purification steps applying different methods were performed.

Table 25. Alkyne-oligonucleotides B1 and B2 sequences

Oligonucleotide	Sequence
B1	(AldU)TTATATATTTATTGGAGACTGACTACCAGATGTAACA
B2	T(AldU)TATATATTTATTGGAGACTGACTACCAGATGTAACA

Table 26. Specifications of oligonucleotide-tethered 2'-deoxynucleotides

ON	OTDN	Yield, %	M (calcd), Da	M (found) [M], Da	Purification method
B1	$d^{B1\text{-}ON\text{-}C2}$dCTP	26	12377.088	12377.123	1 – AM1 2 – AM2
	$d^{B1\text{-}ON\text{-}C4}$dCTP	17	12405.119	12405.142	
	$d^{B1\text{-}ON\text{-}C6}$dCTP	19	12433.127	12433.102	
	$d^{B1\text{-}ON\text{-}PEG2}$dCTP	17	12479.156	12479.188	
	$d^{B1\text{-}ON\text{-}PEG4}$dCTP	22	12567.208	12567.237	
B2	$d^{B2\text{-}ON\text{-}C2}$dCTP	26	12073.020	12073.036	1 – AM3 2 – AM2
	$d^{B2\text{-}ON\text{-}C4}$dCTP	13	12101.067	12101.070	
	$d^{B2\text{-}ON\text{-}C6}$dCTP	14	12129.099	12129.100	
	$d^{B2\text{-}ON\text{-}PEG2}$dCTP	16	12175.104	12175.111	
	$d^{B2\text{-}ON\text{-}PEG4}$dCTP	21	12263.136	12263.170	

HRMS (ESI-): calculated and found monoisotopic mass for [M] are given.

General procedure for synthesis of OTDDNs

All reaction components were applied to the reaction mixture as solutions in water unless specified differently. Corresponding **dd^{N3}NTP** (3 eq., 2-4 mM) solution was added to alkyne-oligonucleotide (100-400 nmol) solution in sodium phosphate buffer (1 ml, 100 mM, pH 7). In a separate vial $CuSO_4$ (100 mM, 5 µmol, 50 µl) and THPTA (5 eq. to $CuSO_4$, 250 mM, 25 µmol, 100 µl) were premixed and added to the reaction mixture, followed by the addition of sodium ascorbate (1 M, 250 µl). Despite the amount of alkyne-oligonucleotide used the reaction volume was kept to 1 ml as well as the

amount of catalyst used. Reaction mixture was stirred at 42 °C and monitored by HPLC applying analytical method indicated in Table 27, after completions of the reaction (20 min – 1.5 h) 0.5 M EDTA-Na_2 solution (0.5 ml, pH 8) was applied. For small scale experiments (8.4 – 16.8 nmol) two purification steps applying different methods were performed by analytical HPLC using TSKgel™ ODS- 100V 3 μm column. The scale-up reactions (84 – 400 nmol) were monitored by HPLC using YMC-Triart C18 150x4.6 mm S – 3 μm column and purified by semi-preparative Waters 2555 Quaternary gradient system using YMC-Actus Triart C18 100x20 mm S – 5 μm column. Products were desalted using DS1 (for preparative system) or DS2 (for analytical system) methods.

Alkyne-oligonucleotides sequences and modifications are given in Table 20. The analytical and purification methods used for reactions monitoring and products' purification are indicated for each OTDDN in Table 27. OTDDNs MS data is provided in Table 27.

Table 27. Specifications of oligonucleotide-tethered 2',3'-dideoxynucleotides

ON	OTDDN	SM, nmol	Yield, %	M (calcd), Da	M (found) [M], Da	Purification method
ON1	dd$^{ON1\text{-}C2}$UTP	8.4	17	6218.044	6240.012	1 – AM4 2 – AM2
ON2	dd$^{ON2\text{-}C2}$UTP	8.4	30	8151.412	8151.405	1 – AM3 2 – AM2
	dd$^{ON2\text{-}C2}$CTP	8.4	41	8150.428	8173.388	1 – AM3 2 – AM2
ON3	dd$^{ON3\text{-}C2}$UTP	8.4	21	7852.272	7852.259	1 – AM5 2 – AM2
ON4	dd$^{ON4\text{-}C2}$CTP	8.4	31	21575.606	21575.625	1 – AM6 2 – AM2
ON5	dd$^{ON5\text{-}C2}$UTP	8.4	42	10419.768	10419.771	1 – AM7 2 – AM2
ON6	dd$^{ON6\text{-}C2}$UTP	8.4	25	7917.329	7917.320	1 – AM8 2 – AM2
	dd$^{ON6\text{-}C2}$UTP	200	34	7917.329	7917.320	AM8 PM1
	dd$^{ON6\text{-}C2}$CTP	84	28	7916.345	7916.342	AM8 PM1
	dd$^{ON6\text{-}C2}$CTP	210	39	7916.345	7916.342	AM8 PM1
	dd$^{ON6\text{-}C4}$CTP	210	24	7944.378	7944.370	AM8 PM1
	dd$^{ON6\text{-}C6}$CTP	210	18	7972.410	7972.404	AM8

						PM1
	dd$^{ON6\text{-}PEG2}$CTP	210	14	8018.415	8018.409	AM8 PM1
	dd$^{ON6\text{-}PEG4}$CTP	210	12	8106.468	8106.458	AM8 PM1
	dd$^{ON6\text{-}C2}$ATP	100	29	7939.354	7939.372	AM9 PM2
	dd$^{ON6\text{-}C2}$GTP	8.4	19	7955.352	7955.359	1 – AM10 2 – AM2
	dd$^{ON6\text{-}C2}$GTP	210	43	7955.352	7955.359	AM10 P3
ON7	dd$^{ON7\text{-}C2}$UTP	16.8	18	7666.136	7666.133	1 – AM8 2 – AM2
	dd$^{ON7\text{-}C2}$UTP	200	12	7666.136	7666.133	AM8 PM1
	dd$^{ON7\text{-}C2}$CTP	210	24	7665.152	7665.151	AM8 PM1
ON8	dd$^{ON8\text{-}C2}$UTP	16.8	16	10127.552	10127.565	1 – AM11 2 – AM2
	dd$^{ON8\text{-}C2}$CTP	8.4	18	10126.568	10142.566	1 – AM11 2 – AM2
ON9	dd$^{ON9\text{-}C2}$UTP	8.4	27	10120.710	10040.924	1 – AM12 2 – AM2
	dd$^{ON9\text{-}C2}$CTP	8.4	11	10119.726	10143.728	1 – AM12 2 – AM2
ON10	dd$^{ON10\text{-}C2}$UTP	100	32	10403.753	10403.764	AM13 PM4
	dd$^{ON10\text{-}C2}$CTP	210	55	10402.769	10386.769	AM14 PM4
ON11	dd$^{ON11\text{-}C2}$UTP	200	20	12744.086	12719.077	AM15 PM5
	dd$^{ON11\text{-}C2}$CTP	210	37	12743.102	12743.101	AM16 PM6
ON12	dd$^{ON12\text{-}C2}$UTP	400	29	7618.203	7618.192	AM17 PM7
	dd$^{ON12\text{-}C2}$CTP	300	48	7617.219	7617.209	AM17 PM7
ON13	dd$^{ON13\text{-}C2}$CTP	210	37	8361.404	8361.424	AM8 PM1
ON14	dd$^{ON14\text{-}C2}$UTP	100	36	13580.214	13580.190	AM8 PM1
	dd$^{ON14\text{-}C2}$CTP	100	32	13579.230	13579.212	AM8 PM1

SM – starting material (alkyne-oligonucleotide). Analytical methods used for reactions monitoring and small scale (8.4 – 16.8 nmol) synthesis purification are indicated as AM. Analytical methods used for scale-up (84 – 400 nmol) reactions purification are indicated as PM.

9. FUNCTIONAL TESTING

9.1 Experimental procedures for incorporation of modified nucleotides

General procedures for incorporation of d^{N3}CTPs and dd^{N3}CTPs (PEX)

Single azido-dCTP and azido-ddCTP incorporation with KOD XL DNA polymerase. For single nucleotide incorporation during primer extension reaction the primer C1 FAM-5'-CATGGGCGGCATGGG-3' and template C1 5'-CCC**G**CCCATGCCGCCCATG-3' were used. Quantity of components used in a single PEX reaction for d^{N3}CTPs incorporation is given in Table 28, for dd^{N3}CTPs incorporation – Table 29. All reactions were incubated for 30 min at 60 ℃. Reaction products were resolved on 12.5% TBE-Urea PAGE. Prior to loading on a gel, samples were mixed in a 1:1 ratio with 2× RNA loading buffer (80% formamide, 0.025 % [w/v] bromophenol blue, 0.025% [w/v] xylene cyanole FF, 0.025% SDS, 20 mM EDTA), heated at 95°C for 5 min and then immediately cooled on ice. Electrophoresis was carried out in 1× TBE buffer at 42 mA for 1 h at 55°C. Gels were imaged with Typhoon™ FLA 9500 system (GE Healthcare).

Multiple azido-dCTP incorporation with KOD XL DNA polymerase. For multiple nucleotide incorporation during primer extension reaction the primer C1 FAM-5'-CATGGGCGGCATGGG-3' and template C4 5'-CTA**G**CAT**G**A**G**CTCA**G**TCCCATGCCGCCCATG-3'were used. Quantity of components used in a single PEX reaction is given in Table 28. The following reactions procedures are exact as described in single azido-dCTP incorporation assay.

Table 28. Reaction components amount used for single and multiple d^{N3}NTP incorporation with KOD XL polymerase

Single d^{N3}CTP incorporation		**Multiple d^{N3}CTP incorporation**	
Reaction component	Quantity, µl	Reaction component	Quantity, µl
Primer C1 (3 µM)	1.5	Primer C1 (3 µM)	1.5
Temp C1 (3 µM)	2.25	Temp C4 (3 µM)	2.25
dGTP (2mM)	0.5	dATP, dGTP, dTTP mix (2 mM)	0.5
Buffer KOD 10x	2	Buffer KOD 10x	2
KOD XL 1:4	0.5	KOD XL 1:4	0.5
H_2O	12.25	H_2O	12.25
d^{N3}CTP (100 µM)	1	d^{N3}CTP (100 µM)	1

Table 29. Reaction components amount used for dd^{N3}NTP incorporation with KOD XL polymerase

dd^{N3}CTP incorporation	
Reaction component	Quantity, µl
Primer C1 (3 µM)	1.5
Temp C1 (3 µM)	2.25
Buffer KOD 10x	2
KOD XL 1:4	0.5
H_2O	12.25
dd^{N3}CTP (2 mM)	1

d^{N3}CTPs incorporation with Phusion (exo-) DNA polymerase

Primer extension reactions were performed in optimal buffers and optimal or near-optimal temperatures for tested polymerase. The primer Cy5-5'-TGCAGACATGGGTAGGCATCCTTGGCGTA-3' and template 3'-ACGTCTGTACCCATCCGTAGGAACCGCATAAAAGAAAA-5' were annealed in 1× annealing buffer (10 mM Tris-HCl pH 8.1 mM EDTA, 50 mM NaCl) to obtain 2 µM final solution. Amount of the components used per reaction are given in Table 30. All reactions were incubated for 1 min at 60 °C. Reaction products were resolved on 15% Urea PAGE. Prior to loading on a gel, samples were mixed in a 1:1 ratio with 2 × DNA loading buffer (98% formamide, 10 mg/mL blue dextran, 10 mM EDTA), heated at 95°C for 5 min and then immediately cooled on ice. Electrophoresis was carried out in 1× TBE buffer at 400 V for 1h at 55 °C. Gels were imaged with Typhoon™ FLA 9500 system (GE Healthcare).

Table 30. Reaction components amount used for d^{N3}NTP incorporation with Phusion (exo-) DNA polymerase

d^{N3}CTP incorporation	
Reaction component	Quantity, µl
Oligonucleotide duplex (2 µM)	1
5× Phire buffer	4
Phusion (exo-) (5U/µl)	0.5
H_2O	12.5
d^{N3}CTP (2 mM)	2

General procedures for incorporation of d^{N3}CTPs (PCR)

The PCR mixtures (10 µl) contained component given in Table 32 for each d^{N3}CTPs independently. The sequences of template, forward and reverse primers utilized in PCR are given in Table 31. Summary of tested PCR

conditions are given in Table 9 (optimal conditions – F). PCR products were purified by commercially available QIAquick® PCR purification kit and analyzed on a 1.3% agarose gel in 0.5× TBE buffer, followed by staining with GelRed.

Table 31.Sequences of primers and template used in PCR

$Primer^{FOR}$	5'-CGTCTTCAAGAATTCTATTTGACA-3'
$Primer^{REV}$	5'-GGAGAGCGTTCACCGACA-3'
$Template^{PCR}$	5'-CGTCTTCAAGAATTCTATTTGACAAAAATGGGCTCGTGTTGT ACAATAAATGTGTCTAAGCTTGGGTCCCACCTGACCCCATGCC GAACTCAGAAGTGAAACGCCGTAGCGCCGATGGTAGTGTGGG GTCTCCCCATGCGAGAGTAGGGAACTGCCAGGCATCAAATAAA ACGAAAGGCTCAGTCGAAAGACTGGGCCTTTCGTTTTATCTGTT GTTTGTCGGTGAACGCTCTCC-3'

Table 32. Reaction components amount used for $d^{N3}NTP$ incorporation during PCR with KOD XL polymerase

PCR with $d^{N3}CTP$	
Reaction component	Quantity, µl
$Primer^{FOR}$ (20 µM)	1
$Primer^{REV}$ (20 µM)	1
$Template^{PCR}$ (15 ng/µl)	1
dATP, dGTP, dTTP mix (2 mM)	1
Buffer KOD 10×	1
KOD XL 1:4	0.5
H_2O	3.5
$dd^{N3}CTP$ (2 mM)	1

General procedures for incorporation of different linkers and oligonucleotides bearing $d^{ON}CTPs$ with various polymerases (PEX)

Primer extension reactions were performed in optimal buffers and optimal or near-optimal temperatures for tested polymerase. The primer Cy5-5'-TGCAGACATGGGTAGGCATCCTTGGCGTA-3' and template 3'-ACGTCTGTACCCATCCGTAGGAACCGCATGACATCGACTCAACTC GCTG-5' were annealed in 1× annealing buffer (10 mM Tris-HCl pH 8.1 mM EDTA, 50 mM NaCl) to obtain 2 µM final solution. 2 pmol of oligonucleotide duplexes per reaction were used alongside with 20 pmol of corresponding $d^{ON}CTP$. Primer extension reactions were performed in optimal buffers and optimal or near-optimal temperatures for each tested polymerase. Tested polymerases and specific reaction conditions are given in. Reaction products were resolved on 15% TBE-Urea PAGE. Prior to loading on a gel, samples

were mixed in a 1:1 ratio with 2× DNA loading buffer (98% formamide, 10 mg/mL blue dextran, 10 mM EDTA), heated at 95 °C for 5 min and then immediately cooled on ice. Electrophoresis was carried out in 1× TBE buffer at 400 V for 1h at 55 °C. Gels were imaged with Typhoon™ FLA 9500 system (GE Healthcare).

Table 33. Polymerases used for incorporation ofd^{ON}CTPs

Family	Polymerase	Amount per reaction	Conditions
B	Phusion (exo-)	2.5 U	**95 °C 10s → 60 °C 10min**
A	Taq	2.5 U	**95 °C 10s → 60 °C 10min**

General procedures for incorporation of $dd^{N3\text{-}C2}$NTPs and $dd^{ON6\text{-}C2}$NTPs with various polymerases (PEX)

Oligonucleotide duplexes with different protruding ends were annealed in 1× annealing buffer (10 mM Tris-HCl pH 8.1 mM EDTA, 50 mM NaCl) to obtain 2 μM final solution. The template and primers sequences used are given in Table 23. Primers were labeled with a fluorescent dye (Cy5) for subsequent detection of PEX products on a gel. Primer extension reactions were performed in optimal buffers and optimal or near-optimal temperatures for each tested polymerase. Control reactions with native dNTPs and dd^{N3}NTPs were conducted to ensure that polymerase of interest is (i) capable to perform conventional primer extension at given conditions and (ii) is able to incorporate ddNTPs with small base modification. Tested polymerases and specific reaction conditions are given in Table 34. In all cases, 2 pmol of oligonucleotide duplex (or single-stranded primer for TdT) and 20 pmol of $dd^{ON6\text{-}C2}$NTP or corresponding native dNTP, or dd^{N3}NTP were used per reaction. Reaction products were resolved on 15% TBE-Urea PAGE. Prior to loading on a gel, samples were mixed in a 1:1 ratio with 2× DNA loading buffer (98% formamide, 10 mg/mL blue dextran, 10 mM EDTA), heated at 95 °C for 5min and then immediately cooled on ice. Electrophoresis was carried out in 1× TBE buffer at 400 V for 1h at 55 °C. Gels were imaged with Typhoon™ FLA 9500 system (GE Healthcare).

Table 34. Polymerases used for $dd^{N3\text{-}C2}$NTPs and $dd^{ON6\text{-}C2}$NTPs incorporation

Family	Polymerase	Amount per reaction	Conditions
A	Thermo Sequenase	40 U	95 °C 1min → 60 °C 30min
A	Klenow fragment (exo-)	5 U	37 °C 30min
B	Phusion (exo-)	20 U	95 °C 1min → 60 °C 30min

X	TdT	20 U	37 °C 30min
RT	SuperScript™ IV	200 U	50 °C 30min

General procedures for incorporation of different linker bearing dd^{ON6}CTPs with various polymerases (PEX)

Template for incorporation testing. The assay was based on the extension of a 3′ recessed end within an oligonucleotide duplex. Template oligonucleotide (5′-GGGGGGGGGGTACGCCAAGGATGCCTACCCATGTCTGCA-3′) and fluorescently labeled primer (5′-Cy5-TGCAGACATGGGTAGGCATCCTTGGCGTA-3′) were annealed in 1× annealing buffer (10 mM Tris-HCl pH 8, 1 mM EDTA, 50 mM NaCl) to obtain 2 μM final solution. Primer extension reactions were performed in optimal buffers and optimal or near-optimal temperatures for each tested polymerase. Control reactions with dCTP were conducted to ensure that polymerase of interest is capable to perform conventional primer extension at given conditions. Tested polymerases and specific reaction conditions are listed in Table 35. In all cases, 2 pmol of oligonucleotide duplex and 20 pmol of dd^{ON6}CTP or dCTP were used per reaction. Primer extension reactions were performed in optimal buffers and optimal or near-optimal temperatures for each tested polymerase. Reaction products were resolved on 15% TBE-Urea PAGE. Prior to loading on a gel, samples were mixed in a 1:1 ratio with 2× DNA loading buffer (98% formamide, 10 mg/mL blue dextran, 10 mM EDTA), heated at 95°C for 10 min and then immediately cooled on ice. Electrophoresis was carried out in 1× TBE buffer at 400 V for 1h at 55 °C. Gels were imaged with Typhoon™ FLA 9500 system (GE Healthcare).

Table 35. Polymerases used for different linker bearing dd^{ON6}CTPs incorporation

Family	Polymerase	Amount per reaction	Conditions
A	Thermo Sequenase	20 U	95 °C 1min → 60 °C 30min
B	Phusion (exo-)	20 U	95 °C 1min → 60 °C 30min
RT	SuperScript™ IV	200 U	50 °C 30min

9.2 Read-through unnatural linker efficiency measurements

Read-through different linkers bearing ON-d^{ON}Cs.

Templates for read-through testing were prepared by incorporation of different d^{ON}CTPs into oligonucleotide Cy5-5'-TGCAGACATGGGTAGGCATCCTTGGCGTA-3' (contains biotin attached

to the nucleobase in between the sequence) which was annealed with 3'-ACGTCTGTACCCATCCGTAGGAACCGCATGACATCGACTCAACTC GCTG-5'. Products were purification using streptavidin magnetic beads Primer extension reaction was then performed with 0.3 pmol of ON-d^{ON}Cs templates and 1.2 pmol of primer Cy3-5'-TGTTACATCTGGTAGTCAGTCTCC-3' using Phusion (exo-) DNA polymerase (2.5 U) in 5× Phire buffer in the presence of dNTP mixture. Primer extension reaction were performed maintaining conditions as follows: 95 °C 1 min → 60°C 1 min → 72°C 1 min (10 linear cycles were applied). Reaction products were resolved on 15% TBE-Urea PAGE. Prior to loading on a gel, samples were mixed in a 1:1 ratio with 2× DNA loading buffer (98% formamide, 10 mg/mL blue dextran, 10 mM EDTA), heated at 95 °C for 10min and then immediately cooled on ice. Electrophoresis was carried out in 1× TBE buffer at 400 V for 50min at 55 °C. Gels were imaged with Typhoon™ FLA 9500 system (GE Healthcare).

Read-through different linkers bearing ON-dd^{ON6}Cs.

Different linkers bearing dd^{ON6}CTPs were incorporated into oligonucleotide 5′-TGCAGACATGGGTAGGCATCCTTGGCGTA-3′ which was annealed with RNA oligonucleotide 5′-guacgccaaggaugccuacccaugucugca-3′ and subjected to PEX by SuperScript IV RT as described above. RNA strand was then digested by RNase H treatment, DNA was purified and target single-stranded ON-dd^{ON}C (**ON-dd$^{C2\text{-}ON}$C**, **ON-dd$^{C4\text{-}ON}$C**, **ON-dd$^{C6\text{-}ON}$C**, **ON-dd$^{PEG2\text{-}ON}$C**, **ON-dd$^{PEG4\text{-}ON}$C** in general **ON-dd^{ON}C**) products were purified by HPLC. Purification was performed applying reverse-phase chromatography (YMC-Actus Triart C18 column) using 100 mM TEAAc/ACN (10-16% in 18 minutes, gradient curve – 7) eluent system for gradient formation and desalted using water/ACN (DS2 method Table 24) gradient via analytical HPLC system. The obtained products molecular masses were approved and are given in Table 36.

Table 36. MS data of different linkers bearing read-through templates (ON-dd^{ON}Cs)

Read-through template	M (calcd), Da	M (found) [M], Da
ON-dd$^{ON6\text{-}C2}$C	16707.909	16707.936
ON-dd$^{ON6\text{-}C4}$C	16735.952	16735.953
ON-dd$^{ON6\text{-}C6}$C	16763.966	16763.981
ON-dd$^{ON6\text{-}PEG2}$C	16809.971	16809.987
ON-dd$^{ON6\text{-}PEG4}$C	16898.020	16898.048

Primer extension reactions were then performed with 2 pmol of primer (5′-Cy5-CAGACGTGTGCTCTTCC-3′) and 2.2 pmol of purified ON-dd^{ON}C templates in optimal buffers for each tested polymerase in the presence of dNTP mixture. Primer and template were pre-annealed before the addition of Klenow (exo-) and SuperScript IV enzymes. Tested polymerases and specific reaction conditions are listed in Table 37. Reaction products were resolved on 15% TBE-Urea PAGE and imaged as described above. Cy5-labelled oligonucleotide (5′-Cy5-TGCAGACATGGGTAGGCATCCTTGGCGTATAGATCGGAAGAGCACACGTCTG-3′) was used as positive control as its length corresponds to the length of expected read-through product. Read-through efficiency was evaluated by densitometry analysis using TotalLab Quant software.

Table 37. Polymerases tested for reading through ON-dd^{ON}C linkers

Family	Polymerase	Amount per reaction	Conditions
A	Thermo Sequenase	20 U	95 °C 1min → 60 °C 1min 72 °C 15min
A	Klenow fragment (exo-)	1 U	30°C 15min
B	Phusion (exo-)	20 U	95 °C 1min → 60 °C 1min 72 °C 15min
RT	SuperScript™ IV	200 U	50 °C 15min

9.3 Experimental procedures for simplified click reaction kinetics measurements

General procedures for SPAAC and CuAAC reactions simplified kinetics measurements with alkyne-oligonucleotide.

Azido-oligonucleotides were prepared during PEX reaction scale-up procedure in exact manner as described in "**Single azido-dCTP and azido-ddCTP incorporation with KOD XL DNA polymerase"** using KOD XL polymerase. The reactions scale was increased 250 times regarding described ones, the incubation was proceeded in thermostatic shaker for 40min at 60 °C and 500 rpm. Obtained products were purified using commercially available QIAquick® Nucleotide Removal Kit (250). The concentrations were determined using NanoDrop™ 2000 spectrophotometer and correspondingly adjusted if required for further SPAAC experiments. The reactions were performed using 70 pmol of corresponding incorporation product azido-oligonucleotide and 1-25 eq. of DBCO-oligonucleotide (structure is given in Scheme 29, oligonucleotide sequence: DBSO-5'-

UTTATATATTTATTGGAGACTGACTACCAGATGTAACA-3')
depending on the reaction. In some cases, spermine tetrahydrochloride was used (55 eq.) as solution in water (2.3 μM) or $MgCl_2$ (55 eq., 4.6 μM). The final reaction concentration was adjusted applying necessary amount of water. The reactions were proceeded in thermomixer (500 rpm) varying reactions temperature in range of 37 – 55 ℃ and time in 10 min – 24 h. Samples of reaction products were taken at the different periods of time and resolved on 12.5% TBE-Urea PAGE. Prior to loading on a gel, samples were mixed in a 1:1 ratio with 2× RNA loading buffer containing N_3-PEG1000 (1000 eq. excess). Electrophoresis was carried out in 1× TBE buffer at 42 mA for 1h at 55 °C. Gels were imaged with Typhoon™ FLA 9500 system (GE Healthcare). The product formation efficiency was evaluated by densitometry analysis using TotalLab Quant software.

CuAAC reactions were proceeded in exact manner as described above, the catalytic system of CuBr and TBTA (5 eq. to Cu) ligand prepared as solution in DMSO:*t*BuOH (3:1) was used alongside with alkyne-oligonucleotide (structure is given in Scheme 29, oligonucleotide sequence: alkyne-5'-UTTATATATTTATTGGAGACTGACTACCAGATGTAACA-3').

General procedures for SPAAC and CuAAC reactions simplified kinetics measurements with alkyne-group bearing fluorescent dyes.

The azido-group bearing oligonucleotides were obtained from primer extension reaction scale-up procedures during incorporation of diverse linker bearing d^{N3}CTPs using KOD XL polymerase (as described above, 250 times larger scale). For CuAAC experiments reactions were performed in DMSO:*t*BuOH 3:1 (6.5 μl) and water (10.5 μl) mixture. To the solution of TAMRA-alkyne (2 eq., 4 μl, 100 μM) a freshly prepared catalyst solution (4 μl) containing: CuBr (10 eq.) and TBTA (5 eq. to Cu) were applied, followed by addition of corresponding incorporation product PEX-d^{N3}C (200 pmol, 25 μl). Reactions were incubated at 37 ℃. After 15 min an aliquot of 25 μl was taken and quenched with 200 μL of PNI buffer (40% (v/v) 5M guanidinium chloride and 60% (v/v) isopropanol) and PEG-azide (2 μl, 10 mM). The rest of the reactions were quenched in the same way after 1 hour. The mixtures were purified with QIAQuick® Nucleotide Removal kit and eluted in 100 μL of milli-Q water. Emission spectra were recorded in a 100 μl quartz cuvette at 25°C. Excitation wavelength was 540 nm, emission window was 550 nm -700 nm. As a control the sample containing all reaction components except TAMRA-alkyne was used.

SPAAC reactions were performed in the exact manner as CuAAC only with absence of CuBr/TBTA catalyst. The reactions were carried out in water media using DBSO-TAMRA alkyne.

9.4 Experimental procedures for semi-targeted NGS library preparation of M13mp118 viral genome

Semi-targeted NGS library preparation approach was developed using M13mp18 bacteriophage single-stranded DNA (Thermo Fisher Scientific) as a sample input. Target-specific primers containing partial Illumina P5 adapter sequence at their 5′ termini were selected: M13-1: 5′-TACACGACGCTCTTCCGATCTAACGGTACGCCAGAATCTTG-3′ and M13-2: 5′-TACACGACGCTCTTCCGATCTAGAGCCACCACCGGAAC-3′. 0.125 pmol of each primer were mixed with 200 ng of M13mp18 DNA in a 20 µl reaction mixture containing 2 pmol of $dd^{ON6\text{-}C2}UTP$, 20 pmol of dNTP mix and 40 U of Thermo Sequenase with thermostable pyrophosphatase in 1× Thermo Sequenase reaction buffer. Primer extension was executed as follows: denaturation at 95 °C for 30s, followed by 15 cycles of denaturation at 95 °C for 30s, annealing/extension at 60 °C for 2min and final extension at 60 °C for 30min.

Half of the primer extension reaction was used directly for indexing PCR. 10 µl of primer extension reaction mixture was combined with 25 µl of Invitrogen Collibri Library Amplification Master Mix (Thermo Fisher Scientific), 20 U of Phusion exo- (1 µl), 5 µl of indexing primers P5 and P7 (50 pmol each) of the sequences given below and 9 µl of nuclease-free water. P5 primer: 5′-AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGAC GCTCTTCCGATCT-3′ and P7 primer: 5′-CAAGCAGAAGACGGCATACGAGAT[8nt index]GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT-3′.

Cycling was performed as follows: denaturation at 98 °C for 30s, followed by 20 cycles of denaturation at 98 °C for 10s, annealing at 60 °C for 30s, extension at 72 °C for 1min, and final extension at 72 °C for 1min. PCR products were then purified using Dynabeads™ Cleanup Beads (Thermo Fisher Scientific). DNA binding to the beads was performed by mixing 45 µl of bead suspension with 50 µl of sample and subsequent incubation at room temperature for 5 min. Sample was then placed on magnet, supernatant was removed, and beads were resuspended in 50 µl of elution buffer containing 10 mM Tris-HCl (pH 8). 50 µl of fresh beads were added again to the sample and binding was repeated. After room temperature incubation, sample was placed on magnet, supernatant was removed, and beads were washed twice with 85% ethanol. To elute libraries, beads were resuspended in 22 µl of elution buffer and incubated for 1min at room temperature. Fragment size distribution was then assessed by Agilent™ 2100 Bioanalyzer™ system

(Agilent Technologies) with High Sensitivity DNA Kit. Quantification of sequenceable molecules was performed with Invitrogen Collibri Library Quantification Kit (Thermo Fisher Scientific). Libraries were sequenced on the Illumina™ MiSeq™ instrument using the MiSeq Reagent Kit v2 (300-cycle) at 2 × 150 bp paired-end mode. [240]

The developed and above described semi-targeted sequencing approach was adapted for preparation of NGS libraries for other samples of interest. Regarding the investigated application of interest, the methodologies were adjusted independently and can be in details found in published articles: microbiome analysis [241], prostate cancer samples transcripts fusion (Drazdauskienė, U. *et.al.* manuscript under review), transcriptome-wide differential expression profiling [242].

CONCLUSIONS

1. The simplified, three steps synthetic strategy was developed and optimized for azido-group bearing 2'-deoxy- and 2',3'-dideoxynucleotides. During the optimization of monophosphorylation reaction condition the reactivity similarities were observed between (d)d^{I}C and (d)d^{I}A nucleosides, as well as dd^{I}U and dd^{I}G. High conversions to (d)d^{I}CMP and (d)d^{I}AMP were obtained when $P_2O_3Cl_4$ was applied. For the synthesis of dd^{I}UMP and dd^{I}GMP, $POCl_3$ was more suitable phosphorylation agent while utility of base resulted in higher conversions to the products. The aqueous-phase Sonogashira cross-coupling reactions between propargylamine and dd^{I}NTPs were successfully performed applying mild reaction conditions. It was observed that for efficient synthesis 40 °C reaction temperature was required (except for dd^{I}ATP).

2. The five and three steps strategies for nucleobase modified nucleotides synthesis were compared. During the synthesis of 5-(3-aminoprop-1-ynyl)-2'-deoxycytidine 5'-triphosphate it was shown that three steps synthesis approach is more efficient, nearly doubling overall yield of the target material. The developed alternative synthesis strategy for nucleobase-modified 2',3'-dideoxynucleotides proved to be robust and rapid, allowing 10 – 20 times scale-up with only minor modifications.

3. Azido-group bearing 2'-deoxy- and 2',3'-dideoxynucleotides designed in this work were substrates for DNA polymerases, namely KOD XL, Taq, Phusion exo-, Klenow exo-, Thermo Sequenase, TdT and SuperScript IV, during primer extension and/or polymerase chain reactions.

4. The relationship between CuAAC and SPAAC efficiency and nature of substrates was investigated. Reactions between azido-group bearing oligonucleotides obtained from incorporation of diverse 2'-deoxycytidine 5'-triphosphates and alkynes labeled by oligonucleotide or TAMRA dye were performed. It was found, that CuAAC and SPAAC reactions rates increase with prolongation of linker connecting azido-group with nucleobase (except for N3-PEG4).

5. Series of oligonucleotide-tethered 2'-deoxy- and 2',3'-dideoxynucleotides were obtained during CuAAC reaction. OTDNs and OTDDNs designed in this work were substrates for polymerases Taq, Phusion exo-, Klenow exo-, Thermo Sequenase, TdT and SuperScript IV. Different linkers bearing single-stranded DNA fragments with incorporated oligonucleotide-tethered 2'-deoxycytidine 5'-triphosphates were investigated as templates during enzymatic synthesis of complementary strand. The model experiment showed that Phusion exo- polymerase can read-through all

unnatural linkages. However, the NGS library preparation experiment resulted in complex reaction composition and indicated inapplicability of OTDNs.

6. It was shown that oligonucleotide-tethered 2',3'-dideoxynucleotides simplifies NGS library preparation workflow in the model experiment of the M13mp18 viral genome sequencing. The nucleobase-to-ribose conjugation enabled Phusion exo- to read-through unnatural linkage with fidelity. Moreover, proposed workflow is applicable for DNA and cDNA library preparation of diverse samples and applications: microbiome analysis, determination of prostate cancer RNA transcripts' fusion events and gene expression analysis by 3'-mRNA termini profiling.

7. The efficiency of read-through unnatural linkers was investigated. It was determined that all diverse linkers, connecting nucleobase-to-oligonucleotide, were biocompatible. Tested polymerases SuperScript IV, Phusion exo-, Thermo Sequenase and Klenow exo- were capable of reading-through the artificial backbone, with the highest single cycle read-through efficiency of 75% by Phusion exo-.

SCIENTIFIC PARTICIPATION

Patent applications

- A. Lubys, I. Čikotienė, Ž. Kapustina, A. Berezniakovas, **J. Medžiūnė**, S. Žeimytė. Tethered Oligos And Uses Thereof. International Application No. PCT/US2020/039009.

Publications

- Kapustina, Ž., **Medžiūnė**, **J**., Alzbutas, G., Rokaitis, I., Matjošaitis, K., Mackevičius, G., Žeimytė, S., Karpus, L., Lubys, A., High-resolution microbiome analysis enabled by linking of 16S rRNA gene sequences with adjacent genomic contexts. *Microb Genom*., 2021a. doi: 10.1099/mgen.0.000624. Accepted manuscript.
- **Medžiūnė**, **J**., Kapustina, Ž., Žeimytė, S., Jakubovska, J., Sindikevičienė, R., Čikotienė, I., Lubys, A., Advanced preparation of fragment libraries enabled by oligonucleotide-modified 2′,3′-dideoxynucleotides. *Commun Chem*, 2022, 5 (34). doi: 10.1038/s42004-022-00649-9. Accepted manuscript.
- Kapustina, Ž., **Medžiūnė**, **J**., Dubovskaja, V., Matjošaitis, K., Žeimytė, S., Lubys, A., Sensitive and accurate analysis of gene expression signatures enabled by oligonucleotide-labelled cDNA *RNA Biology*, 2022. doi: 10.1080/15476286.2022.2078093. Accepted manuscript.
- **Medžiūnė**, **J**., Kapustina, Ž., Žeimytė, S., Lubys, A., Čikotienė, I. Oligonucleotide-tethered 2‘,3‘-dideoxynucleotides: scalable and efficient synthesis strategy for preparation of functionalized 2‘,3‘-dideoxynucleotides. Manuscript in review.
- Drazdauskienė, U., Kapustina, Ž., **Medžiūnė**, **J.**, Dubovskaja, V., Sabaliauskaitė, R., Jarmalaitė, S., Lubys, A., Fusion sequencing via terminator-assisted synthesis (FTAS-seq) reveals novel *TMPRSS2* fusion patterns in prostate cancer. Manuscript in review.

Poster presentations

- **Šulgaitė**, **J.**, Taujanskaitė, O., Berezniakovas, A., Čikotienė, I., Šeputienė, V., Lubys, A. Synthesis of functionalized dNTPs for enzymatic applications. Balticum Organicum Syntheticum 2018, Tallinn, Estonia, 2018. **Best poster award**.
- Kapustina, Ž., **Šulgaitė**, **J**., Žeimytė, S., Lubys A. Innovative approach for digital gene expression and alternative polyadenylation profiling. European Human Genetics Virtual Conference, ESHG 2020.
- Kapustina, Ž., **Medžiūnė**, **J.**, Žeimytė, S., Liucvaikytė, A., Lubys A. Innovative approach for digital gene expression and alternative polyadenylation profiling. American Society of Human Genetics Virtual Meeting, ASHG 2020.
- Kapustina, Ž., Alzbutas G, **Šulgaitė**, **J.**, Žeimytė, S., Lubys A. High-precision characterization of microbial communities by the analysis of 16S rRNA gene genomic context. Cold Spring Harbor Laboratory Microbiome Virtual Meeting, 2020.
- Kapustina, Ž., **Medžiūnė**, **J**., Drazdauskienė, U., Sindikevičienė, R., Dubovskaja, V., Lubys A. Exploratory analysis of genetic rearrangements. 2021 Advances in Genome Biology and Technology Virtual General Meeting, AGBT 2021.
- Jakubovska, J., Sindikevičienė, R., **Medžiunė**, **J.**, Kerzhner, M., Kuhn, P., Kapustina, Ž., Lubys A. Oligonucleotide-modified terminators for high-throughput single-stranded DNA sequencing. Revolutionizing Next-Generation Sequencing (Virtual 4th edition), VIB Tools and Technologies conference 2021.
- Drazdauskienė, U., Kapustina, Ž., **Medžiūnė**, **J.**, Gasiunienė, M., Sindikevičienė, R., Dubovskaja, V., Sabaliauskaitė, R., Lubys, A. Semi-targeted sequencing of fusion transcripts in prostate cancer. European Human Genetics Virtual Conference 2021.
- **Medžiūnė**, **J.**, Kapustina, Ž., Žeimytė, S., Jakubovska, J., Sindikevičienė, R., Čikotienė, I., Lubys, A. Introduction of universal priming sites to nucleic acids via oligonucleotide-tethered 2',3'-dideoxynucleotide terminators. Life Sciences Baltics, Virtual Meeting 2021.
- Kapustina, Ž., **Medžiūnė**, **J**., Alzbutas, G., Rokaitis, I., Matjošaitis, K., Mackevičius, G., Žeimytė, S., Karpus, L., Lubys, A. High-precision characterization of microbial communities by

the analysis of 16S rRNA gene genomic context. Life Sciences Baltics, Virtual Meeting 2021.

- **Medžiūnė**, **J.**, Kapustina, Ž., Žeimytė, S., Jakubovska, J., Sindikevičienė, R., Čikotienė, I., Lubys, A. Synthesis of oligonucleotide-tethered nucleotides for simplified NGS library preparation. SCANC – 17th Symposium on Chemistry of Nucleic Acid Components, Česky Krumlov, Czech Republic, 2022. **Best poster award**.
- **Medžiūnė**, **J.**, Kapustina, Ž., Žeimytė, S., Jakubovska, J., Sindikevičienė, R., Čikotienė, I., Lubys, Advanced preparation of fragment libraries enabled by oligonucleotide-tethered 2‘,3‘-dideoxynucleotides. Balticum Organicum Syntheticum 2022, Vilnius, Lithuania, 2022. **Best poster award**.

CURRICULUM VITAE

Name, Surname	Justina Medžiūnė
Date of birth	5th October, 1992
Phone number	+37060281469
Email address	justina.medziune@thermofisher.com justina.sulgaite@gmail.com
Education	
2017 – 2021	Doctoral studies (Chemistry, N 003) Faculty of Chemistry and Geosciences, Vilnius University
2015 – 2017	MSc Chemistry, Faculty of Chemistry and Geosciences, Vilnius University
2011 – 2015	BSc Chemistry, Faculty of Chemistry and Geosciences, Vilnius University
1999 – 2011	Elektrėnai Versmės Gymnasium (first-class certificate)
2009 – 2011	Elektrėnai musical school (guitar)
2004 – 2009	Elektrėnai musical school (piano) (first-class certificate)
Training courses, secondments, and other activities	
2020	"Manager's start" by ISM Executive School, Vilnius, Lithuania
2019 01 – 04	Institute of Organic Chemistry and Biochemistry, Academy of Sciences of the Czech Republic, Michal Hocek group
2016 08 25 – 26	International Electrosynthesis Workshop", Ryga, Latvia, Latvian Institute of Organic Synthesis
2014 – 2015	Lithuanian Science Council project "Student's scientific activities promotion"
2009 – 2010	"Certificates of recognition presented for outstanding performance in Deliberating in a Democracy", it is an international project that facilitates by constitutional rights foundation LA, constitutional rights foundation – and street Law Yns;
2008 – 2009	Certificate in participation in „UNESCO World Heritage sites" project
Work experience	
Since 2022	Staff Scientist at Thermo Fisher Scientific Baltics, UAB
2021 - 2022	Scientist III at Thermo Fisher Scientific Baltics, UAB
2019 – 2021	Scientist II at Thermo Fisher Scientific Baltics, UAB

2017 – 2019	Junior Scientist at Thermo Fisher Scientific Baltics, UAB
2013 – 2014	Sales manager at “Studio Moderna”, UAB
2010 01 – 06	Data analyst at “Reservation partner LT”, UAB
Teaching experience	
2019 – 2021	Supervision of MSc final theses

ACKNOWLEDGEMENTS

I would like to sincerely thank my supervisor prof. dr. Inga Čikotienė for useful advises, patience, support, opportunity to perform doctoral research and start my, as a chemist, carrier. For welcoming me in her laboratory as a bachelor student and leading me all the way, for encouraging me to seek for more and strive for constant self-development.

I am grateful to conceiver of this research idea prof. dr. Arvydas Lubys for his valuable advices, insights, reliance on me and giving the ability to develop this conceptual idea to the matter.

I am grateful to my precious colleague dr. Žana Kapustina for her professionalism and endless creativity in development of the applications of oligonucleotide-tethered 2'-3'-dideoxynucleotides. I am sincerely grateful to you for all your advices, collaboration on joint articles preparation, patience, and huge support during all this time. You are my rock.

I sincerely appreciate Simona Žeimytė collaboration on chemical synthesis of oligonucleotide-tethered 2'-3'-dideoxynucleotides. Thank you for your hard work, precision, all moral support, understanding, carrying, and being friend.

Warmest thanks to Renaldas Rimkus, dr. Aurelija Zlatkienė, dr. Lukas Taujenis and dr. Vytautas Tamošiūnas for being my teachers, helping me grow, sharing your scientific and life experiences. Thank you all for being the authorities I wanted to align with! I also want to thank all OSL group members for all their support. Thanks to my former student Ugnė Šinkevičiūtė for giving me the honor to lead you during your master thesis, for delighting me with your achievements and enabling me to grow not just as scientist but as a mentor and person. I am grateful for your honest carrying and support.

I thank Artūras Berezniakovas and Odeta Taujanskaitė for their contribution and hard work with oligonucleotide-tethered 2'-deoxynucleotides.

I have to express my heartily thanks to Advance Technologies group members, you proved to be the best colleagues despite the short time that we are working together. I am thankful for your honest support, carrying, encouragements and amazing sarcasm that makes me smile no matter what!

I also want to express my gratitude to 113 laboratory members where I had the honor to start my chemist's carrier, I could not wish for more moral support, carrying and love as I was given by you all! Thank you, Linas Labanauskas, Rita Sadzevičienė and colleagues from Institute of Chemistry for, at the beginning of my career, showing me the real chemistry and embracing my confidence as a scientist.

I am grateful to prof. dr. Michal Hocek and Alessandro Panattoni for welcoming me at the Institute of Organic Chemistry and Biochemistry, for teaching me the fundamentals of biochemistry, contributing to my research.

Finally, my deepest gratitude to my husband, my spiritual supporter, my motivator, and believer. Your love, patience and understanding were my rock during all these years. I am also thankful to my parents and my family for believing in me, for being always by my side, all your encouragements and love helped me to fulfill this huge goal of mine.

REFERENCES

[1] P. J. McCown, A. Ruszkowska, C. N. Kunkler, K. Breger, J. P. Hulewicz, M. C. Wang, N. A. Springer and J. A. Brown, "Naturally occurring modified ribonucleosides," *WIREs RNA,* p. e1595, 2020.

[2] S. Gad, A. Aurias, N. Puget, A. Mairal, C. Schurra, M. Montagna, S. Pages, V. Caux, S. Mazoyer, A. Bensimon and D. Stoppa-Lyonnet, "Color bar coding the BRCA1 gene on combed DNA: A useful strategy for detecting large gene rearrangements," *Genes Chromosom. Cancer,* vol. 31, pp. 75-84, 2021.

[3] A. D. Keefe and S. T. Cload, "SELEX with modified nucleotides," *Curr Opin Chem Biol,* vol. 12, pp. 448-456, 2008.

[4] M. L. Metzker, "Sequencing technologies — the next generation," *Nat. Rev.,* vol. 11, p. 31–46 , 2010.

[5] Y. Zhang, Y. Gao, X. Wena and H. Ma, "Current prodrug strategies for improving oral absorption of nucleoside analogues," *Asian J. Pharm. Sci.,* vol. 9, no. 2, p. 65−74, 2014.

[6] C. Pollard, S. D. Koker, X. Saelens, G. Vanham and J. Grooten, "Challenges and advances towards the rational design of mRNA vaccines," *Trends Mol. Med.,* vol. 19, no. 12, p. 705–713, 2013.

[7] E. D. Clercq, "The history of antiretrovirals: key discoveries over the past 25 years," *Rev. Med. Virol.,* vol. 19, p. 287–299, 2009.

[8] D. E. Chapman and G. Powis, "Disposition and metabolism in mice of the potential antitumor and anti-human immunodeficiency virus-1 agent, 2-chloro-2',3'-dideoxyadenosine," *Cancer Chemother. Pharmacol.,* vol. 27, pp. 285-289, 1991.

[9] U. Pradere, E. C. Garnier-Amblard, S. J. Coats and F. Amblard, "Synthesis of Nucleoside Phosphate and Phosphonate Prodrugs," *Chem. Rev.,* vol. 114, pp. 9154-9218, 2014.

[10] A. D. Barone, C. Chen, G. H. McGall, K. Rafii, P. R. Buzby and J. J. Dimeo, "Novel Nucleoside Triphosphate Analogs For The Enzymatic Labeling Of Nucleic Acids," *Nucleosides, Nucleotides and Nucleic Acids,* vol. 20, pp. 1141-1145, 2001.

[11] P. J. Merrill, D. R. Johan, R. C. William, H. F. Worden and T. G. Leonard, "Process and reagents for DNA sequence analysis". Europe Patent EP0252683B1, 13 01 1988.

[12] A. Shivalingam, A. E. S. Tyburn, A. H. El-Sagheer and T. Brown, "Molecular Requirements of High-Fidelity Replication-Competent DNA Backbones for Orthogonal Chemical Ligation," *J. Am. Chem. Soc.,* vol. 139, pp. 1575-1583, 2017.

[13] A. H. El-Sagheer and T. Brown, "Click Nucleic Acid Ligation: Applications in Biology and Nanotechnology," *Acc. Chem. Res.,* vol. 45, no. 8, p. 1258–1267, 2012.

[14] A. H. El-Sagheera, A. P. Sanzonea, R. Gaoa, A. Tavassolia and T. Browna, "Biocompatible artificial DNA linker that is read through by DNA polymerases and is functional in Escherichia coli," *Proc. Natl. Acad. Sci. U.S.A,* vol. 108, no. 28, pp. 11338-11343, 2011.

[15] F. Miura, T. Fujino, K. Kogashi, Y. Shibata, M. Miura, H. Isobe and T. Ito, "Triazole linking for preparation of a next-generation sequencing library from single-stranded DNA," *Nucleic Acids Res.,* vol. 46, no. 16, p. e95, 2018.

[16] A. Routh, S. R. Head, P. Ordoukhanian and J. E. Johnson, "ClickSeq: Fragmentation-free next-generation sequencing via click-ligation of adaptors to stochastically terminated 3’-azido cDNAs," *J. Mol. Biol.,* vol. 427, no. 16, pp. 2610-2616, 2015.

[17] L. H. Thoresen, G. S. Jiao, W. C. Haaland, M. L. Metzker and K. Burgess, "Rigid, Conjugated, Fluoresceinated Thymidine Triphosphates: Syntheses and Polymerase Mediated Incorporation into DNA Analogues," *Chem. Eur. J.,* vol. 9, pp. 4603-4610, 2003.

[18] M. Hocek and M. Fojta, "Cross-coupling reactions of nucleoside triphosphates followed by polymerase incorporation. Construction and applications of base-functionalized nucleic acids," *Org. Biomol. Chem.,* vol. 6, pp. 2233-2241, 2008.

[19] K. Lohmann, "The pyrophosphate fraction in the muscle," *Naturwiss,* vol. 17, p. 624–625, 1929.

[20] C. H. Fiske and Y. Subbarow, "Phosphorus compounds of muscle and liver," *Science,* vol. 70, p. 381–382, 1929.

[21] B. Roy, A. Depaix, C. Perigaud and S. Peyrottes, "Recent Trends in Nucleotide Synthesis," *Chem. Rev.,* vol. 116, p. 7854–7897, 2016.

[22] M. Guinan, C. Benckendorff, M. Smith and G. J. Miller, "Recent Advances in the Chemical Synthesis and Evaluation of Anticancer Nucleoside Analogues," *Molecules,* vol. 25, no. 9, p. 1–25, 2020.

[23] S. Mahmoud, S. Hasabelnaby, S. F. Hammad and T. M. Sakr, "Antiviral Nucleoside and Nucleotide Analogs: A Review," *J. Adv. Pharm. Res.,* vol. 2, no. 2, p. 73–88, 2018.

[24] R. Kole, A. R. Krainer and S. Altman, "RNA therapeutics: beyond RNA interference and antisense oligonucleotides," *Nat. Rev.,* vol. 11, p. 125–140, 2012.

[25] T. Yamamoto, M. Nakatani, K. Narukawa and S. Obika, "Antisense drug discovery and development," *Future Med. Chem.,* vol. 3, no. 3, p. 339−365, 2011.

[26] M. M. Zhang, R. Bahal, T. P.Rasmussen, J. E. Manautou and X.-b. Zhong, "The growth of siRNA-based therapeutics: Updated clinical studies," *Biochem. Pharmacol.,* vol. 189, pp. 1-12, 2021.

[27] C. Cui, W. Shu and P. Li, "Fluorescence In situ Hybridization: Cell-Based Genetic Diagnostic and Research Applications," *Front. Cell Dev. Biol.,* vol. 4, no. 89, pp. 1-11, 2016.

[28] T. J. Aitman, "DNA microarrays in medical practice," *BMJ,* vol. 323, no. 7313, p. 611–615, 2001.

[29] J. M. Heather and B. Chain, "The sequence of sequencers: The history of sequencing DNA," *Genomics,* vol. 107, no. 1, p. 1–8, 2016.

[30] M. Hocek, "Synthesis of Base-Modified 2′-Deoxyribonucleoside Triphosphates and Their Use in Enzymatic Synthesis of Modified DNA for Applications in Bioanalysis and Chemical Biology," *J. Org. Chem.,* vol. 79, no. 21, p. 9914–9921, 2014.

[31] K. Burgess and D. Cook, "Syntheses of Nucleoside Triphosphates," *Chem. Rev.,* vol. 100, pp. 2047-2059, 2000.

[32] L. Weinschenk and C. Meier, "Chemical Syntheses of Nucleoside Triphosphates," in *Chemical Synthesis of Nucleoside Analogues*, John Wiley & Sons, Inc., 2013, p. 209–227.

[33] J. .. G. Moffatt, "A General Synthesis of Nucleoside-5' Triphosphates," *Can. J. Chem.,* vol. 42, pp. 599-604, 1964.

[34] D. E. Hoard and D. G. Ott, "Conversion of Mono- and Oligodeoxyribonucleotides to 5’-Triphosphates," *J. Am. Chem. Soc.,* vol. 87, no. 8, pp. 1785-1788, 1965.

[35] J. Liudeig, "A New Route to Nucleoside 5'-triphosphates," *Acta Biochim. Biophys. Sin.,* vol. 16, pp. 131-133, 1981.

[36] V. J. Davisson, D. R. Davis, V. M. Dixit and C. D. Poulter, "Synthesis of Nucleotide 5’-Diphosphates from 5’- 0 -Tosyl Nucleosides," *J. Am. Chem. Soc.,* vol. 52, pp. 1794-1801, 1987.

[37] M. Yoshikawa, T. Kato and T. Takenishi, "A Novel Method For Phosphorylation of Nucleosides to 5'-Nucleotides," *Tetrahedron Lett.,* vol. 50, pp. 5065-5068, 1967.

[38] A. El-Tayeb, A. Qi and C. E. Muller, "Synthesis and Structure-Activity Relationships of Uracil Nucleotide Derivatives and Analogues as Agonists at Human P2Y2, P2Y4, and P2Y6 Receptors," *J. Med. Chem.,* vol. 49, pp. 7076-7087, 2006.

[39] C. Crauste, C. Perigaud and S. Peyrottes, "Synthesis of 2',3'-Dideoxynucleoside Phosphoesters Using H-Phosphonate Chemistry on Soluble Polymer Support," *J. Org. Chem.,* vol. 76, pp. 997-1000, 2011.

[40] T. Ikemoto, A. Haze, H. Hatano, Y. Kitamoto, M. Ishida and K. Nara, "Phosphorylation of Nucleotides with Phosphorus Oxychloride in Trialkyl Phosphate," *Chem. Pharm. Bull.,* vol. 43, no. 2, pp. 210-215, 1995.

[41] M. Yoshikawa and T. K. a. T. Takenishi, "Studies of Phosphorylation. III. Selective Phosphorylation of Unprotected Nucleosides," *Bull. Chem. Soc. Jpn. ,* vol. 42, pp. 3505-3508, 1969.

[42] K. I. Imai, S. Fujii, K. Takanohashi, Y. Furukawa, T. Masuda and M. Honjo, "Studies on Phosphorylation. IV. Selective Phosphorylation of the Primary Hydroxyl Group in Nucleosides," *J. Org. Chem.,* vol. 34, no. 6, pp. 1547-1550, 1969.

[43] A. Simoncsits and J. Tomasz, "A Simple One-Step Synthesis of Ribonucleoside 2',3'-Cyclic Phosphate 5'-Phosphates," *Biochim. Biophys. Acta,* vol. 395, pp. 74-79, 1975.

[44] J. Ye, X. Liao and J. A. Piccirilli, "Synthesis of 2‘-C-Difluoromethylribonucleosides and Their Enzymatic Incorporation into Oligonucleotides," *J. Org. Chem.,* vol. 70, no. 20, p. 7902–7910, 2005.

[45] A. V. Manuilov, S. S. Hixson and R. A. Zimmermann, "New photoreactive tRNA derivatives for probing the peptidyl transferase center of the ribosome," *RNA*, vol. 13, no. 5, pp. 793-800, 2007.

[46] T. Kovacs and L. Otvos, "Simple Synthesis of 5-vinyl- and 5-ethynyl-2'-deoxyuridine-5'-triphosphates," *Tetrahedron Lett.,* vol. 29, no. 36, pp. 4525-4528, 1988.

[47] F. Seela, H. P. Mnth and A. Roling, "Syntheses of Pyrrolo(2,3-d]pyrimidine 2',3'-Dideoxyribonucleosides Related to 2',3'-Dideoxyadenosine and 2',3'-Dideoxyguanosine and Inhibitory Activity of 5'-Triphosphates on HIV-1 Reverse Transcriptase," *Helv. Chim. Acta,* vol. 74, pp. 554-564, 1991.

[48] I. Gillerman and B. Fischer, "An Improved One-Pot Synthesis of Nucleoside 5′-Triphosphate Analogues," *Nucleosides, Nucleotides and Nucleic,* vol. 29, pp. 245-256, 2010.

[49] G. Jones and S. P. Stanforth, "The Vilsmeier Reaction of Fully Conjugated Carbocycles and Heterocycles," in *Organic Reactions*, vol. 49, England, John Wiley and Sons, Inc., 1997, pp. 1-67.

[50] G. M. Tener, "2-Cyanoethyl Phosphate and its Use in the Synthesis of Phosphate Esters," *J. Am. Chem. Soc.,* vol. 83, no. 1, pp. 159-168, 1961.

[51] A. M. Michelson and A. R. Todd, "Nucleotides. Part XXIII. Mononucleotides derived from deoxycytidine. Note on the structure of cytidylic acids a and b," *J. Chem. Soc.*, pp. 34-40, 1954.

[52] C. Zhu, C. Tang, Z. Cao, W. He, Y. Chen, X. Chen, K. Guo and H. Ying, "Fully Automated Continuous Meso-flow Synthesis of 5′-Nucleotides and Deoxynucleotides," *Org. Process Res. Dev.,* vol. 18, no. 11, pp. 1575-1581, 2014.

[53] M. Kuwahara, J. Nagashima, M. Hasegawa, T. Tamura, R. Kitagata, K. Hanawa, S. Hososhima, T. Kasamatsu, H. Ozaki and H. Sawai, "Systematic characterization of 2′-deoxynucleoside- 5′-triphosphate analogs as substrates for DNA polymerases by polymerase chain reaction and kinetic studies on enzymatic production of modified DNA," *Nucleic Acids Res.,* vol. 34, no. 19, pp. 5383-5394, 2006.

[54] A. Colliera and G. Wagner, "A facile two-step synthesis of 8-arylated guanosine mono- and triphosphates (8-aryl GXPs)," *Org. Biomol. Chem.,* vol. 4, pp. 4526-4532, 2006.

[55] S. Jager, G. Rasched, H. Kornreich-Leshem, M. Engeser, O. Thum and M. Famulok, "A Versatile Toolbox for Variable DNA Functionalization at High Density," *J. Am. Chem. Soc.,* vol. 127, no. 43, p. 15071–15082, 2005.

[56] J. P. Peters, S. P. Yelgaonkar, S. Srivatsan, Y. Tor and L. J. Maher, "Mechanical properties of DNA-like polymers," *Nucleic Acids Res.,* vol. 41, no. 22, p. 10593–10604, 2013.

[57] A. R. Kore, A. Senthilvelan, B. Srinivasan and M. Shanmugasundaram, "Facile protection-free one-pot synthesis of 7-deaza-2′-deoxyguanosine-5′-triphosphate - A versatile molecular biology probe," *Can. J. Chem.,* vol. 91, no. 8, pp. 718-720, 2013.

[58] A. Krause, A. Hertl, F. Muttach and A. Jaschke, "Phosphine-Free Stille–Migita Chemistry for the Mild and Orthogonal Modification of DNA and RNA," *Chem. Eur. J.,* vol. 20, no. 50, pp. 16613-16619, 2014.

[59] H. Rao, A. A. Sawant, A. A. Tanpure and S. G. Srivatsan, "Posttranscriptional chemical functionalization of azide-modified oligoribonucleotides by bioorthogonal click and Staudinger reactions," *Chem. Commun.,* vol. 48, pp. 498-500, 20112.

[60] A. Panattoni, R. Pohl and M. Hocek, "Flexible Alkyne-Linked Thymidine Phosphoramidites and Triphosphates for Chemical or Polymerase Synthesis and Fast Postsynthetic DNA Functionalization through Copper-Catalyzed Alkyne–Azide 1,3-Dipolar Cycloaddition," *Org. Lett.,* vol. 20, no. 13, p. 3962–3965, 2018.

[61] N. Milisavljevic, P. Perlikova, R. Pohl and M. Hocek, "Enzymatic synthesis of base-modified RNA by T7 RNA polymerase. A systematic study and comparison of 5-substituted pyrimidine and 7-substituted 7-deazapurine nucleoside triphosphates as substrates," *Org. Biomol. Chem.,* vol. 16, pp. 5800-5807, 2018.

[62] S. Wolf, T. Zismann, N. Lunau, S. Warnecke, S. Wendicke and C. Meier, "A convenient synthesis of nucleoside diphosphate glycopyranoses and other polyphosphorylated bioconjugates," *Eur. J. Cell Biol,* vol. 89, pp. 63-75, 2010.

[63] J. Ludwig and F. Eckstein, "Rapid and Efficient Synthesis of Nucleoside 5'-O -(1-Thiotriphosphates), 5'-Triphosphates and 2',3'-Cyclophosphorothioates Using 2-Chloro-4H- 1,3,2-

benzodioxaphosphorin-4-one," *J. Org. Chem.*, vol. 54, pp. 631-635, 1989.

[64] J. Caton-Williams, L. Lin, M. Smith and Z. Huang, "Convenient synthesis of nucleoside 5'-triphosphates for RNA transcription," *Chem. Commun.*, vol. 47, pp. 8142-8144, 2011.

[65] J. Caton-Williams, M. Smith, N. Carrasco and Z. Huang, "Protection-Free One-Pot Synthesis of 2'-Deoxynucleoside 5'-Triphosphates and DNA Polymerization," *Org. Lett.*, vol. 13, no. 16, pp. 4156-4159, 2011.

[66] M. Mahler, B. Reichardt, P. Hartjen, J. v. Lunzen and C. Meier, "Stereoselective Synthesis of D- and L-Carbocyclic Nucleosides by Enzymatically Catalyzed Kinetic Resolution," *Chem. Eur. J.*, vol. 18, no. 35, pp. 11046-11062, 2012.

[67] M. F. Matsuura, R. W. Shaw, J. D. Moses, H. J. Kim, M. J. Kim, M. S. Kim, S. Hoshika, N.Karalkar and S. A. Benner, "Assays To Detect the Formation of Triphosphates of Unnatural Nucleotides: Application to Escherichia coli Nucleoside Diphosphate Kinase," *ACS Synth. Biol.*, vol. 5, no. 3, pp. 234-240, 2016.

[68] J. Caton-Williams, R. Hoxhaj, B. Fiaz and Z. Huang, "Use of a novel 5'-regioselective phosphitylating reagent for one-pot synthesis of nucleoside 5'-triphosphates from unprotected nucleosides," *Curr Protoc Nucleic Acid Chem*, pp. 1-21, 2013.

[69] N. A. Boyle, P. Fagan, J. L. Brooks, M. Prhavc and J. Lambert, "2′,3′-Dideoxynucleoside 5′-β,γ-(Difluoromethylene) Triphosphates With α-P-Thio or α-P-Seleno Modifications: Synthesis and Their Inhibition of HIV-1 Reverse Transcriptase," *Nucleosides, Nucleotides, and Nucleic Acids*, vol. 24, pp. 1651-1664, 2005.

[70] W. R. Chambers and H. G. Khorana, "Nucleoside Polyphosphates. V. Syntheses of Guanosine 5'-Di- and Triphosphates," *J. Am. Chem. Soc.*, vol. 79, pp. 3752-3755, 1957.

[71] G. Weimann and H. G. Khorana, "Studies on Polynucleotides. XVII. On the Mechanism of Internucleotide Bond Synthesis by Carbodiimide Method," *J. Am. Chem. Soc.*, vol. 84, no. 22, pp. 4329-4341, 1962.

[72] J. G. Moffat and H. G. Khorana, "Nucleoside Polyphosphates. X. The Synthesis and Some Reactions of Nucleoside-5’

Phosphoromorpholidates and Related Compounds. Improved Methods for the Preparation of Nucleoside-5’ Polyphosphates," *J. Am. Chem. Soc.,* vol. 83, no. 3, pp. 649-658, 1961.

[73] Z. Xu, "A review on the chemical synthesis of pyrophosphate bonds in bioactive nucleoside diphosphate analogs," *Bioorganic Med. Chem. Lett.,* vol. 25, no. 18, pp. 3777-3783, 2015.

[74] D. E. Hoard and D. G. Ott, "Conversion of Mono- and Oligodeoxyribonucleotides to 5’-Triphosphates," *J. Am. Chem. Soc.,* vol. 87, no. 8, pp. 1785-1788, 1965.

[75] M. Maeda, A. D. Patel and A. Hampton, "Formation of ribonucleotide 2', 3'-cyclic carbonates during conversion of ribonucleoside 5'-phosphates to diphosphates and triphosphates by the phosphorimidazoddate procedure," *Nucleic Acids Res.,* vol. 4, no. 8, pp. 2843-2853, 1977.

[76] J. W. Kozarich, A. C. Chinault and S. M. Hecht, "Ribonucleoside phosphates via phosphorimidazolidate intermediates. Synthesis of pseudoadenosine 5'-triphosphate," *Biochemistry,* vol. 12, no. 22, pp. 4458-4463, 1973.

[77] T. Yamaguchi, M. Saneyoshi, H. Takahashi, S. Hirokawa, R. Amano, X. Liu, M. Inomata and T. Maruyama, "Synthetic Nucleosides and Nucleotides. 43. Inhibition of Vertebrate Telomerases by Carbocyclic Oxetanocin G (C.OXT-G) Triphosphate Analogues and Influence of C.OXT-G Treatment on Telomere Length in Human HL60 Cells," *Nucleosides, Nucleotides, and Nucleic Acids,* vol. 25, pp. 539-551, 2006.

[78] H. S. Kim, R. G. Ravi, V. E. Marquez, S. Maddileti, A. K. Wihlborg, D. Erlinge, M. Malmsjo, J. L. Boyer, T. K. Harden and K. A. Jacobson, "Methanocarba Modification of Uracil and Adenine Nucleotides: High Potency of Northern Ring Conformation at P2Y1, P2Y2, P2Y4, and P2Y11 but Not P2Y6 Receptors," *J. Med. Chem.,* vol. 45, no. 1, pp. 208-218, 2002.

[79] A. R. Kore and G. Parmar, "Convenient Synthesis of Nucleoside-5′-Diphosphates from the Corresponding Ribonucleoside-5′-phosphoroimidazole," *Synth. Commun.,* vol. 36, no. 22, pp. 3393-3399, 2006.

[80] W. Wu, C. L. F. Meyers and R. F. Borch, "A Novel Method for the Preparation of Nucleoside Triphosphates from Activated

Nucleoside Phosphoramidates," *Org. Lett.*, vol. 6, no. 13, pp. 2257-2260, 2004.

[81] S. Mohamady, A.Desoky and S. D. Taylor, "Sulfonyl Imidazolium Salts as Reagents for the Rapid and Efficient Synthesis of Nucleoside Polyphosphates and Their Conjugates," *Org. Lett.*, vol. 14, no. 1, pp. 402-405, 2012.

[82] J. Booml, R. Crea, W. C. Luytan and A. B. Vink, "2,2,2-Tribromoethyl phosphoromorpholinochloridate: a convenient reagent for the synthesis of ribonucleoside mono-, di- and tri-phosphates," *Tetrahedron Lett.*, vol. 32, pp. 2779-2782, 1975.

[83] T. Hiroshi, K. Toshimi and H. Tsujiaki, "8-Quinolyl nucleoside 5'-phosphates as auseful intermediate for the synthesis of nucleoside 5'-di- and 5'-tri-phosphates," *Chem. Lett.*, vol. 6, no. 6, pp. 655-658, 1977.

[84] M. Honcharenko, M. Zytekc, B. Bestas, P. Moreno, J. Jemielity, E. Darzynkiewicz, C. I. E. Smith and R. Stromberg, "Synthesis and evaluation of stability of m3G-CAP analogues in serum-supplemented medium and cytosolic extract," *Bioorg. Med. Chem.*, vol. 21, no. 24, pp. 7921-7928, 2013.

[85] S. V. Vasilyeva, I. R. Grin, B. P. Chelobanov and D. A. Stetsenko, "2′,3′-Dideoxyuridine triphosphate conjugated to SiO2 nanoparticles: Synthesis and evaluation of antiproliferative activity," *Bioorganic Med. Chem. Lett.* , vol. 28, no. 7, pp. 1248-1251, 2018.

[86] J. A. Stock, "Synthesis of Phosphonate Analogues of Thymidine Di- and Triphosphate from 5’- 0 -Toluenesulfonylthymidine," *J. Org. Chem.*, vol. 44, no. 22, pp. 3997-4000, 1979.

[87] V. M. Dixit and C. D. Poulter, "Convenient synthesis of adenosine 5'-diphosphate, adenosine 5'-methylenediphosphate, and adenosine 5'-triphosphate," *Tetrahedron Lett.*, vol. 25, no. 37, pp. 4055-4058, 1984.

[88] S. Mohamady and S. D. Taylor, "Synthesis of Nucleoside Tetraphosphates and Dinucleoside Pentaphosphates via Activation of Cyclic Trimetaphosphate," *Org. Lett.*, vol. 15, no. 11, pp. 2612-2615, 2013.

[89] S. Mohamady and S. D. Taylor, "Synthesis of Nucleoside Triphosphates from 2′-3′-Protected Nucleosides Using Trimetaphosphate," *Org. Lett.,* vol. 18, no. 3, pp. 580-583, 2016.

[90] J. Singh, A. Ripp, T. M. Haas, D. Qiu, M. Keller and P. A. Wender, "Synthesis of Modified Nucleoside Oligophosphates Simplified: Fast, Pure, and Protecting Group Free," *J. Am. Chem. Soc.,* vol. 141, pp. 15013-15017, 2019.

[91] V. K. Sharma and J. K. Watts, "Oligonucleotide therapeutics: chemistry, delivery and clinical progress," *Future Med. Chem.,* vol. 7, no. 16, pp. 1-22, 2015.

[92] K. H. Shaughnessy, "Palladium-Catalyzed Modification of Unprotected Nucleosides, Nucleotides, and Oligonucleotides," *Molecules,* vol. 20, pp. 9419-9454, 2015.

[93] A. G. Firth, K. Wilson, C. G. Baumann and I. J. S. Fairlamb, "Interrogation of a Sonogashira Cross-Coupling of 8-Bromoguanosine with Phenylacetylene on Amberlite: Evidence for Pd/Cu Ion Binding and Propagation of Pd/Cu Nanoparticles," *Nucleosides, Nucleotides and Nucleic Acids,* vol. 30, pp. 168-184, 2011.

[94] J. H. Cho, C. D. Prickett and K. H. Shaughnessy, "Efficient Sonogashira Coupling of Unprotected Halonucleosides in Aqueous Solvents Using Water-Soluble Palladium Catalysts," *Eur. J. Org. Chem.,* vol. 2010, no. 19, pp. 3678-3683, 2010.

[95] Y. C. Shih and T. C. Chien, "Practical synthesis of 6-aryluridines via palladium(II) acetate catalyzed Suzukie-Miyaura cross-coupling reaction," *Tetrahedron,* vol. 67, pp. 3915-3923, 2011.

[96] S. o. M. C. N. T. f. R. Labelling, "E. Trevisiol; E. Defrancq; J. Lhomme; A. Laayoun; P. Cros," *Tetrahedron,* vol. 56, no. 35, pp. 6501-6510, 2000.

[97] G. Herve and C. Len, "Heck and Sonogashira couplings in aqueous media – application to unprotected nucleosides and nucleotides," *Sustain. Chem. Process.,* vol. 3, no. 3, pp. 1-14, 2015.

[98] S. Bhilare, H. Shet, Y. S. Sanghvi and A. R. Kapdi, "Discovery, Synthesis, and Scale-up of Effcient Palladium Catalysts Useful for the Modification of Nucleosides and Heteroarenes," *Molecules,* vol. 7, no. 1645, p. 25, 2020.

[99] F. W. Hobbs, "Palladium-catalyzed synthesis of alkynylamino nucleosides. A universal linker for nucleic acids," *J. Org. Chem.,* vol. 54, no. 14, pp. 3420-3422, 1989.

[100] G. Sagi, L. Otvos, S. Ikeda, G. Andrei, R. Snoeck and E. Clercq, "Synthesis and Antiviral Activities of 8-Alkynyl-, 8-Alkenyl-, and 8-Alkyl-2‘-deoxyadenosine Analogues," *J. Med. Chem.*, vol. 37, pp. 1307-1311, 1994.

[101] S. E. Lee, A. Sidorov, T. Gourlain, N. Mignet, S. J. Thorpe, J. A. Brazier, M. J. Dickman, D. P. Hornby, J. A. Grasby and D. M. Williams, "Enhancing the catalytic repertoire of nucleic acids: a systematic study of linker length and rigidity," *Nucleic Acids Res.,* vol. 29, no. 7, pp. 1565-1573, 2001.

[102] J. Riedl, P. Horakova, P. Šebest, R. Pohl, L. Havran, M. Fojta and M. Hocek, "Tetrathiafulvalene-Labelled Nucleosides and Nucleoside Triphosphates: Synthesis, Electrochemistry and the Scope of Their Polymerase Incorporation into DNA," *Eur. J. Org. Chem.* , vol. 2009, no. 21, pp. 3519-3525, 2009.

[103] Y. Saito, K. Matsumoto, S. S. Bag, S. Ogasawara, K. Fujimoto, K. Hanawa and I. Saito, "C8-alkynyl- and alkylamino substituted 2′-deoxyguanosines: a universal linker for nucleic acids modification," *Tetrahedron,* vol. 64, no. 16, pp. 3578-3588, 2008.

[104] A. Olejniczak, B. Wojtczak and Z. J. Lesnikowski, "2′-Deoxyadenosine Bearing Hydrophobic Carborane Pharmacophore," *Nucleosides, Nucleotides, and Nucleic Acids,* vol. 26, pp. 1611-1613, 2007.

[105] F. Seela and S. A. Ingale, "“Double Click” Reaction on 7-Deazaguanine DNA: Synthesis and Excimer Fluorescence of Nucleosides and Oligonucleotides with Branched Side Chains Decorated with Proximal Pyrenes," *J. Org. Chem.,* vol. 75, pp. 284-295, 2010.

[106] P. Kumar, M. E. Ostergaard, B. Baral, B. A. Anderson, D. C. Guenther, M. Kaura, D. J. Raible, P. K. Sharma and P. J. Hrdlicka, "Synthesis and Biophysical Properties of C5-Functionalized LNA (Locked Nucleic Acid)," *J. Org. Chem.,* vol. 79, no. 11, pp. 5047-5061, 2014.

[107] K. H. Shaughnessy and R. B. Vasher, "Palladium-Catalyzed Cross-Coupling in Aqueous Media: Recent Progress and Current Applications," *Curr. Org. Chem.,* vol. 9, no. 7, pp. 585-604, 2005.

[108] K. H. Shaughnessy, "Beyond TPPTS: New Approaches to the Development of Efficient Palladium-Catalyzed Aqueous-Phase Cross-Coupling Reactions," *Eur. J. Org. Chem.*, vol. 2006, no. 8, pp. 1827-1835, 2006.

[109] B. Cornilsa and E. G. Kuntz, "Introducing TPPTS and related ligands for industrial biphasic processes," *J. Organomet. Chem.*, vol. 502, no. 1-2, pp. 177-186, 1995.

[110] E. Fache, C. Santini, F. Senocq and J. M. Basset, "Homogeneous catalysis in water Part III. The catalytic hydrogenation of propionaldehyde with $(RuCl_2L_2)_2$, $RuHClL_3$, $RuH(OAc)L_3$, RuH_2L_4, $RuHIL_3$, $RuCl_2(CO)_2L_2$ and $[Ru(OAc)(CO)_2L]_2$, $(LP(C_6H_4–mSO_3Na)_3 \cdot 3H_2O)$: A kinetic investigation of the salt effect in water," *J. Mol. Catal*, vol. 72, no. 3, pp. 337-350, 1992.

[111] A. Andriollo, J. Carrasquel, J. Marino, F. A. Lopez, D. E. Paez, I. Rojas and N. Valencia, "On the synthesis and characterization of two ruthenium water-soluble complexes: preliminary results on the hydrogenation of cinnamaldehyde in a biphasic system. Surface activity of the ligands TPPMS and TPPTS," *J. Mol. Catal. A Chem.*, vol. 116, no. 1-2, pp. 157-165, 1997.

[112] J. P. Genet, E. Blart and M. Savignac, "Palladium-catalyzed cross-coupling reactions in homogeneous aqueous medium," *Synlett*, vol. 9, pp. 715-717, 1992.

[113] C. Amatore, E. Blart, J. P. Genet, A. Jutand, S. Lemaire-Audoire and M. Savignac, "New synthetic applications of water-soluble acetate Pd/TPPTS catalyst generated in situ. evidence for a true Pd(0) species intermediate," *J. Org. Chem.*, vol. 60, no. 21, pp. 6829-6839, 1995.

[114] E. G. Kuntza and O. M.Vittori, "Redox chemistry of Pd2+, Pt2+, Rh3+–TPPTS systems in water: pH influence on the preparation of low valent TPPTS complexes," *J. Mol. Catal. A Chem.*, vol. 129, no. 2-3, pp. 159-171, 1998.

[115] A. L. Casalnuovo and J. C. Calabrese, "Palladium-Catalyzed Alkylations in Aqueous Media," *J. Am. Chem. Sol.*, vol. 112, no. 11, pp. 4324-4330, 1990.

[116] M. Vrabel, R. Pohl, B. Klepetarova, I. Votruba and M. Hocek, "Synthesis of 2'-deoxyadenosine nucleosides bearing bipyridine-type ligands and their Ru-complexes in position 8 through cross-

coupling reactions," *Org. Biomol. Chem.,* vol. 5, pp. 2849-2857, 2007.

[117] M. Vrabel, R. Pohl, I. Votruba, M. Sajadi, S. A. Kovalenko, N. P. Ernstingb and M. Hocek, "Synthesis and photophysical properties of 7-deaza-2′-deoxyadenosines bearing bipyridineligands and their Ru(II)-complexes in position 7," *Org. Biomol. Chem.,* vol. 6, pp. 2852-2860, 2008.

[118] L. Kalachova, R. Pohl and M. Hocek, "Synthesis of 2'-Deoxyuridine and 2'-Deoxycytidine Nucleosides Bearing Bipyridine and Terpyridine Ligands at Position 5," *Synthesis,* vol. 1, pp. 105-112, 2009.

[119] S. Ikonen, H. Macıckova-Cahova, R. Pohl, M. Sanda and M. Hocek, "Synthesis of nucleoside and nucleotide conjugates of bile acids, and polymerase construction of bile acid-functionalized DNA," *Org. Biomol. Chem.,* vol. 8, pp. 1194-1201, 2010.

[120] P. Capek, H. Cahov, R. Pohl, M. Hocek, C. Gloeckner and A. Marx, "An Efficient Method for the Construction of Functionalized DNA Bearing Amino Acid Groups through Cross-Coupling Reactions of Nucleoside Triphosphates Followed by Primer Extension or PCR," *Chem. Eur. J.,* vol. 13, pp. 6196-6203, 2007.

[121] M. Ondrus, V. Sykorova, L. Bednarova, R. Pohl and M. Hocek, "Enzymatic synthesis of hypermodified DNA polymers for sequence-specific display of four different hydrophobic groups," *Nucleic Acids Res.* , vol. 48, no. 21, pp. 11982-11993, 2020.

[122] A. R. Kore, A. Senthilvelan and M. Shanmugasundaram, "Highly chemoselective palladium-catalyzed Sonogashira coupling of 5-iodouridine-5′-triphosphates with propargylamine: a new efficient method for the synthesis of 5-aminopropargyl-uridine-5′-triphosphates," *Tetrahedron Lett.,* vol. 53, no. 24, pp. 3070-3072, 2012.

[123] P. Brazdilova, V. M, R. Pohl, H. Pivonkova, L. Havran, M. Hocek and M. Fojta, "Ferrocenylethynyl Derivatives of Nucleoside Triphosphates: Synthesis, Incorporation, Electrochemistry, and Bioanalytical Applications," *Chem. Eur. J.,* vol. 13, pp. 9527-9533, 2007.

[124] M. Vrabel, P. Horakova, H. Pivonkova, L. Kalachova, H. Cernocka, H. Cahova, R. Pohl, P. Sebest, L. Havran, M. Hocek and M. Fojta, "Base-Modified DNA Labeled by [Ru(bpy)3]2+ and

[Os(bpy)3]2+ Complexes: Construction by Polymerase Incorporation of Modified Nucleoside Triphosphates, Electrochemical and Luminescent Properties, and Applications," *Chem. Eur. J.*, vol. 15, no. 5, pp. 1144-1154, 2009.

[125] J. Balintova, M. Plucnara, P. Vidlakova, R. Pohl, L. Havran, M. Fojta and M. Hocek, "Benzofurazane as a New Redox Label for Electrochemical Detection of DNA: Towards Multipotential Redox Coding of DNA Bases," *Chem. Eur. J.*, vol. 19, pp. 12720-12731, 2013.

[126] A. Simonova, L. Havran, R. Pohl, M. Fojta and M. Hocek, "Phenothiazine-linked Nucleosides and Nucleotides for Redox Labelling of DNA," *Org. Biomol. Chem.*, vol. 15, pp. 6984-6996, 2017.

[127] M. Kromer, K. Bartova, V. Raindlova and M. Hocek, "Synthesis of Dihydroxyalkynyl and Dihydroxyalkyl Nucleotides as Building Blocks or Precursors for Introduction of Diol or Aldehyde Groups to DNA for Bioconjugations," *Chem. Eur. J.*, vol. 24, no. 46, pp. 11890-11894, 2018.

[128] P. Havranova-Vidlakova, M. Kromer, V. Sykorova, M. Trefulka, M. Fojta, L. Havran and M. Hocek, "Vicinal Diol-Tethered Nucleobases as Targets for DNA Redox Labeling with Osmate Complexes," *ChemBioChem*, vol. 21, no. 1-2, pp. 171-180, 2019.

[129] P. Kielkowski, R. Pohl and M. Hocek, "Synthesis of Acetylene Linked Double-Nucleobase Nucleos(t)ide Building Blocks and Polymerase Construction of DNA Containing Cytosines in the Major Groove," *J. Org. Chem.*, vol. 76, pp. 3457-3462, 2011.

[130] P. Kielkowski, H. Cahova, R. Pohl and M. Hocek, "Flexible double-headed cytosine-linked 2'-deoxycytidine nucleotides. Synthesis, polymerase incorporation to DNA and interaction with DNA methyltransferases," *Bioorg. Med. Chem.*, vol. 24, pp. 1268-1276, 2016.

[131] H. Lonnberg, "Solid-Phase Synthesis of Oligonucleotide Conjugates Useful for Delivery and Targeting of Potential Nucleic Acid Therapeutics," *Bioconjugate Chem.*, vol. 20, no. 6, pp. 1065-1094, 2009.

[132] J. Dadova, H. C. and M. Hocek, "Polymerase Synthesis of Base-Modified DNA," in *Modified Nucleic Acids. Nucleic Acids and*

Molecular Biology, vol. 31, Switzerland, Springer, Cham, 2016, pp. 123-144.

[133] P. R. Langer, A. A. Waldrop and D. C. Ward, "Enzymatic synthesis of biotin-labeled polynucleotides: novel nucleic acid affinity probes," *Proc. NatL Acad. Sci. USA*, vol. 78, no. 11, pp. 6633-6637, 1981.

[134] M. Kuwahara and N. Sugimoto, "Molecular Evolution of Functional Nucleic Acids with Chemical Modifications," *Molecules*, vol. 15, pp. 5423-5444, 2010.

[135] M. Hocek, "Enzymatic Synthesis of Base-Functionalized Nucleic Acids for Sensing, Cross-linking, and Modulation of Protein−DNA Binding and Transcription," *Acc. Chem. Res.*, vol. 52, pp. 1730-1737, 2019.

[136] H. Cahova, R. Pohl, L. Bednarova, K. Novakova, J. Cvacka and M. Hocek, "Synthesis of 8-bromo-, 8-methyl- and 8-phenyl-dATP and their polymerase incorporation into DNA," *Org. Biomol. Chem.*, vol. 6, pp. 3657-3660, 2008.

[137] A. S. P. Gowda, M. Lee and T. E. Spratt, "N2-Substituted 2′-Deoxyguanosine Triphosphate Derivatives as Selective Substrates for Human DNA Polymerase κ," *Angew. Chem. Int. Ed.*, vol. 56, no. 10, pp. 2628-2631, 2017.

[138] S. Matsuda, A. M. Leconte and F. E. Romesberg, "Minor Groove Hydrogen Bonds and the Replication of Unnatural Base Pairs," *J. Am. Chem. Soc.*, vol. 129, no. 17, pp. 5551-5557, 2007.

[139] J. P. Peters, L. S. Mogil, M. J. McCauley, M. C. Williams and J. Maher, "Mechanical Properties of Base-Modified DNA Are Not Strictly Determined by Base Stacking or Electrostatic Interactions," *Biophys. J.*, vol. 107, no. 2, pp. 448-459, 2014.

[140] J. Matyasovsky, R. Pohl and M. Hocek, "2-Allyl- and Propargylamino-dATPs for Site-Specific Enzymatic Introduction of a Single Modification in the Minor Groove of DNA," *Chem. Eur. J.*, vol. 24, no. 56, pp. 14938-14941, 2018.

[141] J. Matyasovsky and M. Hocek, "2-Substituted 2′-deoxyinosine 5′-triphosphates as substrates for polymerase synthesis of minor-groove-modified DNA and effects on restriction endonuclease cleavage," *Org. Biomol. Chem.*, vol. 18, pp. 255-262, 2020.

[142] M. Kromer, M. Brunderova, I. Ivancova, L. P. Slavetinska and M. Hocek, "2-Formyl-dATP as Substrate for Polymerase Synthesis of Reactive DNA Bearing an Aldehyde Group in the Minor Groove," *ChemPlusChem,* vol. 85, no. 6, pp. 1164-1170, 2020.

[143] A. Simonova, I. Magrina, V. Sykorova, R. Pohl, M. Ortiz, L. Havran, M. Fojta, C. O'Sullivan and M. Hocek, "Tuning of Oxidation Potential of Ferrocene for Ratiometric Redox Labelling and Coding of Nucleotides and DNA," *Chem. Eur. J.* , vol. 26, no. 6, pp. 1286-1291, 2020.

[144] I. Ivancova, D. L. Leone and M. Hocek, "Reactive modifications of DNA nucleobases for labelling, bioconjugations, and cross-linking," *Curr. Opin. Chem. Biol.* , vol. 52, pp. 136-144, 2019.

[145] T. Carell, M. Q. Kurz, M. Muller, M. Rossa and F. Spada, "Non-canonical Bases in the Genome: The Regulatory Information Layer in DNA," *Angew. Chem. Int. Ed.* , vol. 57, no. 16, pp. 4296-4312, 2018.

[146] P. Guo, S. Yan, J. Hu, X. Xing, C. Wang and X. Xu, "Selective Detection of 5-Formyl-2'-deoxycytidine in DNA Using a Fluorogenic Hydroxylamine Reagent," *Org. Lett.,* vol. 15, no. 13, pp. 3266-3269, 2013.

[147] R. E. Hardisty, F. Kawasaki, A. B. Sahakyan and S. Balasubramanian, "Selective Chemical Labeling of Natural T Modifications in DNA," *J. Am. Chem. Soc.,* vol. 137, no. 29, pp. 9270-9272, 2015.

[148] C. Liu, Y. Chen, Y. Wang, F. Wu, X. Zhang, W. Yang, J. Wang, Y. Chen, Z. He, G. Zou, S. Wang and X. Zhou, "A highly efficient fluorescence-based switch-on detection method of 5-formyluracil in DNA," *Nano Res.,* vol. 10, pp. 2449-2458, 2017.

[149] E. A. Raiber, G. Portella, S. M. Cuesta, R. Hardisty, P. Murat, Z. Li, M. Iurlaro, W. Dean, J. Spindel, D. Beraldi, Z. Liu, M. A. Dawson, W. Reik and S. Balasubramanian, "5-Formylcytosine organizes nucleosomes and forms Schiff base interactions with histones in mouse embryonic stem cells," *Nat. Chem.,* vol. 10, pp. 1258-1266, 2018.

[150] W. Hirose, K. Sato and A. Matsuda, "Selective Detection of 5-Formyl-2′-deoxyuridine, an Oxidative Lesion of Thymidine, in DNA by a Fluorogenic Reagent," *Angew. Chem. Int. Ed.,* vol. 49, no. 45, pp. 8392-8394, 2010.

[151] Q. Zhou, K. Li, Y. H. Liu, L. L. Li, K. K. Yu and H. Zhang, "Fluorescent Wittig reagent as a novel ratiometric probe for the quantification of 5-formyluracil and its application in cell imaging," *Chem. Commun.*, vol. 54, pp. 13722-13725, 2018.

[152] B. Samanta, J. Seikowski and C. Hobartner, "Fluorogenic Labeling of 5-Formylpyrimidine Nucleotides in DNA and RNA," *Angew. Chem. Int. Ed.*, vol. 55, pp. 1912-1916, 2016.

[153] S. Wickramaratne, S. Mukherjee, P. W. Villalta, O. D. Scharer and N. Y. Tretykova, "Synthesis of Sequence-Specific DNA−Protein Conjugates via a Reductive Amination Strategy," *Bioconjugate Chem.*, vol. 24, pp. 1496-1506, 2013.

[154] V. Raindlov, R. Pohl, M. Sanda and M. Hocek, "Direct Polymerase Synthesis of Reactive Aldehyde-Functionalized DNA and Its Conjugation and Staining with Hydrazines," *Angew. Chem. Int. Ed.*, vol. 49, pp. 1064-1066, 2010.

[155] J. Dadova, P. Orsag, R. Pohl, M. Brzdov, M. Fojta and M. Hocek, "Vinylsulfonamide and Acrylamide Modification of DNA for Cross-linking with Proteins," *Angew. Chem. Int. Ed.*, vol. 52, pp. 1-5, 2013.

[156] J. Matyasovsky, P. Perlkov, V. Malnuit, R. Pohl and M. Hocek, "2-Substituted dATP Derivatives as Building Blocks for PolymeraseCatalyzed Synthesis of DNA Modified in the Minor Groove," *Angew. Chem. Int. Ed*, vol. 55, pp. 1-5, 2016.

[157] J. M. Prober, R. J. Dam, C. W. Robertson, F. W. Hobbs and G. L. Trainor, "Process and reagents for DNA sequence analysis". Europe Patent EP0252683B1, 18 01 1995.

[158] B. B. Rosenblum, L. G. Lee, S. L. Spurgeon, S. H. Khan, S. M. Menchen, C. R. Heiner and S. M. Chen, "New dye-labeled terminators for improved DNA sequencing patterns," *Nucleic Acids Res.*, vol. 25, no. 22, pp. 4500-4504, 1997.

[159] S. Nampalli, M. Khot, J. R. Nelson, P. K. Flick, C. W. Fuller and S. Kumar, "Fluorescence resonance energy transfer dye nucleotide terminators: a new synthetic approach for high-throughout DNA sequencing," *Nucleosides, Nucleotides and Nucleic Acids*, vol. 20, no. 4-7, pp. 361-367, 2001.

[160] D. Dziuba, R. Pohla and M. Hocek, "Polymerase synthesis of DNA labelled with benzylidene cyanoacetamide-based fluorescent

molecular rotors: fluorescent light-up probes for DNA-binding proteins," *Chem. Commun.*, vol. 51, pp. 4880-4882, 2015.

[161] M. Augustin, W. Ankenbauer and B. Angerer, "Progress towards single-molecule sequencing: enzymatic synthesis of nucleotide-specifically labeled DNA," *J. Biotechnol.*, vol. 86, pp. 289-301, 2001.

[162] M. Kuwahara, K. Hanawa, K. Ohsawa, R. Kitagata, H. Ozakia and H. Sawai, "Direct PCR amplification of various modified DNAs having amino acids: Convenient preparation of DNA libraries with high-potential activities for in vitro selection," *Bioorg. Med. Chem.*, vol. 14, pp. 2518-2526, 2006.

[163] R. S. Duthie, I. M. Kalve, S. B. Samols, S. Hamilton, I. Livshin, M. Khot, S. Nampalli, S. Kumar and C. W. Fuller, "Novel Cyanine Dye-Labeled Dideoxynucleoside Triphosphates for DNA Sequencing," *Bioconjugate Chem.*, vol. 13, no. 4, pp. 699-706, 2002.

[164] R. Huisgen, "Kinetics and Mechanism of 1,3-Dipolar Cycloadditions," *Angew. Chem. Int. Ed.*, vol. 2, no. 11, pp. 633-696, 1963.

[165] V. K. Tiwari, B. B. Mishra, K. B. Mishra, N. Mishra, A. S. Singh and X. Chen, "Cu-Catalyzed Click Reaction in Carbohydrate Chemistry," *Chem. Rev.*, vol. 116, pp. 3086-3240, 2016.

[166] C. W. Tornoe, C. Christensen and M. Meldal, "Peptidotriazoles on Solid Phase: [1,2,3]-Triazoles by Regiospecific Copper(I)-Catalyzed 1,3-Dipolar Cycloadditions of Terminal Alkynes to Azides," *J. Org. Chem.*, vol. 67, no. 9, pp. 3057-3064, 2002.

[167] V. V. Rostovtsev, L. G. Green, V. V. Fokin and K. B. Sharpless, "A Stepwise Huisgen Cycloaddition Process: Copper(I)-Catalyzed Regioselective "Ligation" of Azides and Terminal Alkynes," *Angew. Chem.*, vol. 114, no. 14, pp. 2708-2711, 2002.

[168] J. E. Hein and V. V. Fokin, "Copper-catalyzed azide–alkyne cycloaddition (CuAAC) and beyond: new reactivity of copper(I) acetylides," *Chem. Soc. Rev.*, vol. 39, pp. 1302-1315, 2010.

[169] L. Zhang, X. Chen, P. Xue, H. H. Y. Sun, I. D. Williams, K. B. Sharpless, V. V. Fokin and G. Jia, "Ruthenium-Catalyzed Cycloaddition of Alkynes and Organic Azides," *J. AM. CHEM. SOC.*, vol. 127, no. 46, pp. 15998-15999, 2005.

[170] L. Liang and D. Astruc, "The copper(I)-catalyzed alkyne-azide cycloaddition (CuAAC) “click” reaction and its applications. An overview," *Coord. Chem. Rev.*, vol. 255, pp. 2933-2945, 2011.

[171] T. R. Chan, R. Hilgraf, K. B. Sharpless and V. V. Fokin, "Polytriazoles as Copper(I)-Stabilizing Ligands in Catalysis," *Org. Lett.*, vol. 6, no. 7, pp. 2853-2855, 2004.

[172] Q. Wang, T. R. Chan, R. Hilgraf, V. V. Fokin, K. B. Sharpless and M. G. Finn, "Bioconjugation by Copper(I)-Catalyzed Azide-Alkyne [3 + 2] Cycloaddition," *J. Am. Chem. Soc.*, vol. 125, no. 11, pp. 3192-3193, 2003.

[173] V. Hong, S. I. Presolski, C. Ma and M. G. Finn, "Analysis and Optimization of Copper-Catalyzed Azide–Alkyne Cycloaddition for Bioconjugation," *Angew. Chem.*, vol. 121, pp. 10063-10067, 2009.

[174] S. S. Gupta, J. Kuzelka, P. Singh, W. G. Lewis, M. Manchester and M. G. Finn, "Accelerated Bioorthogonal Conjugation: A Practical Method for the Ligation of Diverse Functional Molecules to a Polyvalent Virus Scaffold," *Bioconjugate Chem.* , vol. 16, pp. 1572-1579, 2005.

[175] V. O. Rodionov, S. I. Presolski, S. Gardinier, Y. H. Lim and M. G. Finn, "Benzimidazole and Related Ligands for Cu-Catalyzed Azide-Alkyne Cycloaddition," *J. Am. Chem. Soc.*, vol. 129, pp. 12696-12704, 2007.

[176] C. Freidel, S. Kaloyanova and K. Peneva, "Chemical tags for site-specific fluorescent labeling of biomolecules," *Amino Acids*, vol. 48, pp. 1357-1372, 2016.

[177] N. J. Agard, J, A. Prescher and C. R. Bertozzi, "A Strain-Promoted [3 + 2] Azide-Alkyne Cycloaddition for Covalent Modification of Biomolecules in Living Systems," *J. Am. Chem. Soc.*, vol. 126, pp. 15046-15047, 2004.

[178] D. C. Kennedy, C. S. McKay, M. C. B. Legault, D. C. Danielson, J. A. Blake, A. F. Pegoraro, A. Stolow, Z. Mester and J. P. Pezacki, "Cellular Consequences of Copper Complexes Used To Catalyze Bioorthogonal Click Reactions," *J. Am. Chem. Soc.*, vol. 133, no. 44, pp. 17993-18001, 2011.

[179] J. M. B. Nicholas J. Agard, J. A. Prescher, A. Lo and C. R. Bertozzi, "A Comparative Study of Bioorthogonal Reactions with Azides," *ACS Chem. Biol.*, vol. 1, no. 10, pp. 644-648, 2006.

[180] J. M. Baskin, J. A. Prescher, S. T. Laughlin, N. J. Agard, P. V. Chang, I. A. Miller, A. Lo, J. A. Codelli and C. R. Bertozzi, "Copper-free click chemistry for dynamic in vivo imaging," *Proc. Natl. Acad. Sci. U. S. A.*, vol. 104, pp. 16793-16797, 2007.

[181] J. Dommerholt, S. Schmidt, R. P. Temming, L. J. A. Hendriks, F. P. J. T. Rutjes, J. C. M. Hest, D. J. Lefeber, P. Friedl and F. L. Delft, "Readily Accessible Bicyclononynes for Bioorthogonal Labeling and Three-Dimensional Imaging of Living Cells," *Angew. Chem.*, vol. 49, pp. 9422-9425, 2010.

[182] X. Ning, J. Guo, M. A. Wolfert and G.-J. Boons, "Visualizing Metabolically Labeled Glycoconjugates of Living Cells by Copper-Free and Fast Huisgen Cycloadditions," *Angew. Chem. Int. Ed.*, vol. 47, pp. 2253-2255, 2008.

[183] V. Hong, N. F. Steinmetz, M. Manchester and M. G. Finn, "Labeling Live Cells by Copper-Catalyzed Alkyne−Azide Click Chemistry," *Bioconjugate Chem.*, vol. 21, no. 10, pp. 1912-1916, 2010.

[184] D. Kondhare, A. Zhang, P. Leonard and F. Seela, "Alkynylated and Dendronized 5-Aza-7-deazaguanine Nucleosides: Cross-Coupling with Tripropargylamine and Linear Alkynes, Click Functionalization, and Fluorescence of Pyrene Adducts," *J. Org. Chem.*, vol. 85, pp. 10525-10538, 2020.

[185] J. Zayas, M. Annoual, J. K. Das, Q. Felty, W, G. Gonzalez, J. Miksovska, N. Sharifai, A. Chiba and S. F. Wnuk, "Strain Promoted Click Chemistry of 2- or 8-Azidopurine and 5-Azidopyrimidine Nucleosides and 8-Azidoadenosine Triphosphate with Cyclooctynes. Application to Living Cell Fluorescent Imaging," *Bioconjugate Chem.*, vol. 26, no. 8, pp. 1519-1532, 2015.

[186] D. Kodr, C. P. Yenice, A. Simonova, D. P. Saftic, R. Pohl, V. Sykorova, M. Ortiz, L. Havran, M. Fojta, Z. J. Lesnikowski, C. K. O'Sullivan and M. Hocek, "Carborane- or Metallacarborane-Linked Nucleotides for Redox Labeling. Orthogonal Multipotential Coding of all Four DNA Bases for Electrochemical Analysis and Sequencing," *J. Am. Chem. Soc.*, vol. 143, pp. 7124-7134, 2021.

[187] R. S. Sorensen, A. H. Okholm, D. Schaffert, A. L. B. Kodal, K. V. Gothelf and J. Kjems, "Enzymatic Ligation of Large Biomolecules to DNA," *ACS Nano,* vol. 7, no. 9, pp. 8098-8104, 2013.

[188] P. M. E. Gramlich, S. Warncke, J. Gierlich and T. Carell, "Click–Click–Click: Single to Triple Modification of DNA," *Angew. Chem. Int. Ed.,* vol. 47, pp. 3442-3444, 2008.

[189] L. Massaad-Massade, S. Boutary, M. Caillaud, C. Gracia, B. Parola, S. B. Gnaouiya, B. Stella, S. Arpicco, E. Buchy, D. Desmaele, P. Couvreur and G. Urbinati, "New Formulation for the Delivery of Oligonucleotides Using “Clickable” siRNA-Polyisoprenoid-Conjugated Nanoparticles: Application to Cancers Harboring Fusion Oncogenes," *Bioconjugate Chem.,* vol. 29, no. 6, pp. 1961-1972, 2018.

[190] M. M. Haque and X. Peng, "DNA-associated click chemistry," *Sci. China Chem.* , vol. 57, pp. 215-231, 2014.

[191] H. Xiong, P. Leonard and F. Seela, "Construction and Assembly of Branched Y-Shaped DNA: “Click” Chemistry Performed on Dendronized 8-Aza-7-deazaguanine Oligonucleotides," *Bioconjugate Chem.,* vol. 23, no. 4, pp. 856-870, 2012.

[192] S. R. Gerrard, C. Hardiman, M. Shelbourne, I. Nandhakumar, B. Norden and T. Brown, "A New Modular Approach to Nanoassembly: Stable and Addressable DNA Nanoconstructs via Orthogonal Click Chemistries," *ACS Nano,* vol. 6, no. 10, pp. 9221-9228, 2012.

[193] C. K. L. Gordon, D. Wu, A. Pusuluri, T. A. Feagin, A. T. Csordas, M. S. Eisenstein, C. J. Hawker, J. Niu and H. T. Soh, "Click-Particle Display for Base-Modified Aptamer Discovery," *ACS Chem. Biol.,* vol. 14, no. 12, pp. 2652-2662, 2019.

[194] A. Hottin and A. Marx, "Structural Insights into the Processing of Nucleobase-Modified Nucleotides by DNA Polymerases," *Acc. Chem. Res.,* vol. 49, no. 3, pp. 418-427, 2016.

[195] S. Obeid, A. Baccaro, W. Welte, K. Diederichs and A. Marx, "Structural basis for the synthesis of nucleobase modified DNA by Thermus aquaticus DNA polymerase," *Proc. Natl. Acad. Sci. U.S.A.,* vol. 107, no. 50, pp. 21327-21331, 2010.

[196] S. Obeid, H. Bubkamp, W. Welte, K. Diederichs and A. Marx, "Interactions of non-polar and “Click-able” nucleotides in the confines of a DNA polymerase active site," *Chem. Commun.*, vol. 48, no. 67, pp. 8320-8322, 2012.

[197] K. Bergen, A. L. Steck, S. Strutt, A. Baccaro, W. Welte, K. Diederichs and A. Marx, "Structures of KlenTaq DNA Polymerase Caught While Incorporating C5-Modified Pyrimidine and C7-Modified 7-Deazapurine Nucleoside Triphosphates," *J. Am. Chem. Soc.*, vol. 134, no. 29, pp. 11840-11843, 2012.

[198] A. Hottin, K. Betz, K. Diederichs and A. Marx, "Structural Basis for the KlenTaq DNA Polymerase Catalysed Incorporation of Alkene- versus Alkyne-Modified Nucleotides," *Chem. Eur. J.*, vol. 23, no. 9, pp. 2109-2118, 2017.

[199] G. Gebeyehu, P. Y. Rao, P. SooChan, D. A. Simms and L. Klevan, "Novel biotinylated nucleotide - analogs for labeling and colorimetric detection of DNA," *Nucleic Acids Res.*, vol. 15, no. 11, pp. 4513-4534, 1987.

[200] M. Vaghefi, Nucleoside Triphosphates and their Analogs, Boca Raton: Taylor & Francis, 2005.

[201] K. J. Livak, F. W. Hobbs and R. J. Zagursky, "Detection of single base differences using biotinylated nucleotides with very long linker arms," *Nucleic Acids Res.*, vol. 20, no. 18, pp. 4831-4837, 1992.

[202] T. Gourlain, S. A, N. Mignet, A. J. Thorpe, S. E. Lee, J. A. Grasby and D. M. Williams, "Enhancing the catalytic repertoire of nucleic acids. II. Simultaneous incorporation of amino and imidazolyl functionalities by two modified triphosphates during PCR," *Nucleic Acids Res.*, vol. 29, no. 9, pp. 1898-1905, 2001.

[203] T. S. Godovikova, D. M. Kolpashchikov, T. N. Orlova, V. A. Richter, T. M. Ivanova, S. L. Grochovsky, T. V. Nasedkina, L. S. Victorova and A. I. Poletaev, "5-[3-(E)-(4-Azido-2,3,5,6-tetrafluorobenzamido)propenyl-1]-2'-deoxyuridine-uridine-5'-triphosphate Substitutes for Thymidine-5'-triphosphate in the Polymerase Chain Reaction," *Bioconjugate Chem.*, vol. 10, pp. 529-537, 1999.

[204] G. T. Pauly, E. V. Bobst, D. Bruckman and A. M. Bobst, "Template-Dependent Incorporation of Spin-Labeled Thymidine Analogs into Viral DNA," *Helv. Chim.*, vol. 72, pp. 110-116, 1989.

[205] B. C. Froehler, S. Wadwani, T. J. Terhorst and S. R. Gerrard, "Oligodeoxynucleotides Containing C-5 Propyne Analogs of 2'-Deoxyuridine and 2'-Deoxycytidine," *Tetrahedron Lett.,* vol. 33, no. 37, pp. 5307-5310, 1992.

[206] C. A. Buhr, R. W. Wagner, D. Grant and B. C. Froehler, "Oligodeoxynucleotides containing C-7 propyne analogs of 7-deaza-24-deoxyguanosine and 7-deaza-24-deoxyadenosine," *Nucleic Acids Res.,* vol. 24, no. 15, pp. 2974-2980, 1996.

[207] J. Hel and F. Seela, "Propynyl groups in duplex DNA: stability of base pairs incorporating 7-substituted 8-aza-7-deazapurines or 5-substituted pyrimidines," *Nucleic Acids Res. ,* vol. 30, no. 24, pp. 5485-5496, 2002.

[208] F. Sanger, S. Nicklen and A. R. Coulson, "DNA sequencing with chain-terminating inhibitors," *Proc. Nati. Acad. Sci. USA,* vol. 74, no. 12, pp. 5463-5467, 1977.

[209] W. Ansorge, B. S. Sproat, J. Stegemann and C. Schwager, "A non-radioactive automated method for DNA sequence determination," *J. Biochem. Bioph. Meth.,* vol. 13, no. 6, pp. 315-323, 1986.

[210] L. M. Smith, S. Fung, M. W. Hunkapiller, T. J. Hunkapiller and L. E. Hood, "The synthesis of oligonucleotides containing an aliphatic amino group at the 5′ terminus: synthesis of fluorescent DNA primers for use in DNA sequence analysis," *Nucleic Acids Res.,* vol. 13, no. 7, pp. 2399-2412, 1985.

[211] W. Ansorge, B. Sproat, J. Stegemann, C. Schwager and M. Zenke, "Automated DNA sequencing: ultrasensitive detection of fluorescent bands during electrophoresis," *Nucleic Acids Res.,* vol. 15, no. 11, pp. 4593-4602, 1987.

[212] J. M. Prober, G. L. Trainor, R. J. Dam, F. W. Hobbs, C. W. Robertson, R. J. Zagursky, A. J. Cocuzza, M. A. Jensen and K. Baumeister, "A System for Rapid DNA Sequencing with Fluorescent Chain-Terminating Dideoxynucleotides," *Science,* vol. 238, no. 4825, pp. 336-341, 1987.

[213] S. Kumar and C. W. Fuller, "Advances in Dye-Nucleotide Conjugate Chemistry for DNA Sequencing," in *New High Throughput Technologies for DNA Sequencing and Genomics*, Elsevier B.V., 2007, pp. 119-149.

[214] M. J. Levene, J. Korlach, S. W. Turner, H. G. Craighead and W. W. Webb, "Zero-Mode Waveguides for Single-Molecule Analysis at High Concentrations," *Science*, vol. 299, no. 5607, pp. 682-686, 2003.

[215] S. Hardin, X. Gao, J. Briggs, R. Willson and S. C. Tu, "Methods for real-time single molecule sequence determination". United States Patent US7329492, 07 07 2000.

[216] J. G. K. Williams, "System and methods for nucleic acid sequencing of single molecules by polymerase synthesis". United States Patent US6255083, 14 12 1998.

[217] V. A. Litosh, B. P. Stupi, M. L. Metzker and W. Wu, "Photocleavable labeled nucleotides and nucleosides and labeled nucleotides and nucleosides and methods for their use in dna sequencing". United States Patent WO2008070749, 05 12 2006.

[218] J. Bowers, J. Mitchell, E. Beer, P. R. Buzby, M. Causey, J. W. Efcavitch, M. Jarosz, E. Krzymanska-Olejnik, L. Kung, D. Lipson, G. M. Lowman, S. Marappan, P. McInerney, A. Platt, A. Roy, S. M. Siddiqi, K. Steinmann and J. F. Thompson, "Virtual terminator nucleotides for next-generation DNA sequencing," *Nat Methods*, vol. 6, pp. 593-595, 2009.

[219] D. C. Knapp, S. Serva, J. D'Onofrio, A. Keller, A. Lubys, A. Kurg, M. Remm and J. W. Engels, "Fluoride-Cleavable, Fluorescently Labelled Reversible Terminators: Synthesis and Use in Primer Extension," *Chem. Eur. J.*, vol. 17, pp. 2903-2915, 2011.

[220] J. Ju, D. H. Kim, L. Bi, Q. Meng, X. Bai, Z. Li, X. Li, M. S. Marma, S. Shi, J. Wu, J. R. Edwards, A. Romu and N. J. Turro, "Four-color DNA sequencing by synthesis using cleavable fluorescent nucleotide reversible terminators," *Proc. Natl Acad. Sci. USA*, vol. 103, no. 52, pp. 19635-19640, 2006.

[221] J. Guo, N. Xu, Z. Li, S. Zhang, J. Wu, D. H. Kim, M. S. Marma, Q. Meng, H. Cao, X. Li, S. Shi, L. Yu, S. Kalachikov, J. J. Russo, N. J. Turro and J. Ju, "Four-color DNA sequencing with 3'-O-modified nucleotide reversible terminators and chemically cleavable fluorescent dideoxynucleotides," *Proc. Natl Acad. Sci. USA*, vol. 105, no. 27, pp. 9145-9150, 2008.

[222] F. Chen, M. Dong, M. Ge, L. Zhu, L. Ren, G. Liu and R. Mu, "The History and Advances of Reversible Terminators Used in New

Generations of Sequencing Technology," *Genom. Proteom. Bioinform.*, vol. 11, pp. 34-40, 2013.

[223] S. R. Head, H. K. Komori, S. A. LaMere, T. Whisenant, F. V. Nieuwerburgh, D. R. Salomon and P. Ordoukhanian, "Library construction for next-generation sequencing: Overviews and challenges," *Biotechniques,* vol. 56, pp. 61-77, 2014.

[224] P. Gyarmati, Y. Song, J. Hallman and M. Kaller, "Chemical fragmentation for massively parallel sequencing library preparation," *J. Biotechnol.*, vol. 168, no. 1, pp. 95-100, 2013.

[225] E. Knierim, B. Lucke, J. M. Schwarz, M. Schuelke and D. Seelow, "Systematic Comparison of Three Methods for Fragmentation of Long-Range PCR Products for Next Generation Sequencing," *PLoS One,* vol. 6, no. 11, p. e28240, 2011.

[226] L. Aigrain, Y. Gu and M. A. Quail, "Quantitation of next generation sequencing library preparation protocol efficiencies using droplet digital PCR assays - a systematic comparison of DNA library preparation kits for Illumina sequencing," *BMC Genomics,* vol. 17, pp. 1-11, 2016.

[227] A. Adey, H. G. Morrison, Asan, X. Xun, J. O. Kitzman, E. H. Turner, B. Stackhouse, A. P. MacKenzie, N. C. Caruccio, X. Zhang and J. Shendure, "Rapid, low-input, low-bias construction of shotgun fragment libraries by high-density in vitro transposition," *Genome Biol.,* vol. 11, p. R119, 2010.

[228] F. Vigneault, A. M. Sismour and G. M. Church, "Efficient microRNA capture and bar-coding via enzymatic oligonucleotide adenylation," *Nat. Methods,* vol. 5, pp. 777-779, 2008.

[229] A. H. El-Sagheer and T. Brown, "Synthesis and Polymerase Chain Reaction Amplification of DNA Strands Containing an Unnatural Triazole Linkage," *J. Am. Chem. Soc.,* vol. 131, no. 11, pp. 3958-3964, 2009.

[230] A. Routh, P. Ji, E. Jaworski, Z. Xia, W. Li and E. J. Wagner, "Poly(A)-ClickSeq: click-chemistry for next-generation 3'-end sequencing without RNA enrichment or fragmentation," *Nucleic Acids Res.,* vol. 45, no. 12, p. e112, 2017.

[231] E. Jaworski, R. M. Langsjoen, B. Mitchell, B. Judy, P. Newman, J. A. Plante, K. S. Plante, A. L. Miller, Y. Zhou, D. Swetnam, S. Sotcheff, V. Morris, N. Saada, R. R. Machado, A. McConnell, S.

G. Widen, J. Thompson, J. Dong, P. Ren and A. L. Routh, "Tiled-ClickSeq for targeted sequencing of complete coronavirus genomes with simultaneous capture of RNA recombination and minority variants," *eLife,* vol. 10, p. e68479, 2021.

[232] S. Mikutis, M. Gu, E. Sendinc, M. E. Hazemi, H. Kiely-Collins, D. Aspris, G. S. Vassiliou, Y. Shi, K. Tzelepis and G. J. L. Bernardes, "meCLICK-Seq, a Substrate-Hijacking and RNA Degradation Strategy for the Study of RNA Methylation," *ACS Cent. Sci.,* vol. 6, pp. 2196-2208, 2020.

[233] Z. Stasevskij, P. Gibas, J. Gordevicius, E. Kriukiene and S. Klimasauskas, "Tethered Oligonucleotide-Primed Sequencing, TOP-Seq: A High-Resolution Economical Approach for DNA Epigenome Profiling," *Mol. Cell,* vol. 65, pp. 554-564, 2016.

[234] W. Lohmann, "Halogen-Substitution Effect on the Optical Absorption Bands of Uracil," *Z. Naturforsch ,* vol. 29c, pp. 493-495, 1974.

[235] G. R. Abel, J. Z. Calabrese, J. Ayco, J. E. Hein and T. Ye, "Measuring and suppressing the oxidative damage to DNA during Cu(I)-catalyzed azide-alkyne cycloaddition," *Bioconjugate Chem.,* vol. 27, no. 3, p. 698–704, 2016.

[236] M. Dizdaroglu, P. Jaruga and H. R. M. Birincioglu, "Free Radical-Induced Damage To DNA: Mechanisms and Measurements," *Free Radic. Biol. Med.,* vol. 32, no. 11, pp. 1102-1115, 2002.

[237] P. Y. Liu, N. Jiang, J. Zhang, X. Wei, H. H. Lin and X. Q. Yu, "The Oxidative Damage of Plasmid DNA by Ascorbic Acid Derivatives in vitro: The First Research on the Relationship between the Structure of Ascorbic Acid and the Oxidative Damage of Plasmid DNA," *Chem. Biodivers.,* vol. 3, pp. 958-966, 2006.

[238] P. J. Rothwell and G. Waksman, "Structure and mechanism of DNA polymerases," *Adv Protein Chem,* vol. 71, pp. 401-440, 2005.

[239] M. A. Reeve and C. W. Fuller, "A novel thermostable polymerase for DNA sequencing," *Nature,* vol. 376, pp. 796-797, 1995.

[240] J. Medžiūnė, Ž. Kapustina, S. Žeimytė, J. Jakubovska, R. Sindikevičienė, I. Čikotienė and A. Lubys, "Advanced preparation

of fragment libraries enabled by oligonucleotide-modified 2′,3′-dideoxynucleotides," *Commun Chem* , vol. 5, no. 34, pp. 1-8, 2022.

[241] Ž. Kapustina, J. Medžiūnė, G. Alzbutas, I. Rokaitis, K. Matjošaitis, G. Mackevičius, S. Žeimytė, L. Karpus and A. Lubys, "High-resolution microbiome analysis enabled by linking of 16S rRNA gene sequences with adjacent genomic contexts," *Microbial Genomics,* vol. 7, p. 000624, 2021.

[242] Ž. Kapustina, J. Medžiūnė, V. Dubovskaja, K. Matjošaitis, S. Žeimytė and A. Lubys, "Sensitive and accurate analysis of gene expression signatures enabled by oligonucleotide-labeled cDNA," *RNA Biology,* p. DOI: 10.1080/15476286.2022.2078093, 2022.

[243] I. s. a. a. o. t. h. genome, "International Human Genome Sequencing Consortium," *Nature,* vol. 409, p. 860–921, 2001.

SANTRAUKA

ĮVADAS

Nukleotidai yra vienų svarbiausių biopolimerų nukleorūgščių statybiniai blokai. Natūralios nukleotidų modifikacijos turi didelę svarbą bioprocesų reguliavime ar sąsajų su ligų atsiradimu ir tolimesne jų progresija. [1] Diagnostikos, bioprocesų stebėjimo ir nukleorūgščių terapijos vystymasis lėmė modifikavimo ir žymėjimo nukleorūgščių vis didėjantį poreikį. Modifikuoti nukleotidai tapo pagrindu šiuolaikinėse molekulinės biologijos taikymo srityse, tokiose kaip fluorescencinė *in-situ* hibridizacija (FISH) [2], sisteminga ligandų evoliucija eksponentiškai praturtinant (SELEX) [3], naujos kartos sekoskaita (NKS) [4] ir daugelyje kitų, neatsiejant vaistų [5] ir RNR vakcinų [6] kūrimo. Plačios panaudojimo galimybės ir jų augimas sukūrė poreikį modifikuotų nukleotidų įvairovei, modifikacijas siejant su konkrečiu taikymu.

Modifikuoti 2’-deoksinukleozidų 5’-trifosfatai (dNTP) tapo diagnostikos ir bioprocesų eigos stebėjimo pagrindu. 2’,3’-Dideoksinukleozidų 5‘-trifosfatai (ddNTP) sulaukė mažiau dėmesio ir yra menkai ištyrinėti. Modifikuotų 2’,3’-dideoksinukleozidų (ddN) ir ddNTP-ų pritaikymo pradžia siejama su antivirusinių [7], priešvėžinių [8] vaistų kūrimu, o vėliau ir nukleorūgščių fermentiniu žymėjimu [10]. ddNTP-ų pritaikymo revoliuciniu proveržiu laikomas Sanger sekoskaitos metodo sukūrimas, kuriame fluorescencinėmis žymėmis modifikuoti ddNTP-ai buvo naudojami kaip DNR replikacijos terminatoriai. [11] Sanger sekoskaita leido nustatyti žmogaus genomo seką, šis tyrimas užtruko 10 metų ir reikalavo didelių finansinių investicijų. [243] Poreikis pigiau, greičiau ir tiksliai analizuoti nukleorūgščių sekas paskatino esamų sekos nustatymo metodų tobulinimo tyrimus. Naujos kartos sekoskaitos (NKS) sukūrimas buvo revoliucinis proveržis nukleorūgščių sekų analizėje ir tapo kasdieniu įrankiu šiuolaikiniuose biologijos ir biomedicinos moksluose.

Nepriklausomai nuo naudojamos sekos nustatymo platformos, visos nukleorūgštys turi būti atitinkamai paruoštos sukonstruojant fragmentų bibliotekas. Toks išankstinis nukleorūgščių apdorojimas yra daugiapakopė procedūra. Siekdami gerinti mėginių paruošimo NKS procesus mokslininkai kuria alternatyvius metodus, vienas tokių pavyzdžių – fermentinių etapų pakeitimas cheminiais pritaikant „click“ reakciją. 3’-Azido grupe žymėtų nukleotidų fermentinis įterpimas ir tolimesnė variu (I) katalizuojama „click“ (CuAAC) reakcija su alkino grupę turinčiu oligonukleotidu išsivystė į cheminio ligavimo metodą. [12, 13] Paprastai modifikuotos nukleorūgštys

konstruojamos taip, kad imituotų natūralias, tačiau tuo pačiu pranoktų specifinėmis funkcijomis kaip: atsparumu nukleazėms, lydymosi temperatūros padidėjimu ir kt. Triazolas, suformuojamas CuAAC reakcijos metu, yra gerai žinoma nukleorūgščių fosfatinės jungties mimikinė grupė. [14] Sukurta cheminio ligavimo alternatyva buvo sėkmingai pritaikyta siekiant supaprastinti NKS bibliotekų ruošimą, tačiau turėjo tam tikrų trūkumų. Replikacija per nenatūraliai sujungtą nukleorūgšties grandinę komplementarios grandinės sintezės metu nebuvo tiksli dėl atsirandančių delecijų. Be to, buvo stebimas vario katalizuojamas DNR skilimas, sąlyginai mažas CuAAC reakcijos ir perskaitymo per nenatūralią jungtį efektyvumas. [15, 16] Nepaisant to, buvo įrodytas technologinis įgyvendinamumas, atveriantis tolesnio tobulinimo perspektyvas.

Didėjant modifikuotų nukleotidų ir nukleorūgščių pritaikymų įvairovei, kartu didėja efektyvių, sąlyginai paprastų ir universalių sintezės metodikų poreikis. Nesudėtingam įvairių žymių įvedimui į modifikuotų dNTP ir ddNTP struktūras neretai yra suformuojamos arba prijungiamos atitinkamos funkcinės grupės, pavyzdžiui, amino ar azido, kurios toliau gali būti lengvai modifikuojamos. Siekiant išlaikyti modifikuotų nukleotidų, kaip substratų polimerazėms savybes, svarbu tinkamai parinkti molekulės dizainą. Tyrimai rodo, kad didelę įtaką modifikuotų nukleotidų efektyviam įterpimui turi jungtukas, siejantis nukleobazę su žyme. [17] Dėl plataus modifikuotų dNTP panaudojimo, yra sukurta nemažai jų sintezės metodikų. [18] Tuo tarpu modifikuoti ddNTP yra siauriau tyrinėjami, o minėti metodai nėra pritaikyti jų paruošimui. Dažniausiai žymėti dNTP ir ddNTP yra sintetinami atliekant penkių stadijų procedūrą. Toks kelias yra pakankamai ilgas ir tikslinių produktų išeigos nėra didelės. [11]

Darbo tikslas

Disertacijos tikslas buvo sukurti ir susintetinti oligonukleotidais modifikuotus 2’-deoksi- ir 2’,3’-dideoksinukleozidų 5’-trifosfatus (OTDN ir OTDDN), kurie būtų substratai polimerazėms ir pasižymėtų perskaitymo per nenatūralią jungtį galimybėmis.

Tikslui įgyvendinti išsikelti šie uždaviniai:

1. Išvystyti supaprastintą, alternatyvų sintezės metodą tiksliniams, azido grupę turintiems, 2’-deoksi-, 2’,3’-dideoksinukleotidams (d^{N3}NTP ir dd^{N3}NTP) gauti ir metodą palyginti su klasikiniu sintezės keliu.
2. Pritaikyti dd^{N3}NTP-ų reakcijų sąlygas didesnės skalės sintezei.

3. Ištirti $dN^{N3}TP$-ų ir $dd^{N3}NTP$-ų kaip substratų savybes fermentiniais nukleorūgščių sintezės būdais (PGR ir/arba PEX).

4. Ištirti CuAAC ir SPAAC reakcijų efektyvumo priklausomybę nuo substrato prigimties.

5. Susintetinti oligonukleotidais modifikuotus 2'-deoksi-, 2',3'-dideoksinukleotidus (OTDN ir OTDDN) ir įvertinti jų, kaip substratų fermentams savybes pradmens pratęsimo reakcijos ir nuskaitymo per nenatūralią jungtį metu.

6. Siekiant supaprastinti NKS bibliotekų paruošimą, pritaikyti OTDDN-us nukleorūgščių žymėjimui.

7. Įvertinti fermentinio perskaitymo per nenatūralias jungtis efektyvumą.

Iš gautų rezultatų buvo suformuluoti šie **ginamieji teiginiai**:

1. Azido grupe žymėti nukleotidai gali būti gauti naudojant tradicinę penkių stadijų, arba alternatyvią trijų stadijų, sintezės strategijas.

2. Trijų stadijų sintezės strategija gali būti taikoma ir yra efektyvi azido-grupę turinčių 2',3'-dideoksinukleotidų sintezei geromis išeigomis.

3. Azido grupę turintys 2'-deoksi- ir 2',3'-dideoksinukleotidai yra substratai įvairioms DNR polimerazėms pradmenų pratęsimo ir polimerazės grandininės reakcijų metu.

4. CuAAC ir SPAAC reakcijų greičiai priklauso nuo azido grupę turinčių substratų prigimties.

5. Oligonukleotidais modifikuoti 2'-deoksi- ir 2',3'-dideoksinukleotidai yra DNR polimerazių substratai.

6. Empiriškai parinktas jungtukų, siejančių nukleobazę su oligonukleotidu, dizainas yra biologiškai suderinamas, o taip sujungtos nukleorūgštys gali būti naudojamos kaip matricos komplementarios grandinės sintezei.

7. Nukleorūgščių sintezės terminavimas oligonukleotidu modifikuotais nukleotidais supaprastina NKS bibliotekų paruošimo procesą, apjungiant fragmentavimo ir adapterių ligavimo žingsnius į vieną ir įgalina šių junginių panaudojimą DNR ir kDNR sekoskaitoje.

Siekiant oligonukleotidais modifikuotus nukleotidus (OTDN ir OTDDN) pritaikyti NKS bibliotekų ruošime, jie turi atitikti du pagrindinius kriterijus. Pirmiausiai, nepaisant steriškai didelės modifikacijos, parinktas molekulės dizainas turi įgalinti fermentinį OTDN ir OTDDN įterpimą PEX metu. Susintetinti fragmentai turi tarnauti kaip matrica leidžianti tikslų ir efektyvų perskaitymą per nenatūralią jungtį. Abiem kriterijams pasiekti jungtuko, tarp nukleobazės ir oligonukleotido, struktūra yra labai svarbi. Parenkant tinkamą jungtuko dizainą reikia atsižvelgti į kelis aspektus: modifikacijos padėtis (C5 – pirimidinuose ir C7 – 7-deazapurinuose); formuojamo C-C ryšio prigimtis (trigubą ryšį prie nukleobazės geriau atpažįsta polimerazės); reaktyvių funkcinių grupių parinkimas ir įvedimas; galiausiai jungtuko ilgio ir struktūros parinkimas ir suformavimas. Remiantis minėtais aspektais buvo sukurti ir susintetinti įvairiais jungtukais modifikuoti d(d)NTPai, kurie CuAAC reakcijos metu sujungti su oligonukleotidais. Šių oligonukleotidais modifikuotų nukleotidų jungtuko formavimui buvo pasirinktas propargilaminas. Azido grupė įvesta naudojant įvairaus ilgio ir cheminės prigimties azido-NHS esterius. Fermentinio įterpimo ir perskaitymo per nenatūralią jungtį efektyvumo priklausomybės nuo substrato prigimties įvertinimui buvo naudojami skirtingus jungtukus turintys OTDN ir OTDDN.

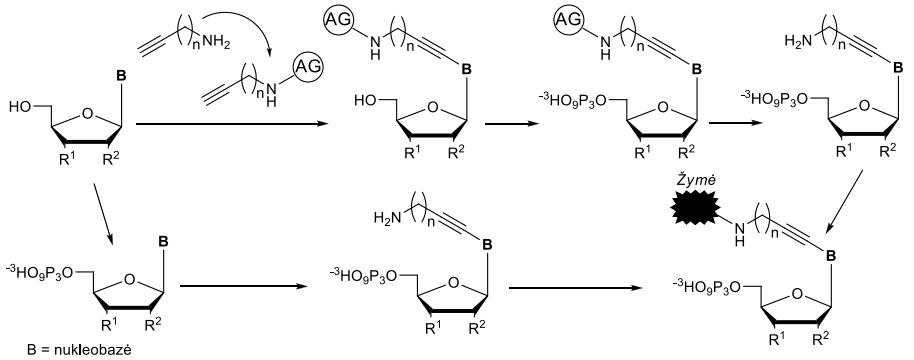
Dešimtame skyriuje bus aprašoma azido grupę turinčių 2’-nukleotidų ir oligonukleotidais modifikuotų 2’-nukleotidų sintezė ir panaudojimas.

10. Oligonukleotidais modifikuotų 2’-deoksinukleotidų sintezė ir panaudojimas

10.1 Azido grupę turinčių 2’-deoksinukleotidų sintezė klasikiniu metodu

Yra dvi žinomos sintezės strategijos, skirtos nukleobazėje modifikaciją turintiems nukleotidams gauti. Pirmoji – labai plačiai taikoma, klasikinė, penkių stadijų metodika (Schema 1a). Pirmajame sintezės etape dažnai naudojamos paladžio katalizuojamos C-C ryšio sudarymo reakcijos tarp funkcionalizuotų alkanų, alkenų arba alkinų ir heterociklinėje bazėje halogeną turinčių nukleozidų. Čia svarbu žinoti, kad jungtuko struktūroje esančios nukleofilinės prigimties funkcinės grupės (tokios, kaip amino- ar hidroksi-) turi būti blokuojamos apsauginėmis grupės. Toliau, atliekama trifosforilinimo reakcija ir sintetinami atitinkami nukleotidai. Paskutiniame etape, atlikus funkcinių grupių deblokavimą, yra įvedama tikslinė žymė (Schema 1a). Šios

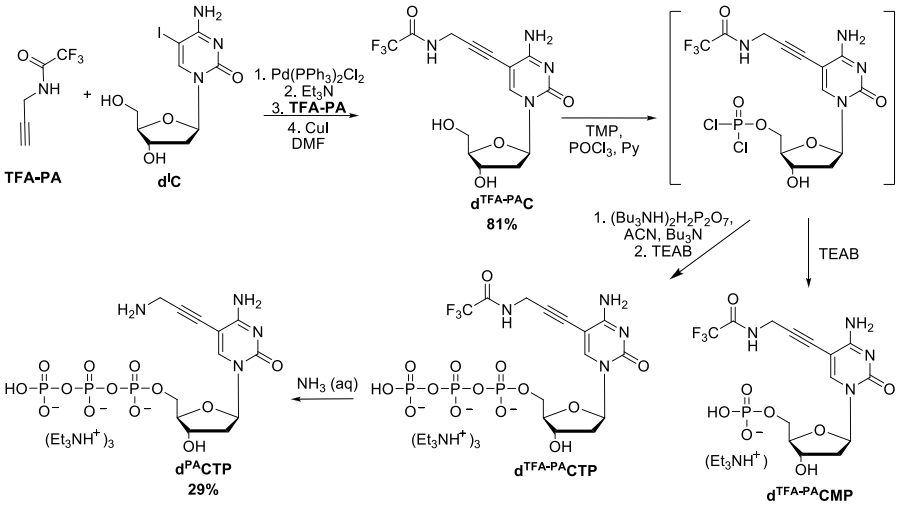
sintezės strategijos limituojanti stadija yra modifikuoto nukleozido trifosforilinimas, kurio efektyvumas priklauso nuo modifikacijos prigimties. Atliekant šiuo keliu nukleobazėje modifikuotų nukleotidų sintezę, dažnai gaunamos mažos tikslinių produktų išeigos. Tačiau yra žinoma ir alternatyvi, trijų stadijų metodologija. Pirmiausiai, halogeninti nukleozidai yra trifosforilinami, toliau seka jungtuko formavimas pritaikant vandeninėje terpėje atliekamas Pd-katalizuojamas C-C ryšio formavimo reakcijas (pvz., Sonogashira) ir galiausiai yra įvedama tikslinė žymė (Schema 1b). Dėl daugeliu atveju gaunamų didesnių tikslinių produktų suminių išeigų ir laiko prasme greitesnės žymėtų nukleotidų sintezės, ši metodologija, yra vis plačiau taikoma. Azido grupę turintiems nukleotidams gauti buvo išbandytos ir palygintos abi minėtos metodikos.



Schema 1. Nukleobazėje modifikuotų nukleotidų sintezės strategijos, a) klasikinė penkių stadijų sintezė, b) alternatyvi trijų stadijų sintezė. AG – apsauginė grupė.

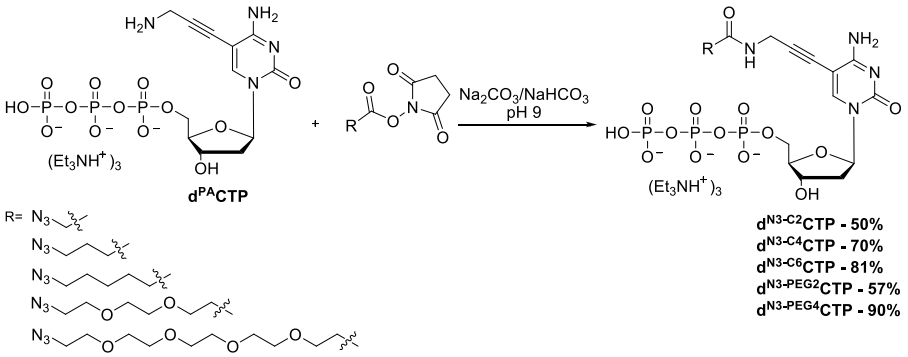
Klasikinė sintezės strategija buvo pritaikyta propargilaminu žymėto dCTP sintezėje. Atlikus Sonogashira reakciją tarp 5-I-dC ($d^{I}C$) ir 2,2,2-trifluoro-N-(prop-2-in-1-il)acetamido (TFA-PA) buvo gautas tikslinis nukleozidas $d^{TFA\text{-}PA}C$ 81% išeiga (Schema 2). Tačiau pastebėtas ir galimas produkto degradavimas. Atitinkamo trifosfato sintezei ($d^{TFA\text{-}PA}CTP$) buvo išbandytos kelios skirtingos reakcijų sąlygos, keičiant fosforo oksichlorido ($POCl_3$) ir bazės kiekius, reakcijos temperatūrą. Geriausi rezultatai gauti reakciją atliekant -20 – -15 °C temperatūroje, esant 2 ekv. piridino (Py) ir 4 ekv. $POCl_3$. Konversija į tikslinį $d^{TFA\text{-}PA}CMP$ siekė 64%, suminis priemaišų kiekis – 30%. Tributilamoniopirofosfato (TBAPF) mišinys ($(Bu_3NH)_2H_2P_2O_7$ + Bu_3N + ACN) buvo naudojamas antroje, trifosforilinimo reakcijos stadijoje. Neatliekant tarpinio produkto gryninimo, toliau buvo atliekama amino grupės deblokavimo reakcija naudojant amoniako tirpalą. Tikslinis produktas $d^{TFA\text{-}PA}CTP$ išgrynintas pritaikant jonų mainų chromatografiją 29% išeiga (Schema 2). Atlikus kelis sintezės metodologijos pakartojimus ir skalės didinimo

eksperimentus pastebėta, kad gauti rezultatai neatsikartoja (skirtumai stebimi trifosforilinimo reakcijoje).



Schema 2. d^{PA}CTP sintezė pritaikant klasikinę penkių stadijų metodologiją.

Azido grupė buvo įvedama reaguojant d^{PA}CTP su įvairiais azido-NHS esteriais. Susintetinti įvairaus ilgio ir cheminės prigimties jungtukus turintys d^{N3}CTP (Schema 3).



Schema 3. Azido grupę turinčių dCTP sintezė.

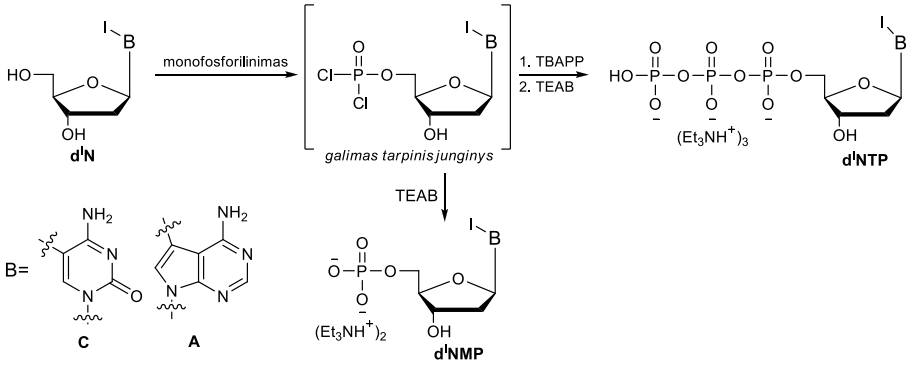
10.2 Azido grupę turinčių 2'-deoksinukleotidų sintezė alternatyviu sintezės metodu

Alternatyvi sintezės strategija nukleobazėje modifikuotiems nukleotidams gauti (Schema 1b) siejama su vandeninėje terpėje tirpių trifenilfosfino ligandų sukūrimu ir pritaikymu Pd-katalizuojamose reakcijose (Sonogashira, Heck,

Suzuki, kt.). Šio metodo pranašumas – universalesnė ir trumpesnė trijų stadijų sintezė. Trifosforilinimo reakcija atliekama su halogenintais nukleozidais, kurie yra inertiškesni lyginant su nukleofilinės prigimties modifikacijas turinčiais junginiais. Sekanti Pd-katalizuojama reakcija atliekama vandeninėje terpėje. Tokiu būdu, vos dviejų sintezės stadijų metu, gaunami reaktyvią grupę turintys nukleotidai tinkami tolimesniam žymės įvedimui.

10.2.1 Jodintų 2'-deoksinukleozidų fosforilinimo reakcijos sąlygų optimizavimas

Atlikus literatūros analizę, halogenintų nukleozidų trifosforilinimo stadija buvo identifikuota kaip limituojanti. Dažniausiai yra taikomos standartinės sintezės sąlygos nepriklausomai nuo nukleozido prigimties ar gaunamos išeigos. Trifosfatų sintezė susideda iš dviejų etapų: aktyvuotų monofosfatų ar tarpinių junginių suformavimo ir trifosfato sintezės. Limituojanti stadija yra monofosforilintų nukleozidų suformavimas. Atitinkamas trifosfatas yra sėkmingai gaunamas tolimesniame žingsnyje naudojant pirofosfato druskas. Šio darbo tikslas buvo optimizuoti paprastą, nereikalaujančią specifinių ar brangių reagentų, vieno žingsnio arba kitaip dar vadinamą vieno-indo sintezės metodiką (Schema 4).



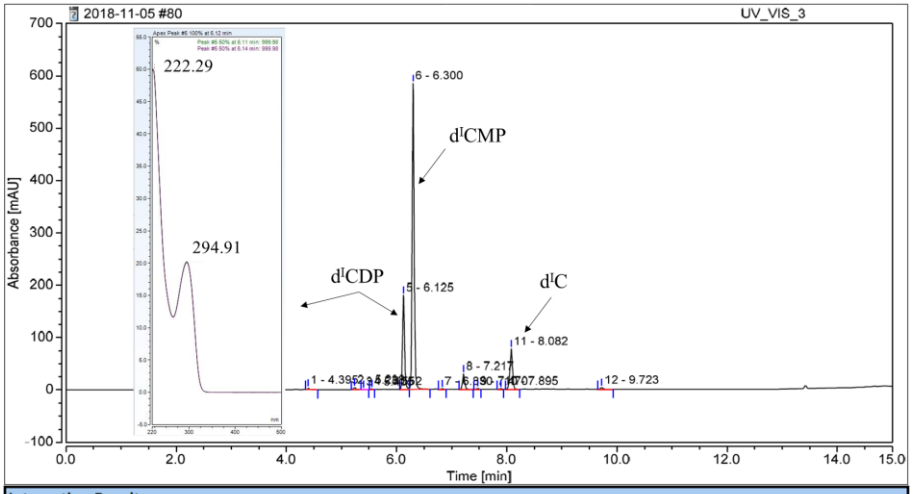
Schema 4. $d^{I}NMP$ ir $d^{I}NTP$ sintezė.

Kaip jau minėta anksčiau, nukleozidų 5'-trifosfatų sintezėje limituojanti stadija yra aktyvaus monofosfato suformavimas. Tik pasiekus gerą konversiją į aktyvų tarpinį junginį gali būti atliekamas antrasis sintezės žingsnis – atitinkamo trifosfato formavimas (Schema 4). Atsižvelgiant į tai, toliau buvo optimizuojamos monofosforilinimo reakcijų sąlygos $d^{I}CMP$ ir $d^{I}AMP$ gauti.

Pirmiausiai buvo optimizuojamos $d^{I}CMP$ sintezės sąlygos. Naudojant fosforo oksichloridą ($POCl_3$) ir reakciją atliekant kambario temperatūroje buvo stebimas lėtas produkto ir įvairių priemaišų susidarymas (Lentelė 1, Nr. 1 – 2). Ilginant reakcijos laiką, konversija į tikslinį monofosfatą nekito. Siekiant sumažinti priemaišų susidarymą tolimesnės reakcijos buvo

atliekamos žemesnėje temperatūroje (0 – 5 °C). Padidinus $POCl_3$ kiekį (iki 3 ekv.) buvo gauta vidutiniška (44%) konversija į tikslinį produktą (Lentelė 1, Nr. 3 – 4).

Iš mokslinės literatūros duomenų yra žinoma, kad bazės priedas neretai sąlygoja geresnes konversijas į tikslinius monofosfatus. Dažniausiai fosforilinimo reakcijose naudojamos bazės yra tributilaminas (TBA) ir protonų kempinė (PS), rečiau - piridinas (Py). Taip pat yra žinoma, kad bazės priedas sumažina tam tikrų priemaišų, kaip glikozidinio ryšio skilimo produktų, susidarymą. Reakcijoje naudojant po 2 ekv. TBA ir $POCl_3$ buvo stebimas greitas specifinių priemaišų susidarymas (Lentelė 1, Nr. 6). Tuo tarpu padidinus abiejų komponentų kiekius produkto susidarymas buvo greitesnis nei prieš tai minėtų priemaišų, o po 30 min buvo gauta 56% konversija į tikslinį d^{I}CMP (Lentelė 1, Nr. 7 ir Paveikslas 1). Atliekant reakcijos mišinio masių spektrometrijos (MS) analizę be produkto masės buvo identifikuotas 3',5'-difosfatas (3',5'-d^{I}CDP), susidarantis dėl naudojamo $POCl_3$ pertekliaus.



Integration Results

No.	Peak Name	Retention Time min	Area mAU*min	Height mAU	Relative Area %	Relative Height %	Amount n.a.
5		6.125	7.420	180.249	19.26	20.29	n.a.
6		6.300	25.089	584.627	65.11	65.82	n.a.
11		8.082	4.020	77.060	10.43	8.68	n.a.

Paveikslas 1. d^{I}C monofosforilinimo reakcijos mišinio HPLC analizė po 1 val. Reakcija atlikta pagal sąlygas Nr. 7 (Lentelė 1).

Kadangi panaudojant fosforo oksichloridą buvo pasiekta tik vidutinė (65%, Lentelė 1, Nr. 7) konversija į tikslinį monofosfatą, vėliau buvo išbandytas kitas fosforilinimo agentas – pirofosforilchloridas ($P_2O_3Cl_4$).

Lentelė 1. $d^{I}C$ monofosforilinimo reakcijos sąlygų optimizavimas.

Nr.	Fosforilinimo agentas	Fosforilinomo agento kiekis, ekv.	Tirpiklis	Bazė, ekv.	Koncentracija, M	Laikas	Konversija į $d^{I}CMP$, %	$d^{I}C$, %
1[a]	$POCl_3$	1.2	TMP	-	0,1	1 val.	19	63
2[a,b]		1.2	TMP	-	0,1	1 val.	19	68
3[c]		1.2	TMP	-	0,1	2 val. 4 val.	33 35	63 59
4		2	TMP	-	0,2	1 val. 3 val.	22 38	74 56
5		3	TMP	-	0,2	1 val. 3 val.	31 44	62 40
6		2	TMP	TBA, 2	0,2	1 val. 3 val.	27 29	26 22
7		3	TMP	TBA, 3	0,2	30 min. 1 val. 3 val.	56 65 58	28 11 2
8		3	TMP	TBA, 3	0,3	3 val.	≤ 3	97
9		3	TMP	TBA, 3	0,4	3 val.	≤ 3	98
10	$P_2O_3Cl_4$	2	TMP	-	0,1	3 val.	geliavo	99
11		3	TMP	TBA, 2	0,1	3 val.	geliavo	98
12		2	ACN	-	0,1	30 min. 1 val.	56 70	31 9
13[d]		3	ACN	-	0,1	1 val. 3 val.	53 48	5 2
14		2	ACN	TBA, 1.3	0,1	10 min. 1 val.	59 60	26 25
15		3	ACN	TBA, 2	0,1	10 min. 1 val.	67 70	16 3
16		3	ACN	TBA, 3	0,1	10 min.	*	-
17		3	ACN	Py, 2	0,1	10 min. 3 val.	57 60	24 4
18		3	ACN	PS, 2	0,1	3 val.	66	5
19[e]		**2**	**ACN**	**-**	0,1	**15 min.**	**68**	**7**

Visos reakcijos atlikto 0 °C temperatūroje, nebent nurodyta kitaip. a – reakcija atlikta kambario temperatūroje; b – reakcijoje pridėtas 0,1 ekv. vandens; c – reakcija atlikta -5 – (-10) °C temperatūrą; d – specifinės priemaišos susidarymas; e – sintezės didesnėje skalėje rezultatai (reakcijoje naudojamas 1 g pradinio nukleozido), * – stebimas įvairių priemaišų gausus susidarymas, apsunkinantis tikslinio junginio identifikavimą.

Minėtas fosforilinimo agentas buvo panaudotas nukleozidų polifosfatų Sintezėje dar 1960 metais, tačiau nesusilaukė didelio populiarumo. Atliekant reakcijas trimetilfosfate (TMP), pastebėtas greitas reakcijos mišinių

geliavimas (Lentelė 1, Nr. 10 – 11). Šis efektas nebuvo stebimas reakcijas atliekant ACN. Naudojant 2 ekv. $P_2O_3Cl_4$ po valandos gauta 70% konversija į tikslinį d^ICMP (Lentelė 1, Nr. 12) kartu su 15% šalutinio 3',5'-d^ICDP. Didinant fosforilinimo agento kiekį reakcija vyko greičiau, tačiau susidarė daugiau priemaišos (po 1 val. – 53% d^ICMP ir 42% - priemaišos) (Lentelė 1, Nr. 13). Siekiant išsiaiškinti bazės pridėjimo įtaka reakcijos eigai išbandytos trys bazės (TBA, Py ir PS) ir įvairūs $P_2O_3Cl_4$, bazės kiekiai (Lentelė 1, Nr. 14 – 18). Daugeliu atveju bazės pridėjimas greitino reakciją, jau po 10 min gauta 67% konversija į tikslinį monofosfatą (Lentelė 1, Nr. 15). Tačiau kartu su reakcijos greičio padidėjimu, susidarė daugiau priemaišų (Lentelė 1, Nr. 16). Tuo tarpu pačios bazės prigimtis neturėjo didelės įtakos nei konversijai į produktą, nei priemaišų įvairovei (Lentelė 1, Nr. 17 – 18).

Kadangi tikslas buvo susintetinti 5-I-2'-deoksicitidino 5'-trifosfatą, optimaliomis monofosforilinimo reakcijų sąlygomis (pasirinktos Nr. 12, Lentelė 1), buvo suformuotas aktyvus monofosfatas. Reakciją atliekant didesnėje skalėje po 15 min konversija į d^ICMP siekė 68% (Lentelė 1, Nr. 19). Vieno-indo d^ICTP sintezės antroje stadijoje buvo naudojamas TBAPP mišinys ($(Bu_3NH)_2H_2P_2O_7$ + Bu_3N + ACN), reakcija atliekama 20 min kambario temperatūroje. Konversija į tikslinį d^ICTP siekė 63%, o produktas išgrynintas jonu mainų chromatografijos metodu 54% išeiga.

Optimizavus d^ICMP sintezės sąlygas ir susintetinus d^ICTP vėliau buvo optimizuojamos d^IAMP reakcijos sąlygos. d^IA monofosforilinimo reakcijai, kaip ir d^IC atveju, $POCl_3$ nebuvo efektyvus fosforilinimo agentas. Nepriklausomi nuo naudoto $POCl_3$ kiekio, bazės ar reakcijos temperatūros konversija į d^IAMP geriausiomis sąlygomis siekė tik 38% (Lentelė 2 Nr. 1 – 4). Todėl tolimesni eksperimentai buvo atliekami su pirofosforilchloridu. Kaip ir d^ICMP sintezės atveju, reakcijas atliekant TMP tirpiklyje, nepriklausomai nuo kitų naudotų reakcijos komponentų ar jų kiekių, buvo stebimas reakcijos mišinių geliavimas (Lentelė 2 Nr. 5 – 9). Reakcijoje naudojant 3 ekv. $P_2O_3Cl_4$ su papildomu 1 ekv. priedu po valandos, konversija į tikslinį produktą siekė 66% (Lentelė 2 Nr. 10). Priešingai nei d^IC monofosforilino atveju, d^IAMP sintezei bazė neturėjo reikšmingos įtakos konversijai į produktą ar reakcijos greičiui. (Lentelė 2 Nr. 11 – 15).

d^IATP vieno-indo reakcijai buvo pasirinktos reakcijos sąlygos Nr. 10 (Lentelė 2). Atliekant d^IA monofosforilinimo reakciją ir formuojant aktyvų monofosfatą buvo pastebėtas reakcijos greičio padidėjimas, kaip ir d^ICMP sintezės atveju, ir po valandos konversija į d^IAMP siekė 76% (Lentelė 2 Nr. 16). d^IATP sintezės antroje stadijoje buvo naudojamas TBAPP mišinys, reakcija atliekama 30 min kambario temperatūroje. Konversija į d^IATP siekė 71%, o po gryninimo gauto produkto išeiga buvo 65%.

Lentelė 2. d^lA monofosforilinimo reakcijos sąlygų optimizavimas.

Nr.	Fosforilinimo agenas	Fosforilinimo agento kiekis, ekv.	Tirpiklis	Bazė, ekv.	Reakcijos laikas	Konversija į d^lAMP, %	d^lA, %
1	$POCl_3$	1.2 + 1[a]	TMP	-	2 val.	18	77
2		1.2 + 1[a]	TMP	TBA, 2	1 val.	29	65
					2 val.	38	43
3[b]		1.2 + 2[c]	TMP	TBA, 2	3 val.	26	71
4		1.2 + 1[a]	TMP	PS, 1.5	2 val.	≤ 3	89
5	$P_2O_3Cl_4$	1.2	TMP	TBA, 2	2 val.	geliavo	95
6		1.2	TMP	Py, 2	2 val.	geliavo	89
7		1.2	TMP	PS, 1	2 val.	geliavo	99
8[d]		1.2 + 1[a]	TMP	TBA, 2	2 val.	geliavo	97
9		1.2 + 1[a]	TMP	-	2 val.	geliavo	99
10		3 + 1[a]	ACN	-	5 min.	48	50
					1 val.	56	40
					2 val.	66	29
11		3	ACN	TBA, 2	5 min.	14	86
12		3 + 1[e]	ACN	Py, 2	5 min.	26	59
					3.5 val.	62	11
13		3	ACN	Py, 4	5 min.	20	78
14		3 + 3[e]	ACN	PS, 1	5 min.	42	57
					3.5 val.	53	42
15		3 + 3[f]	ACN	PS, 1 + 1[f]	30 min.	40	58
					1.5 val.	47	46
16[g]		**3**	**ACN**	**-**	**1 val.**	**76**	**17**

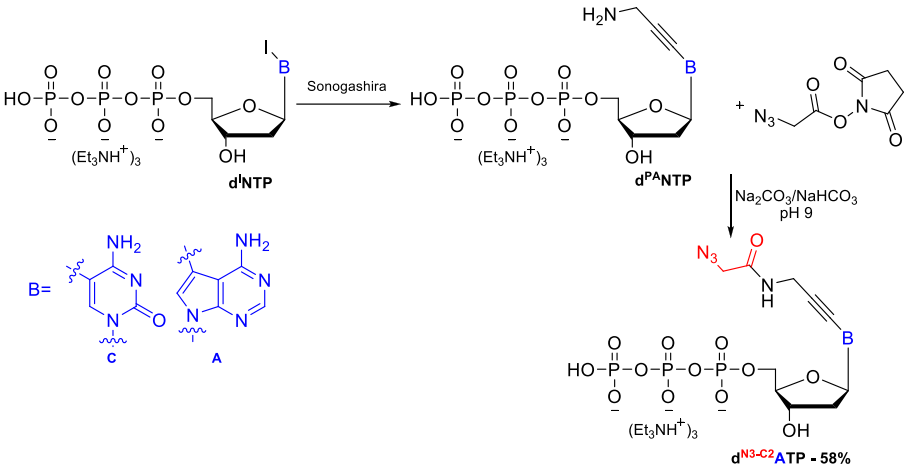
Visų reakcijų koncentracija buvo 0,1 M, reakcijos atliktos 0 °C temperatūroje, nebent nurodyta kitaip. a – po valandos pridėtas papildomas fosforilinimo agento kiekis; b – reakcija atlikta -11 – (-15) °C temperatūroje; c – kas valandą buvo pridedamas papildomas fosforilinimo agento kiekis; d – reakcija atlikta kambario temperatūroje; e – papildomas fosforilinimo agento kiekis pridėtas po 3 val.; f – papildomi reagentų kiekiai pridėti po 30 min.; g – sintezės didesnėje skalėje rezultatai (reakcijoje naudojamas 1 g pradinio nukleozido).

Optimizuojant d^lNMP reakcijos sąlygas buvo pastebėtos tam tikros tendencijos. Fosforo oksichloridas buvo prastas fosforilinimo agentas abiem nukleozidams. Reakcijos koncentracijos didinimas arba temperatūros mažinimas lėtino reakciją, iš dalies tai gali būti siejama su prastu d^lN tirpumu. TMP buvo netinkamas tirpiklis kai reakcijose naudojamas pirofosforilchloridas. Gerai konversijai į tikslinius monofosfatus pasiekti $P_2O_3Cl_4$ buvo tinkamesnis fosforilinimo agentas. d^lCMP sintezėje naudojant bazę buvo greičiau pasiekiama gera konversija į produktą, tačiau buvo

stebimas ir greitesnis priemaišų formavimasis. Gera konversija į tikslinį $d^{I}CMP$ pasiekta reikėjo naudojant 2 ekv. fosforilinimo agento, tuo tarpu $d^{I}AMP$ – 3 ekv. Reakcijas atliekant didesnėje skalėje abiem atvejais buvo stebimas greitesnis produkto susidarymas.

10.2.2 Amino ir azido grupę turinčių 2'-deoksinukleotidų sintezė

Po sėkmingos $d^{I}CTP$ ir $d^{I}ATP$ sintezės tolimesniam nukleotidų modifikavimui įvedant propargilaminą buvo atliekama Sonogashira reakcija (Schema 5). Literatūroje aprašomose metodikose tiksliniams junginiams susintetinti neretai naudojamos aukštos reakcijos temperatūros (kaip 70 °C). Tačiau tokiomis reakcijų sąlygomis trifosfatai skyla, taip apsunkindami tikslinių junginių gryninimą ir sąlygodami vidutines produktų išeigas. Tiksliniams $d^{PA}NTP$ susintetinti išbandytos kelios reakcijų sąlygos. Siekiant išvengti trifosfatų skilimo, eksperimentai atlikti 30 °C temperatūroje. Pagal reakcijų sąlygas Nr. 2 (Lentelė 3) $d^{PA}ATP$ buvo išskirtas 77% išeiga. Susintetintas $d^{PA}ATP$ toliau dalyvavo reakcijoje su azido-NHS esteriu, kur tikslinis produktas $d^{N3\text{-}C2}ATP$ buvo gautas 58% išeiga (Schema 5).



Schema 5. Amino ir azido grupę turinčių dNTP sintezė.

Sonogashira reakcija su $d^{I}CTP$ nebuvo efektyvi ir po 15 val. buvo gauta tik 51% konversija į produktą. $d^{PA}CTP$ išgrynintas naudojant jonų mainų chromatografiją 45% išeiga. Klasikiniu sintezės metodu suminė $d^{PA}CTP$ išeiga siekė 13% (keturių stadijų). Tuo tarpu ši, alternatyvi, strategija įgalino susintetinti $d^{PA}CTP$ esant 24% suminei išeigai (dvi stadijos).

Lentelė 3. Išbandytos vandeninėje terpėje atliekamos Sonogashira reakcijos sąlygos tiksliniams $d^{PA}NTP$ gauti.

Nr.	Nukleotidas	Propargilaminas, ekv.	CuI, ekv.	Katalizatorius	Et_3N, ekv.	laikas	Konversija į $d^{PA}NTP$, %
1[a]	$d^{I}ATP$	2	0,1	$[(C_6H_5)_3P]_2PdCl_2$	3	1 val.	70
2	**$d^{I}ATP$**	**2**	**0,1**	**$Pd(OAc)_2$ (5 mol%) /TPPTS (25 mol%)**	**2**	**1 val.** **3 val.** **16 val.**	**86** **84** **57**
3	$d^{I}ATP$	2	0,1	$Pd(OAc)_2$ (5 mol%) /TPPTS (25 mol%)	10	1 val. 3 val. 16 val.	90 73 34
4	$d^{I}CTP$	2	0,1	$Pd(OAc)_2$ (5 mol%) /TPPTS (25 mol%)	10	6 val. 16 val.	40 51

Reakcijų koncentracija 0,06 M. Visos reakcijos atliktos palaikant 30 °C temperatūrą, ACN:H_2O (1:2) mišinyje, nebent nurodyta kitaip. a – reakcija atliekama dimetilformamide.

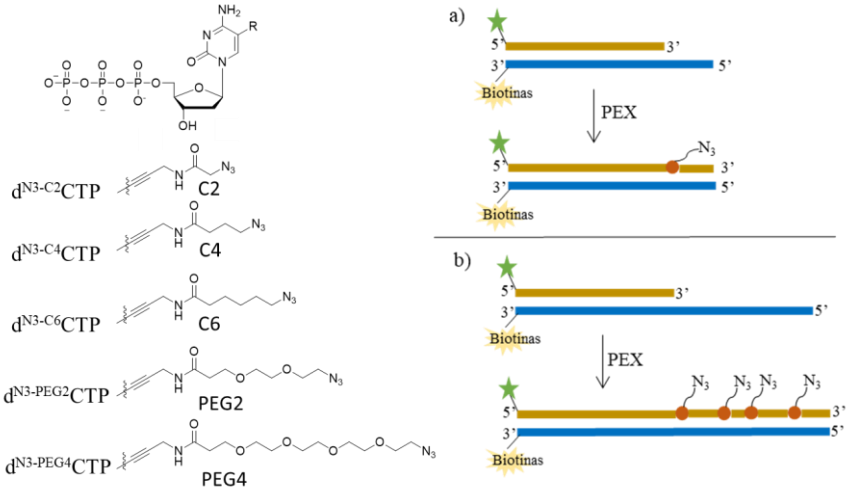
10.3 Azido grupę turinčių 2'-deoksinukleotidų funkcinis testavimas fermentinėse reakcijose ir reaktyvumo „click" reakcijos metu nustatymas

Nukleotidai, savo modifikacijose turintys terminalines funkcines grupes, tokias kaip amino, azido, alkino ir kt., turi didelį poreikį ir plačias pritaikymo galimybes. Taip funkcionalizuoti nukleotidai gali būti toliau žymimi tiek prieš jų fermentinį įterpimą į nukleorūgštis, tiek po jo. Nors darbo tikslas buvo susintetinti oligonukleotidais modifikuotus nukleotidus, jų pirmtakų, kaip substratų polimerazėms, savybės yra labai svarbios. Azido-alkino ciklo prijungimo reakcija („click" reakcija), kaip ir žymėtų NHS esterių reakcija su aminais, yra laikomos paprastais ir greitais metodais, kurie gali būti atliekami gyvose ląstelėse. Susintetinti įvairūs $d^{N3}CTP$ galėtų būti naudojami tolimesniame nukleorūgščių žymėjime pritaikant tiek vario katalizuojamą (CuAAC), tiek žiedo įtempties (SPAAC) "click" reakcijas. Atsižvelgiant į visą tai, $d^{N3}CTP$ buvo testuoti fermentinėse nukleorūgščių sintezės reakcijose, o azido grupe žymėti oligonukleotidai – reakcijose su įvairiais alkinais.

10.3.1 Azido grupę turinčių 2'-deoksicitidino 5'-trifosfatų funkcinis testavimas fermentinėse reakcijose

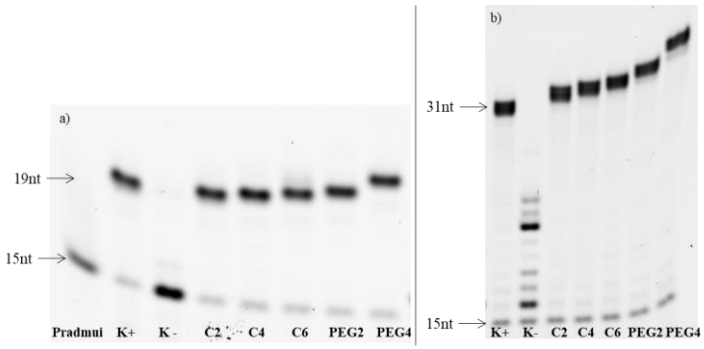
KOD XL polimerazė yra laukinio tipo KOD polimerazės ir mutantinės formos, kuriai trūksta 3'→5' egzonukleazės aktyvumo, mišinys. Ši

polimerazė yra plačiai taikoma modifikuotų nukleotidų įterpimui PEX ir PGR reakcijų metu ir yra gerai žinoma dėl tolerancijos nukleobazėje modifikuotiems 2'-deoksinukleotidams. Susintetintų d^{N3}CTPs ($d^{N3\text{-}C2}$CTP, $d^{N3\text{-}C4}$CTP, $d^{N3\text{-}C6}$CTP, $d^{N3\text{-}PEG2}$CTP, $d^{N3\text{-}PEG4}$CTP) fermentinis įterpimas testuotas dvejose sistemose (Paveikslas 2 a ir b). Abiem sistemoms buvo atitinkamai parinktos matricos, besiskiriančios ilgiu ir G nukleotidų kiekiu kopijuojamoje grandinėje.

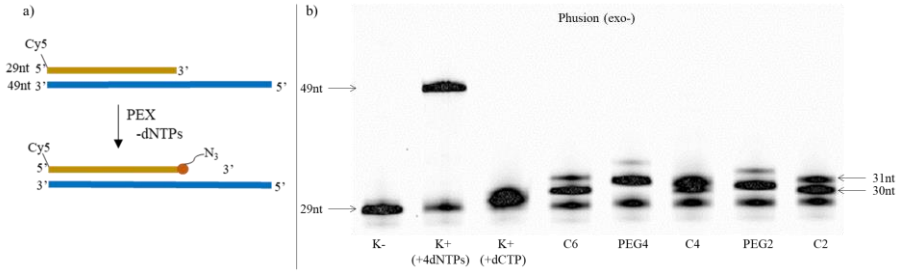


Paveikslas 2. Skirtingų d^{N3}CTP įterpimo sistemų grafinis reprezentavimas; a) vieno modifikuoto d^{N3}CTP įterpimas, b) daugkartinis modifikuoto d^{N3}CTP įterpimas.

Atlikus PEX reakcijas abejose sistemose, rezultatai analizuoti poliakrilamidiniuose geliuose (PAGE). Gautus rezultatus lyginant su kontrolėmis nustatyta, kad visi testuoti substratai yra sėkmingai įterpiami KOD XL polimerazės nepriklausomai nuo naudotos sistemos (Paveikslas 3a ir b). Kita labai plačiai naudojama polimerazė Phusion (exo-) taip pat buvo tiriama d^{N3}CTP įterpimo reakcijose. Pasirinktoje testavimo sistemoje buvo stebimas vieno modifikuoto d^{N3}CTP įterpimas (Paveikslas 4a). Reakcijos mišinių PAGE analizė patvirtino sėkmingą įvairių d^{N3}CTP įterpimą (Paveikslas 4b).



Paveikslas 3. Fermentinis d^{N3}CTP įterpimas pradmens pratęsimo reakcijos (PEX) metu naudojant KOD XL polimerazę; a) PEX eksperimentai atlikti naudojant matricą C1 ir pradmenį 1, K+ – teigiama kontrolė, naudojami natūralūs dCTP ir dGTP, K- – neigiama kontrolė, nenaudojamas natūralus dCTP, b) – PEX eksperimentai atlikti naudojant matricą C4 ir pradmenį 1, K+ – teigiama kontrolė, naudojami visi 4 natūralūs dNTP, K- – neigiama kontrolė, nenaudojamas natūralus dCTP. C2, C4, C6, PEG2, PEG4 – pažymi jungtukų struktūrą tarp nukleobazės ir azido grupės.



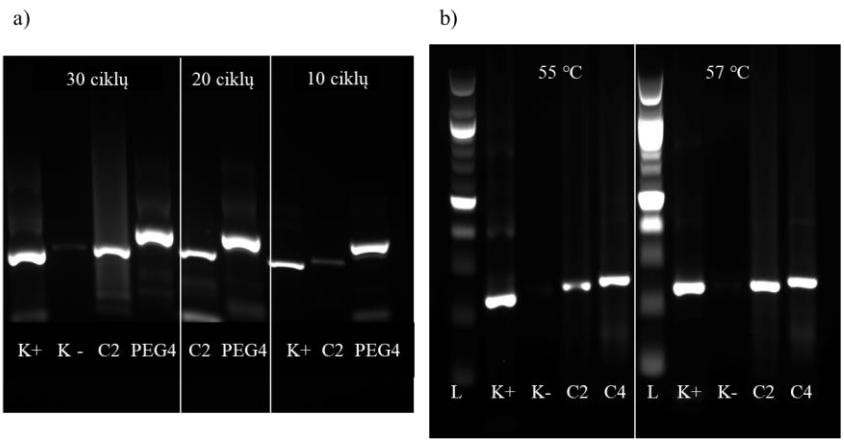
Paveikslas 4. Fermentinis d^{N3}CTP įterpimas pradmens pratęsimo reakcijos (PEX) metu naudojant Phusion (exo-) polimerazę; a) sistema naudota d^{N3}CTP įterpimui, b) PEX reakcijų PAGE analizė, kur K- – neigiama kontrolė, reakcijoje nenaudojami dNTP, K+ (+4dNTP) – teigiama kontrolė, reakcijoje naudojami visi keturi natūralūs dNTP, K+ (+dCTP) – teigiama kontrolė, reakcijoje naudojamas tik dCTP.

d^{N3}CTP buvo testuoti PGR reakcijose pilnai pakeičiant natūralų dCTP. Kadangi susintetinti d^{N3}CTP buvo sėkmingai įterpiami PEX reakcijų metu, vėliau buvo siekiama išsiaiškinti, ar galimas šių modifikuotų nukleotidų dalyvavimas PGR, pilnai pakeičiant natūralų dCTP. Tokiu būdu būtų gaunamos didelio modifikacijos laipsnio nukleorūgštys. Atliktų PGR eksperimentų reakcijų sąlygos pateiktos Lentelė 4, o rezultatai analizuoti atliekant elektroforezę agaroziniame gelyje (Paveikslas 5). PGR eksperimentus atlikus su visais d^{N3}CTP pagal reakcijų sąlygas Nr. 1 (Lentelė 4) pastebėtas susidariusių produktų kiekių skirtumas. Optimizuojant PGR sąlygas, reakcijose su $d^{N3\text{-}C6}$CTP, $d^{N3\text{-}PEG2}$CTP, $d^{N3\text{-}PEG4}$CTP esant 10 PGR ciklų buvo stebimas produkto susidarymas, tuo tarpu $d^{N3\text{-}C2}$CTP ir $d^{N3\text{-}C4}$CTP atvejais produktai beveik nesusidarė. Ciklų skaičiaus didinimas turėjo įtakos

PGR su $d^{N3\text{-}C2}$CTP produkto susidarymui, tačiau vis tiek buvo stebimas ženklus kiekių skirtumas (Paveikslas 5a). Padidinus elongacijos stadijos laiką bei sulydymo temperatūrą PGR su $d^{N3\text{-}C2}$CTP ir $d^{N3\text{-}C4}$CTP produktų kiekiai atitiko teigiamos kontrolės rezultatus (Paveikslas 5b).

Lentelė 4. Testuotos PGR reakcijų sąlygos d^{N3}CTP įterpimui.

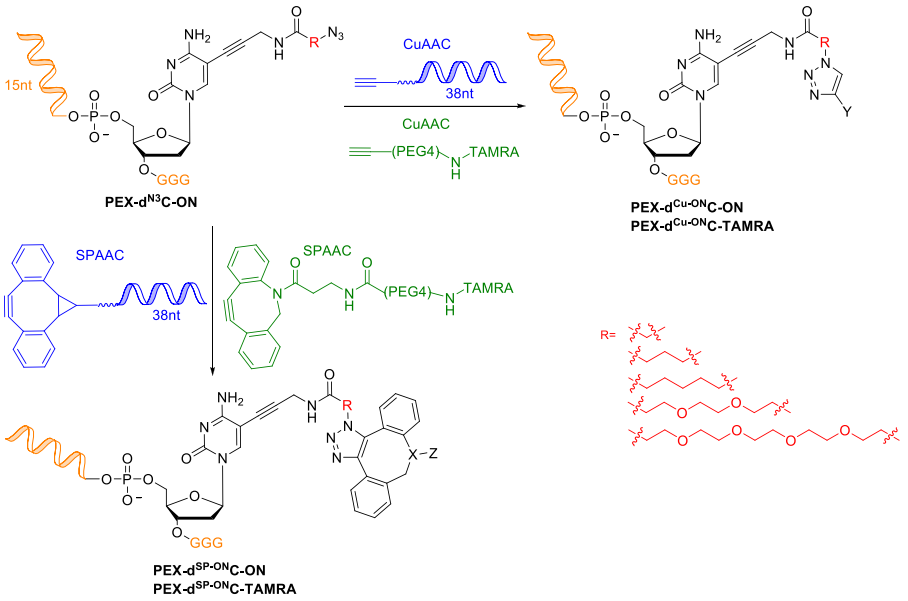
Nr.	Reakcijos sąlygos	Reakcijos stadijos	Temperatūra	Laikas	Ciklai
1	A, B, C, D	Pakaitinimas	94 °C	3 min.	1
		Denatūravimas	94 °C	1 min.	A - 39 B - 9 C- 19 D - 29
		Sulydimas	55 °C	1 min.	
		Elongacija	72 °C	1,5 min.	
		Elongacija	72 °C	5 min.	1
2	E	Pakaitinimas	94 °C	3 min.	1
		Denatūravimas	94 °C	1 min.	39
		Denatūravimas	55 °C	1 min.	
		Elongacija	72 °C	**2 min.**	
		Elongacija	72 °C	5 min.	1
3	F	Pakaitinimas	94 °C	3 min.	1
		Denatūravimas	94 °C	1 min.	39
		Sulydimas	**57 °C**	1 min.	
		Elongacija	72 °C	**2 min.**	
		Elongacija	72 °C	5 min.	1



Paveikslas 5. Fermentinis d^{N3}CTP įterpimas PGR metu; a) PGR rezultatai pagal reakcijų sąlygas B, C, D (Lentelė 4), b) PGR rezultatai pagal reakcijų sąlygas E ir F (Lentelė 4). K+ – teigiama kontrolė, reakcijoje naudojami visi natūralūs dNTP, K- – neigiama kontrolė, reakcijoje nenaudojami natūralūs dNTP, L – produktų ilgio standartas.

10.3.2 Azido-alkino ciklo prijungimo reakcijos greičio priklausomybės nuo substratų prigimties tyrimas

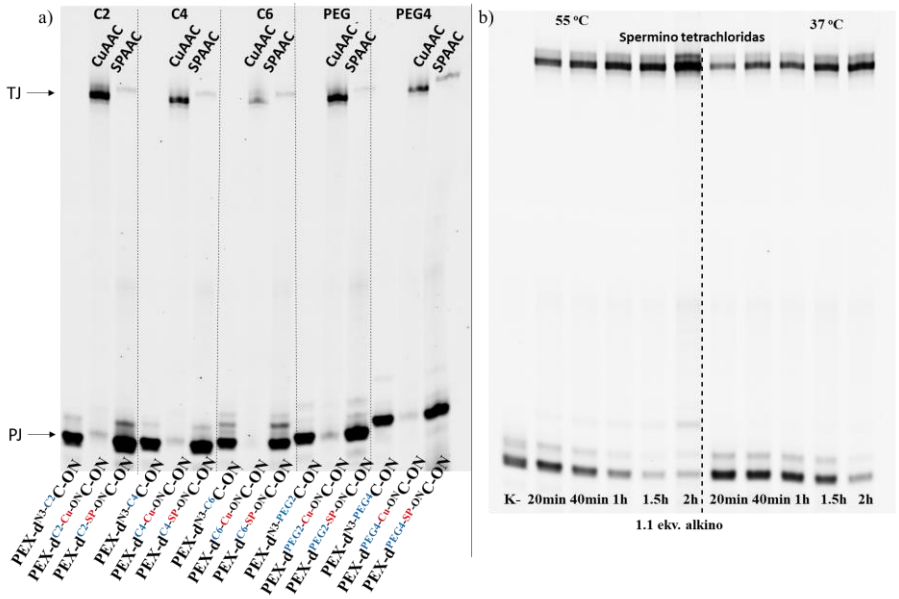
Oligonukleotinė modifikacija nukleorūgščių struktūroje gali būti įvedama dviem būdais: įterpiant oligonukleotidais (ON) modifikuotus nukleotidus arba įterpiant azido grupę turinčius nukleotidus ir tuomet atliekant „click“ reakciją su alkino-oligonukleotidais. Kadangi ON yra steriškai didelė molekulė ir tai gali turėti neigiamą įtaką tokias žymes turinčių nukleotidų įterpimui, buvo nuspręsta panagrinėti „click“ reakcijas tarp oligonukleotidų. Ankščiau aprašytomis sąlygomis naudojant KOD XL polimerazę buvo pakartoti PEX eksperimentai didesnėje skalėje. Įterpimo produktai (PEX-d^{N3}C-ON) su įvairiais jungtukais tarp nukleobazės ir azido grupės buvo išgryninti ir toliau naudoti „click“ reakcijose su alkinais. Pasirinktoje modelinėje sistemoje, CuAAC ir SPAAC reakcijų greičio priklausomybei nuo substrato prigimties tirti, naudoti alkinai su oligonukleotidine ir TAMRA dažo žymėmis (Schema 6). Reakcijos buvo stebimos laike atliekant PAGE analizę.



Schema 6. Pasirinkta modelinė sistema CuAAC ir SPAAC reakcijų greičiams nuo substrato prigimties tirti.

Pirmiausiai buvo atliktos CuAAC ir SPAAC reakcijos su skirtingais PEX-d^{N3}C-ON ir aklino-oligonukleotidais (Schema 6 ir Paveikslas 6a). CuAAC reakcijose nepriklausomai nuo PEX-d^{N3}C-ON cheminės struktūros buvo pasiekta pilna konversija. SPAAC reakcijų atveju buvo stebimas tik labai mažas tikslinių produktų susidarymas. Optimizuojant „click“ reakcijų sąlygas,

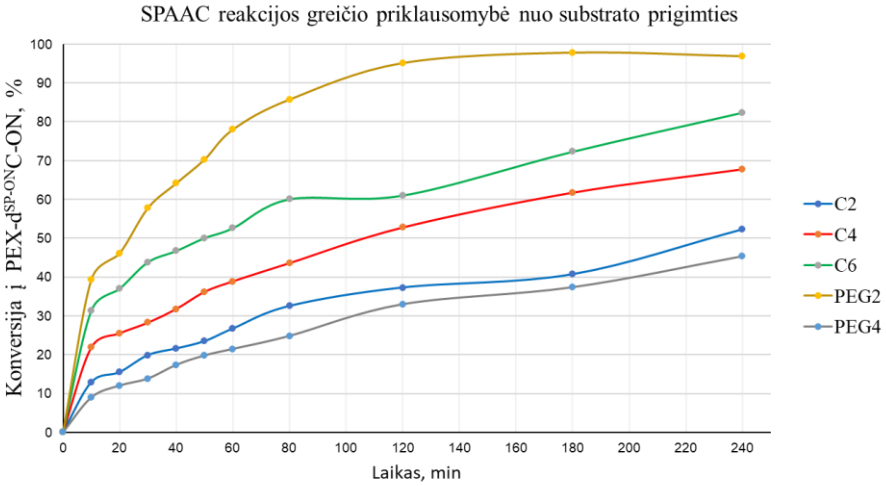
nustatyta, kad naudojant spermino tetrachloridą SPAAC reakcijose, alkino kiekį galima sumažinti iki 1.1 ekv. Reakcijas atliekant 37 ar 55 °C temperatūrose po dviejų valandų buvo stebima beveik pilna pradinio junginio konversija (Paveikslas 6b). Tikslesniam reakcijos greičio priklausomybės nuo substratų prigimties tyrimui atlikti buvo naudojamos SPAAC sąlygos, kurioms esant reakcija vyksta lėčiau (37 °C). Optimizuojant CuAAC reakcijų sąlygas, neretai buvo gaunami neatsikartojantys rezultatai. Tai gali būti siejama su vario katalizuojama oligonukleotų degradaciją. Atsižvelgiant į tai CuAAC reakcijų su alkino-oligonukleotidu greičio priklausomybė nuo substratų prigimties nebuvo nagrinėjama pasirinktoje sistemoje.



Paveikslas 6. CuAAC ir SPAAC reakcijų su PEX-d^{N3}C-ON rezultatai; a) CuAAC reakcijos buvo atliekamos 37 °C, 4 val. esant 5 ekv. alkino-oligonukleotido, naudota CuBr/TBATA (1:4) katalitinė sistema, reakcijos atliktos DMSO:*t*-BuOH (3:1) mišinyje. SPAAC reakcijos atliktos 37 °C temperatūroje, 4 val., esant 5 ekv. alkino-oligonukleotido, vandenyje, b) SPAAC reakcijos rezultatai optimaliomis sąlygomis reakcijas atliekant su PEX-d$^{C6\text{-}SP\text{-}ON}$C-ON. K- – neigiama kontrolė, įterpimo produktas PEX-d$^{N3\text{-}C6}$C-ON. PJ – pradinis junginys, TJ – tikslinis junginys.

Optimaliomis SPAAC reakcijos sąlygomis (Paveikslas 6b esant 37 °C) buvo atlikti eksperimentai su visais įterpimo substratais (PEX-d^{N3}C-ON). Iš gautų PAGE analizės duomenų densitometriškai įvertintos konversijos į produktus. Trijų nepriklausomų pakartojimų duomenys naudoti procentinėms vertėms gauti, o reakcijų rezultatai atvaizduoti grafiškai (Paveikslas 7). Gauti rezultatai parodė, kad jungtuko prigimtis turi ženklią įtaką SPAAC reakcijos

greičiui. Substratai pagal reaktyvumą išsidėsto tokia tvarka: PEX-d$^{N3\text{-}PEG2}$C-ON>PEX-d$^{N3\text{-}C6}$C-ON>PEX-d$^{N3\text{-}C4}$C-ON>PEX-d$^{N3\text{-}C2}$C-ON>PEX-d$^{N3\text{-}PEG4}$C-ON. Galima daryti išvadą, kad kuo ilgesnis jungtukas tarp nukleobazės ir azido grupės tuo didesnis SPAAC reakcijos greitis, išimtis buvo stebima PEG4 jungtuko atveju.

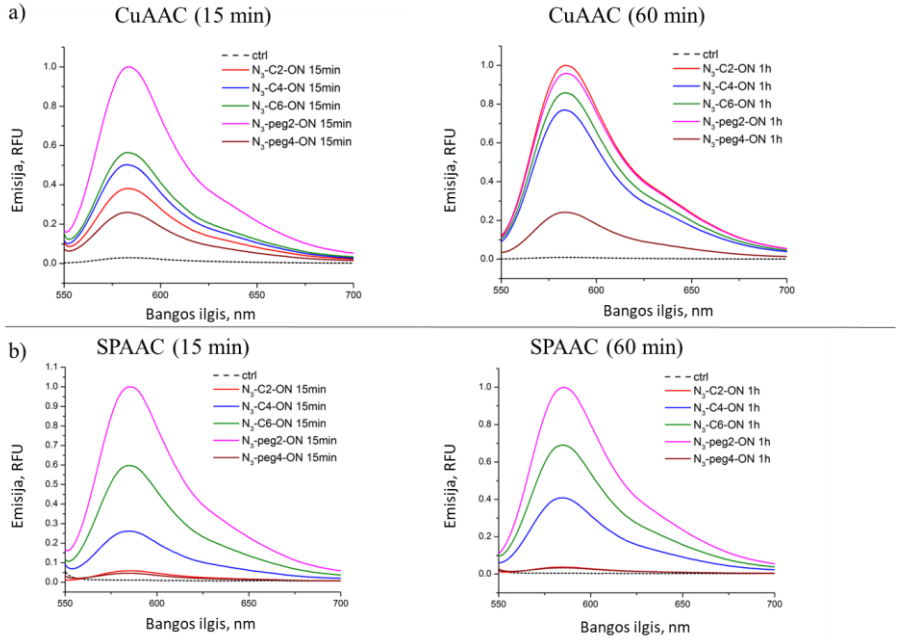


Paveikslas 7. SPAAC reakcijos greičio priklausomybė nuo PEX-d^{N3}C-ON cheminės struktūros (standartinis nuokrypis 10 %).

Panašūs „click“ reakcijų greičio priklausomybės nuo substratų prigimties tyrimai atlikti su PEX-d^{N3}C-ON ir TAMRA dažu žymėtais alkinais (Schema 6). Šiuo atveju buvo analizuojami mėginiai po 15 ir 60 min nuo reakcijos pradžios. Išmatavus emisiją, gauti duomenys normalizuoti pagal didžiausią vertę ir atvaizduoti grafiškai (Paveikslas 8). Iš gautų duomenų nustatyta, kad PEX-d^{N3}C-ON reaktyvumo tendencijos nepriklausė nei nuo „click“ reakcijos nei nuo alkino žymės. Buvo stebimos tos pačios tendencijos, kaip ir SPAAC reakcijos su alkino-oligonukleotidais atveju. Substratų reaktyvumas išsidėsto tokia tvarka: PEX-d$^{N3\text{-}PEG2}$C-ON>PEX-d$^{N3\text{-}C6}$C-ON>PEX-d$^{N3\text{-}C4}$C-ON>PEX-d$^{N3\text{-}C2}$C-ON>PEX-d$^{N3\text{-}PEG4}$C-ON. Tačiau reaktyvumų skirtumai tarp jungtukų CuAAC atveju nėra tokie ženklūs kaip SPAAC atveju (Paveikslas 8a ir b, 15 min). To priežastis gali būti didesnis CuAAC reakcijos greitis lyginant su SPAAC. Bendras reakcijų greičių skirtumas gali būti stebimas palyginant mėginių po valandos grafikus (Paveikslas 8a ir b, 60 min).

Atlikti CuAAC ir SPAAC reakcijų greičio priklausomybės tyrimai nuo substratų cheminės prigimties. Nepriklausomai nuo alkino žymės (ON ar TAMRA) ar PEX-d^{N3}C-ON struktūros buvo gautos tos pačios reaktyvumo tendencijos. Nustatyta, kad kuo ilgesnė jungtis tarp azido grupės ir

nukleobazės, tuo didesnis yra reakcijos greitis. Visais atvejais buvo stebima išimtis su PEG4 jungtuku, kuris reagavo lėčiausiai.

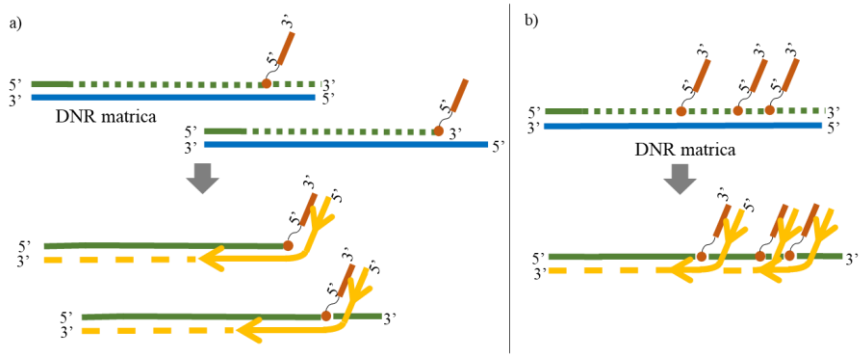


Paveikslas 8. CuAAC ir SPAAC reakcijų greičio priklausomybės nuo substrato prigimties tyrimas reaguojant įvairiems PEX-d^{N3}C-ON ir TAMRA dažu žymėtiems alkinams.

10.4 Oligonukleotidais modifikuotų nukleotidų sintezė ir panaudojimas fermentinėse reakcijose

Naujos kartos sekoskaitos (NKS) bibliotekų paruošimas yra ilgas, daugiastadijinis procesas, turintis tam tikrų iššūkių (literatūros apžvalga 4-as skyrius). Visoms dabartinėms sekos nustatymo platformoms reikalingas specifinis nukleorūgščių apdorojimas, kad būtų sukurta sekoskaitai tinkama fragmentų biblioteka. Paprastai tai apima DNR arba RNR fragmentavimą iki platformai būdingo dydžio diapazono, tada 3', 5' galų paruošimą ir sekoskaitos platformai specifinių adapterių ligavimą. Fermentinis adapterių ligavimas yra neefektyvus ir komplikuotas etapas. Adapterių tarpusio ligavimasis arba dalinis fragmentų galų pažymėjimas (tik vienas iš adapterių yra priliguojamas) sumažina originalios bibliotekos kompleksiškumą. Siekiant išvengti kritinių NKS bibliotekų ruošimo etapų ir supaprastinti patį procesą yra kuriamos alternatyvios metodologijos. Galimybė panaudoti nukleotidus, turinčius oligonukleotido modifikaciją kaip universalią pradmens lydymosi vietą, leistų išvengti fermentinio adapterių ligavimo

žingsnio (Paveikslas 9). Tai ženkliai supaprastintų NKS bibliotekų paruošimo procesą ir pašalintų limituojantį jo žingsnį.

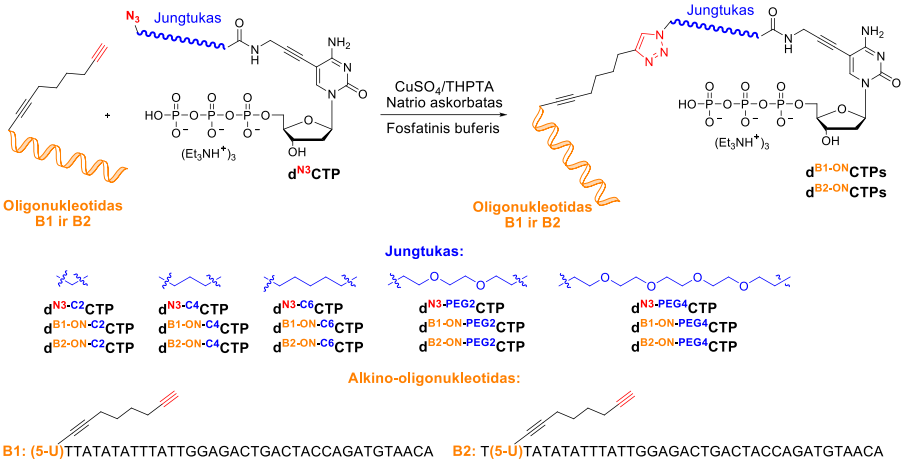


Paveikslas 9. Oligonukleotidais modifikuotų 2'-deoksinukleotidų panaudojimo principas NKS bibliotekų ruošime. a) vieno OTDN įterpimas ir atitinkamų produktų susidarymas, b) kelių OTDN įterpimas ir atitinkamų produktų susidarymas.

Buvo iškeltos dvi hipotezės apie oligonukleotidais modifikuotų 2'-deoksinukleotidų (OTDN) panaudojimą. Pirmiausiai, OTDN įterpimo įvykių skaičius būtų kontroliuojamas varijuojant jo koncentracija. Kadangi oligonukleotidas yra steriškai didelė modifikacija, tai taip pat galėtų turėti įtakos OTDN įstatymo dažniui ar net sąlygoti grandinės terminavimą (Paveikslas 9a).

Hipotezėms patvirtinti buvo susintetinti įvairūs OTDN. Technologijos tinkamumui įrodyti buvo susintetinti OTDN turintys skirtingo ilgio jungtukus tarp oligonukleotido ir nukleobazės. Tiksliniams produktams gauti pasirinkti du oligonukleotidai: B1 – terminalinio alkino modifikacija įvesta į pirmą heterociklinę oligonukleotido bazę sekoje (5'-gale), B2 - terminalinio alkino modifikacija įvesta į antrą heterociklinę oligonukleotido bazę sekoje (5'-gale) (Schema 7). Visi CuAAC eksperimentai buvo atliekami labai mažoje skalėje (8.4 nmol alkino-ON) esant azido pertekliui (2 ekv.).

Reakcijų eigą stebint HPLC metodu, konversija į tikslinius OTDN siekė nuo 51% iki 82% (Lentelė 5). Analizuojant HPLC chromatogramas buvo pastebėta, kad po tam tikro laiko produkto smailės intensyvumas pradeda mažėti, kas indikavo galimą oligonukleotido degradaciją (tai buvo patvirtinta atlikus LC/MS analizę). Dėl pradinių junginių, produkto ir priemaišų tarpusavio panašumo, OTDN paruošimą apsunkino gryninimo metodų parinkimas. Galiausiai produktai buvo išgryninti analitiniu HPLC, pakartotinai naudojant kelis skirtingus metodus, kas galėjo sąlygoti mažas produktų išeigas (Lentelė 5).



Schema 7. Įvairių OTDN sintezė atliekant CuAAC reakciją.

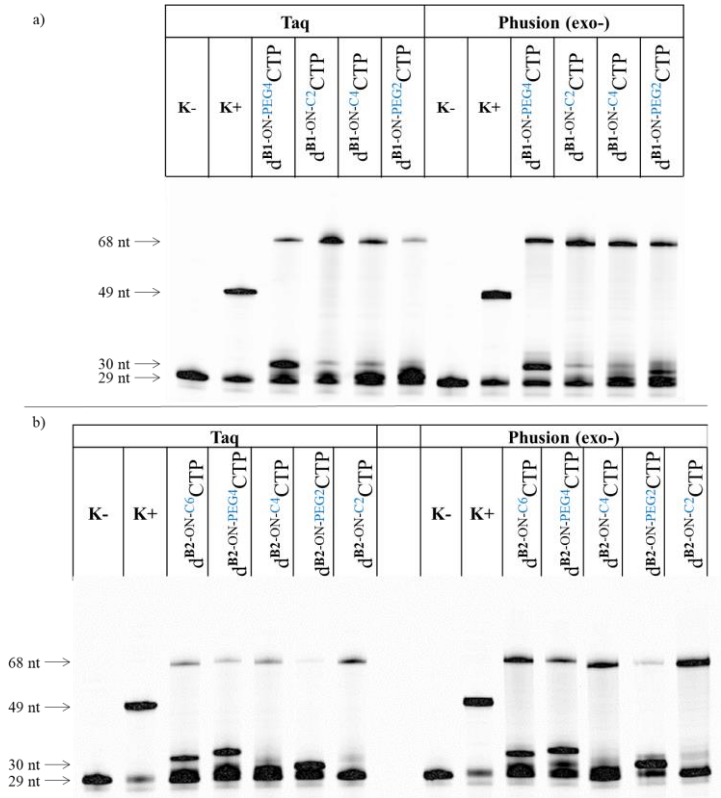
Susintetinti OTDN produktai (Lentelė 5) vėliau buvo testuojami PEX reakcijoje naudojant Taq ir Phusion (exo-) polimerazes (Paveikslas 10). Pasirinkta pradmens ir matricos sistema turėjo 20 nukleotidų (nt) iškyšą. Vieno OTDN įterpimui stebėti, PEX reakcijoje nebuvo naudojami natūralūs dNTP.

Lentelė 5. Susintetinti OTDN.

Nr.	ON	$d^{N3}CTP$	Laikas	OTDN	Konversija į OTDN, %	Išeiga, %
1	B1 (38nt)	$d^{N3\text{-}C2}CTP$	1 val.	$d^{B1\text{-}ON\text{-}C2}CTP$	81	26
2		$d^{N3\text{-}C4}CTP$	1.5 val.	$d^{B1\text{-}ON\text{-}C4}CTP$	76	17
3		$d^{N3\text{-}C6}CTP$	1 val.	$d^{B1\text{-}ON\text{-}C6}CTP$	70	19
4		$d^{N3\text{-}PEG2}CTP$	1 val.	$d^{B1\text{-}ON\text{-}PEG2}CTP$	72	17
5		$d^{N3\text{-}PEG4}CTP$	1.5 val.	$d^{B1\text{-}ON\text{-}PEG4}CTP$	80	22
6	B2 (38nt)	$d^{N3\text{-}C2}CTP$	1 val.	$d^{B2\text{-}ON\text{-}C2}CTP$	82	26
7		$d^{N3\text{-}C4}CTP$	3 val.	$d^{B2\text{-}ON\text{-}C4}CTP$	51	13
8		$d^{N3\text{-}C6}CTP$	1 val.	$d^{B2\text{-}ON\text{-}C6}CTP$	47	14
9		$d^{N3\text{-}PEG2}CTP$	2 val.	$d^{B2\text{-}ON\text{-}PEG2}CTP$	66	16
10		$d^{N3\text{-}PEG4}CTP$	1.5 val.	$d^{B2\text{-}ON\text{-}PEG4}CTP$	68	21

Atlikus įterpimo reakcijos mėginių PAGE analizę buvo stebimas 68 nt produkto susidarymas, patvirtinantis sėkmingą OTDN įterpimą. Abi testuotos polimerazės sėkmingai įterpė OTDN nepriklausomai nuo oligonukleotido (B1 ar B2) ar jungtuko struktūros. Tačiau gelyje buvo stebimas ir produkto, atitinkančio 30 nt ilgį, susidarymas (Paveikslas 10). Kadangi natūralūs nukleotidai šiose reakcijose nebuvo naudojami, tokių produktų susidarymas galėtų būti siejamas su prastu gryninimu ir likutiniu $d^{N3}CTP$ kiekiu. Todėl

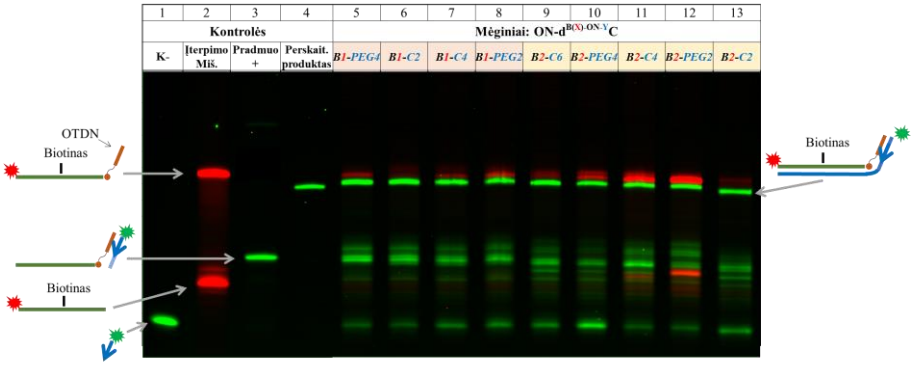
OTDN įterpimo efektyvumas negalėjo būti vertinamas kiekybiškai. Atsižvelgiant į šiuos rezultatus buvo optimizuojami ir kuriami nauji HPLC metodai.



Paveikslas 10. PEX reakcijos mišinių su Taq ir Phusion (exo-) polimerazėmis PAGE analizės; a) OTDN su B1 oligonukleotidų įterpimas, b) OTDN su B2 oligonukleotidu įterpimas. K- – neigiama kontrolė, nenaudojami dNTP, K+ – teigiama kontrolė, naudojami visi keturi dNTP. Visi OTDN įterpimo eksperimentai atlikti nenaudojant natūralių dNTP.

Susintetinus ir išgryninus tikslinius OTDN be azidų likučių, buvo atliktos pradmes pratęsimo reakcijos OTDN įterpimui. Tiksliniai įterpimo produktai (ON-d^{ON}C) buvo išgryninti ir testuoti perskaitymo per nenatūralią jungtį eksperimente. Buvo pasirinkta modelinė sistema ir pasiruoštos kontrolės, padėsiančios identifikuoti reakcijos komponentus ir, svarbiausia, tikslinį produktą (Paveikslas , eilutės 1 – 4). Pasirinktoje sistemoje tiek OTDN įterpimo produktas, tiek pradmuo, naudotas perskaitymo eksperimentui, turėjo fluorescencines žymes. Išgryninti OTDN įterpimo produktai ON-d^{ON}C buvo naudojami kaip matricos PEX eksperimentuose. Atlikus reakcijos mišinių PAGE analizę, nustatyta, kad nepriklausomai nuo oligonukleotido

(B1 ar B2) ar jungtuko struktūrų visais atvejais buvo stebimas tikslinio produkto susidarymas. Trumpesnių fragmentų susidarymas, leidžia daryti prielaidą, kad polimerazė perskaitymo metu stringa ties nenatūralia jungtimi.



Paveikslas Įvairių ON-d^{ON}C perskaitymo per nenatūralią jungtį testavimas naudojant Phusion (exo-) polimerazę. K- – neigiama kontrolė, pradmuo naudotas komplementarios grandinės sintezėje; Įterpimo Miš. – OTDN įterpimo reakcijos mišinys: pradmuo ir įterpimo produktas; Pradmuo+ – pradmuo, naudotas perskaitymo eksperimente, pratęstas iki nenatūralios jungties; Perskait. produktas – produkto ilgio standartas. Reakcijų sąlygos: 95 °C – 1 min. → 61 °C – 1 min. → 72 °C – 1 min. (10 ciklų).

Apibendrinant, gauti rezultatai įrodo, kad visi jungtukai yra biologiškai suderinami ir Phusion (exo-) polimerazė geba nuskaityti ir susintetinti komplementarią grandinę. Tačiau šis eksperimentas neleidžia tiksliai įvertinti kuris iš jungtukų yra efektyviausias. Nepaisant to, parodyta, kad pasiūlytas metodas įgyvendinamas pritaikant OTDN. Panaudojus OTDN realiose sistemose buvo pastebėti keli technologiniai iššūkiai. OTDN buvo geri substratai ir kartotinai įterpiami PEX reakcijos metu. Atliekant tokių fragmentų perskaitymo eksperimentus gauti labai kompleksiški fragmentų mišiniai, apsunkinantys jų identifikavimą. Tuo tarpu fermentinis perskaitymas esant šakotoms matricos struktūroms buvo dar mažesnio efektyvumo. Kadangi sekoskaitos pagrindiniai aspektai remiasi specifinio ilgio fragmentų analize, padengiant visą tiriamo mėginio seką, tokie OTDN negali būti naudojami NKS bibliotekų paruošime.

11. Oligonukleotidais modifikuotų 2’,3’-dideoksinukleotidų sintezė ir panaudojimas

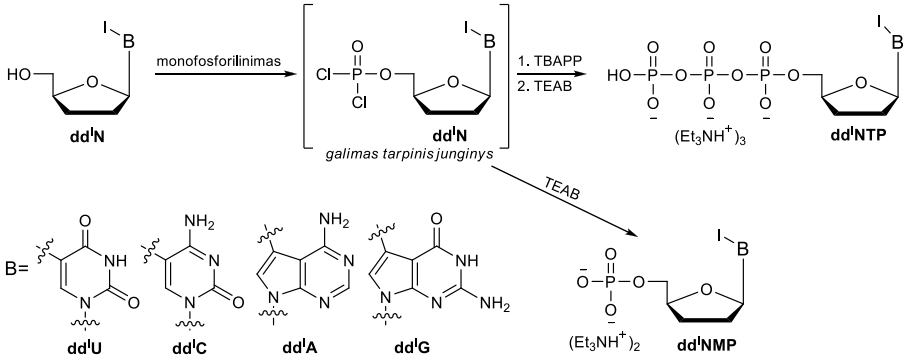
10.4 skyriuje parodėme, kad pasiūlytas metodas grįstas OTDN įterpimu yra įmanomas, tačiau turintis esminių iššūkių. Oligonukleotidais modifikuotų 2’,3’-dideoksinukleotidų (OTDDN) panaudojimas išspręstų esamas technines

kliūtis ir žymiai supaprastintų patį NKS bibliotekų ruošimo protokolą. Tolimesniame skyriuje bus aprašoma optimizuota nukleobazėje modifikuotų 2',3'-dideoksinukleotidų ir OTDDN sintezė ir panaudojimas.

11.1 Jodintų 2',3'-dideoksinukleozidų fosforilinimo reakcijų sąlygų optimizavimas

Modifikuotų nukleotidų paskirtis priklauso nuo jų modifikacijos prigimtiems. Todėl mūsų tikslas buvo optimizuoti sintezės strategiją, kuri būtų universali, ir pritaikoma didesnės skalės sintezei. Buvo nuspręsta pritaikyti trijų stadijų strategiją modifikuotų ddNTP sintezei. Šis metodas yra pranašesnis laiko, sąnaudų atžvilgiu ir sąlygoja didesnes produktų išeigas, o funkcionalizuoti nukleotidai, tokie kaip d(d)^{PA}NTP, gaunami vos per du sintezės etapus.

Pirmasis šios sintezės strategijos etapas yra dd^{I}NTP sintezė, kur limituojanti stadija yra aktyvaus tarpinio monofosfato (arba, kaip schemoje vaizduojama, fosforodichloridato) suformavimas (Schema 8). Pasiekus gerą konversiją į aktyvų tarpinį junginį trifosfoforilinimo reakcija toliau yra atliekama panaudojant pirofosfato druskas. Reakcijų eiga buvo sekama HPLC, stebint fosforodichloridato hidrolizės produkto (dd^{I}NMP) susidarymą ir procentinį kiekį reakcijos mišinyje.



Schema 8. dd^{I}NMP ir dd^{I}NTP sintezė.

Pirmiausiai buvo optimizuojamos dd^{I}C monofosforilinimo reakcijos sąlygos. Reakcijose su $POCl_3$, nepriklausomai nuo agento ar bazės kiekio, tikslinis produktas praktiškai nesusidarė (≤3%) (Lentelė 6, Nr. 1 – 4). Monofosforilinimo reakcijoje naudojant 2 ekv. $P_2O_3Cl_4$ formavosi specifinė priemaiša (Lentelė 6, Nr. 5 – 6), kurios absorbcijos maksimumas pasislinkęs (priemaišos – 288 nm, dd^{I}C – 295 nm) (Paveikslas 11). Šie duomenys reiškia, kad vyksta molekulės chromoforo (nukleobazės) struktūros pokyčiai.

Reakcijos mišiniai įgavo geltoną spalvą, kuri su laiku intensyvėjo, kas galbūt galėtų reikšti vykstantį halogenų apsikeitimą nukleobazėje. Panašūs pastebėjimai buvo apžvelgiami ir kitų mokslininkų. A. Colier *et.al.* [54] ir C. Zhu *et.al.* [52] stebėjo fosforilinimo reakcijos metu vykstantį halogenų nukleobazėje apsikeitimą ir patvirtino tai atlikdami susidariusios priemaišos BMR analizę. Padidinus bazės kiekį iki 2 ekv., reakcija nevyko (Lentelė 6, Nr. 7).

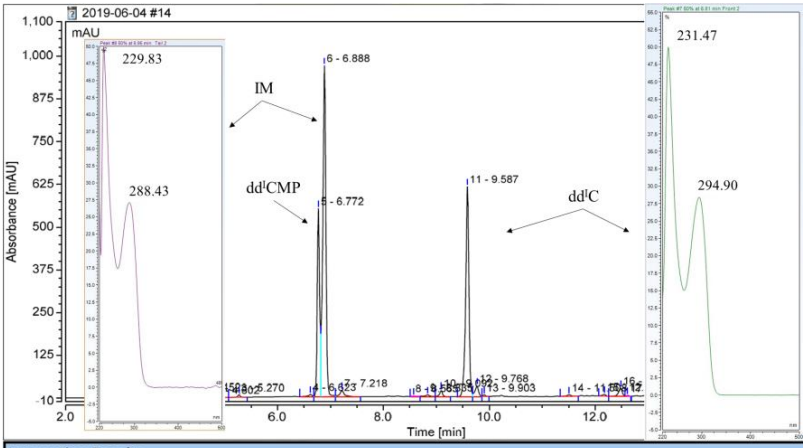
Lentelė 6. dd^{I}C monofosforilinimo reakcijos sąlygų optimizavimas.

Nr.	Fosforilinimo agentas	Fosforilinimo agento kiekis, ekv.	Tirpiklis	TBA, ekv.	Reakcijos laikas	Konversija į dd^{I}CMP, %	Konversija į priemaišą, %	dd^{I}C, %
1	$POCl_3$	2	TMP	-	2 val.	≤3	-	98
2		2	TMP	2	2 val.	≤3	-	97
3		3	TMP	-	2 val.	5	-	92
4		3	TMP	2	2 val.	≤3	-	98
5	$P_2O_3Cl_4$	2	ACN	-	0 min.	8	1	90
					30 min.	14	23	59
					1 val.	15	33	47
					2 val.	14	52	29
6		2	ACN	0,5	30 min.	19	24	50
					1 val.	19	45	29
					2 val.	18	66	10
7		2	ACN	2	2 val.	≤3	≤3	95
8		3	ACN	-	0 min.	20	1	78
					30 min.	63	13	16
					1 val.	68	18	7
					2 val.	67	25	4
9		3	ACN	0,25	30 min.	82	3	7
					1 val.	86	4	2
					2 val.	86	6	1
10		3	ACN	0,5	30 min.	51	4	37
					1 val.	59	4	26
					2 val.	57	29	13
11		3	ACN	1	30 min.	86	1	4
					1 val.	88	1	2
					2 val.	89	2	1
12		3	ACN	3	2 val.	10	1	84
13[a]		**3**	**ACN**	**-**	**20 min.**	**87**	**2**	**5**

Reakcijos buvo atliktos 0 – 5 °C temperatūroje, koncentracija – 0,1 M. a – sintezės didesnėje skalėje rezultatai (reakcijoje naudojamas 1 g pradinio nukleozido). Reakcija buvo atlikta -10 – (-15) °C temperatūroje.

Naudojant didesnį kiekį (3 ekv.) $P_2O_3Cl_4$ jau po 30 min. buvo gauta 63% pradinio nukleozido konversija į tikslinį produktą (Lentelė 6, Nr. 8). Keičiant bazės kiekį reakcijoje (0,25 – 3 ekv.) buvo stebimas priemaišos susidarymo sumažėjimas ir pasiekta gera konversija (>80%) į tikslinį dd^{I}CMP (Lentelė 6, Nr. 9 ir 11). Tačiau didinant bazės kiekį nuo 0,25 ekv. buvo gauti kontraversiški rezultatai, tai galėtų būti siejama su TBA prastu išsimaišymu acetonitrile.

Vieno-indo dd^{I}CTP sintezės monofosforilinimo stadijai pasirinktos sąlygos Nr. 9 (Lentelė 6). Atliekant reakciją didesnėje skalėje, susidarė daugiau priemaišų. Todėl buvo nuspręsta reakciją pakartoti pagal sąlygas Nr. 8 (Lentelė 6), o siekiant sumažinti priemaišų susidarymą reakcija buvo atlikta -10 – (-15) °C temperatūroje. Šiuo atveju po 20 min. buvo pasiekta 87% konversija į tarpinį dd^{I}CMP (Lentelė 6, Nr. 13). Antrame reakcijos žingsnyje panaudojus TBAPF mišinį (TBAPF tirpalas ACN + TBA) po 30 min. buvo gauta 79% konversija į tikslinį dd^{I}CTP. Produktas išgrynintas jonu mainų chromatografijos metodu 74% išeiga.



Integration Results

No.	Peak Name	Retention Time min	Area mAU*min	Height mAU	Relative Area %	Relative Height %
5		6.772	27.814	550.502	18.74	24.20
6		6.888	66.225	968.311	44.63	42.57
11		9.587	43.276	614.481	29.16	27.02

Paveikslas 11. Reakcijos mišinio HPLC chromatograma. Mėginys paruoštas po valandos, reakciją atliekant sąlygomis aprašytomis Nr. 6 (Lentelė 6.). IM – priemaiša.

Optimizuojant dd^{I}A monofosforilinimo reakcijos sąlygas nustatyta, kad $POCl_3$ nebuvo efektyvus fosforilinimo agentas, nepriklausomai nuo keistų reakcijos parametrų (Lentelė 7, Nr. 1 – 6). Todėl reakcijų sąlygoms optimizuoti tolimesniuose eksperimentuose buvo naudojamas $P_2O_3Cl_4$. Dviejų ekvivalentų $P_2O_3Cl_4$ perteklius buvo nepakankamas, o bazės priedas, nepriklausomai nuo kiekio, neturėjo įtakos produkto susidarymui (Lentelė 7,

Nr. 7 – 11). Gera konversija į tikslinį dd^{I}AMP buvo pasiekta naudojant 3 ekv. $P_2O_3Cl_4$ su papildomu agento priedu po vienos valandos (1 ekv.). Esant minėtoms reakcijos sąlygoms konversija į produktą siekė 77 – 85% (Lentelė 7, Nr. 12 – 13).

Lentelė 7. dd^{I}A monofosforilinimo reakcijos sąlygų optimizavimas.

Nr.	**Fosforilinimo agentas**	**Fosforilinimo agento kiekis, ekv.**	**Tirpiklis**	**TBA, ekv.**	**Reakcijos laikas**	**Konversija į dd^{I}AMP, %**	**dd^{I}A, %**
1[a]	$POCl_3$	2 + 1[b]	TMP	-	2 val.	4	90
2		2 + 1[b]	TMP	-	2 val.	≤3	91
3		3	TMP	2	2 val.	3	93
4		3 + 1[b]	TMP	-	2 val.	4	90
5		3 + 1[b]	TMP	-	2 val.	3	90
6		3	TMP	3	2 val.	4	92
7	$P_2O_3Cl_4$	2	ACN	0,25	2 val.	5	92
8		2	ACN	0,5	2 val.	3	95
9		2	ACN	2	2 val.	4	93
10 [a]		2 + 1[b]	ACN	-	0 min.	4	90
					30 min.	15	79
					1 val.	17	77
					2 val.	48	46
11		2 + 1[b]	ACN	-	0 min.	4	90
					30 min.	7	86
					1 val.	7	86
					2 val.	37	55
12[a]		3 + 1[b]	ACN	-	0 min.	7	86
					30 min.	50	44
					1 val.	51	42
					2 val.	85	6
13		3 + 1[b]	ACN	-	0 min.	11	83
					30 min.	34	59
					1 val.	35	57
					2 val.	77	14
14[c]		**3**	**ACN**	**-**	**30 min.**	**96**	**-**
					1 val.	**95**	**-**

Reakcijos buvo atliktos 5 – 10 °C temperatūroje, koncentracija – 0,1 M. a – reakcija atlikta 0 – 5 °C temperatūroje, b – po valandos pridėtas papildomas fosforilinimo agento kiekis, c – sintezės didesnėje skalėje rezultatai (reakcijoje naudojamas 1 g pradinio nukleozido).

Vieno-indo dd^{I}ATP sintezei buvo pasirinktos reakcijos sąlygos Nr. 12 (Lentelė 7). Monofosforilinomo reakcijos didesnėje skalėje metu konversija į

dd^{I}AMP siekė net 96% (Lentelė 7, Nr. 14). Pridėjus TBAPP ir TBA mišinį acetonitrile, buvo gauta 79 % konversija į dd^{I}ATP, o produktas išgrynintas 73% išeiga.

Priešingai nei dd^{I}A ir dd^{I}C atvejais, dd^{I}U monofosforilinimas buvo sėkmingai atliktas reakcijoje naudojant $POCl_3$. Gerai konversijai į produktą pasiekti buvo reikalingas 3 ekvivalentų fosforilinimo agento perteklius ir bazės priedas (Lentelė 8, Nr. 1 – 2). Keičiant TBA kiekį reakcijoje buvo nustatyta, kad geriausia konversija (93%) į dd^{I}UMP gauta esant 1.5 ekv. bazės (Lentelė 8, Nr. 3 – 5). Įdomu tai, kad monofosforilinimą atliekant su $P_2O_3Cl_4$ buvo stebimas specifinės priemaišos susidarymas. Analizuojant HPLC chromatogramą pastebėta, kad priemaišos absorbcijos maksimumas nepakito, tai reiškia, kad molekulės struktūriniai pokyčiai vyko ne nukleobazėje. Atlikus

Lentelė 8. dd^{I}U monofosforilinimo reakcijos sąlygų optimizavimas.

Nr.	Foforilinimo agentas	Fosforilinimo agento kiekis, ekv.	Tirpiklis	TBA, ekv.	Reakcijos laikas	Konversija į dd^{I}UMP, %	Konversija į priemaišą, %	dd^{I}U, %
1	$POCl_3$	1,5	TMP	-	30 min.	≤3	1	97
					2 val.	6	1	92
2		1,5 + 1,5[a]	TMP	1	30 min.	16	-	77
					3 val.	60	-	34
					4.5 val.	78	-	17
3		3	TMP	1	30 min.	55	-	41
					2.5 val.	84	-	11
					4 val.	89	-	4
4[b]		**3**	**TMP**	**1,5**	**30 min.**	**68**	**-**	**28**
					1 val.	**84**	**-**	**10**
					2 val.	**93**	**-**	**2**
5		3	TMP	2	30 min.	55	-	31
					1 val.	84	-	2
					1.5 val.	89	-	2
6	$P_2O_3Cl_4$	3	ACN	-	30 min.	22	64	6
					1 val.	20	75	1
					3 val.	13	82	-
					5 val.	8	88	-
7		3	ACN	3	30 min.	63	17	-
					1 val.	62	18	-

Reakcijos buvo atliktos 0 – 5 °C temperatūroje, reakcijų koncentracija – 0,1 M. a – po valandos pridėtas papildomas fosforilinimo agento kiekis, b – sintezės didesnėje skalėje rezultatai (reakcijoje naudojamas 1 g pradinio nukleozido).

reakcijos mišinio gryninimą jonų mainų chromatografijos metodu, priemaiša eliavo nesulaikoma reakcijos mišinio užnešimo metu. Tokia junginių eliucija rodo, kad priemaišos struktūroje nėra fosfatinių grupių, kurios sąveikautų su sorbentu. Atsižvelgiant į minėtus aspektus ir literatūros duomenis galima daryti prielaidą, kad susidariusi priemaiša yra glikozidinio ryšio skilimo pasekmė. Literatūroje plačiai aprašomas glikozidinio ryšio skilimas vykstantis dėl rūgščios monofosforilinimo reakcijos mišinio terpės, o bazės priedas reakcijoje sumažina šią degradaciją. [30, 54, 52] Tokia pati tendencija buvo stebima ir mūsų eksperimentuose (Lentelė 8, Nr. 6 – 7).

dd^{l}UTP sintezei didesnėje skalėje buvo pasirinktos reakcijos sąlygos Nr. 4 (Lentelė 8), kurioms esant konversija į tikslini produktą siekė 93%. Trifosforilinimo stadijoje esant TBAPP buvo pasiekta 89% konversija į dd^{l}UTP, o produktas išgrynintas 82% išeiga.

Monofosforilinimo reakcijos eksperimentai toliau buvo atliekami su dd^{l}G. dd^{l}G monofosforilinimui, kaip ir dd^{l}U, $POCl_3$ buvo tinkamas fosforilinimo agentas. Nenaudojant bazės reakcija vyko lėtai, o po 5 valandų buvo gauta tik 42% konversija į monofosfatą. Didinant bazės kiekį reakcija vyko greičiau (2 val.), o konversija į dd^{l}GMP siekė 90% (Lentelė 9, Nr. 2 – 6). Atliekant sintezę didesnėje skalėje po valandos konversija į dd^{l}GMP siekė 92%. Po trifosforilinimo su TBAPP konversija į dd^{l}GTP buvo 82%. Produktas išgrynintas 57% išeiga.

Lentelė 9. dd^{l}G monofosforilinimo reakcijos sąlygų optimizavimas.

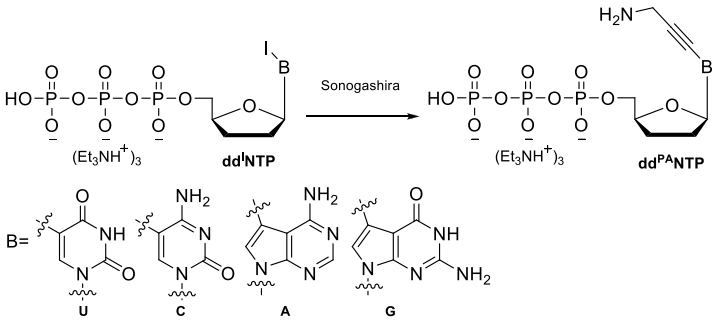
Nr.	**Fosforilinimo agentas**	**Fosforilinimo agento kiekis, ekv.**	**Tirpiklis**	**TBA, ekv.**	**Reakcijos laikas**	**Konversija į dd^{l}GMP, %**	**dd^{l}G, %**
1	$POCl_3$	3	TMP	-	5 val.	42	55
2		3	TMP	1	2,5 val.	62	36
3		3	TMP	2	3,5 val.	87	11
4[a]		**3**	**TMP**	**2,5**	**1** val. **1.5** val. **2** val.	**92** **92** **90**	**3** **1** **-**
5		3	TMP	2,8	1 val. 2 val.	93 81	1 -
6		3	TMP	3	1,5 val.	92	2

Reakcijos buvo atliktos 0 – 5 °C temperatūroje, reakcijų koncentracija – 0,1 M. a – sintezės didesnėje skalėje rezultatai (reakcijoje naudojamas 1 g pradinio nukleozido).

Monofosforilinimo reakcijų sąlygų optimizavimo metu buvo pastebėtos tam tikros tendencijos. Buvo stebimi reaktyvumų panašumai tarp dd^{I}A ir dd^{I}C, taip pat tarp dd^{I}G ir dd^{I}U. dd^{I}A ir dd^{I}C monofosforilinimui $P_2O_3Cl_4$ buvo efektyvesnis fosforilinimo agentas, reakcijos vyko trumpai (20 – 30 min). Tuo tarpu dd^{I}G ir dd^{I}U monofosforilinimui $POCl_3$ buvo tinkamesnis fosforilinimo agentas. Gerai konversijai į produktus pasiekti reakcijas reikėjo vykdyti 1,5 – 2 val. esant bazės priedui.

11.2 Jodintų-2',3'-dudeoksinukleotidų vandeninėje terpėje atliekamų Sonogashira reakcijų sąlygų optimizavimas

Antrasis svarbus pasirinktos sintezės strategijos žingsnis yra jungtuko formavimas įvedant funkcinę grupę tolimesniam žymėjimui. Ankstesniame skyriuje (10.2.2) buvo pritaikyta vandeninėje terpėje atliekama Sonogashira reakcija tiksliniams propargilaminą turintiems 2'-deoksinukleotidams gauti. Optimizavus dd^{I}NTP sintezės sąlygas ir susintetinus produktus didesnėje skalėje, toliau buvo optimizuojamos Sonogashira reakcijos sąlygos tiksliniams dd^{PA}NTP gauti (Schema 9).



Schema 9. Amino grupę turinčių dd^{PA}NTP sintezė.

Sonogashira reakcijas atliekant su dd^{I}CTP buvo išbandyti du katalizatoriai. Dėl prasto dd^{I}CTP tirpumo dimetilformamide reakcijose su $[(C_6H_5)_3P]_2PdCl_2$ buvo gautos mažos konversijos į produktą (Lentelė 10, Nr. 1 – 2). Kiti eksperimentai atlikti naudojant $Pd(OAc)_2$/TPPTS katalitinę sistemą. Esant 0,05 M pradinio junginio koncentracijai, nepriklausomai nuo propargilamino, bazės kiekio ar reakcijos temperatūros, konversija į dd^{PA}CTP geriausiu atveju siekė vos 16% (Lentelė 10, Nr. 3 – 5). Reikšmingas konversijos pokytis buvo pasiektas padidinus reakcijos koncentraciją iki 0,14 M (Lentelė 10, Nr. 6). Geros konversijos į tikslinį produktą buvo gautos reakcijas atliekant kambario temperatūroje ir nenaudojant bazės (Lentelė 10, Nr. 6 – 8). Atliekant Sonogashira reakciją didesnėje skalėje buvo gauta 75% konversija į dd^{PA}CTP

(Lentelė 10 Nr. 8). Pakartojus eksperimentą švelniai šildant (40 °C) po valandos konversija į dd^{PA}CTP siekė net 92% (Lentelė 10, Nr. 9). Produktas išgrynintas jonų mainų chromatografijos metodu, gauta 88% išeiga.

Lentelė 10. Sonogashira reakcijos sąlygų optimizavimas dd^{PA}CTP sintezei.

Nr.	Propargilaminas, ekv.	CuI, ekv.	Katalizatorius	Et_3N, ekv.	Koncentracija, M	Laikas	Konversija į dd^{PA}CTP, %
1[a]	2	0,2	$[(C_6H_5)_3P]_2PdCl_2$ (10 mol%)	3	0,05	16 val.	45
2[b]	2	0,2	$[(C_6H_5)_3P]_2PdCl_2$ (10 mol%)	3	0,05	16 val.	29
3	2	0,1	$Pd(OAc)_2$ (5 mol%) /TPPTS (25 mol%)	2	0,05	16 val.	16
4	3	0,1	$Pd(OAc)_2$ (5 mol%) /TPPTS (25 mol%)	10	0,05	16 val.	≤3
5 [c]	5	0,1	$Pd(OAc)_2$ (5 mol%) /TPPTS (25 mol%)	10	0,05	1 val.	9
6	2	0,1	$Pd(OAc)_2$ (5 mol%) /TPPTS (25 mol%)	2	0,14	1 val. 4 val. 16 val.	52 64 76
7	2	0,1	$Pd(OAc)_2$ (5 mol%) /TPPTS (25 mol%)	-	0,14	1 val. 4 val. 16 val.	69 72 79
8[d]	2	0,1	$Pd(OAc)_2$ (5 mol%) /TPPTS (25 mol%)	-	0,14	30 min. 1.5 val.	45 75
9[e]	**2**	**0,1**	**$Pd(OAc)_2$ (5 mol%) /TPPTS (25 mol%)**	**-**	**0,14**	**1** val.	**92**

Visos reakcijos buvo atliekamos kambario temperatūroje ir ACN:H_2O (1:2)mišinyje, nebent nurodyta kitaip. a – reakcija atlikta DMF, b – reakcija atlikta DMF:H_2O (2:1)mišinyje, c – reakcija atlikta 70 °C temperatūroje, d – sintezės didesnėje skalėje rezultatai, e – sintezės didesnėje skalėje rezultatai, reakciją atliekant 40 °C temperatūroje.

Monofosforilinimo reakcijų metu buvo pastebėtas reaktingumo panašumas tarp dd^{I}C ir dd^{I}A. Todėl optimizuojant Sonogashira reakcijos sąlygas dd^{PA}ATP sintezei pirmiausiai buvo išbandytos geriausios dd^{PA}CTP sintezės sąlygos. Konversija į produktą po valandos siekė 72% (Lentelė 11, Nr. 1). Reakcijoje naudojant 2 ekv. trietilamino (Et_3N), jau po 1 val. konversija į dd^{PA}ATP siekė 85% (Lentelė 11, Nr. 2). Tačiau ilginant reakcijos laiką buvo stebima ir produkto degradacija. Atliekant dd^{PA}ATP sintezę didesnėje skalėje

jau po 30 min konversija į produktą siekė 86% (Lentelė 11, Nr. 3). Po produkto gryninimo buvo gauta 69% išeiga.

Lentelė 11. Sonogashira reakcijos sąlygų optimizavimas $dd^{PA}ATP$ sintezei.

Nr.	Propargilaminas, ekv.	CuI, ekv.	Katalizatorius	Et_3N, eqkv	Laikas	Konversija į $dd^{PA}ATP$, %
1	2	0,1	$Pd(OAc)_2$ (5 mol%) /TPPTS (25 mol%)	-	1 val. 4 val. 16 val.	72 78 83
2	2	0,1	$Pd(OAc)_2$ (5 mol%) /TPPTS (25 mol%)	2	1 val. 4 val. 16 val.	85 79 61
3[a]	**2**	**0,1**	**$Pd(OAc)_2$ (5 mol%) /TPPTS (25 mol%)**	**2**	**30 min.**	**86**

Visos reakcijos buvo atliekamos kambario temperatūroje, ACN:H_2O (1:2) mišinyje. a – sintezės didesnėje skalėje rezultatai.

Lentelė 12. Sonogashira reakcijos sąlygų optimizavimas $dd^{PA}UTP$ sintezei.

Nr.	Propargilaminas, ekv.	CuI, ekv.	Katalizatorius	Bazė, ekv.	Koncentracija, M	Laikas	Konversija į $dd^{PA}UTP$, %
1	2	0,1	$Pd(OAc)_2$ (5 mol%) /TPPTS (25 mol%)	TEA, 2	0,07	1 val. 4 val. 16 val.	67 74 69
2	2	0,1	$Pd(OAc)_2$ (5 mol%) /TPPTS (25 mol%)	TEA, 10	0,07	1 val. 4 val. 16 val.	7 43 41
3	2	0,1	$Pd(OAc)_2$ (5 mol%) /TPPTS (25 mol%)	DIPEA, 2	0,07	1 val. 4 val. 16 val.	59 72 60
4	2	0,1	$Pd(OAc)_2$ (5 mol%) /TPPTS (25 mol%)	-	0,07	1 val. 4 val.	67 68
5[a]	**2**	**0,1**	**$Pd(OAc)_2$ (5 mol%) /TPPTS (25 mol%)**	**-**	**0,1**	**1 val.**	**92**

Visos reakcijos buvo atliekamos 40 °C temperatūroje, ACN:H_2O (1:2) mišinyje. a – sintezės didesnėje skalėje rezultatai.

Optimizuojant dd^{PA}UTP sintezės sąlygas buvo pastebėta, kad reakcijoje didinant bazės kiekį susidaro daugiau priemaišų, o esant 10 ekv. pertekliui gaunama vidutinė konversija į tikslinį junginį (Lentelė 12 Nr. 1 – 2 ir 4). Bazės prigimtis didelės įtakos reakcijos konversijai neturėjo (Lentelė 12 Nr. 3). Sintezę atliekant didesnėje skalėje buvo padidinta ir pradinio nukleotido koncentracija reakcijos mišinyje, dėl to po valandos konversija į dd^{PA}UTP siekė net 92%. Produktas išgrynintas 73% išeiga.

dd^{PA}GTP sintezės atveju kaip ir dd^{PA}UTP, Sonogashira reakciją reikėjo atlikti 40 °C temperatūroje, o geresnė konversija į produktą gauta, kai reakcijoje nebuvo naudojama bazė (Lentelė 13 Nr. 1 – 3). Atliekant dd^{PA}GTP sintezę didesnėje skalėje po 3 val. buvo gauta 72% konversija į produktą, po gryninimo gauta 41% išeiga (Lentelė 13 Nr. 4).

Lentelė 13. Sonogashira reakcijos sąlygų optimizavimas dd^{PA}GTP sintezei.

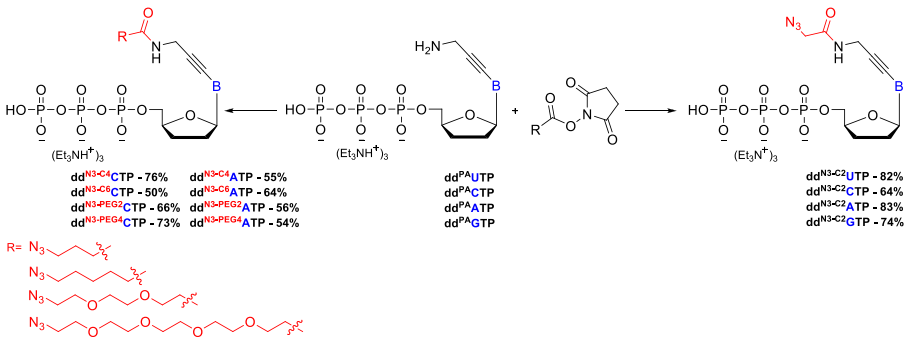
Nr.	Propargilaminas, ekv.	CuI, ekv.	Katalizatorius	Et_3N, ekv.	Koncentracija, M	Laikas	Konversija į dd^{PA}GTP, %
1[a]	2	0,1	$Pd(OAc)_2$ (5 mol%) /TPPTS (25 mol%)	-	0,07	4 val.	44
2	2	0,1	$Pd(OAc)_2$ (5 mol%) /TPPTS (25 mol%)	2	0,07	4 val.	56
3	2	0,1	$Pd(OAc)_2$ (5 mol%) /TPPTS (25 mol%)	-	0,07	4 val.	61
4[b]	**2**	**0,1**	**$Pd(OAc)_2$ (5 mol%) /TPPTS (25 mol%)**	**-**	**0,1**	**3 val.**	**72**

Visos reakcijos buvo atliekamos palaikant 40 °C temperatūrą ir ACN:H_2O (1:2) mišinyje. a – reakcija atlikta kambario temperatūroje, b – sintezės didesnėje skalėje rezultatai.

Optimizuojant vandeninėje terpėje atliekamos Sonogashira reakcijos sąlygas, pastebėtos bendros tendencijos. Gerai konversijai į tikslinius produktus pasiekti reakcijos turėjo būti atliekamos 40 °C temperatūroje, be papildomo bazės kiekio. Išimtis buvo stebima dd^{PA}ATP sintezės atveju – reakcija buvo atliekama kambario temperatūroje esant 2 ekv. bazės. Reakcijos koncentracija turėjo didelę įtaką produkto susidarymui ir reakcijos laikui. Visi, amino grupe funkcionalizuoti nukleotidai buvo sėkmingai susintetinti didesnėje skalėje geromis išeigomis (41 – 88%).

11.3 Azido grupę turinčių 2’,3’-dideoksinukleotidų ir oligonukleotidais modifikuotų 2’,3’-dideoksinukleotidų sintezė

Susintetinus amino grupe funkcionalizuotus dd^{PA}NTP, azido grupė buvo įvedama reakcijos su įvairiais azido-NHS esteriais metu (Paveikslas 12). Visi keturi N3-C2 jungtuką turintys nukleotidai buvo susintetinti geromis išeigomis (64 – 83%). Taip pat buvo susintetinta eilė dd^{N3}CTP ir dd^{N3}ATP su įvairaus ilgio ir cheminės prigimties jungtukais (išeigos 50 – 76%).



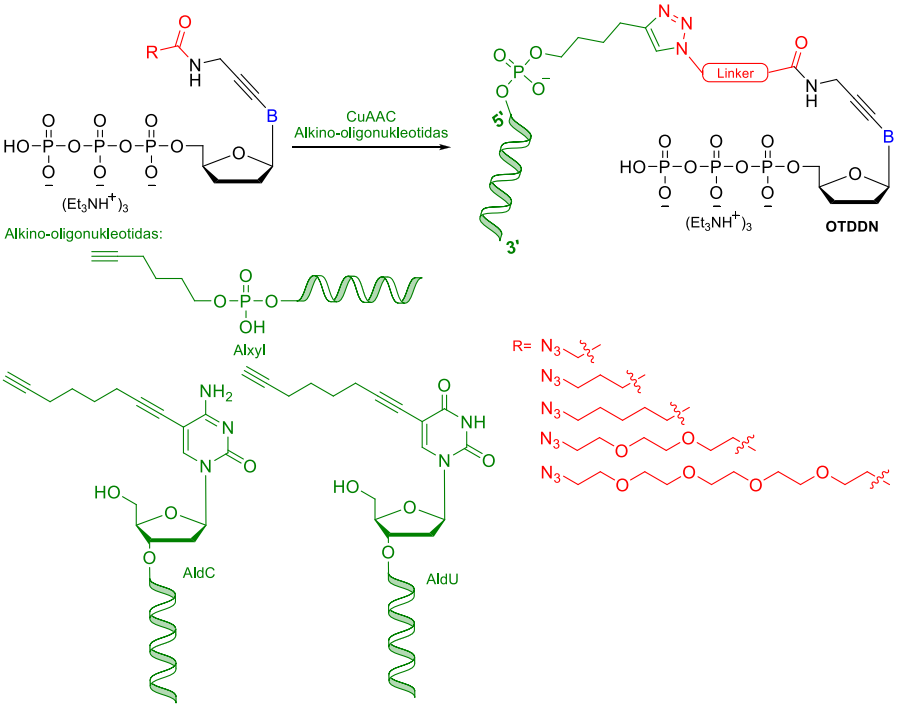
Paveikslas 12. Azido grupę turinčių 2’,3’-dideoksinukleotidų sintezė.

Skirtingoms taikymo sritims buvo susintetinta eilė OTDDN analogų, besiskiriančių oligonukleotidų sekomis ir modifikacijomis. CuAAC reakcijose buvo naudojami įvairūs azido-nukleotidai ir oligonukleotidai, turintys alkino grupę. Naudotų alkinų terminalinis trigubas ryšys buvo prijungtas arba prie 5’-fosfato, arba prie pirmos nukleobazės oligonukleotido sekoje (Schema 10). Kadangi Al modifikacija gali būti įvesta tik pirimidinuose (dC ir dU), o Alxyl bet kuriame nukleotide, tolimesni eksperimentai atlikti su Alxyl modifikaciją turinčiais oligonukleotidias.

Unikaliems molekuliniams identifikatoriams į transkriptus įvesti (UMI) buvo naudojami oligonukleotidai, turintys dalį atsitiktinių sekų (pažymėtos N). Tam buvo naudojami keli skirtingi alkino-ON (Lentelė 14, ON5 ir ON8 – 11). ON stabilumui nuo nukleazių poveikio pagerinti buvo naudojami alkino-ON su 3’-gale esančiomis tiofosfatinėmis modifikacijomis (pažymėtos *) (Lentelė 14). Fragmentų gryninimui su streptavidinu dengtomis dalelėmis atlikti, kai kurie alkino-ON turėjo biotino žymę 3’-gale. Siekiant padidinti atsparumą egzonukleazių poveikiui naudoti 3’-fosfatą turintys alkino-ON (Lentelė 14).

Pirmieji CuAAC reakcijos eksperimentai su alkino-ON buvo atliekami pagal sąlygas, nurodytas komerciškai prieinamame Jena Biocience “CuAAC

Biomolecule Reaction Buffer Kit" rinkinyje. Pirminė reakcijų skalė – 8,4 nmol alikino-ON. Atliekant reakcijos mišinių HPLC analizes buvo stebima oligonukleotidų degradacija. Vario (I) katalizuojamas oligonukleotidų skilimas CuAAC reakcijose yra plačiai aprašomas literatūroje. [235, 236, 237]



Schema 10. Oligonukleotidais modifikuotų 2',3'-dideoksinukleotidų (OTDDN) sintezė.

Atlikus eilę eksperimentų 8,4 nmol skalėje, OTDDN sintezė buvo padidinta iki 84 – 400 nmol (alkino-oligonukleotido). Ši sintezės skalė leido padidinti reaguojančių komponentų koncentraciją reakcijos mišinyje, o tuo pačiu dvigubai sumažinti Cu kiekį. Susintetinti įvairūs OTDDN yra pateikti lentelėje, kartu su tam tikromis charakteristikomis (Lentelė 15, ir Schema 10). Didžioji dalis susintetintų OTDDN buvo citidino ir uridino analogai. Reakcijas atliekant su ON6 (Lentelė 14) buvo susintetinti visi keturi nukleotidų analogai (Lentelė 15, Nr. 8, 10, 15 ir 17). Darant hipotezę, kad jungtuko ilgis turės didelę įtaką perskaitymo per nenatūralią jungtį efektyvumui tačiau neturint tinkamo metodo tam išmatuoti, trumpiausias jungtukas buvo pasirinktas pirmiesiems OTDDN taikymo eksperimentams atlikti. Sukūrus metodą perskaitymo efektyvumui įvertinti buvo susintetinti citidino OTDDN analogai su visais jungtukais (Lentelė 15, Nr. 10 – 14).

Lentelė 14. OTDDN sintezei naudoti oligonukleotidai.

Nr.	Pav.	Seka	5’-galas	3’-galas
1	ON1	**C**TATAGTGAGTCGTATTA	AldC	-
2	ON2	**U**AGATCGGAAGAGCACACGTCTG	AldU	Biotinas
3	ON3	**U**AGATCGGAAGAGCACACGTCTG	AldU	Fosfatas
4	ON4	**U**AGATCGGAAGAGCACACGTCTGA ACTCCAGTCACATGCCTAAATCTCG TATGCCGTCTTCTGCTTG	AldU	Biotinas
5	ON5	**U**NNNNNNNNAGATCGGAAGAGCGT CGTGTA	AldU	Biotinas
6	ON6	**A**GATCGGAAGAGCACACGTCTG	Alxyl	Biotinas
7	ON7	**A**GATCGGAAGAGCACACGT*C*T*G	Alxyl	Fosfatas
8	ON8	**N**NNNNNNNAGATCGGAAGAGCACA CGT*C*T*G	Alxyl	Fosfatas
9	ON9	**N**NNNNNNNAGATCGGAAGAGCGTC GTGTA	Alxyl	Biotinas
10	ON10	**N**NNNNNNNAGATCGGAAGAGCACA CGTCTG	Alxyl	Biotinas
11	ON11	**N**NNNNNNNAGATCGGAAGAGCGTC GTGTAGGGAAAGAG	Alxyl	Fosfatas
12	ON12	**A**GATCGGAAGAGCACACGTCTG	Alxyl	Fosfatas
13	ON13	**C**TGTCTCTTATACACATCTCCGAG	Alxyl	Biotinas
14	ON14	**AT**CACCGACTGCCCATAGAGAGGA AAGCGGAGGCGTAGTGG	Alxyl	Fosfatas

* – indikuoja tiofosfatines jungtis, N – indikuoja atsitiktinio nukleobazių sąstato sekas, kur N – bet kuris nukleotidas, AldU – alkino modifikacija įvesta į uridino C5-poziciją (okta-1,7-diinas), AldC – alkino modifikacija įvesta į citidino C5-poziciją (okta-1,7-diinas), Alxyl – heksinil- modifikacija, prijungta prie oligonukleotido 5’-fosfato.

Didžiausias iššūkis OTDDN ruošime, kaip ir OTDN atveju, buvo gryninimas. Kadangi alkino-ON nuo OTDDN struktūriškai skiriasi tik per vieną nukleotidą, jų atskyrimas HPLC, dėl panašių sulaikymo trukmių, yra sudėtingas. Azido-ddNTP, o ypač purinų analogai, dėl savo modifikacijos pasižymi ilga sulaikymo trukme. Tuo tarpu likutinis kiekis bet kurio reakcijos komponento yra kritiškas tolimesniame jų panaudojime. Varis inhibuoja polimerazes, azido-ddNTP konkuruoja su OTDDN fermentinio įtepimo metu, o likutinis alkino-ON kiekis eikvoja pradmenį tolimesniame nuskaitymo per nenatūralią jungtį eksperimente. Siekiant eksperimentus su OTDDN atlikti tiksliai, tinkami HPLC metodai turėjo būti sukurti užtikrinant reikiamą OTDDN grynumą. Tyrimo metu buvo susintetinti 28 skirtingi OTDDN, jų gryninimui atlikti ir reakcijoms stebėti buvo sukurti 18 analitinių ir 8 preparatyviniai HPLC metodai. Visų produktų grynumas siekė ≥95%, o kritinės priemaišos – nedetektuotos.

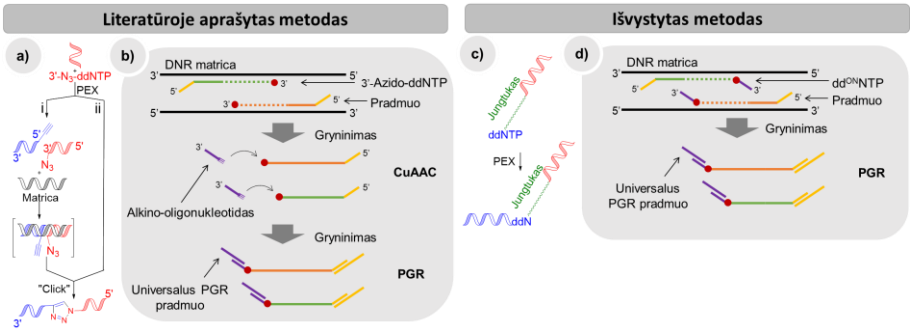
Lentelė 15. Susintetinti OTDDN.

Nr.	ON	Azidas	OTDDN	PJ, nmol	TJ, nmol	Išeigos, %
1	ON1	dd$^{N3\text{-}C2}$UTP	dd$^{ON1\text{-}C2}$UTP	8.4	1.4	17
2	ON2	dd$^{N3\text{-}C2}$UTP	dd$^{ON2\text{-}C2}$UTP	8.4	2.5	30
3		dd$^{N3\text{-}C2}$CTP	dd$^{ON2\text{-}C2}$CTP	8.4	3.4	41
4	ON3	dd$^{N3\text{-}C2}$UTP	dd$^{ON3\text{-}C2}$UTP	8.4	1.8	21
5	ON4	dd$^{N3\text{-}C2}$CTP	dd$^{ON4\text{-}C2}$CTP	8.4	2.6	31
6	ON5	dd$^{N3\text{-}C2}$UTP	dd$^{ON5\text{-}C2}$UTP	8.4	3.5	42
7	ON6	dd$^{N3\text{-}C2}$UTP	dd$^{ON6\text{-}C2}$UTP	8.4	2.1	25
8		dd$^{N3\text{-}C2}$UTP	dd$^{ON6\text{-}C2}$UTP	200	68	34
9		dd$^{N3\text{-}C2}$CTP	dd$^{ON6\text{-}C2}$CTP	84	24	28
10		dd$^{N3\text{-}C2}$CTP	dd$^{ON6\text{-}C2}$CTP	210	82	39
11		dd$^{N3\text{-}C4}$CTP	dd$^{ON6\text{-}C4}$CTP	210	51	24
12		dd$^{N3\text{-}C6}$CTP	dd$^{ON6\text{-}C6}$CTP	210	39	18
13		dd$^{N3\text{-}PEG2}$CTP	dd$^{ON6\text{-}PEG2}$CTP	210	30	14
14		dd$^{N3\text{-}PEG4}$CTP	dd$^{ON6\text{-}PEG4}$CTP	210	25	12
15		dd$^{N3\text{-}C2}$ATP	dd$^{ON6\text{-}C2}$ATP	100	29	29
16		dd$^{N3\text{-}C2}$GTP	dd$^{ON6\text{-}C2}$GTP	8.4	1.6	19
17		dd$^{N3\text{-}C2}$GTP	dd$^{ON6\text{-}C2}$GTP	210	90	43
18	ON7	dd$^{N3\text{-}C2}$UTP	dd$^{ON7\text{-}C2}$UTP	16.8	3.1	18
19		dd$^{N3\text{-}C2}$UTP	dd$^{ON7\text{-}C2}$UTP	200	24	12
20		dd$^{N3\text{-}C2}$CTP	dd$^{ON7\text{-}C2}$CTP	210	50	24
21	ON8	dd$^{N3\text{-}C2}$UTP	dd$^{ON8\text{-}C2}$UTP	16.8	2.7	16
22		dd$^{N3\text{-}C2}$CTP	dd$^{ON8\text{-}C2}$CTP	8.4	1.5	18
23	ON9	dd$^{N3\text{-}C2}$UTP	dd$^{ON9\text{-}C2}$UTP	8.4	2.2	27
24		dd$^{N3\text{-}C2}$CTP	dd$^{ON9\text{-}C2}$CTP	8.4	0.9	11
25	ON10	dd$^{N3\text{-}C2}$UTP	dd$^{ON10\text{-}C2}$UTP	100	32	32
26		dd$^{N3\text{-}C2}$CTP	dd$^{ON10\text{-}C2}$CTP	210	115	55
27	ON11	dd$^{N3\text{-}C2}$UTP	dd$^{ON11\text{-}C2}$UTP	200	40	20
28		dd$^{N3\text{-}C2}$CTP	dd$^{ON11\text{-}C2}$CTP	210	78	37
29	ON12	dd$^{N3\text{-}C2}$UTP	dd$^{ON12\text{-}C2}$UTP	400	116	29
30		dd$^{N3\text{-}C2}$CTP	dd$^{ON12\text{-}C2}$CTP	300	143	48
31	ON13	dd$^{N3\text{-}C2}$CTP	dd$^{ON13\text{-}C2}$CTP	210	77	37
32	ON14	dd$^{N3\text{-}C2}$UTP	dd$^{ON14\text{-}C2}$CTP	100	36	36
33		dd$^{N3\text{-}C2}$CTP	dd$^{ON14\text{-}C2}$CTP	100	32	32

Visose CuAAC reakcijose buvo naudojami 2 – 3 ekv. atitinkamo azido, reakcijos atliktos 37 °C temperatūroje. PJ – pradinis junginys (alkino-oligonukleotidas), TJ – tikslinis junginys (atitinkamas OTDDN). Visi reakcijos komponentai buvo naudojami pasiruošiant jų atitinkamus vandeninius tirpalus. Alkino-oligonukleotido tirpalas buvo ruošiamas fosfatiniame buferyje. Atliekant reakcijas 8,4 nmol skalėje reakcijos tūris buvo 200 uL, 84 – 400 nmol skalėje – 1 ml.

12. Oligonukleotidais modifikuotų 2’,3’-dideoksinukleotidų panaudojimas NKS bibliotekų ruošime

Gebėjimas nustatyti nukleorūgščių sekas įgalino revoliucinį proveržį šiuolaikinės biologijos ir biomedicinos mokslų galimybių plėtroje. Per pastaruosius du dešimtmečius buvo sukurtos naujos, alternatyvios NKS technologijos, įgalinančios didelio našumo, greitą ir efektyvų sekos analizių atlikimą. Dėl ženkliai sumažėjusių kaštų, reikalingų analizėms atlikti, NKS tapo kasdieniu įrankiu daugelyje gamtos mokslų sričių. Visgi, klasika tapusi NKS technologija vis dar turi iššūkių, susijusių su mėginių paruošimu. Sukurti alternatyvūs NKS bibliotekų ruošimo metodai, kaip Click-Seq, ženkliai supaprastina procesą. Nors fermentinis adapterių ligavimas yra pakeičiamas CuAAC reakcija grįsta alternatyva, technologija vis tiek turi tam tikrų iššūkių: prieš cheminį ligavimą („click“ reakciją) turi būti atliekamas fragmentų gryninimas azido-ddNTP pašalinti; cheminis ligavimas yra neefektyvus, jei nenaudojama matrica abiem oligonukleotidams (matricos panaudojimas nežinomų sekų analizei neįmanomas) (Paveikslas 13a); cheminio ligavimo metu vykstanti oligonukleotidų degradacija mažina NKS bibliotekos kompleksiškumą; po adapterių prijungimo būtinas gryninimas ir efektyvus Cu pašalinimas (Cu slopina polimerazių aktyvumą); labai mažas perskaitymo per nenatūralią jungtį efektyvumas (< 4%); žemo tikslumo komplementarios grandinės sintezė (Paveikslas 13a ir b). [12, 13, 229, 14] Nepaisant to, šie



Paveikslas 13. NKS bibliotekų paruošimo metodų palyginimas: literatūroje aprašyto metodo a) schema reprezentuojanti cheminį ribozė-ribozė ligavimą, i) “click” reakcija su papildomai naudojama matrica, ii) “click” reakcija be matricos, b) NKS bibliotekų paruošimo procesas pritaikant ribozė-ribozė cheminį ligavimą; Išvystytas metodas: c) ON įvedimas PEX terminavimo su OTDDN metu, d) NKS bibliotekų paruošimo procesas panaudojant OTDDN.

tyrimai pademonstravo nukleorūgščių su nenatūraliais jungtukais tarp nukleotidų panaudojimą molekulinėje biologijoje, tuo pačiu atverdami galimybes naujų dirbtinių DNR karkasų kūrimui.

NKS bibliotekų ruošimo kontekste OTDDN įterpimas leistų apjungti du pagrindinius – fragmentų generavimo ir adapterių pridėjimo – etapus į vieną (Paveikslas 13c ir d). OTDDN panaudojimas vietoje azido-ddNTP PEX reakcijoje generuotų ne tik specifinio diapazono DNR fragmentus, bet tuo pačiu metu būtų įvestas žinomos sekos adapteris. Toks metodas dar labiau sutrumpintų NKS bibliotekų paruošimą, apjungiant tris etapus į vieną (fragmentavimą, galų paruošimą ir adapterių ligavimą) ir išvengtų prieš tai minėtų iššūkių (Paveikslas 13c ir d).

12.2 Azido grupę turinčių ir oligonukleotidais modifikuotų 2',3'-dideoksinukleotidų fermentinis įterpimas

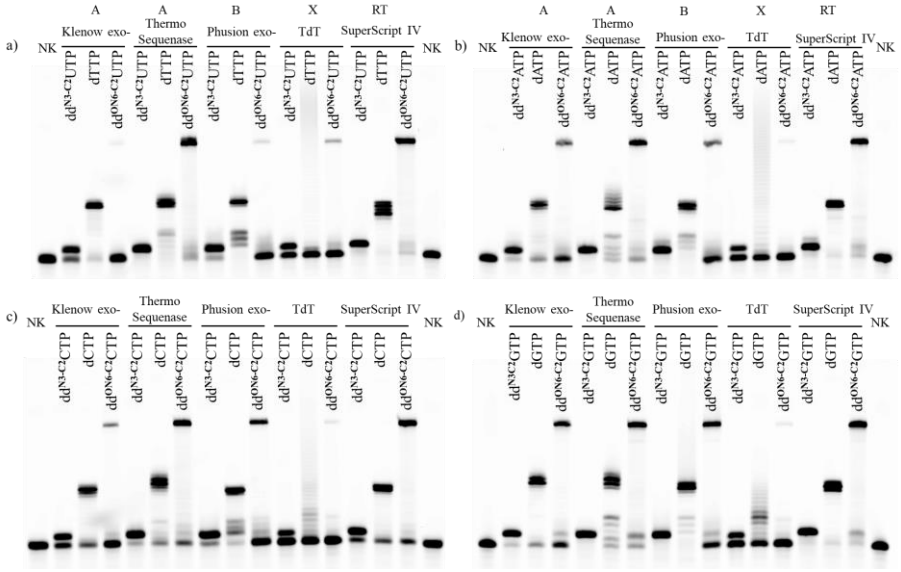
Susintetinti ir išgryninti $dd^{N3}NTP$ ir $dd^{ON6\text{-}C2}NTP$ buvo testuojami pradmens pratęsimo reakcijose naudojant A, B, X ir AT šeimų polimerazes. Bendrai, DNR polimerazės yra skirstomos į septynias šeimas (A, B, C, D, X, Y ir AT) pagal kristalografinę struktūrą ir sekos homologiją, dėl kurių gali skirtis jų funkcionalumas su tiriamais substratais. Pasirinktoje sistemoje buvo naudojami oligonukleotidų dupleksai su 10 nt iškyšomis (Lentelė 16). Visos išbandytos polimerazės neturėjo 3'-5' egzonukleazinio aktyvumo ir įterpė $dd^{N3}NTP$ ir $dd^{ON6\text{-}C2}NTP$. Pastebėti $dd^{ON6\text{-}C2}NTP$ įterpimo efektyvumo skirtumai tarp polimerazių, kaip ir pirmenybės teikimas specifinėms terminatorių nukleobazėms. Klenow exo- pasižymėjo geresniu inkorporavimo efektyvumu naudojant oligonukleotidais modifikuotus 2',3'-dideoksipurinus, o ne pirimidinus. Panaši tendencija buvo pastebėta $dd^{N3}NTP$ darinių atveju.

Lentelė 16. Oligonukleotidų dupleksai naudoti $dd^{N3\text{-}C2}NTP$ ir $dd^{ON6\text{-}C2}NTP$ įterpimo eksperimentuose.

Pav.	Oligonukleotidų dupleksas
Dup^{A}	5'-AAAAAAAAAATACGCCAAGGATGCCTACCCATGTCTGCA-3' 3'-ATGCGGTTCCTACGGATGGGTACAGACGT-Cy5-5'
Dup^{T}	5'-TTTTTTTTTTTACGCCAAGGATGCCTACCCATGTCTGCA-3' 3'-ATGCGGTTCCTACGGATGGGTACAGACGT-Cy5-5'
Dup^{G}	5'-GGGGGGGGGGTACGCCAAGGATGCCTACCCATGTCTGCA-3' 3'-ATGCGGTTCCTACGGATGGGTACAGACGT-Cy5-5'
Dup^{C}	5'-CCCCCCCCCCTACGCCAAGGATGCCTACCCATGTCTGCA-3' 3'-ATGCGGTTCCTACGGATGGGTACAGACGT-Cy5-5'

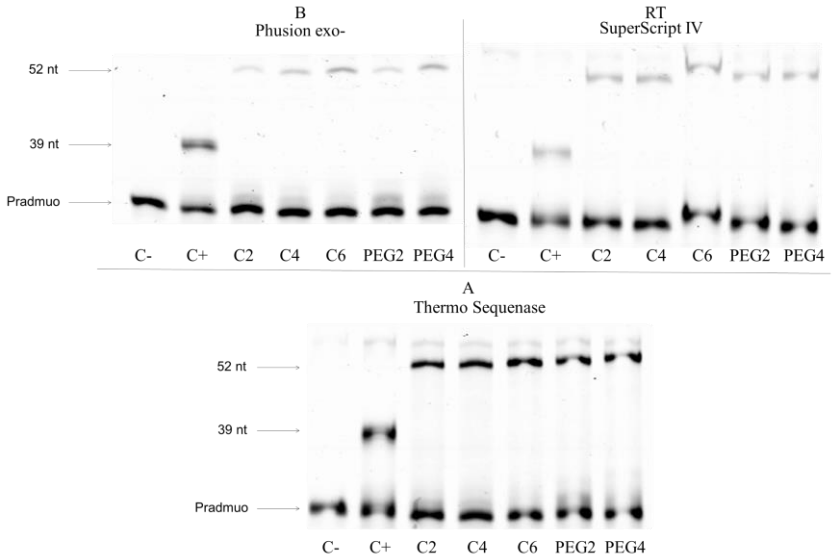
Kita A šeimos polimerazė, Thermo Sequenase, puikiai įterpė visus dd^{N3}NTP ir dd$^{ON6\text{-}C2}$NTP substratus (Paveikslas 14). Tai dalinai galėtų būti siejama su jos žinoma savybe nediskriminuoti ddNTP. [239]

B šeimos DNR polimerazė Phusion exo- dd$^{ON6\text{-}C2}$NTP įterpė prasčiau nei Thermo Sequenase, ypač dd$^{ON6\text{-}C2}$UTP atveju (Paveikslas 14a). Mažiausiu įterpimo efektyvumu pasižymėjo terminalinė deoksinukleotidiltransferazė (TdT), nors ir yra žinoma, kad TdT toleruoja steriškai dideles modifikacijas nukleobazėse. [187] Gauti rezultatai galėtų būti dalinai siejami su neoptimaliomis reakcijų sąlygomis (dd$^{ON6\text{-}C2}$NTP ir matricos santykis). Atvirkštinė transkriptazė (AT) SuperScript™ savo įterpimo efektyvumu buvo labai panaši į Thermo Sequenase (Paveikslas 14).



Paveikslas 14. Fermentinis dd^{N3}NTP ir dd$^{ON6\text{-}C2}$NTP įterpimas naudojant įvairias DNR polimerazes. PEX reakcijų su įvairiomis DNR polimerazėmis, naudojant dd^{N3}NTP, dNTP arba dd^{ON}NTP, elektroferogramos. Eksperimentuose naudotos matricos: a) Dup^{A}, b) Dup^{T}, c) Dup^{G}, d) Dup^{C}. NK – neigiama kontrolė.

Įvairius jungtukus tarp nukleobazės ir oligonukleotido turinčių dd^{ON6}CTP įterpimas buvo taip pat įvertintas. Pastebėtos tos pačios polimerazių tendencijos, be reikšmingų skirtumų, susijusių su jungtuko struktūra. Thermo Sequenase pasižymėjo geriausiu įterpimo efektyvumu, kaip buvo pastebėta ir anksčiau (Paveikslas 15).



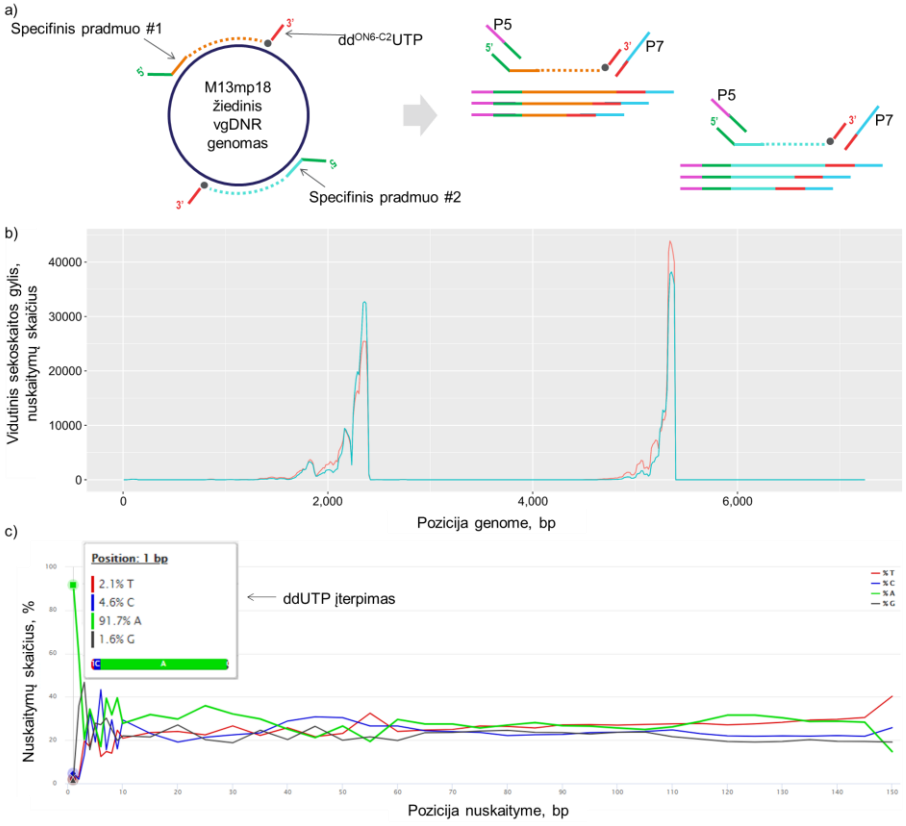
Paveikslas 15. Skirtingus jungtukus turinčių $dd^{ON6}CTP$ įterpimas naudojant skirtingas DNR polimerazes. K- – neigiama kontrolė, nenaudojami natūralūs dNTP, K+ – teigiama kontrolė, reakcijoje naudojant dCTP; C2, C4, C6, PEG2, PEG4 – indikuoja skirtingu jungtukus.

12.3 DNR bibliotekų ruošimas panaudojant oligonukleotidais modifikuotus 2',3'-dideoksinukleotidus

12.3.1 Pusiau taikininė M13mp18 viruso genomo sekoskaita

Siekiant įrodyti metodo pritaikomumo galimybes, kaip modelinė sistema NKS bibliotekos ruošimui naudojant OTDDN, buvo pasirinktas mažo kompleksiškumo viengrandis M13mp18 bakteriofago genomas. Buvo parinkti specifiniai pradmenys nukreipti į du M13mp18 genomo lokusus, su daliniais P5 Illumina™ platformos adapteriais (Paveikslas 16a – #1 ir #2 – specifiniai pradmenys). $dd^{ON6\text{-}C2}UTP$ įterpimui buvo naudojama Thermo Sequenase, atliekant 15 linijinių pradmens pratęsimo ciklų. OTDDN įterpimo metu terminuojama DNR sintezė ir tuo pačiu metu susintetinta nukleorūgštis pažymima daliniu Illumina adapteriu. Šiuo etapu sugeneruoti tinkamo ilgio fragmentai, kurie vėliau buvo PGR amplifikuojami, taip įvedant pilnus P5 ir P7 Illumina adapterius (Paveikslas 16a). Phusion exo- polimerazė buvo naudojama bibliotekos amplifikacijos reakcijoje, siekiant nuskaityti per nenatūralią OTDDN jungtį, o paruošta biblioteka buvo sekvenuojama. Gauti duomenys parodė, kad nusekvenuoti du genomo lokusai su būdinga fragmentų struktūra: fiksuotas intarpo galas atitiko PEX naudotų pradmenų sekas, o atsitiktinai pasiskirstę galai atitiko atsitiktinį $dd^{ON6\text{-}C2}UTP$ įterpimą (Paveikslas 16b). Analizuojant nuskaitymų sekas, nebuvo pastebėta jokių

sisteminių klaidų ar artefaktų, kurie galėtų būti susiję su nenatūralaus DNR karkaso buvimu fragmentų bibliotekoje. Nukleobazių pasiskirstymo analizė parodė ženklų adenino dominavimą (>90%) pirmoje atvirkštinio nuskaitymo padėtyje (Paveikslas 16c), o tai atitinka ddUTP įterpimo vietą. Tai rodo, kad absoliuti dauguma sekvenuotų molekulių susidarė be perskaitymo trikdžių, kitaip sakant, komplementarią grandinę sintetinanti polimerazė įterpė nukleotidą komplementarų OTDDN-ui iškart už nenatūralios jungties.



Paveikslas 16. Pusiau-taikininė M13mp18 viruso genomo sekoskaita. a) NKS bibliotekos paruošimo schema panaudojant dd$^{ON6\text{-}C2}$UTP. b) Nuskaitymai koncentruoti dvejuose lokusuose, kurių vienas sekvenuotas intarpo galas užfiksuotas specifiniais pradmenimis konkrečiose genomo vietose. Kitas galas yra pažymimas atsitiktiniu dd$^{ON6\text{-}C2}$UTP įterpimu. Oranžinė ir mėlynos linijos žymi techninius pakartojimus. c) Nukleobazių pasiskirstymo nusekvenuotuose fragmentuose grafikas.

Palyginimui, minėtu ribozė-ribozė cheminio ligavimo atveju, buvo detektuojamas 69% nuskaitymų tikslumas. Autoriai klaidingų sekų atvejus aiškino kaip nukleobazės pakeitimo įvykius, atsirandančius dėl nenatūralios jungties. [15] Kadangi atvirkštinių nuskaitymų galai mūsų sekos nustatymo eksperimente buvo prilyginti prie referentinio genomo net ir tuo atveju, kai

pirmoji bazė nėra A, netobulas nuskaitymas greičiausiai atsirado dėl bazės praleidimo, sąlygojančio nukleotido deleciją. Svarbu tai, kad šis retas įvykis nemažina sekoskaitos duomenų kokybės ir neapsunkina jų analizės. [240]

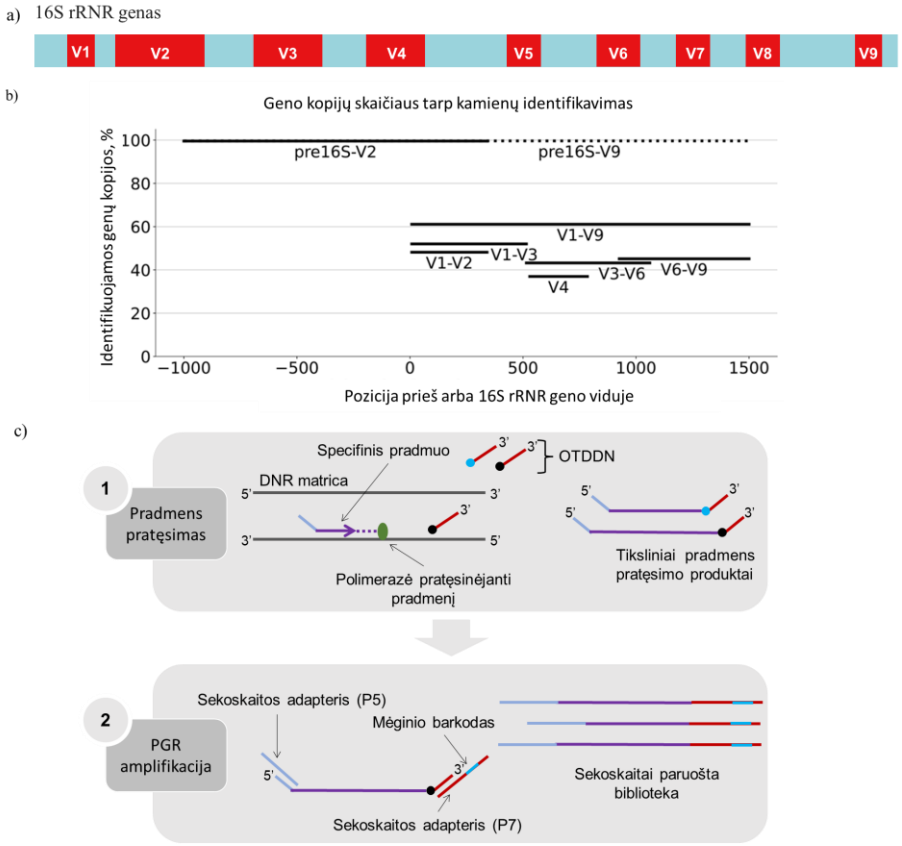
Sukurtą NKS bibliotekų ruošimo technologiją pavadinome pusiau taikinine sekoskaita, pabrėžiant, kad norimam lokusui užfiksuoti reikalingas tik vienas pradmuo, o sintetinamos grandinės terminacija OTDDN vyksta nepriklausomai nuo matricos sekos pasroviui nuo pradmens.

12.3.2 Mikrobiomo analizė pasitelkiant pusiau taikininę 16S rRNR geno sekoskaitą

Mikrobiomas yra mikroorganizmų populiacija arba bendra mikroorganizmų genetinė informacija tam tikroje aplinkoje (pvz., dirvožemyje, konkrečios vietovės vandenyje, žmonių, gyvūnų ekskrementuose). Mikrobiomų profiliavimas yra labai svarbus ir turi didelį potencialą nustatyti mikroorganizmus ir atskleisti jų populiacijos raidą konkrečioje buveinėje, taip pat išsiaiškinti funkcinius bendruomenių skirtumus arba ištirti, kaip mikrobiomų populiacijų įvairovę veikia išoriniai veiksniai (pvz., dirvožemio tręšimas). Viena iš galimybių charakterizuoti mikrobų populiaciją yra paremta specifinės bakterijų genomo srities – 16S rRNR geno – sekos nustatymu. Šis genas yra kiekvienos bakterijos genome ir yra būtinas pastarųjų išgyvenimui. 16S rRNR geno seka susideda iš devynių variabilių regionų (V1 – V9), besiribojančių su konservatyviomis sekomis (Paveikslas 17a). Glaudžiai susijusių bakterinių rūšių nustatymas gali būti sudėtingas sekvenuojant tik 16S rRNR geną, kadangi geno sekos skirtumai gali būti tokie maži kaip vieno nukleotido polimorfizmas (VNP), kuris nebūtinai bus variabiliuose regionuose. Tuo tarpu kiekybinį identifikavimą apsunkina galimas kelių 16S RNR geno kopijų buvimas viename bakterijos genome. Iškėlėme hipotezę, kad labai konservatyvių 16S rRNR geno regionų sekoskaita kartu su mažesnio konservatyvumo regionais už geno ribų gali suteikti vertingų duomenų, kad būtų galima tiksliau nustatyti mikrobų bendruomenės sudėtį rūšies lygmeniu. [241]

Pirmiausiai buvo atlikta viešai prieinamų bakterinių genomų *in silico* analizė apimant 16S rRNR geno sekas kartu su 1 kb regionais prieš srovę nuo šio geno. Gauta duomenų bazė parodė bakterijų teorinio identifikuojamumo procentą, priklausomai nuo sekvenavimui pasirinktų genų regionų (Paveikslas 17b). Pavyzdžiui, sekvenuojant V1 – V9 regioną butų identifikuojama 60% populiacijos, o V1 – V2 apie 50%. Tuo tarpu sekvenavimas prieš srovę nuo 16S rRNR geno leistų identifikuoti beveik 100% populiacijos. Praktiniam šių duomenų patikrinimui NKS bibliotekos buvo paruoštos naudojant DNR, išgrynintą iš dirvožemio mėginių. NKS biblioteka buvo ruošta apimant V1 – V2 16S rRNR geno regionus ir genomo dalį prieš srovę nuo 16S rRNR geno.

Tam pasiekti buvo parinkti specifiniai pradmenys, orientuoti į konservatyvų regioną tarp V2 ir V3 variabilių sričių bei turintys dalines Illumina adapterio sekas 5’ gale. PEX reakcijos su šiais pradmenimis buvo atsitiktinai terminuojamos OTDDN, tokiu būdu įvedant antrą dalinę adapterio seką. Manipuliuojant OTDDN ir natūralių dNTP santykiu buvo reguliuojamas susidarančių fragmentų ilgis. Gauti pradmens pratęsimo produktai buvo amplifikuojami indeksų įvedimo PGR metu, kur buvo gauti pilnus P5 ir P7 Illumina adapterius turintys fragmentai (Paveikslas 17c).



Paveikslas 17. Mikrobiomo analizė pusiau taikininės sekoskaitos metodu , a) 16S rRNR genas: V1 – V9 indikuoja variabilius regionus, mėlyni intarpai reprezentuoja konservatyvius regionus, b) teoriškai identifikuojamų 16S rRNR geno kopijų skaičiaus procentas, nustatytas pagal įvairius 16S rRNR geno regionus, c) 16S rRNR geno NKS bibliotekos paruošimas pusiau taikininei sekoskaitai.

Kaip buvo prognozuojama iš *in silico* analizės duomenų, bibliotekos paruošimas, apimant dalį šalia 16S rRNR sekų regiono, pagerino klasifikavimo tikslumą rūšių lygmeniu. Taikant sukurtą metodą buvo identifikuotos bakterijų rūšys, kurių atskyrimas buvo neįmanomas naudojant

kitus šiuo metu plačiai naudojamus metodus. OTDDN pagrįstas metodas taip pat leido identifikuoti 16S rRNR geno kopijų skaičių, ko kiti metodai, grįsti tik 16S rRNR vidugeninių sričių sekoskaita, negalėjo įvertinti. Tai įrodo metodo pranašumą ir didelę diagnostinę vertę. [241]

12.4 RNR bibliotekų ruošimas panaudojant oligonukleotidais modifikuotus 2',3'-dideoksinukleotidus

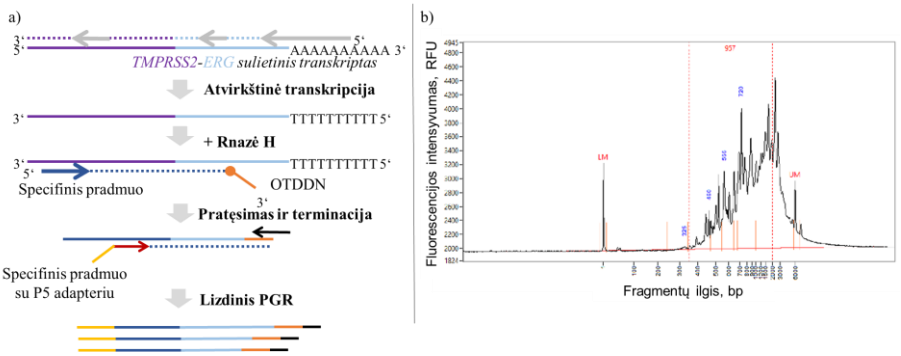
12.4.1 Prostatos vėžio sulietinių RNR transkriptų sekoskaita

Navikiniai susirgimai yra šių dienų iššūkis medicinos ir kitų biomokslų bendruomenei. Skiriamas didelis dėmesys siekiant suprasti pastarųjų atsiradimo priežastis, gebėti to išvengi ir pasiūlyti efektyvius gydymo metodus.

Genetiniai persitvarkymai, kurių metu yra sujungiami nesusiję, skirtingą genetinę informaciją koduojantys genai, yra dažni prostatos vėžio reiškiniai. TMPRSS2 yra baltymas, reaguojantis į vyriškųjų hormonų pokyčius. ERG – transkripcijos veiksnys, reguliuojantis ląstelių dalijimąsi ir metabolizmą, įjungdamas tam tikrų genų ekspresiją. Esant pastarojo pertekliui ląstelės yra linkusios supiktybėti. *TMPRSS2* ir *ERG* genų susiliejimas buvo pastebėtas prostatos vėžio atveju. Genų suliejimai gali sąlygoti ne tik pilnų genų sujungimą, bet ir nepilnų jų dalių. Tokių susiformavusių iRNR tolimesnė transliacijos išdava gali būti visiškai arba dalinai neaktyvių atitinkamų baltymų sintezė. *TMPRSS2-ERG* sulietinių genų įvairovės identifikavimas bei nustatymas, kas yra ekspresuojama navikinėse ląstelėse, yra vienas iš būdų leidžiantis parinkti tinkamą gydymą. Buvo sukurta nauja, pusiau taikininė RNR sekoskaitos metodika, grįsta OTDDN panaudojimu. Metodas įgalina praturtinti tikslinį geną (mūsų atveju *TMPRSS2*) ir tuo pačiu metu profiliuoti pastarojo 3'-galo liejinių įvairovę (Paveikslas 18).

NKS bibliotekų paruošimui iš visuminės RNR pirmiausiai buvo susintetinta kDNR atvirkštinės transkripcijos metu. Panaudojus RNazę H, buvo hidrolizuota RNR grandinė. Toliau buvo atlikta PEX reakcija su specifiniais pradmenimis, komplementariais *TMPRSS2* geno egzonui 1. Reakcijos produktų sintezė buvo terminuota OTDDN. Gauti fragmentai toliau buvo amplifikuojami, o gauta biblioteka sekvenuojama Illumina platforma (Paveikslas 18).

Apibendrinant, pasiūlytas metodas įgalino praturtinti dominantį geną ir tuo pačiu metu profiliuoti visą įvairovę su juo susiliejusių genų, net ir neturint pirminės informacijos apie šių genų-partnerių sekas. Atlikus sekoskaitą buvo detekduota plati įvairovė *TMPRSS2* geno 3'-galo liejinių izoformų ir net 11 ankščiau neaprašytų genų-partnerių.



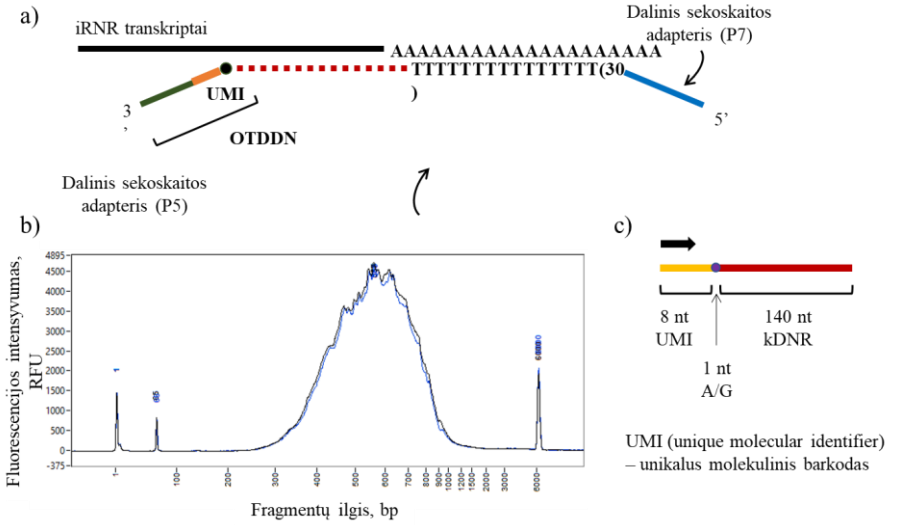
Paveikslas 18. NKS bibliotekos paruošimas sulietinių transkriptų analizei, a) bibliotekos paruošimo procesas, b) gautos bibliotekos fragmentų dydžio pasiskirstymas.

12.4.2 Genų raiškos analizė

Transkriptomas yra RNR visuma specifinėje ląstelėje tam tikru jos vystymosi laikotarpiu ir tam tikromis aplinkos sąlygomis. Kiekviena daugialąsčio organizmo ląstelė turi tą pačią genetinę informaciją, tačiau ląstelių funkcijų diversifikavimą lemia tai, kaip ši genetinė informacija išreiškiama į iRNR, o vėliau į baltymus. Transkriptomo analizė suteikia išsamią informaciją apie ląstelės reguliacinius veiksnius ir turi didelį potencialą ligų diagnostikoje. Pavyzdžiui, navikinių ląstelių identifikavimas, jų funkcijų ir baltymų ekspresijos pokyčių nustatymas įgalintų parinkti efektyvų gydymą.

Transkriptomo analizei buvo sukurtas paprastas ir greitas vieno mėgintuvėlio iRNR sekos nustatymo metodas, grįstas OTDDN panaudojimu. Kadangi visos iRNR turi poliA uodegėlę 3’-gale, atvirkštinei transkripcijai (AT) buvo atitinkamai parinktas komplementarus oligo(dT) pradmuo. Čia, kaip ir anksčiau aprašytuose metoduose, pradmens pratęsimo reakcija yra atsitiktinai terminuojama OTDDN, taip suformuojant kDNR fragmentus, pasiskirsčiusius specifinio ilgio intervale (Paveikslas 19). Siekiant nustatyti ląstelių funkciją būtina gebėti kiekybiškai įvertinti iRNR sąstatą. Šiam tikslui pasiekti eksperimentuose buvo naudojami OTDDN, savo sekoje turintys 8 nt atsitiktines sekas, kurios veikė kaip unikalūs molekuliniai identifikatoriai (UMI) (Paveikslas 19a ir c). Siekiant metodą supaprastinti dar labiau, tie patys eksperimentai buvo atlikti ne tik su išgrynintais iRNR mėginiais, bet ir tiesiogiai su ląstelių lizatais. Reakcija buvo atlikta viename mėgintuvėlyje esant tiriamoms ląstelėms, lizės buferiui ir visiems atvirkštinės transkripcijos komponentams bei OTDDN. Taip paruošti mėginiai dalyvavo indeksų įvedimo PGR amplifikacijoje, o paruošta biblioteka toliau buvo sekvenuojama.

Sukurtas transkriptomo analizės metodas parodė, kad OTDDN panaudojimas įgalino greitai ir paprastai generuoti kDNR bibliotekas, praturtintas 3’-galo sekomis, ir ženkliai sutrumpino visą NKS bibliotekų paruošimo procesą. Tuo tarpu atsitiktinių sekų buvimas OTDDN struktūroje įgalino kiekvieną fragmentą pažymėti unikaliu molekuliniu identifikatoriumi, suteikiant tikslumą kiekybiniam iRNR vertinimui (unikalus žymėjimas leido tiksliai atskirti biologines transkriptų kopijas nuo techninių). [242]



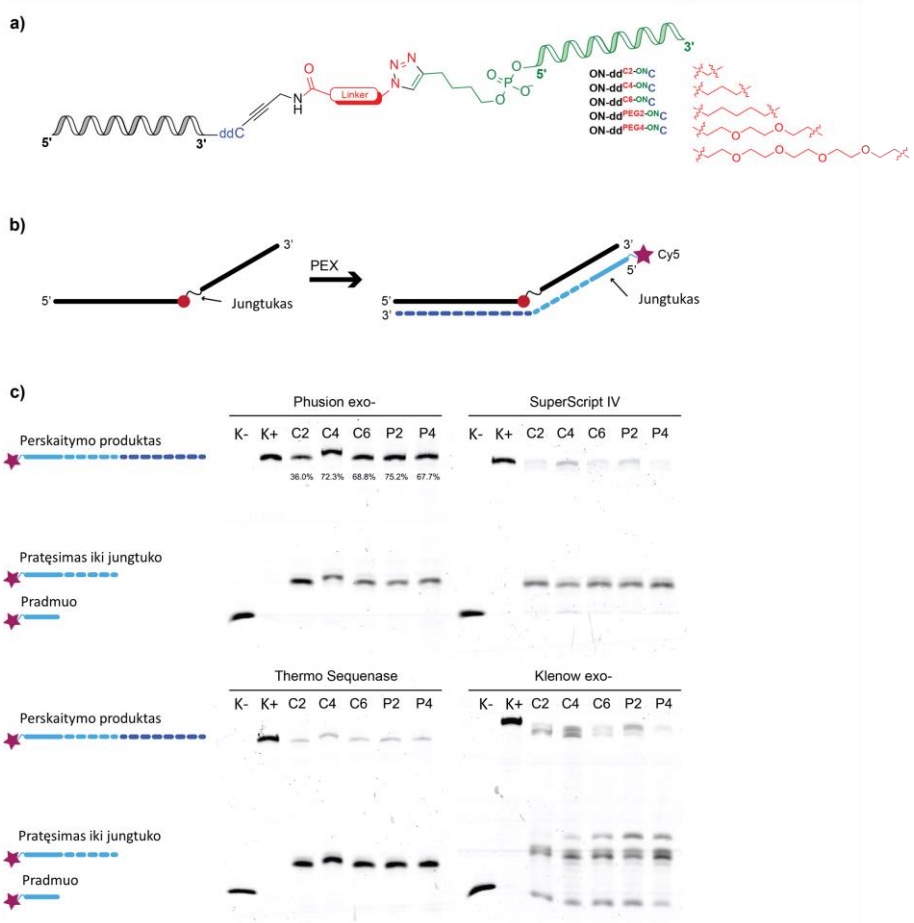
Paveikslas 19. NKS bibliotekų paruošimo procesas iRNR 3’-galų profiliavimui. a) bibliotekos paruošimo procesas, b) gautos bibliotekos fragmentų dydžio pasiskirstymas, c) nuskaitomo fragmento principinė struktūra.

12.5 Perskaitymo per nenatūralią jungtį efektyvumo įvertinimas

Įrodžius OTDDN pritaikomumą NKS bibliotekų ruošime tiek DNR, tiek RNR sekvenavimui, buvo svarbu išsiaiškinti perskaitymo per nenatūralią jungtį efektyvumą. Aptarti žinomi ribozė-ribozė oligonukleotidų sujungimo metodai pasižymėjo labai mažu perskaitymo efektyvumu, kuris buvo ≤4%. [16] Todėl buvo labai svarbu išsiaiškinti, ar pasiūlytas alternatyvus nukleobazė-fosfatas sujungimas yra pranašesnis komplementarios grandinės sintezės efektyvumui.

Tikslui pasiekti buvo pasirinkta modelinė sistema ir susintetinti skirtingus jungtukus turintys OTDDN įterpimo produktai (Paveikslas 20a). Tiksliam perskaitymo įvertinimui produktai turėjo būti tinkamai išgryninti, nepaliekant neįterpto OTDDN. Tam buvo optimizuoti HPLC metodai ir tiksliniai įterpimo produktai išgryninti. Tolimesnėje pradmens pratęsimo reakcijoje buvo naudojamas Cy5 žymėtas pradmuo, o reakcijos atliktos su keturiomis

polimerazėmis. Visos testuotos polimerazės gebėjo perskaityti per nenatūralią jungti, tačiau Phusion exo- pasižymėjo didžiausiu efektyvumu. Iš gautų rezultatų matyti, kad visi jungtukai, ilgesni už C2, kuris buvo naudojamas prieš tai aprašytuose NKS bibliotekų paruošimo metoduose, buvo efektyvesni. C4 ir PEG2 pasižymėjo didesniu nei 70% perskaitymo per nenatūralią jungtį efektyvumu.



Paveikslas 20. Perskaitymo per nenatūralią jungtį efektyvumo nustatymas. a) perskaitymo per nenatūralią jungtį įvertinimui naudotos matricos struktūra (ON-dd^{ON6}C), b) PEX eksperimento schema, c) PEX reakcijos mišinių elektroforegramos. K- – neigiama kontrolė, nepratęstas pradmuo, K+ – teigiama kontrolė, oligonukleotidas tokio pačio ilgio, kaip ir laukiamas PEX produktas, C2, C4, C6, P2 ir P4 – identifikuoja jungtukus, kur P2 – PEG2, o P4 – PEG4.

Gauti rezultatai parodo, kad nukleobazė-fosfatas oligonukleotidų sujungimo būdas perskaitymo efektyvumu yra pranašesnis už ribozė-ribozė

sujungimą. Perskaitymo per nenatūralią jungtį efektyvumo padidinimas yra ypač svarbus dirbant su mažais tiriamų nukleorūgščių kiekiais arba atliekant transkriptomų analizes, kur reikalingas maksimalus tyrimo jautrumas, siekiant identifikuoti silpnai išreikštus genus.

IŠVADOS

1. Supaprastinta, trijų stadijų, strategija buvo pritaikyta ir optimizuota azido grupę turinčių 2’-deoksinukleotidų, 2’,3’-dideoksinukleotidų sintezei. Optimizuojant monofosforilinimo reakcijų sąlygas, buvo pastebėti reaktyvumo panašumai tarp nukleozidų (d)d^{I}C ir (d)d^{I}A, taip pat dd^{I}U ir dd^{I}G. Gera reakcijos konversija į (d)d^{I}CMP ir (d)d^{I}AMP buvo pasiekta naudojant $P_2O_3Cl_4$. dd^{I}UMP ir dd^{I}GMP sintezei $POCl_3$ buvo labiau tinkamas fosforilinimo agentas, o bazė pagerino pradinio junginio konversiją į nukleozido monofosfatą. Optimizuojant vandeninėje terpėje atliekamos Sonogashira reakcijos sąlygas nustatyta, kad geros konversijos į dd^{PA}NTP pasiekiamos reakcijas atliekant 40 ^{o}C temperatūroje (išimtis dd^{PA}ATP).
2. Palygintos trijų ir penkių stadijų sintezės strategijos nukleobazėje modifikuotiems nukleotidams gauti. Atliekant 5-(3-aminoprop-1-inil)-2’-deoksicitidino 5’-trifosfato sintezę parodyta, kad pirmoji metodika yra efektyvesnė, o tikslinis junginys gautas beveik du kartus didesne sumine išeiga. Optimizuota alternatyvi sintezės strategija buvo sėkmingai pritaikyta didesnės skalės (10 – 20 kartų) azido-2’,3’-dideoksinukleotidų sintezei.
3. Šiame darbe sukurti azido grupę turintys 2’-deoksi- ir 2’,3’-dideoksinukleotidai buvo panaudoti DNR polimerazių KOD XL, Taq, Phusion exo-, Klenow exo-, Thermo Sequenase, TdT ir SuperScript IV substratais polimerazės grandininės ir/arba pradmens pratęsimo reakcijose.
4. Ištirta CuAAC ir SPAAC reakcijų greičio priklausomybė nuo įvairių azido grupę turinčių 2’-deoksicitidin-5’-trifosfatų įterpimo produktų (PEX-d^{N3}-C-ON) ir alkinų su oligonukleotido bei TAMRA žymėmis. Nustatyta, kad CuAAC ir SPAAC reakcijų greitis didėja ilgėjant jungtukui tarp azido grupės ir nukleobazės (išskyrus d$^{N3\text{-}PEG4}$CTP). Analogiškos reaktyvumo tendencijos buvo stebimos nepriklausomai nuo alkino struktūros.
5. Pasitelkiant CuAAC reakciją buvo susintetinta plati įvairovė oligonukleotidais modifikuotų 2’-deoksi- ir 2’,3’-dideoksinukleotidų. Šiame darbe sukurti OTDN ir OTDDN buvo substratai DNR polimerazėms Taq, Phusion exo-, Klenow exo-, Thermo Sequenase, TdT ir SuperScript IV. Oligonukleotidais-modifikuoti 2’-deoksinukleozidų-5’-trifosfatų įterpimo produktai buvo tiriami kaip matricos fermentinio perskaitymo per nenatūralią jungtį metu. Principinio eksperimento metu buvo parodyta, kad Phusion exo-

polimerazė gali perskaityti per nenatūralią jungtį. Tačiau dėl galimo pakartotinio OTDN įterpimo, realiuose NKS bibliotekų paruošimo eksperimentuose, gaunama kompleksiška perskaitymo reakcijos mišinio sudėtis, ribojanti OTDN pritaikymą.

6. Buvo parodyta, kad oligonukleotidais-modifikuoti 2',3'-dideoksinukleotidai supaprastina ir sutrumpina NKS bibliotekų paruošimą principiniame M13mp18 viruso genomo sekos nustatymo eksperimente. Pasirinktas nukleobazė-ribozė oligonukleotidų sujungimas įgalina Phusion exo- perskaityti per nenatūralią jungtį neprarandant tikslumo. Buvo pademonstruota, kad pasiūlytas NKS bibliotekų paruošimo metodas panaudojant OTDDN yra tinkamas ruošiant DNR ir kDNR bibliotekas įvairiems mėginiams ir taikymo sritims: mikrobiomo, prostatos vėžio sulietinių transkriptų ir genų ekspresijos, grįstos iRNR 3'-galo profiliavimu, analizėms.
7. Ištirtas fermentinio perskaitymo per nenatūralias jungtis efektyvumas. Nustatyta, kad visi pasiūlyti jungtukai tarp nukleobazės ir oligonukleotido buvo biologiškai suderinami. Tirtos polimerazės SuperScript IV, Phusion exo-, Thermo Sequenase ir Klenow exo- gebėjo susintetinti komplementarią DNR grandinę per nenatūralią OTDDN jungtį, geriausio jungtuko vieno ciklo perskaitymas siekė 75%.

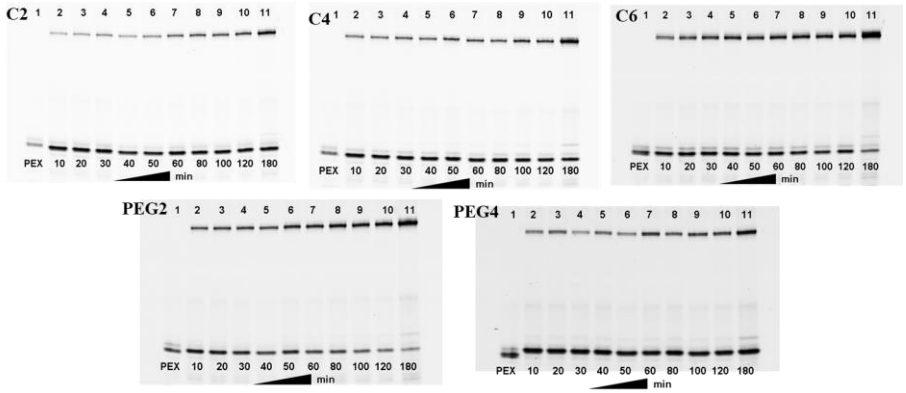


Figure S1. PAGE analysis of SPAAC reaction rate dependency on substrate structure determination.

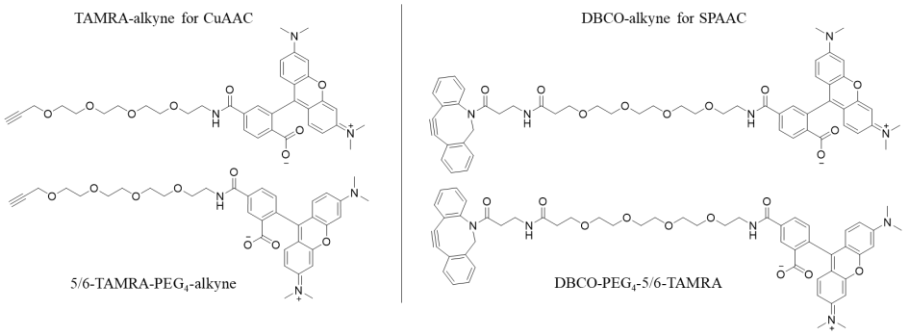


Figure S2. Structure of fluorescently labeled TAMRA-alkynes for CuAAC and SPAAC reactions.

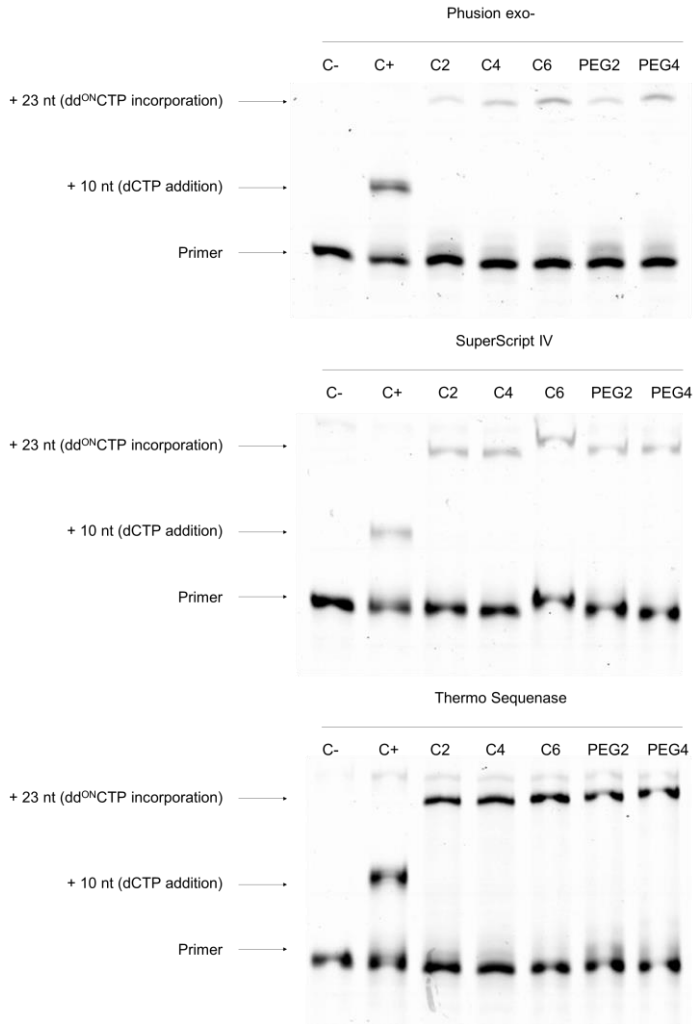


Figure S3. Incorporation of dd^{ON}CTPs bearing different linkers by three DNA polymerases. C- – negative control, C+ – positive control (PEX in the presence of dCTPs), C2 – PEX in the presence of dd$^{C2\text{-}ON}$CTP, C4 – PEX in the presence of dd$^{C4\text{-}ON}$CTP, C6 – PEX in the presence of dd$^{C6\text{-}ON}$CTP, PEG2 – PEX in the presence of dd$^{PEG2\text{-}ON}$CTP, PEG4 – PEX in the presence of dd$^{PEG4\text{-}ON}$CTP.

NOTES

NOTES

Vilniaus universiteto leidykla
Saulėtekio al. 9, III rūmai, LT-10222 Vilnius
El. p. info@leidykla.vu.lt, www.leidykla.vu.lt
bookshop.vu.lt, journals.vu.lt
Tiražas 20 egz.