thermoscientific

Advanced preparation of fragment libraries enabled by oligonucleotide-modified 2',3'-dideoxynucleotides

Justina Medžiūnė[a,b], Žana Kapustina[a], Simona Žeimytė[a], Jevgenija Jakubovska[a], Rūta Sindikevičienė[a], Inga Čikotienė[a], Arvydas Lubys[a]

[a]Thermo Fisher Scientific Baltics, V. A. Graičiūno 8, Vilnius, Lithuania, LT-02241;

[b]Institute of Chemistry, Faculty of Chemistry and Geosciences, Vilnius University, Naugarduko 24, Vilnius, Lithuania, LT-03225.

ABSTRACT

The ever-growing demand for inexpensive, rapid, and accurate exploration of genomes calls for refinement of existing sequencing techniques. The development of next-generation sequencing (NGS) was a revolutionary milestone in genome analysis. While modified nucleotides already were inherent tools in sequencing and imaging, further modification of nucleotides enabled the expansion into even more diverse applications. Herein we describe the design and synthesis of oligonucleotide-tethered (2′,)3′-(di)deoxynucleotide (d^{ON}NTP and dd^{ON}NTP) bearing universal priming sites attached to the nucleobase, as well as their enzymatic incorporation and performance in read-through assays. In the context of NGS library preparation, the incorporation of dd^{ON}NTP fulfills two requirements at once: the fragmentation step is integrated into the workflow and the obtained fragments are readily labeled by platform-specific adapters. DNA polymerases can incorporate dd^{ON}NTP nucleotides, as shown by primer extension assays. More importantly, reading through the unnatural linkage during DNA synthesis was demonstrated, with 25-30% efficiency in single-cycle extension.

INTRODUCTION

The establishment of next-generation sequencing (NGS) techniques revolutionized research abilities in modern biology and biomedical sciences. 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 original library and impoverishment of sequencing results. Consequently, novel library preparation techniques need to be developed to improve the conversion efficiency and simplify the workflow. Some alternative chemical ligation techniques based on the formation of ribose-to-ribose connection were reported (Figure 1a)[1,2]. However, this strategy exhibited several drawbacks: numerous purification steps, only templated click reaction results in efficient synthesis, Cu mediated degradation of oligonucleotides is observed, more than half of the obtained reads are shorter than expected, caused by Cu-mediated DNA degradation, the read-through efficiency very low (Figure 1a and b). [3,4]

We hypothesized that the proper design of the linker connecting nucleotide's nucleobase with oligonucleotide's ribose ring or nucleobase would maintain its substrate properties during incorporation, moreover, might provide the ability for polymerase to read-through it during the synthesis of complementary strand. Such approach would substantially simplify the NGS library preparation workflow (Figure 1c and d).

Results: Synthesis of OTDNs, OTDDNs and library preparation for NGS

Figure 1. Comparison of library preparation strategies

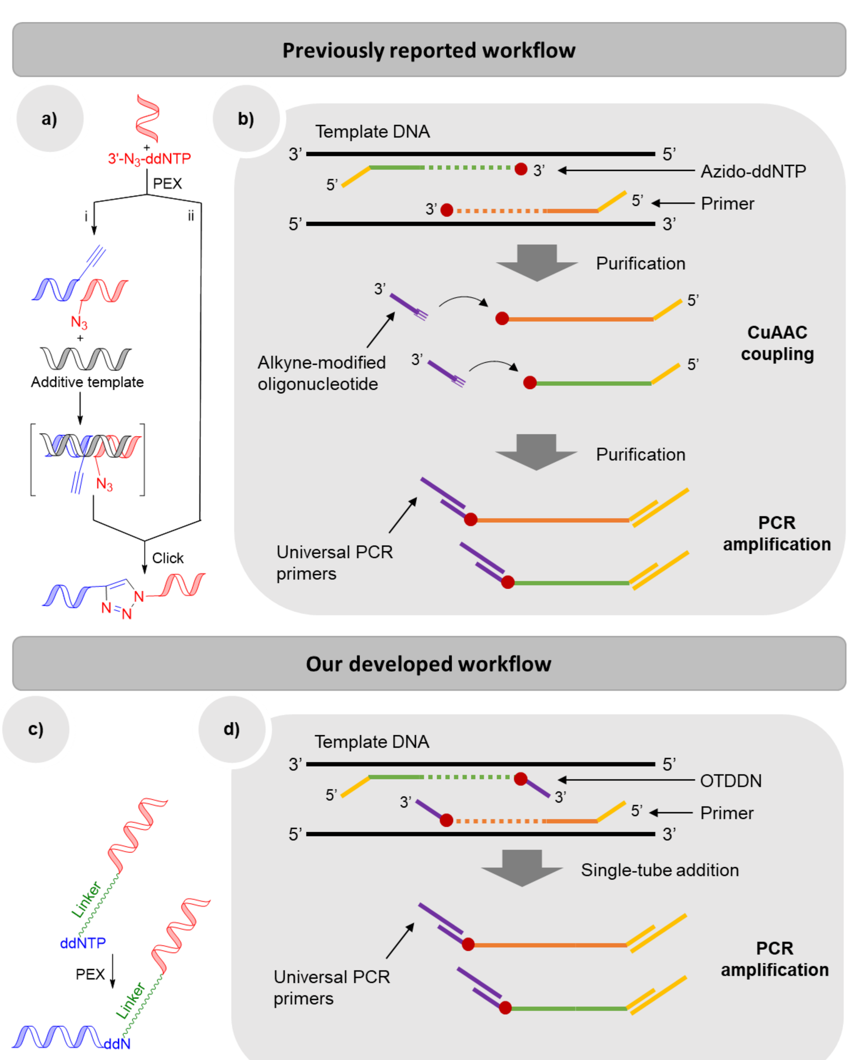
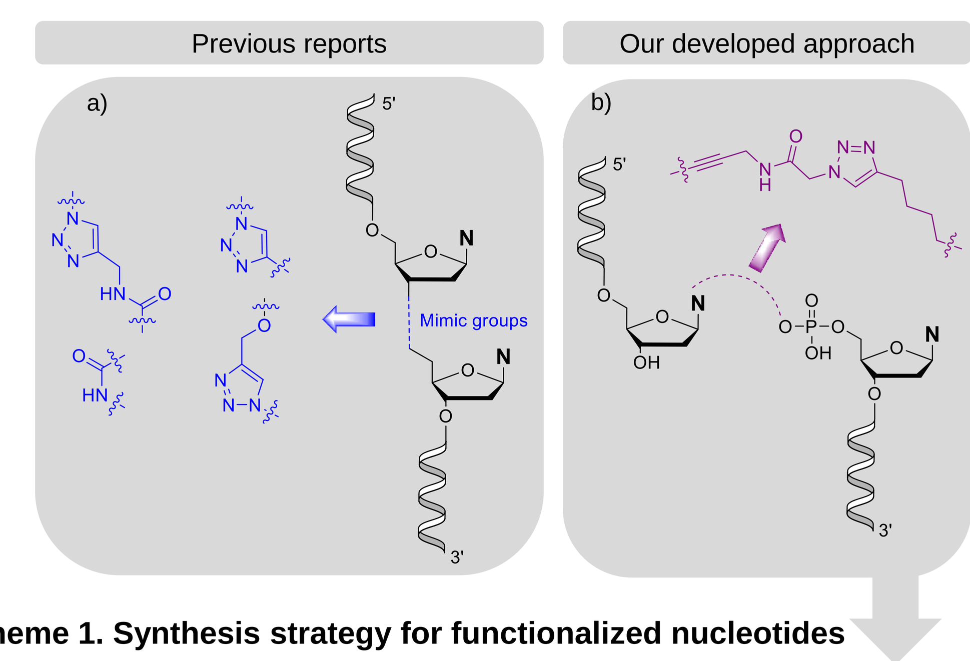
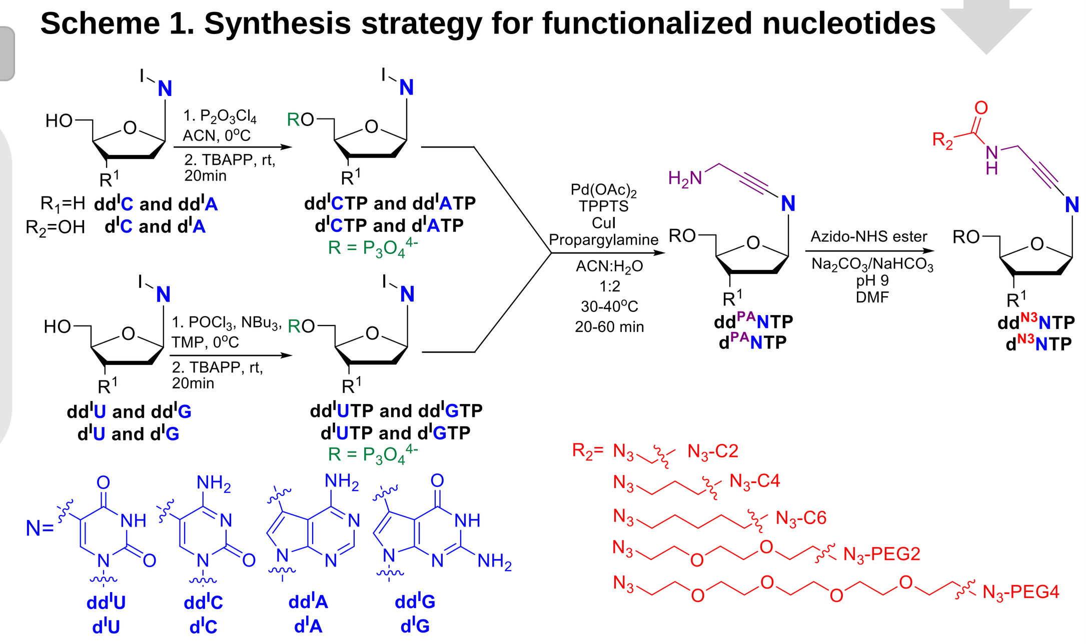


Figure 2. Comparison of conjugation approaches



Scheme 1. Synthesis strategy for functionalized nucleotides



Scheme 2. Synthesis of OTDNs and OTDDNs

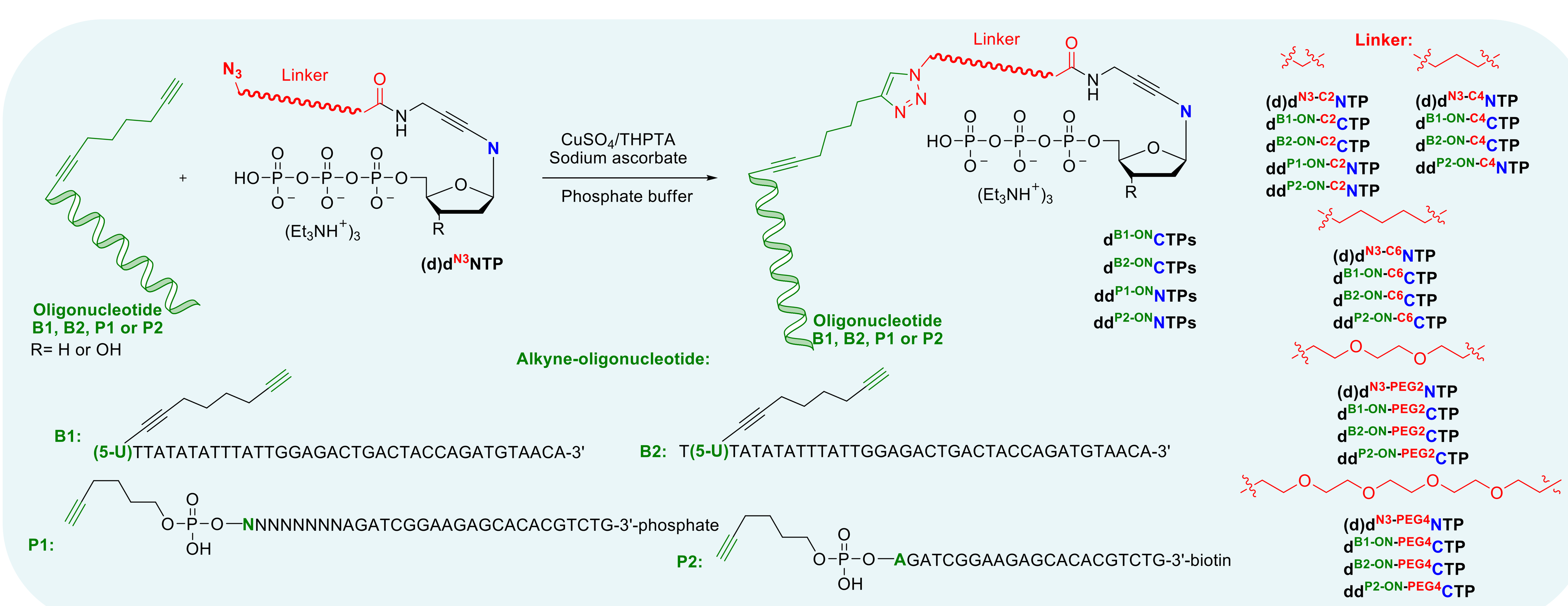
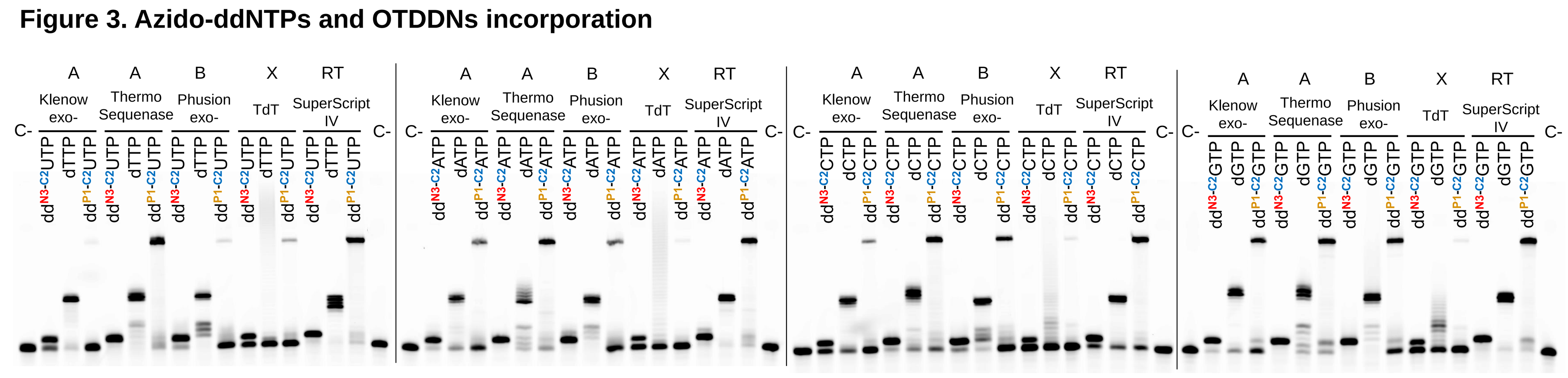


Figure 3. Azido-ddNTPs and OTDDNs incorporation



Scheme 3. Read-though product formation ON-dd^{ON}NTP

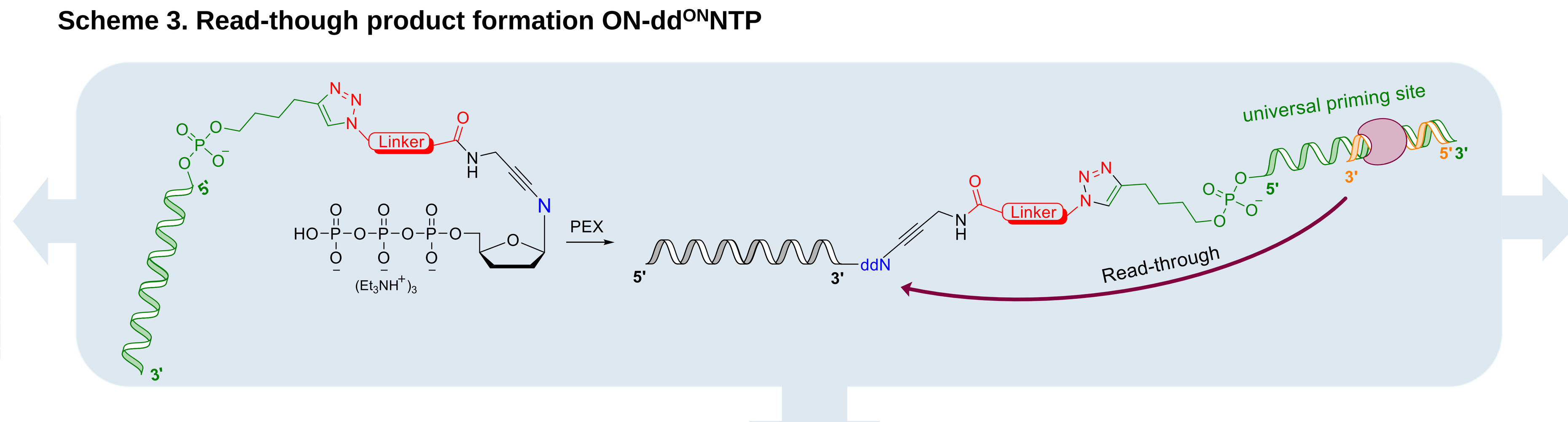
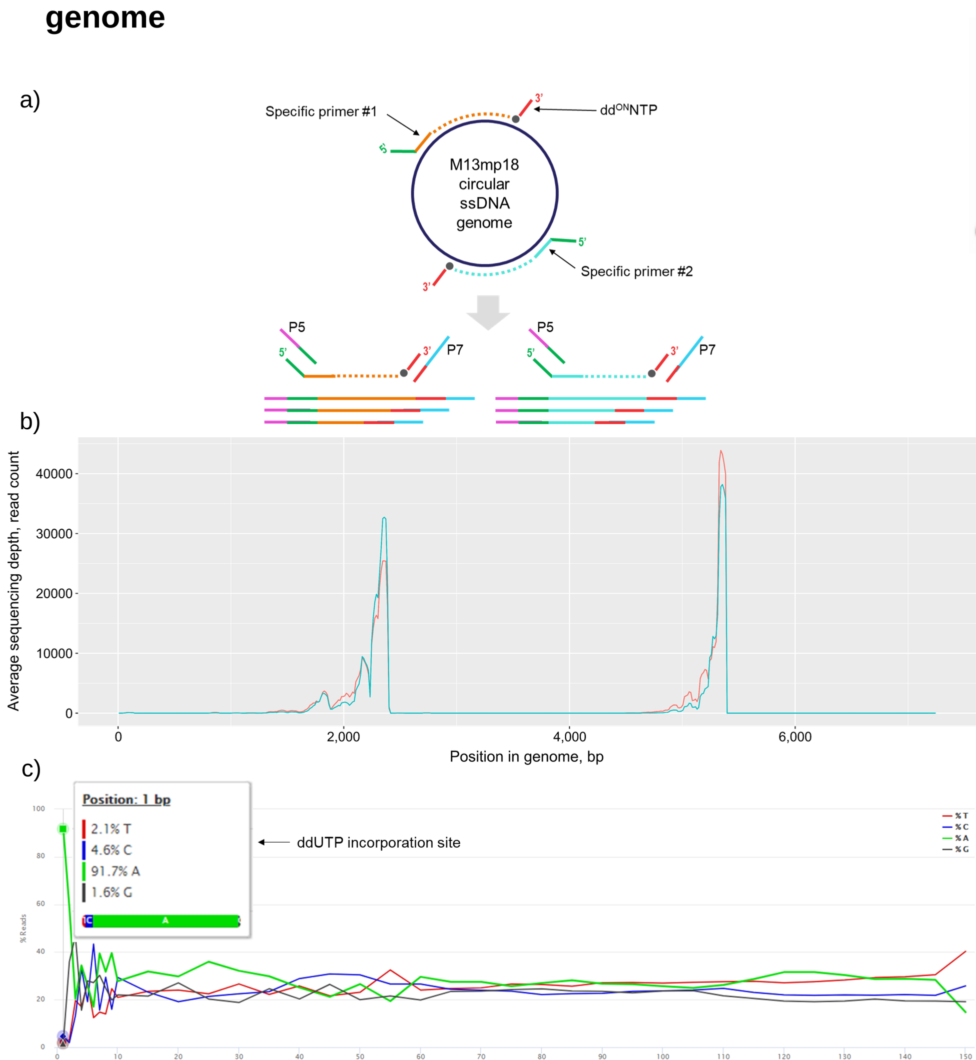
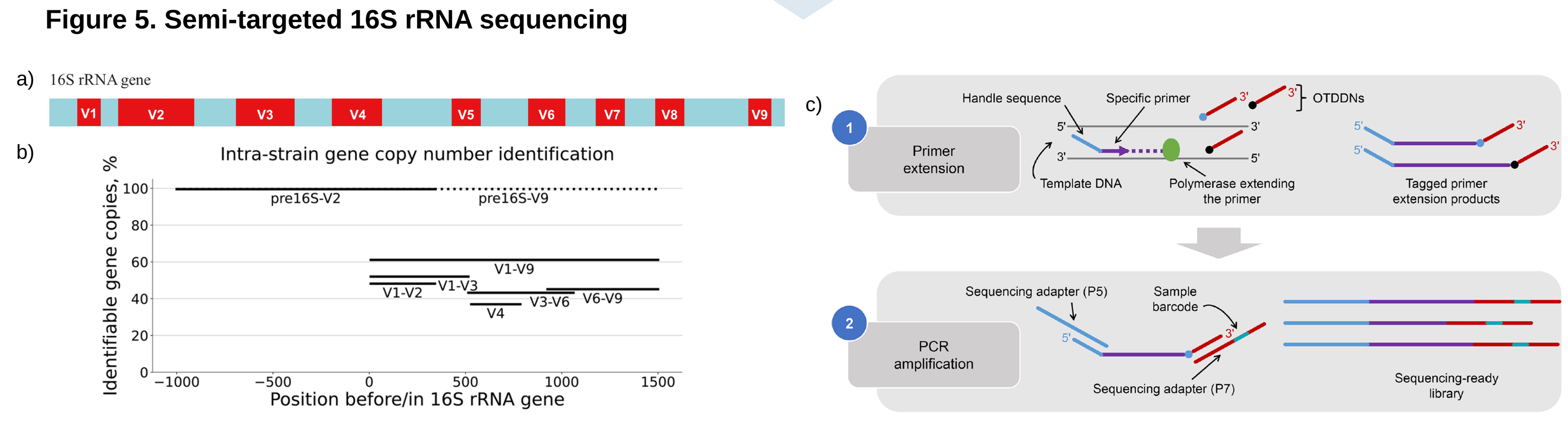


Figure 4. Semi-targeted sequencing of the M13mp18 viral genome



a) Library preparation with dd^{ON}NTPs workflow. 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^{ON}UTP incorporation. 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.

Figure 5. Semi-targeted 16S rRNA sequencing



Determination of microbial communities in soil samples, a) representation of 16S rRNA gene: V1 – V9 indicate 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. [6]

CONCLUSIONS

In conclusion, we have designed and synthesized oligonucleotide-modified 2',3'-dideoxynucleotide terminators, with properties that go beyond commonly labelled nucleotides. Represented efficient dd^{ON}NTPs synthesis and purification strategy enables to obtain these complex molecules with necessary qualities for enzymatic processes. We showed that, even though, their structure possesses bulky oligonucleotide label,[6] the linker design empowers both the enzymatic incorporation and read-through. This enabled us to integrate the fragmentation and adapter addition into a single enzymatic step, thus substantially simplifying sample preparation workflow for high-throughput sequencing. The biocompatibility artificial backbone paves the way for numerous applications, such as whole-genome and whole-transcriptome fragment library preparation for NGS analysis, semi-targeted library preparation with great potential to investigate gene fusion events or characterize microbial communities[6], nucleic acid labelling for subsequent detection employing universal priming site, and others that benefit from the addition of a known artificial sequence to DNA or cDNA.

REFERENCES

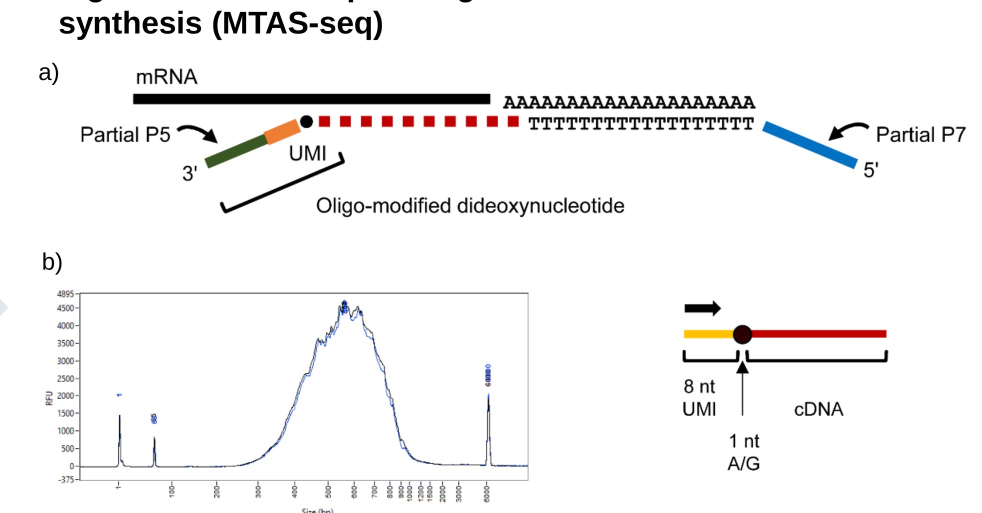
1. El-Sagheer, A. H., Brown, T. Click Nucleic Acid Ligation: Applications in Biology and Nanotechnology. *Acc. Chem. Res.* **45** (8), 1258-1267 (2012).
2. Kumar, R., El-Sagheer, A., Tumpane, J., Lincoln, P., Wilhelmsson, L. M., Brown, T. Template-Directed Oligonucleotide Strand Ligation, Covalent Intramolecular DNA Circularization and Catenation Using Click Chemistry. *J. Am. Chem. Soc.* **129** (21), 6859-6864 (2007).
3. Routh, A., Head, S. R., Ordoukhanian, P., Johnson, J. E. ClickSeq: Fragmentation-Free Next-Generation Sequencing via Click Ligation of Adaptors to Stochastically Terminated 3'-Azido cDNAs. *J. Mol. Biol.* **427** (16), 2610-2616 (2015).
4. Miura, F., Fujino, T., Kogashi, K., Shibata, Y., Miura, M., Isobe, H., Ito, T. Triazole linking for preparation of a next-generation sequencing library from single-stranded DNA. *Nucleic Acids Res.* **46** (16), e95 (2018).
5. J. Medžiūnė, Ž. Kapustina, S. Žeimytė, J. Jakubovska, R. Sindikevičienė, I. Čikotienė, A. Lubys, Advanced preparation of fragment libraries enabled by oligonucleotide-modified 2′,3′-dideoxynucleotides *Commun Chem*, **5**, 34 (2022).
6. Ž. Kapustina, J. Medžiūnė, G. Alzbutas, K. Matijošaitis, G. Mackevičius, S. Žeimytė, L. Karpus, A. Lubys, High-resolution microbiome analysis enabled by linking of 16S rRNA gene sequences with adjacent genomic contexts *Microb. Genom.*, **7**(9), 000624 (2021).
7. Kapustina, Ž., Medžiūnė, J., Dubovskaja, V., Matjošaitis, K., Žeimytė, S., Lubys, A., Sensitive and accurate analysis of gene expression signatures enabled by oligonucleotide-labeled cDNA *RNA Biology*, DOI: 10.1080/15476286.2022.2078093 (2022).

ACKNOWLEDGEMENTS

We are grateful Dr. Lukas Taujenis, Dr. Vytautas Tamošiūnas and Martynas Melinskas for cooperation in finding suitable method for LC-MS analysis and data processing.

TRADEMARKS/LICENSING

Figure 6. mRNA sequencing via terminator-assisted synthesis (MTAS-seq)



Reverse transcription starts from an oligo(dT) primer containing a portion of the Illumina P7 adapter sequence. Primer extension terminates upon the incorporation of OTDDN bearing a portion of the Illumina P5 adapter sequence. This yields cDNA fragments which can be PCR-amplified using standard Illumina indexing primers.

We developed MTAS-seq (Figure 7a) for rapid and simple transcriptome-wide differential expression profiling and 3′UTR detection. Reverse transcription primer targets polyA tails of eukaryotic mRNAs and is extended by reverse transcriptase. Primer extension is terminated by stochastic incorporation of OTDDNs yielding oligonucleotide-labeled cDNA fragments (Figure 7b) whose average length is determined by the ratio of OTDDNs to respective dNTPs.

ThermoFisher SCIENTIFIC