De novo DNA synthesis using polymerase-nucleotide conjugates

TECHNISCHE UNIVERSITÄT DARMSTADT

Fachbereich Biologie der
Technischen Universität Darmstadt

zur Erlangung des Grades
Doktor rerum naturalium
(Dr. rer. nat)

Dissertation
von Sebastian Palluk

Erstgutachterin: Prof. Dr. Beatrix Süß

Zweitgutachter: Prof. Dr. Johannes Kabisch

Darmstadt 2018

Palluk, Sebastian: *De novo* DNA synthesis using polymerase-nucleotide conjugates
Darmstadt, Technische Universität Darmstadt
Jahr der Veröffentlichung der Dissertation: 2019
Tag der mündlichen Prüfung: 17.12.2018

Summary

The terminal deoxynucleotidyl transferase (TdT) is the key enzyme proposed for enzymatic DNA synthesis, based on its ability to extend single stranded DNA rapidly using all four different deoxynucleoside triphosphates (dNTPs). Proposals to employ TdT for the *de novo* synthesis of defined DNA sequences date back to at least 1962, and typically involve using the polymerase together with 3'-modified reversible terminator dNTPS (RTdNTPs), analogous to Sequencing by Synthesis (SBS) schemes. However, polymerases usually show a low tolerance for 3'-modified RTdNTPs, and the catalytic site of TdT seems particularly difficult to engineer in order to enable fast incorporation kinetics for such modified dNTPs. Until today, no practical enzymatic DNA synthesis method based on this strategy has been published.

Here, we developed a novel approach to achieve single nucleotide extension of a DNA molecule by a polymerase. By tethering a single dNTP to the polymerase in a way that it can be incorporated by the polymerase moiety, we generate so called polymerase-nucleotide conjugates. Once a polymerase-nucleotide conjugate extends a DNA molecule by its tethered dNTP, the polymerase moiety stays covalently attached to the extended DNA via the linkage to the incorporated nucleotide, and blocks other polymerase-nucleotide conjugates from accessing the 3'-end of the DNA molecule. At the same time, all dNTPs available in the system are tethered to a polymerase-moiety, which prevents them from accessing the enzyme that has extended the DNA molecule. The system therefore only allows single nucleotide extension to occur, until the termination is reversed by cleaving the linker between the extended DNA molecule and the polymerase to enable subsequent cycles.

We tested the hypothesis that polymerase-nucleotide conjugates can be used for single nucleotide extension of a DNA molecule, by generating TdT-nucleotide conjugates using a linker based on a short polyethylene glycol structure. We coupled the linker to an aminoallyl-modified dNTP, and then used a cysteine residue of TdT to attach the linker to the protein. We found that nucleotides tethered to TdT could be incorporated quickly, and that TdT-dNTP conjugates with a single tethered dNTP had the ability to extend a DNA molecule almost exclusively by one nucleotide. TdT-dNTP conjugates with the linker attached to four different cysteine positions were tested, and all attachment positions enabled the generation of functional TdT-dNTP conjugates that could quickly perform the single extension of a DNA molecule.

We then generated polymerase-nucleotide conjugates using a linker that could be cleaved close to the dNTP, to reduce the size of the modification ("scar") that remains on the nucleobase after cleavage of the linker. These conjugates employed fast-cleaving photocleavable moieties for release of the extended DNA molecule. We demonstrated that conjugates based on this attachment strategy can extend a primer by all four dif-

ferent dNTPs quickly, and that multiple cycles of nucleotide extension and deprotection can be performed, fulfilling the main requirement of an enzymatic DNA synthesis system. Finally, based on the TdT-dNTP conjugates generated, we synthesized a defined 10-mer DNA sequence, achieving an average stepwise yield of 97.7 %.

Using polymerase-nucleotide conjugates, single nucleotide extensions of a DNA molecule by a polymerase can be achieved in a reversible manner. This novel reversible termination scheme solves certain problems of previous technologies based on RTdNTPs, and might also be implemented with template-dependent polymerases for sequencing applications. However, in this demonstration, it was used together with TdT to synthesize a defined DNA sequence. We believe that the presented scheme offers a promising starting point for the development of a practical enzymatic DNA synthesis technology.

Zusammenfassung

Die Template-unabhängige Polymerase Terminale Deoxynucleotidyl Transferase (TdT) wurde bereits 1962 im Kontext enzymatischer *de novo* Synthese von DNA genannt. TdT katalysiert den Einbau der vier Nucleosidtriphosphate (dNTPs) in einzelsträngige DNA mit sehr hoher Geschwindigkeit und erfüllt damit die grundlegenden Voraussetzungen für den enzymatischen Syntheseprozess. Bisher wurde für die gezielte Synthese von Sequenzen mit TdT vor allem die Nutzung von 3'-modifizierten reversiblen Terminatoren (RTdNTPs) vorgeschlagen, wie sie auch in Sequencing By Synthesis (SBS) Technologien verwendet werden. Allerdings stellen 3'-modifizierte dNTPs oft schlechte Substrate für Polymerasen dar, und im speziellen TdT hat nur eine geringe Toleranz für diese Art von RTdNTPs. Die Kristallstruktur der Polymerase lässt zudem erahnen, dass Proteinengineering für eine höhere Toleranz gegenüber den modifizierten dNTPs eine schwierige Aufgabe darstellt. Bis zum heutigen Tag gibt es keine Publikation eines praktikablen DNA-Syntheseverfahrens basierend auf dieser Strategie.

In dieser Arbeit wird ein neuer Ansatz für die Synthese von DNA mit TdT - und im Allgemeinen für den wiederholten Einbau von einzelnen Nukleotiden mit einer Polymerase (reversible Termination) - vorgestellt. Hierbei wird ein einzelnes dNTP Molekül über einen Linker mit einer Polymerase verbunden, um sogenannte Polymerase-dNTP Konjugate herzustellen. Wenn die Polymerase eines Konjugats das verlinkte dNTP in ein DNA Molekül einbaut, bleibt sie durch den Linker an die DNA gekoppelt, so dass anderen Konjugaten der Zugang zum 3'-Ende des DNA Moleküls verwehrt wird. Durch Spaltung des Linkers kann das DNA Molekül anschließend für weitere Verlängerungsschritte freigegeben werden.

Um zu testen, ob Polymerase-dNTP Konjugate für den Einbau einzelner Nukleotide genutzt werden können, wurden zunächst Konjugate mit einem kurzen Polyethylenglycol-Linker hergestellt. Der Linker wurde an ein Nukleosidtriphosphat mit einer Aminoallyl Basenmodifikation gekoppelt und anschließend über ein Cystein an TdT angehängt. In den Versuchen stellte sich heraus, dass TdT verlinkte dNTPs schnell einbauen kann, und dass TdT-dNTP Konjugate mit nur einem verknüpften dNTP einzelnen Nukleotideinbau ermöglichten. Vier verschiedene Ansatzpunkte des Linkers an der Polymerase wurden getestet; alle ermöglichten einzelnen Nukleotideinbau mit hoher Geschwindigkeit.

Im Folgenden wurden Polymerase-dNTP Konjugate erstellt, deren Linker nahe der Nukleobase gespalten werden kann, so dass nur eine kleine Modifikation am eingebauten Nukleotid bestehen bleibt. Dafür wurden Linker mit photolabilen Gruppen verwendet, die sehr schnell durch Lichteinwirkung gespalten werden können. Konjugate aller vier Nukleosidtriphosphate wurden hergestellt, und hatten alle die Fähigkeit, einen Primer schnell um ein einzelnes Nukleotid zu verlängern. Zudem ermöglichten die Konjugate die Durchführung mehrerer Zyklen von Nukelotideinbau und Entschützung, und erfüllten

damit die grundlegende Voraussetzung für ein System zur DNA Synthese. Die Konjugate wurden schliesslich genutzt, um eine DNA Sequenz von 10 Nukleotiden mit einer durchschnittliche Ausbeute von 97.7 % pro Zyklus zu schreiben.

Polymerase-dNTP Konjugate ermöglichen den wiederholten und gezielten Einbau einzelner Nukleotide in ein DNA Molekül. Dieses Prinzip für reversible Termination könnte daher möglicherweise auch für das Sequenzieren von DNA Molekülen verwendet werden, wenn es mit Template-abhängigen Polymerasen implementiert wird. In dieser Demonstration wurde die Technik genutzt, um enzymatisch eine definierte DNA Sequenz zu schreiben. Die Methode stellt einen vielversprechenden Ansatzpunkt für die Entwicklung einer neuen, praktikablen Technik für die Synthese von DNA Sequenzen dar.

Publications and commentary

Parts of this thesis have already been presented in the following publications:

- Palluk S*, Arlow DH*, de Rond T, Barthel S, Kang JS, Bector R, Baghdassarian HM, Truong AN, Kim PW, Singh AK, Hillson NJ, Keasling JD (**2018**) De novo DNA synthesis using polymerase-nucleotide conjugates. *Nature Biotechnology* 36(7):645650 [74]

  I planned, performed and analyzed experiments, supervised students working on the project, and wrote the manuscript together with DH Arlow. "*" denotes shared authorship.

- Arlow DH, Palluk S (**2017**) Nucleic acid synthesis and sequencing using tethered nucleoside triphosphates. *US Patent Application WO2017223517A1* [4]

The following *News and Views* and *Research Highlight* articles on the work provide a historical context of the research and comment on the promises and implications of an enzymatic approach for *de novo* DNA synthesis:

- Clore, A (**2018**) A new route to synthetic DNA. *Nature Biotechnology* 36(7):593 595 [19]

- Tang, L (**2018**) An enzymatic oligonucleotide synthesizer. *Nature Methods* 15(8):568568 [92]

Contents

1 Introduction

1.1 *De novo* synthesis of DNA sequences

Today, synthetic DNA is an essential part of biological research and bioengineering. DNA oligonucleotides are routinely manufactured by companies such as Integrated DNA Technologies (IDT) and used on a daily basis in laboratories, be it in PCR reactions to amplify specific sequences [71], to introduce mutations [34, 75, 97], or to read DNA using Sanger sequencing [83]. Synthetic DNA is used to introduce indices and perform target capturing for Next Generation Sequencing methods [2, 7, 67, 68, 88], or to provide fixing templates for a cell, once synthetic RNA has guided CRISPR-CAS9 to the right location [5,82]. Longer synthetic sequences can code for synthetic genes used in metabolic engineering [39, 44, 45] or even to synthesize and redesign entire chromosomes [26, 80]. Also non-biological applications of manufactured DNA sequences exist, such as DNA nanotechnology [76, 85] or DNA-based data archiving [17, 29, 91, 102]. In addition, a variety of modifications for DNA oligos such as certain functional groups, varying sugar substitutions, fluorophores, and cleavable moieties are available, enabling the use of oligos as a scaffold to build functional units, place molecules in a defined distance, and design a large variety of experiments.

The list of use cases for synthetic DNA presented above may seem extensive, but in fact only covers some of the applications that exist. Synthetic DNA is omnipresent in biology, and the last decades of biological research as well as our current understanding of the field are highly influenced by our ability to manufacture custom DNA sequences. Interestingly, only few people had foreseen the role that custom DNA manufacturing should play, and the researchers who established oligonucleotide synthesis were initially met with a large amount of skepticism, as described by Marvin Caruthers, one of the most important figures in the development of chemical DNA synthesis [11, 12].

When Har Gobind Khorana demonstrated the "total synthesis of the structural gene for an alanine transfer ribonucleic acid from yeast" [49], Nature New Biology commented on the editorial page as follows: "... like NASA with its Apollo Programme, Khorana's group has shown it can be done, and both feats may well never be repeated" [11]. Today, about 45 years later, Twist Biosciences alone has the capacity to produce 45,000 gene-length constructs per month[1]. Even that Khorana was awarded the Nobel Prize for the "interpretation of the genetic code and its function in protein synthesis" in 1968 did not change the perception of the field. Marvin Caruthers reports that a colleague approached him afterwards at a conference and said: "Marv, why do you want to learn how to

[1] Twist Bioscience, S-1 Registration Statement for IPO, `https://www.sec.gov/Archives/edgar/data/1581280/000119312518291186/d460243ds1.htm` - accessed Oct 28 2018

synthesize DNA? Khorana used it to solve the genetic code and now he has made a gene. But what else can you do with synthetic DNA? Certainly you can find something more useful to do." [11] Caruthers further reports that "There were no biologists, biochemists or molecular biologists anxiously waiting for us to develop our chemistry. They couldn't care less." [11]

Fortunately, Caruthers persisted and developed the chemical DNA synthesis method based on phosphoramidites that is in place until today. The story of the beginning of the field seems particularly interesting, because research on novel, enzymatic methods for DNA synthesis has experienced somewhat comparable skepticisms for a long time. It seems to be a common perception that phosphoramidite synthesis already reaches the limit of what is possible with any man-made synthesis process, and I was further frequently told that "DNA synthesis today is good enough". Because chemical DNA synthesis works excellent for many applications and is well established by now, it sets a very high bar for new technologies, which makes research on them somewhat difficult to justify. Another aspect for the technology development is the relatively small market for custom DNA synthesis. The global market for synthetic DNA and RNA was estimated to be 1.3 billion in 2016[2], and the market for gene-length constructs, the presumably main beneficiary of improved oligosynthesis techniques, makes up only for a small fraction of that market.

However, over the last years, the situation around novel DNA synthesis technologies has changed. Various companies are now developing new, enzyme based DNA synthesis strategies, such as DNA Script (Paris, FR), Nuclera Nucleics and Evonetix (both Cambridge, UK), Molecular Assemblies (San Diego, USA), and Ansa Biotechnologies (Berkeley, USA), which I am involved in. Large players such as Merck (Darmstadt, GER) have programs working on enyzmatic solutions for DNA synthesis, and many established DNA synthesis providers and leading people in the field have expressed interest in enzymatic approaches (e.g., [19]). Only recently, at the SynBioBeta Global Synthetic Biology Summit (October 2018), DNA Script announced the successful synthesis of a 150-mer with a stepwise yield of 99.2 %, being the first enzymatic DNA synthesis company that reports synthesis characteristics comparable to phosphoramidite chemistry. I believe that more efforts across academia and industry will be revealed over the next years, and am looking forward to the developments in the field. Interestingly, besides the approach presented in this thesis based on polymerase-tethered nucleoside triphosphates, all other efforts I am aware of follow the same classic approach based on RTdNTPs that has been discussed for decades, as described in Section 1.3.3.

This chapter gives a short introduction of chemical oligonucleotide synthesis and then describes conventional approaches towards enzymatic DNA synthesis. Finally, the novel enzymatic DNA synthesis concept developed in the framework of this PhD thesis is discussed.

[2]Oligonucleotide Synthesis Market - Forecast to 2021, **Markets and Markets Research, 2017**

1.2 Chemical DNA synthesis

1.2.1 History and general approach

The first coupling of two thymidine residues to generate a dinucleotide was reported in 1955 by Michelson and Todd [69], which started the field of chemical *de novo* DNA synthesis, usually referred to as oligonucleotide synthesis. The basic concept of the synthesis procedure has not changed since then. The synthesis contains two essential steps: 1) A mononucleotide with an activated phosphate reacts with the hydroxy group of another nucleotide, typically the end of a growing DNA molecule attached to a solid support. The activated phosphate can either be at the 3' or 5'-position of the mononucleotide, and then reacts with the respective opposite position. The mononucleotide further contains a protection group that blocks the hydroxy group at which the next incorporation will occur. 2) The protection group of the added nucleotide is removed in a "deprotection step", to generate a free hydroxy group for the next incorporation cycle.

In addition to the protection group cleaved in every cycle, other protection groups that remain attached throughout the whole synthesis prevent side reactions that could occur during the procedure. For example, exocyclic amines of the nucleobases can function as electron donor for the phosphoramidite, which could lead to branching of the growing DNA molecule. Only thymidine and uracil lack an exocyclic amine and don't need respective protection groups on the base. Also, depending on the synthesis scheme, the phosphodiester bond can be protected to increase stability during the procedure. The synthesis process for RNA is very similar, but also requires modification of the hydroxy group in the 2'-position. The general approach to chemical DNA synthesis and its challenges are well summarized by H. Khorana [48].

From the first dinucleotide synthesis in 1955, it took about 17 years until the first report of a chemically synthesized gene, coding for an alanine transfer ribonucleic acid [49]. Khorana et. al synthesized the gene in short sequences of 20 nucleotides that were then enyzmatically joined using T4 ligase. The demonstration was presented in 13 manuscripts, all published in one issue of the *Journal of Molecular Biology*. It took another 9 years until Beaucage and Caruthers presented "deoxynucleoside phosphoramidites as a new class of key intermediates for deoxypolynucleotide synthesis" in 1981 [6], introducing the remarkable technology that has been in place until today, almost 40 years later. The detailed development of chemical DNA synthesis including different activation groups, protection strategies, and solid-support systems is described elsewhere [11, 35, 79].

1.2.2 Phosphoramidite synthesis

Oligos are nowadays almost exclusively produced using the phosphoramidite method. For the synthesis, the growing oligonucleotide is attached to a solid support, often controlled pore glass (CPG) which provides a large surface area [51, 78]. Reagents are then added and removed in a cyclic manner to grow the oligonucleotide by a single monomer per cycle that consists of: 1) Deprotection of the 5' position of the previously added

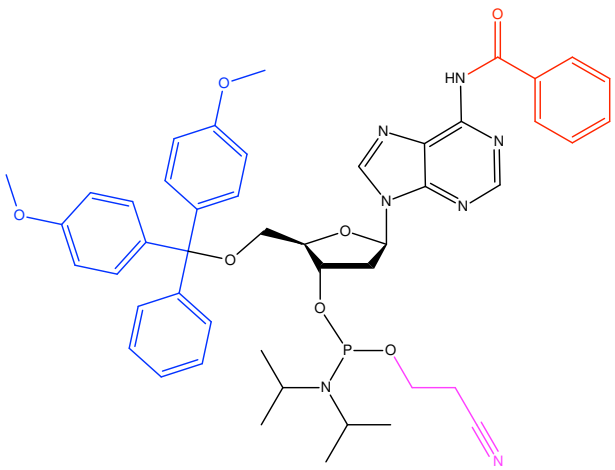


Figure 1.1: Structure of a nucleoside phoshphoramidite monomer (2'-deoxyadenoside phosphoramidite). The DMT group (4,4'-dimethoxytrityl, blue) protects the 5' position, the exocyclic amine is protected by a benzoyl protection group (red). The phosphite group is further protected with a 2-cyanoethyl group (pink). The DMT group is removed in every cycle, the other two protection groups only at the end of the synthesis.

nucleotide. A typical protection group used is DMT (dimethoxytrityl) which can be removed by an acid such as trichloroacetic acid (TCA). During this deprotection step, depurination can occur, which can reduce the overall yield of the synthesized DNA. 2) Coupling of the next phosphoramidite (see Figure 1.1). For the coupling to proceed, the diisopropylamino group of the phosphoramidite gets activated by an acidic catalyst such as ETT. The phosphoramidite monomer is added at a very high concentration, often $> 0.1\,\mathrm{M}$, to drive the reaction close to completion. 3) The new phosphorous connection (phosphite triester) is stabilized by oxidation. Another optional step can be performed and involves the capping of 5'-OH moieties that have not been extended in step 2, e.g. using acetic anhydride. Capping prevents further elongation and therefore prevents oligos that have missed an extension from further growth. Much shorter oligos can be easier separated from the full-length product than those with only a single deletion.

Characteristics of the synthesis procedure

While the exact details of phosphoramidite synthesis are not subject of this thesis, it is very important to understand the performance and characteristics of the process, since they are the benchmark for novel techniques. The yield for a single base addition, which is the combined yield of the steps described above, today consistently exceeds 98 %. Typical yields are above 99 %, and for the synthesis of oligos exceeding 150 nucleotides, yields of 99.5 % are highly desired and often achieved [60]. The cycle time plays an important role for the yield: A cycle can be performed in as little as 2 min, however,

such short steps decrease the synthesis yield strongly. For high-quality oligos, cycle times of 5 min and more are common.

Disadvantages of the phosphoramidite method include the use of toxic reagents and the generation of a significant amount of organic solvent waste. In addition, the synthesis can be finicky and especially the generation of long oligonucleotides requires a certain expertise. Therefore, while phosphoramidite synthesizers were formerly more abundant in laboratories, almost all DNA is nowadays made in facilities with the specific knowhow and equipment.

Generation of longer constructs

Phosphoramidite synthesis enables manufacturing of oligonucleotides up to 200-300 nt in practice [12]. (A yield of 99.5 % over 300 steps leads to 22 % oligos that have the correct sequence, assuming no side-reactions further hamper the yield.) Longer DNA sequences cannot be synthesized directly, but must instead be assembled from multiple oligonucleotides. Various methods for oligonucleotide assembly exist (see, [94, 100, 101]), the most commonly used one is polymerase chain assembly (PCA) [89]. In PCA, oligos coding for the desired construct that partially overlap with each other are repeatedly annealed and extended so that longer constructs and eventually the whole sequence is assembled. One aspect of the generation of constructs that should also be mentioned are error-correction methods that can eliminate erroneous constructs and therefore select for the desired DNA molecules (e.g., [52, 62]).

Assembly of longer constructs has improved over the last years, but is still difficult in many cases and not amenable to all sequences [20], which makes the synthesis of certain constructs impossible. The generation of constructs gets more difficult for 1) oligonucleotides that form secondary structures, preventing them from forming the right interactions with other DNA molecules, 2) sequences with high AT and GC content which impacts the annealing process during PCA, and 3) sequences that contain repeats. Repeats lead to multiple possible options for the assembly, which introduces deletions of certain parts of the DNA. Some of those problems can be mitigated by designing the DNA construct to be assembled in a specific way, in particular if the DNA is supposed to code for a protein, so that multiple codons are available for most positions. If an exact DNA sequence is needed, assembly is more difficult. Further, because some assemblies contain *in vivo* steps, toxicity of genes can lead to restrictions.

Today, gene-length sequences can be ordered for 9 cents/bp from one of the leading gene synthesis providers Twist Bioscience. The delivery time for a construct up to 3.2 kbp is 15-20 business days, and a sequence of this length would cost $270[3]. Importantly, assembly of constructs, independent of the method used, gets much simpler with oligos of higher quality and increased lengths. The most convenient solution to the problem, however, would be the direct synthesis of the desired construct without any need for assembly.

[3] `https://twistbioscience.com/products/genes` - accessed Oct 25 2018

1.3 Enzymatic DNA synthesis

1.3.1 Motivation

Enzymes have always played an important role in *de novo* DNA synthesis. In the early days, the first chemically synthesized dinucleotides were cleaved enzymatically for analysis [69], T4 ligase was used to join the fragments synthesized by Khorana to assemble the tRNA gene [49], and oligo assembly into longer constructs as well as error correction methods in place today are based on enzymes [100]. Proteins also play an essential role in all practiced sequencing methods, be it in Sanger or in NGS techniques. Interestingly, a chemical system was the method of choice for sequencing early on (Maxam-Gilbert sequencing [66]), before it was eclipsed by enzymatic approaches.

The main promise of using enzymes for DNA synthesis is the high specificity of their reactions in combination with the very mild conditions. The core challenge for DNA synthesis is pushing cycle yield closer to 100 %, and the elimination of side-products and damaged molecules is essential for this purpose. Another huge advantage is that enzymes can be engineered in very versatile ways. Many high-throughput screening methods are available, and an almost unlimited space of possible solutions exists. Chemical synthesis, in comparison, is very limited: There is only a number of substituents available, and the most efficient synthesis based on phosphoramidites is likely already in place. One advantage of synthesis in aqueous conditions is further to reduce the amount of hazardous waste, which will enable DNA synthesizers that are easier to handle. In addition, use of other enzymes during synthesis will be possible, opening up the whole toolbox nature has created for handling DNA. Examples include single stranded binding proteins or enzymes that fill in the counter-strand of the synthesized molecule to prevent secondary structures.

1.3.2 Terminal deoxynucleotidyl transferase

Most DNA-dependent DNA polymerases amplify a template sequence to propagate the genetic information and pass it on for cell division. A high evolutionary pressure towards conservation of genetic information exists, and many polymerases have evolved for high fidelity, achieving error rates below 10^{-6}/bp [18] with the help of proof-reading domains that can detect and excise mismatches. Other polymerases are specialized for situations in which perfect amplification is not possible, for example when damaged DNA needs to be copied. They perform replication with lower fidelity, and can be used when "making a mistake is the only way to get ahead" [77]. TdT belongs to neither of those categories, but fulfills an entirely different and very unusual task. TdT writes new, random sequences, thereby actively generating variability in our immune system, that enables adaption to new threats [24]. The polymerase is therefore involved in an active process of evolution that occurs in animals whose adaption to hostile organisms based on random mutations would not be sufficient. In more detail, TdT is involved in genetic recombination during antibody generation, also termed V(D)J recombination (combination of Variable, Diversity, and Joining gene segments) [28, 47].

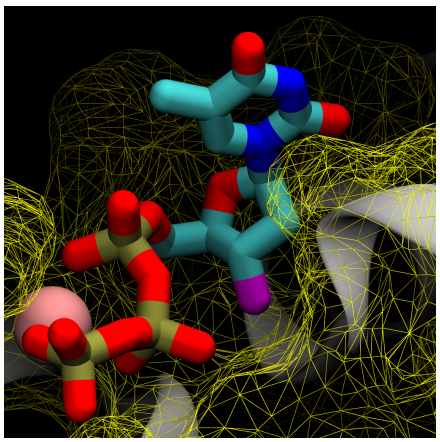


Figure 1.2: A model of the catalytic site of TdT based on PDB structure 4I27 containing dTTP (3'-OH modeled using Maestro) is shown. The 3'-O (purple) directly points towards the backbone of the polymerase (grey). The surface of the polymerase is shown as mesh representation (yellow). The rendering was generated in VMD [37]. This figure was already published in [74].

TdT belongs to the X-family of polymerases and has a pol-beta like folding type [30,57]. It has the same domains as polymerase β, namely the 8 kDa, fingers, palm and thumb domain [24], and contains three asparagine residues that are involved in the two-metal ion binding mechanism of the incoming dNTP across all families of polymerases [42]. TdT has the uncommon ability to work with Co^{2+}, Mg^{2+}, Mn^{2+}, and Zn^{2+}, and shows different dNTP preferences depending on the divalent ions [70]. The incorporation of pyrimidines (dCTP / dTTP) is about 10 x slower than that of purines with Mg^{2+}, but increases about 100 x when Co^{2+} is used as the divalent ion. The incorporation rate of purines only shows a minimal speedup (<1.5 x) when Co^{2+} is used instead of Mg^{2+} [41]. These findings were reported for bovine TdT and only partially match with our observations made with *Mus musculus* TdT. An analysis of the "structures of intermediates along the catalytic cycle of TdT" was published and sheds light on the role of the metal ions, hypothesizing that there might be an additional cation binding position that is of relevance [30]. TdT does not undergo the typical conformational changes most polymerases perform during catalysis [22], does not exhibit processivity, and has a random binding order for the primer and the dNTP [21].

TdT is commonly used in molecular biology laboratories to tail DNA molecules, for example to add a primer-binding site, or to incorporate dNTPs with modified bases (e.g., labeling with fluorescence, radioactive moieties, or attachment of a biotinylated dNTP for purification purposes). TdT can also be employed to detect double strand breaks based on the addition of fluorescent nucleotides, which is used in the so called TUNEL assay to monitor apoptosis [56]. Two detailed reviews on TdT, its role in the immune system, biochemical properties, and usage in different fields of molecular biology are available [24, 70].

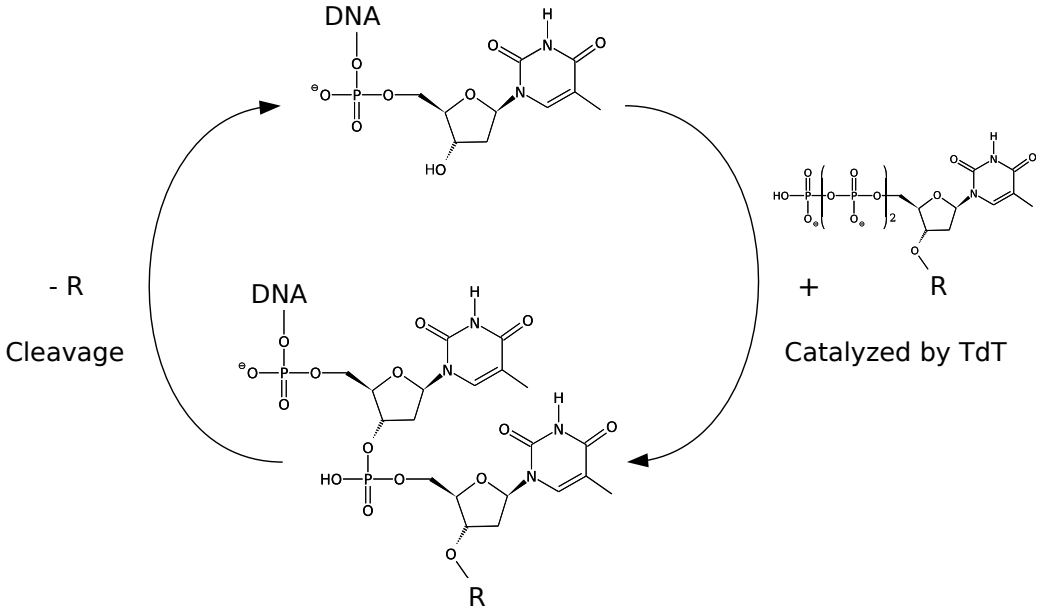


Figure 1.3: The scheme shows how TdT could be used with 3'-modified RTdNTPs for the synthesis of defined DNA sequences. The 3'-end of the growing DNA molecule is first extended by a 3'-blocked RTdNTP. The **R** at the 3'-position could correspond to the various reversible terminator moieties that exist, e.g., to an azidomethyl group. Next, the blocking group is removed, so that a 3'-OH for the next addition cycle is generated. A similar version of this figure was already presented in [73].

1.3.3 Conventional proposals for enzymatic DNA synthesis

TdT fulfills many requirements for an enzymatic DNA synthesis procedure, since it can incorporate all four different dNTPs in a quick manner, without the need for a template. TdT has for a long time been the only known polymerase with template-independent activity [70] and has been part of enzymatic DNA synthesis proposals dating back to 1962 [8,40]. By now, also other polymerases with template-independent activity have been found, in particular Pol Θ [32, 46] that might play a role in the field. Some polymerases with low amounts of template-independent activity are known, such as Pol μ [43], but are likely irrelevant for enzymatic DNA synthesis. Ligases (single-stranded, as well as double-stranded) could in theory play a role, but have limited suitability due to their lower reaction speed.

Almost all proposals for enzymatic DNA synthesis describe usage of TdT together with 3'-modified RTdNTPs, analogous to the procedure performed for sequencing by synthesis (Figure 1.3) [15,95]. Many 3'-modified RTdNTPs with small and fast-cleaving protection groups have been developed and may seem suited for the procedure (e.g., [7,38,50,81,99]). At the same time, TdT was reported to "indiscriminately incorporate ribonucleotides and deoxyribonucleotides" [10], and shows a much smaller preference for dNTPs over NTPs (2-9 x) than typical DNA polymerases (e.g., 2000-6000 x for Pol β) [72]). While one could

conclude from this data that TdT has a high tolerance for sugar modifications in general, this is not the case. The acceptance for 3'-O modifications is in fact small [13,50], which can also be understood in light of the crystal structure (Figure 1.2). The 3'-OH of a dNTP that is bound to the catalytic site directly points towards the backbone of the polymerase, leaving almost no room for any substituents. The earliest reported attempt to use 3'-modified RTdNTPs with TdT dates back to at least 1986 (3'-O acetyl), but the substrate was found not to be incorporated [13]. Over the last years, the group of Montemagno reported the addition of a single nitrobenzole-protected 3'-RTdNTP [64], and later the addition of four nucleotides to a primer with the same strategy. However, a coupling time of 60 min was used and stepwise yields were not reported [65]. Besides 3'-modified RTdNTPs, a different proposal for usage of 3'-O-unblocked RTdNTPs with inhibitory groups attached to the base exists [23]. Using this strategy, it might be very challenging to achieve both, fast incorporation and sufficient termination, since the inhibitory group is present also before the dNTP addition. To the best of my knowledge, except for the 4-mer mentioned above, no scientific report of the *de novo* synthesis of a defined DNA sequence using a template-independent polymerase has been published, besides the work described in this thesis.

1.3.4 Enyzmatic synthesis based on polymerase-nucleotide conjugates

The reversible termination scheme presented here follows a different approach than previous methods for repeated single nucleotide incorporation that are all based on RTdNTPs. Following our strategy, each polymerase molecule is "loaded" with exactly one dNTP that is tethered to the polymerase moiety in a way that enables its incorporation. At the same time, all free dNTPs are removed from the system. When the polymerase moiety extends a primer with its tethered dNTP, it stays attached to the 3'-end of a primer via the tether. In the newly formed complex, the polymerase blocks other polymerase-nucleotide conjugates from accessing the 3'-end of the extended DNA molecule (Figure 1.4). The tethered polymerase itself does not have access to further dNTPs, since these are tethered to other polymerases and are therefore inaccessible. Once the extension step is complete, the excess polymerase-nucleotide conjugates can be removed or inactivated, and the linker can be cleaved to deprotect the extended primer for further extensions. This simple two-step reaction cycle can be iterated to synthesize a defined DNA sequence one nucleotide at a time.

When we perceived the idea for this approach, we were very excited because it promised to circumvent the problem that 3'-blocked RTdNTPs are slowly incorporated, and often require complex enzyme engineering to be accepted by a polymerase at all [16]. At the same time, we thought that the system had a certain advantage for the deprotection step: The linker could in theory accommodate bulky groups, since there is enough space in the cavity of the polymerase. In contrast, to enable incorporation of 3'-RTdNTPs, the modification needs to be small, which limits the scope of potential cleavage groups. 3'-modified RTdNTPs further require the regeneration of an unmodi-

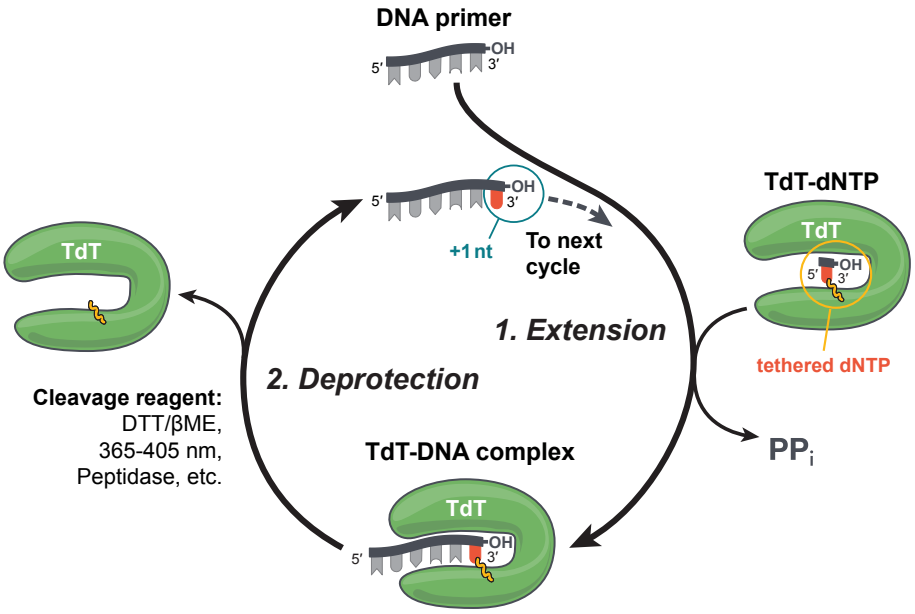


Figure 1.4: The scheme shows how polymerase-nucleotide conjugates can be used for the synthesis of defined DNA sequences. In the first step, the DNA molecule is extended by a polymerase-nucleotide conjugate. Upon the incorporation of its tethered dNTP, the polymerase moiety stays attached and hinders access of other conjugates. In step 2, the linker is cleaved to release the polymerase and generate a free 3'-end of the primer for the next addition step. This figure was already published in [74].

fied 3'-OH upon deprotection, while cleavage of the linker must not necessarily result in a natural base: Part of the linker may stay attached to the incorporated nucleotide as a scar, which promises additional flexibility for the deprotection step. Small scars might be tolerated for follow-up procedures, or could also be eliminated through a "reading" step that produces unmodified DNA.

The scheme presented seemed particularly interesting for template-independent DNA synthesis, since TdT has a high tolerance for base-modifications [53,54]. However, it may be possible to transfer the idea to template-dependent polymerases to perform sequencing based on polymerase-nucleotide conjugates. One promise of employing conjugates for sequencing is the option to label the whole polymerase with reporter molecules, which might enable reliable detection of a single incorporation event and could therefore eliminate the need for cluster-generation.

2 Results

The Result part is separated in five sections. The first section describes the initial investigation to evaluate the concept of polymerase-nucleotides based on the polyethylen glycol (PEG) derived linker PEG_4-SPDP with a cleavable disulfide bond. The second and third section document the development of two polymerase-nucleotide conjugates based on photocleavable linkers, suited for the performance of multiple cycles. In the fourth and fifth section, the use of photocleavable polymerase-nucleotide conjugates for the synthesis of a 10-mer and the 3-mer 5'-CCC-3', as well as the analysis of the reaction products based on next generation sequencing (NGS) are described. Work was performed in collaboration with Daniel H. Arlow, in particular for the experiments described in sections 3-5.

2.1 Conjugates based on PEG_4-SPDP

2.1.1 Linker assembly using PEG_4-SPDP

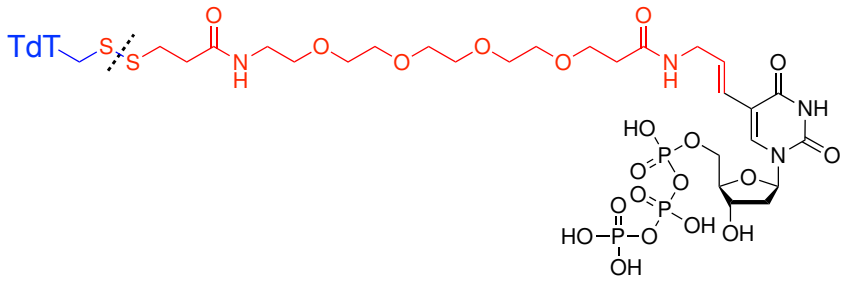


Figure 2.1: Chemical structure of the attachment of aa-dUTP to TdT based on the heterobifunctional crosslinker PEG_4-SPDP. The conjugates based on this linker are also referred to as "TdT-PEG_4-dTTP". The cleavable bond between TdT and the nucleotide is indicated with a black dotted line, the atoms that remain attached to the nucleobase upon cleavage as a "scar" are indicates in red. A similar figure was already published in [74].

To test the functionality of polymerase-nucleotide conjugates for single extension of a DNA molecule, first, a linker system for the attachment of the nucleotide to the polymerase was designed (Figure 2.1). We chose the heterobifunctional crosslinker PEG_4-SPDP that has an orthopyridyl disulfide (OPSS) group reactive towards sulfhydryls and an N-hydroxysuccinimide (NHS) functionality that enables attachment to primary amines (Figure 2.2). Cysteine residues in TdT comprise a sulfhydryl moiety and can

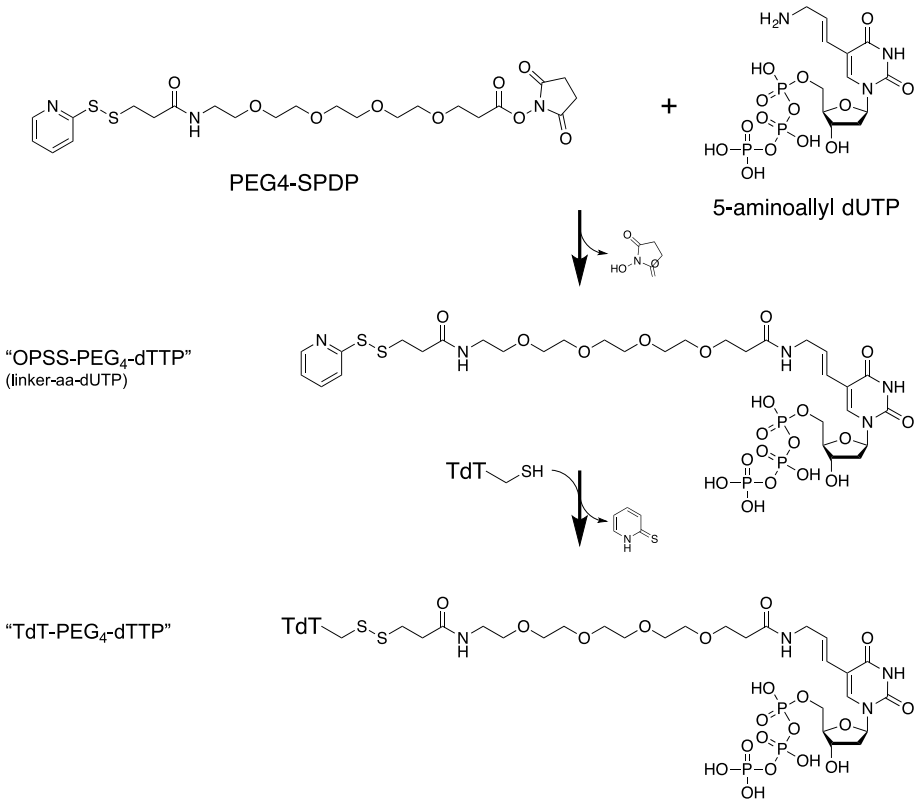


Figure 2.2: The chemical details of the assembly of "TdT-PEG_4-dTTP" are shown. First, 5-aminoallyl dUTP is coupled to PEG_4-SPDP using the NHS-functionality of the linker, generating the linker-dNTP "OPSS-PEG_4-dTTP". In the second step, OPSS-PEG_4-dTTP is attached to TdT by reaction of the OPSS-moiety with a cysteine of the protein, to form "TdT-PEG_4-dTTP". A similar figure was already published in [74].

therefore be used for the attachment to the protein. For dNTPs, modifications of the C5 position of pyrimidines and the C7 position of 7-deazapurines are used to attach fluorophores for next generation sequencing techniques. Modifications in this position do not interfere with base-pairing, and have been well characterized (e.g., [36]). For our system, we chose the dNTP-analog 5-aminoallyl dUTP (aa-dUTP) that contains a primary amine for the attachment to the NHS functionality of the linker.

The coupling of aa-dUTP to TdT was performed in two steps (Figure 2.2). First, cysteine-reactive linker-dTTP molecules were generated by incubating PEG_4-SPDP with a 4-fold excess of aa-dUTP, followed by quenching of unreacted NHS-groups by glycine. In the second step, the crude products were added to TdT to perform the labeling reaction. After proceeding of the labeling-reaction, the conjugates were purified using His-Tag purification to remove untethered nucleotides that could otherwise be incorporated during the following reactions. The success of the coupling reaction was then tested based on the ability of the conjugates to incorporate tethered nucleotides, as described in the next section.

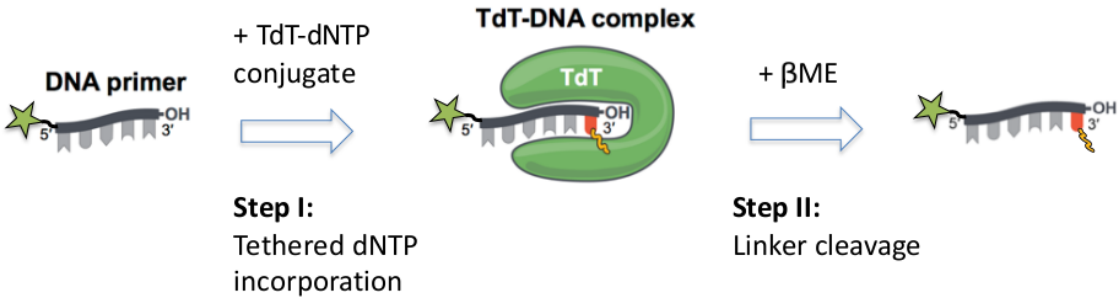


Figure 2.3: Scheme for testing conjugate activity based on SDS-PAGE. A fluorescent primer is incubated with TdT-(PEG_4-)dNTP conjugates to form a covalent TdT-DNA complex that can be resolved on SDS-PAGE. Upon cleavage of the linker, e.g., using βME, the complex is resolved. The released oligo is extended by one or more nucleotides. Depending on the scar, single nucleotide resolution can be achieved during gel electrophoresis.

2.1.2 Tethered incorporation of wildtype TdT

TdT expression for experiments involving the PEG_4-OPSS crosslinker was performed in *E.coli* Rosetta-gami B(DE3)pLysS cells [55] (Novagen). While TdT is usually expressed at lower temperatures to increase solubility [9], we found that incubation at 30 °C produced the most protein using this strain. The short isoform of TdT (TdTs, NCBI Accession number: NP_001036693.1) was expressed from a codon-optimized gene, consisting of residues 132-510 of the wildtype protein, excluding the N-terminal BRCT domain. The same TdT version had already been used in my previous work at TU Darmstadt [73].

During the conjugate reaction, a high molecular-weight complex between the enzyme and the DNA molecule is formed and can then be resolved by cleavage of the linker (Figure 2.3). While SDS-PAGE is usually used to separate denatured proteins, we found that oligos, in particular between 30 and 60 nucleotides, also show an excellent running behavior. Using a fluorescent primer, the migration behavior can be easily traced, and the strong difference in running behavior between the free primer and the TdT-DNA complex enables sufficient separation for analysis in few minutes. When PAGE is run for longer, depending on the size of the scar on the nucleotide added, even single-nucleotide resolution can be achieved.

To investigate tethered incorporation ability, conjugates of wildtype TdT were prepared (TdT*wt*-PEG_4-dTTP) and incubated with the fluorescent primer P1 (5 FAM-dT_{35}) under typical conditions for a TdT extension reaction. Analyzing the reaction products on SDS page, the formation of a high molecular-weight complex was found, indicating that TdT has the ability to incorporate nucleotides tethered to it (Figure 2.4). Upon addition of βME (β-mercaptoethanol), the oligo showed a much faster running behavior, in accordance with cleavage of the tether. The released oligonucleotide migrated in the form of multiple bands, indicating multiple extensions. Using a letter of oligo

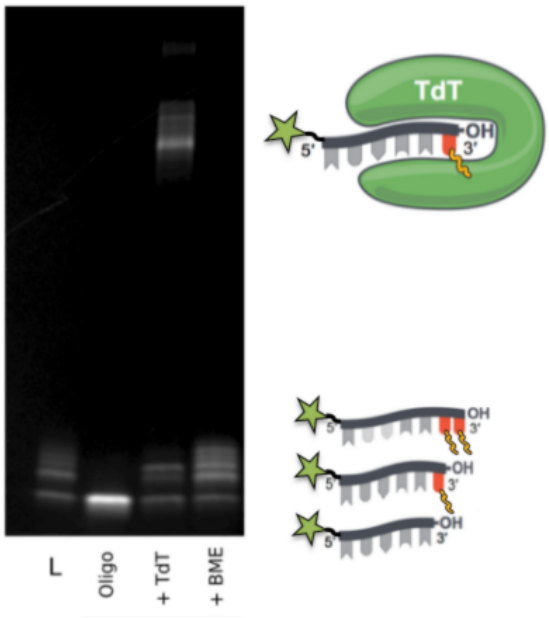
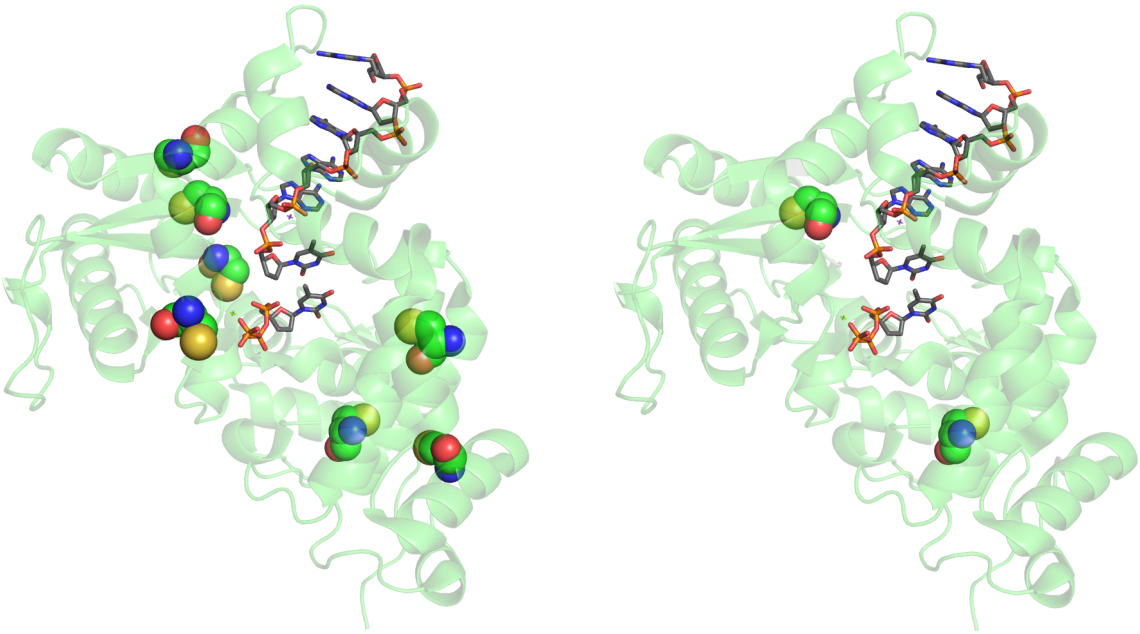


Figure 2.4: Investigation of the ability of wildtype TdT to incorporate tethered dNTPs based on SDS-PAGE. The gel shows the fluorescence of primer P1 (5 FAM-dT_{35}). Lanes correspond to a ladder generated by oligo extensions with free linker-dTTP (L), the oligo before the reaction (Oligo), after the reaction (+TdT), and after cleavage of the linker with βME (+BME). A schematic drawing of the products corresponding to the differend bands is shown on the right hand side.

extension products generated with TdT and free OPSS-PEG_4-dTTP linker-nucleotides as reference, we found that the oligo had been extended by up to 4 nucleotides.

2.1.3 Removal of cysteines

The multiple DNA extensions observed for TdT*wt*-PEG_4-dTTP could potentially be explained by 1) the incorporation of multiple dNTPs tethered to one TdT molecule, or 2) insufficient termination, enabling multiple conjugates to extend the same primer molecule. Wildtype TdT has several cysteines, so it seemed likely that multiple of those could be labeled with a nucleotide-linker molecule. Based on the PDB structure 4I27 [30], we identified cysteines in the positions 188, 216, 302, 378, and 438 as potentially surface accessible and close enough to the catalytic site to enable incorporation of tethered dNTPs (Figure 2.5a). Through multiple rounds of site-directed mutagenesis PCR [34], the cysteines were removed to generate TdT mutant TdTΔ5*cys* with the following mutations: Cys188Ala, Cys216Ser, Cys302Ala, Cys378Ala, and Cys438Ser (Figure 2.5b).



(a) Wildtype TdT. **(b)** TdTΔ5*cys*.

Figure 2.5: Removal of cysteines to generate TdTΔ5*cys*. Through site-directed mutagenesis, five residues of TdT were mutated into alanine or serine, respectively, to generate a TdT variant "silent" for tethered incorporation. The protein is shown in cartoon representation, nucleic acids as sticks, and cysteines as spheres. The rendering was generated in PyMOL [84].

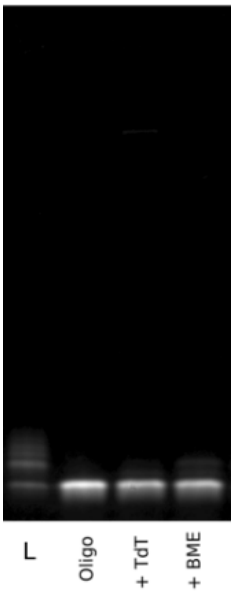


Figure 2.6: Investigation of the ability of TdTΔ5*cys* to incorporate tethered dNTPs based on SDS-PAGE. The gel shows the fluorescence of primer P1 (5 FAM-dT_{35}). Lanes correspond to a ladder generated by oligo extensions with free linker-dTTP (L), the oligo before the reaction (Oligo), after the reaction (+TdT), and after cleavage of the linker with βME (+BME).

2.1.4 Tethered incorporation of TdTΔ5cys

After performing a labeling reaction of TdTΔ5*cys* with linker-nucleotides based on PEG_4-SPDP, activity of the conjugates was tested using the SDS-PAGE incorporation assay. In contrast to wildtype TdT, TdTΔ5*cys* did not show the formation of a covalent complex with the primer, indicating that no nucleotides were tethered to the enzyme in a way that enabled their incorporation. An activity assay of TdTΔ5*cys* using free dNTPs was performed and confirmed that the mutant still had catalytic activity, suggesting that the removed attachment positions for the linker indeed caused the lack of tethered incorporation. A detailed activity comparison of important TdT mutants used in this thesis was performed by the Master student Sebastian Barthel under my supervision and can be found in [74].

2.1.5 Introduction of cysteines

The next step was to develop polymerase-nucleotide conjugates that could only perform a single incorporation. For this purpose, single cysteine residues were inserted into TdTΔ5*cys* using site-directed mutagenesis. Positions for site-directed mutagenesis that were 1) surface accessible and 2) close to the nucleobase of a bound dNTP were identified. Four of these possible attachment points were then selected in order to surround the catalytic site, thereby covering a wide range of the solution space (Figure 2.7). The respective mutants TdT*c*180 (Glu180Cys) and TdT*c*253 (Thr253Cys) with new cysteine positions and TdT*c*188 (Ala188Cys) and TdT*c*302 (Ala302Cys) with re-inserted original cysteines were generated. The distances of the attachment positions for the linker ranged from 12.9 Å to 15.7 Å, measured from the methyl in base position 5 (C5M) to the Cα of the respective amino acid (based on PDB 4I27). The PEG_4-SPDP linker has a length of up to 25.7 Å.

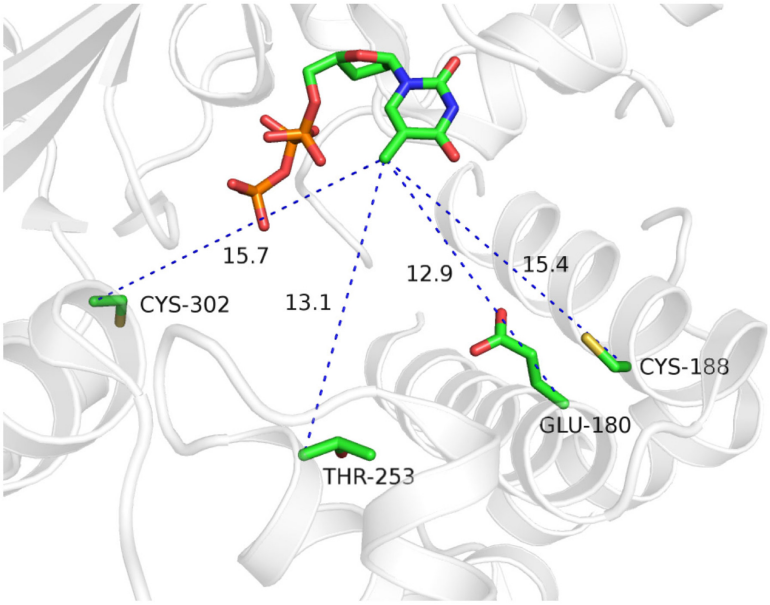


Figure 2.7: The amino acids that were mutated to cysteine in order to generate linker attachment positions are shown in green. The protein is shown in cartoon representation (white), the dTTP bound to the catalytic site is also depicted in green. The distances shown are measured from the Cα position of the respective amino acid to the methyl carbon in the C5 position of dTTP. Distances are stated in Å. This figure was already published in [74]

2.1.6 Tethered incorporation of TdT mutants with one cysteine

Next, we tested if a tethered enzyme moiety would actually block other TdT-dNTP conjugates from approaching an extended DNA molecule, thereby enabling single nucleotide extensions. The four mutants with a single surface-exposed cysteine, TdT*wt*, and TdTΔ5*cys* were subjected to linker-nucleotide labeling reactions with OPSS-PEG_4-dTTP, purified, and then incubated with a fluorescent oligo for 40 s before the reaction was quenched with EDTA. The reaction products were then run on SDS-PAGE before and after cleavage of the linker with βME (Figure 2.8).

All four conjugates with a single surface exposed cysteine showed the formation of a high-molecular weight complex with the oligo during the reaction. The migration behavior of the complex varied depending on the tethering position of the primer, presumably because the incorporation generates a branched DNA-TdT molecule. Nevertheless, upon cleavage of the linker with β-mercaptoethanol, the released primers were predominantly extended by a single linker-nucleotide for all of the four different conjugates with a single attachment position. TdT*c*188 conjugates did not show complete turnover of the primer, which might be caused by slower incorporation or by an insufficient amount of active conjugates to achieve turnover of the DNA. However, the full conversion of the oligo performed by the other conjugates in 40 s indicated that fast turnover based on polymerase-nucleotide conjugates is possible.

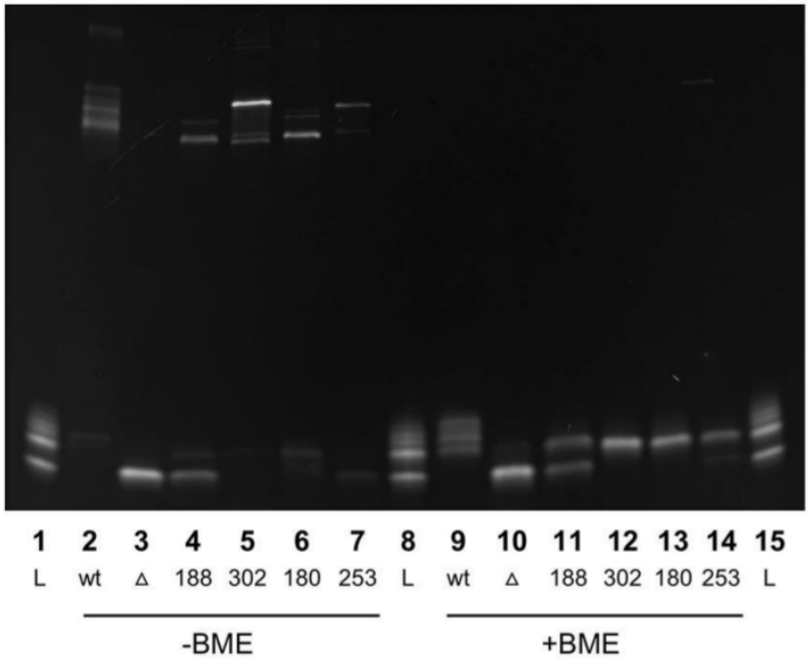


Figure 2.8: TdT-PEG$_4$-dTTP conjugates with a single attachment position were tested for their ability to extend a primer by a single dNTP. The SDS-PAGE gel shows the fluorescent primer P1 (5 FAM-dT$_{35}$). The names correspond to a ladder generated by extensions of the oligo by free linker-dTTP (L), conjugates produced from TdT*wt* (wt), TdTΔ5*cys* (Δ), TdT*c*188 (188), TdT*c*302 (302), and TdT*c*180 (180), and TdT*c*253 (253). The primer extension products are shown on the left side (-BME), the reaction products after cleavage of the linker on the right side (+BME). This figure was already published in [74].

2.1.7 Insertion of MBP fusion protein

In some experiments based on TdT-PEG$_4$-dTTP conjugates with a single attachment position, the formation of a product with two incorporated nucleotides in addition to the singly extended primer was observed. Therefore, the system showed a certain level of imperfect termination, also referred to as "non-termination". Multiple possible explanations for non-termination exist and are discussed in detail (Section 3.1.2) and investigated (Section 2.3.5) later on. However, one straight-forward explanation for non-termination can be insufficient removal of untethered linker-dNTPs after the conjugate labeling reaction.

For the experiments based on PEG$_4$-SPDP, purification of the conjugates was performed based on nickel-NTA (nickel-nitrilotriacetic acid) agarose. This might not be the ideal choice given that nucleoside triphosphates can complex divalent ions like nickel [87] and might therefore bind to the resin. To eliminate potential nucleotide(-linker) carry-over, we switched the purification strategy: By fusing maltose-binding protein (MBP) N-terminally to TdT, affinity chromatography based on amylose resin was enabled. In addition to the improved purification strategy, the fusion also showed increased expres-

sion, a feature commonly observed for MBP-fusion proteins [58]. Therefore, regular *E.coli* BL21(DE3) cells with superior growth compared to Rosetta-gami B(DE3)pLysS could be used for the expression of MBP-TDT (MTdT) variants. Generation of the MTdT expression construct was performed by the undergraduate student Hratch M. Baghdassarian under my supervision.

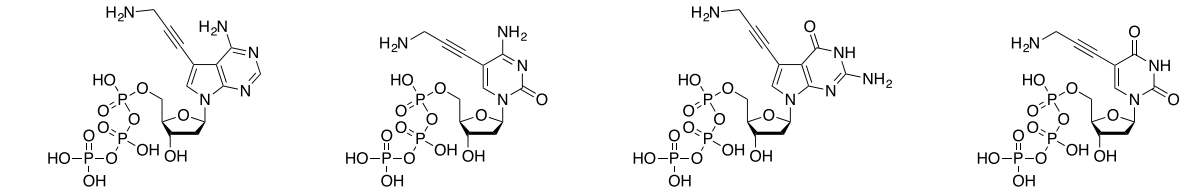


Figure 2.9: Complete set of propargylamino dNTPs (pa-dNTPs) that were used to prepare photocleavable polymerase-nucleotide conjugates. The propargylamino-modification is attached to the C5 position of pyrimidines or the C7 position of 7-deazapurines, respectivly.

2.2 Conjugates based on PC-azido-NHS

2.2.1 Propargylamino-dNTPs for linker attachment

We next aimed to develop polymerase-nucleotide conjugates of all four different bases that would leave a smaller scar upon cleavage of the linker, in order to produce DNA useful for biological applications. Propargylamino-modified dNTPs (pa-dNTPs) of all four bases are commercially available (Figure 2.9), which was not the case for the amino-allyl modification. We therefore used pa-dNTPs in combination with a photolabile linker that can be attached to primary amines, but completely eliminates upon cleavage (Figure 2.10). The nitrobenzole group of the linker can be quickly cleaved using light between wavelengths of 365 nm and 405 nm. The photocleavable linker was more difficult to assemble and bulkier than the PEG_4-OPSS linker, but usage seemed feasible given the promising data on tethered incorporation collected before.

Amplification of propargylamino-DNA

The propargylamino modification is the smallest possible scar for conjugates based on commercially available building blocks. The modification is much smaller than the scar left by TdT-PEG_4-dTTP conjugates and given the compatibility of propargylamino-modifications for sequencing applications, we were optimistic that DNA consisting of pa-dNTPs could be read by a polymerase. However, we still sought to determine the fidelity at which DNA consisting of propargylamino-scared bases can be amplified. An experiment performed by the undergraduate student Rathan Bector under Daniel H Arlow's and my supervision showed that Taq copied the DNA with an error rate of less than $6{*}10^{-4}$/nt (95% CI: $0.2{-}1.4{*}10^{-3}$/nt). The experiment was performed by preparing a scarred DNA molecule using Klenow(exo-) polymerase. The scarred DNA was then isolated and subsequently used as a template for a qPCR with Taq polymerase. The Taq-based PCR product was finally cloned into a plasmid, transformed, and single colonies were sequenced using Sanger and analyzed for errors. The experiment is described in more detail in Palluk et al. 2018 [74].

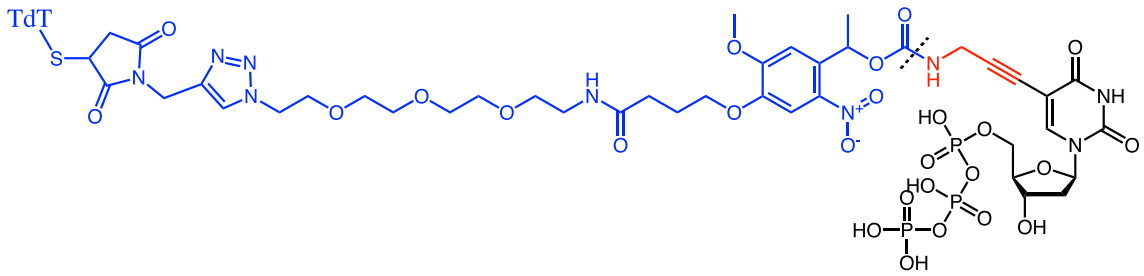


Figure 2.10: Chemical structure of the attachment of pa-dUTP to TdT based on propargyl-maleimide and PC-azido-NHS. The conjugates based on this linker are also referred to as "TdT-triazole-dTTP". The cleavable bond between TdT and the nucleotide is indicated with a black dotted line. The atoms that remain attached to the nucleobase upon cleavage (scar) are indicated in red.

2.2.2 Linker assembly using PC-azido-NHS

The photocleavable linker was assembled from PC-azido-NHS (azido-PEG_3-photocleavable-NHS-carbamate) and propargyl-maleimide (Figure 2.11). The NHS functionality of PC-azido-NHS is reactive towards the primary amine of the pa-dNTP, the azide can be linked to propargyl-maleimide using copper-catalyzed alkyne-azide cycloaddition (Huisgen cycloaddition [33]) to form a triazole, and the maleimide is reactive towards cysteines. The protein used for the assembly was MBP-fused TdT (MTdT) with a cysteine for attachment in position 302 (MTdT*c*302). Nucleotide attachment in this position had resulted in complete primer turnover with PEG_4-SPDP and seemed well-suited also for more bulky linkers based on a computational analysis of the crystal structure of TdT (see linker rendering in Section 2.3.1).

Linker assembly strategy I

In the first iterations of the linker assembly, PC-azido-NHS was coupled to a pa-dNTP, and the propargyl-maleimide was reacted with the protein. Next, the protein was purified to remove excess propargyl-maleimide, and the click reaction between PC-azido-dNTP and MTdT*c*302-propargyl was performed. Using this assembly, conjugates were generated, and tested for activity based on the SDS-PAGE assay. While some activity of the conjugates was found, the majority of the fluorescent oligo ($> 95\,\%$) was not turned over within 1 min of reaction time.

One possible explanation for the lack of conjugate activity was incompatibility of TdT with the bulky photocleavable linker. Another hypothesis was activity loss of TdT during the click reaction. A control experiment was performed, in which MTdT was subjected to the same reaction conditions used in the copper-catalyzed click reaction (in the presence of ascorbate), but without the propargyl-maleimide. The TdT exposed to the click conditions was then tested for its ability to incorporate free dNTPs, and compared to the same protein without the click exposure. We indeed found a strongly reduced activity

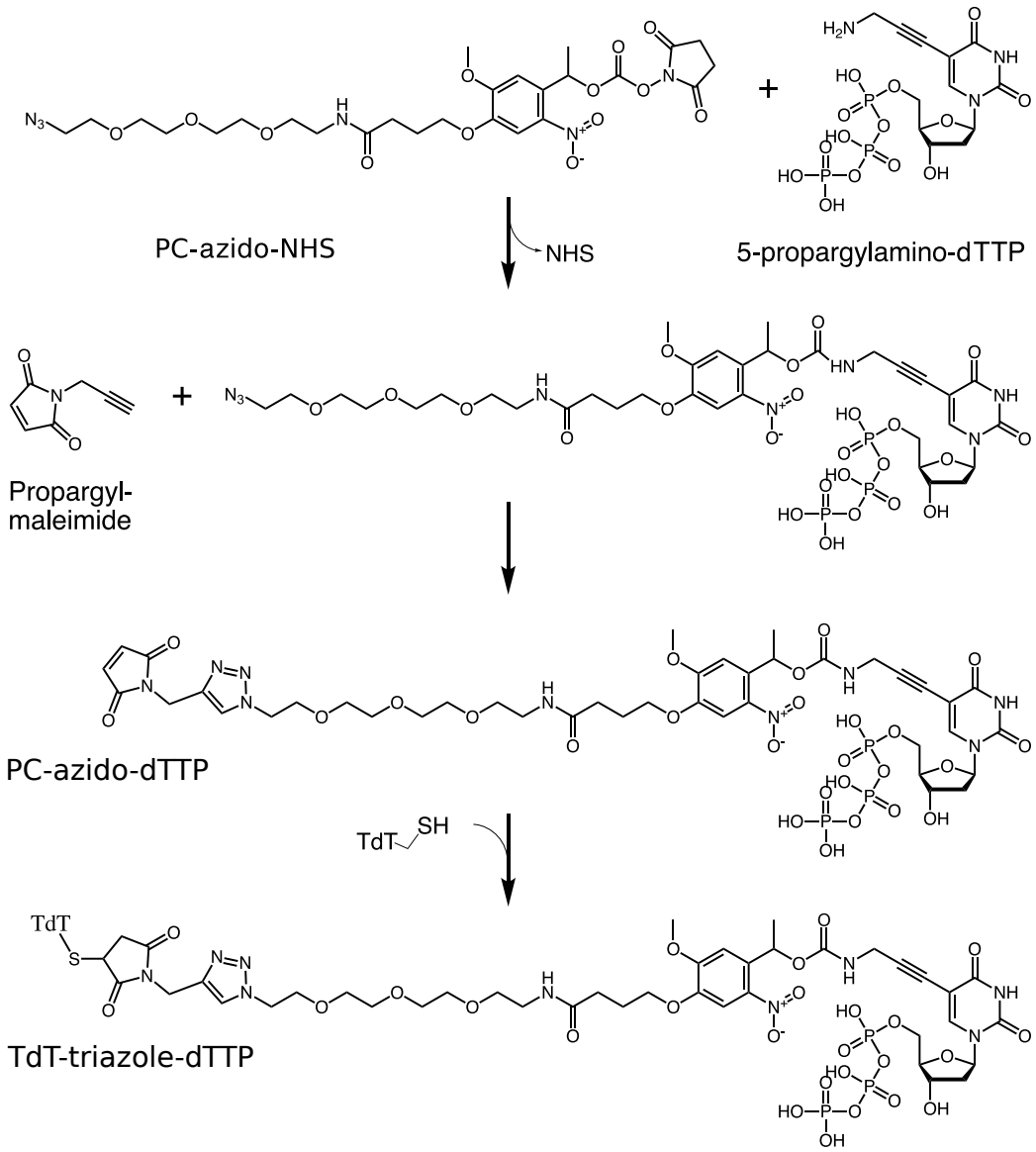


Figure 2.11: Chemical structure of the attachment of pa-dUTP to TdT based on propargyl-maleimide and PC-azido-NHS. The conjugates based on this linker are also referred to as "TdT-triazole-dNTPs". **Linker assembly strategy II** is shown. First, PC-azido-NHS is coupled to pa-dTTP, in the second step, the click reaction with propargyl-maleimide is performed to generate the linker-nucleotide PC-azido-dTTP. In the third step, the linker-nucleotide is coupled to a cysteine of TdT.

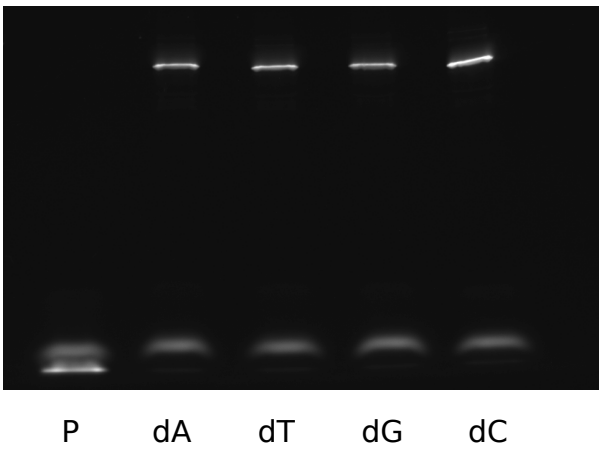


Figure 2.12: Investigation of the ability of TdT-triazole-dNTP conjugates based on MTdTc302 to extend a primer. The fluorescence of primer P3 labeled with LD650 is shown on SDS-PAGE. Lanes correspond to the primer used as starting material (P) and incorporation reactions using TdT-triazole-dATP (dA), -dTTP (dT), -dGTP (dG), and -dCTP (dC). In addition to the fluorescently labeled primer, also the loading dye fluoresces in the wavelengths used for imaging, and is therefore visible in all lanes.

after TdT had undergone the click conditions, indicating that the enzyme was damaged using the assembly procedure described above.

Linker assembly strategy II

To prevent activity loss, the linker assembly was modified and the click reaction was now performed in the absence of the protein (Figure 2.11). First, PC-azido-NHS was coupled to a pa-dNTP based on an NHS reaction. Next, the click reaction between the propargyl-maleimide and the PC-azido-pa-dNTP was performed. To remove the unreacted propargyl-maleimide that could compete with the linker-pa-dNTP during the cysteine attachment, we performed a liquid-liquid extraction based on ethylacetate. The negative charge of the triphosphate keeps dNTPs and linker-dNTPs in the aqueous phase, the propargyl-maleimide instead transitions into the ethylacetate phase and is removed. After aliquoting and drying, linker-dNTP pellets can be stored at $-80\,^{\circ}\mathrm{C}$. Finally, the maleimide reaction between the linker-pa-dNTP and MTdT*c*302 can be performed, and the resulting conjugates can be purified using amylose-resin.

2.2.3 Incorporation of all four nucleotides

Based on the second linker assembly strategy, conjugates of all four pa-dNTPs were generated. Conjugate activity was tested based on the SDS-PAGE assay using primer P3 labeled with the fluorescent dye LD650 (Lumidyne). We found that TdT-triazole-dNTP conjugates of all four pa-dNTPs formed a high-molecular weight complex with the primer and achieved an excellent yield within 1 min (Figure 2.12).

The quick incorporation of all four different nucleotides with minimal scar was an important step towards the synthesis of defined sequences. However, TdT activity loss observed during the copper-catalyzed click reaction made us suspicious about potential off-target effects of the click reaction. We further found a publication stating that DNA bases can be damaged during similar click reactions [1], and therefore sought to develop a novel linker strategy that did not require the internal click-based assembly.

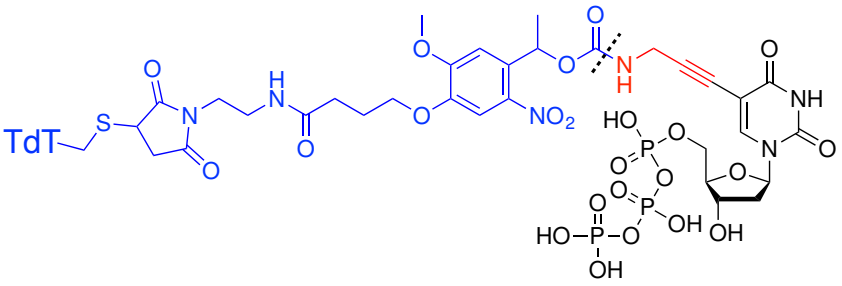


Figure 2.13: Chemical structure of the attachment of pa-dUTP to TdT based on the heterobifunctional crosslinker BP-23354. The conjugates based on this linker are also referred to as "TdT-dNTP". The cleavable bond between TdT and the nucleotide is indicated with a black dotted line, the atoms that remain attached to the nucleobase upon cleavage as a "scar" are indicated in red. A similar figure has been published in [74].

2.3 Conjugates based on BP-23354

Crosslinker BP-23354 [3] has the same attachment moieties as the previously described photocleavable linker for coupling to the amine and the sulfhydryl group, but does not require an internal assembly (Figure 2.13). The linker was ordered as a custom synthesis from BroadPharm (San Diego) and is now also available from their catalog items. A model of BP-23354 attached to MTdT*c*302 was generated to investigate possible steric constraints of the attachment strategy, since the linker is more rigid than PEG_4-SPDP and shorter than PC-azido-NHS. Molecular modeling was performed using Avogadro [31] and Maestro (Figure 2.14). For the modeling, the amino acids 299 (Leu) to 305 (Arg) were set to be flexible, all other atoms of the protein and all atoms of the dNTP and the primer were set as "fixed". The model shows how the linker fits into the protein structure, reaching through a groove between the attachment position on the nucleobase and cysteine residue 302. In the conformation modeled, there are no clashes between the nitrobenzole moiety of the linker with residues of the protein.

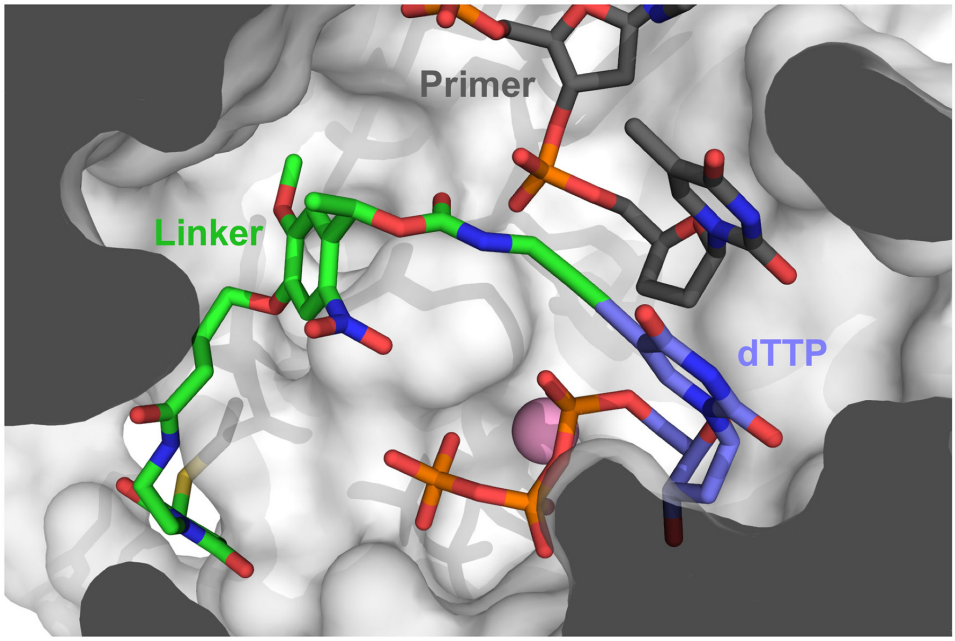


Figure 2.14: Rendering of a model of BP-23354 attached to the 5 position of dTTP and to cysteine 302. The model is based on PDB structure 4I27 and shows mutant (M)TdT*c*302. The 3'-OH of dTTP was inserted using Maestro, the amino acid changes of the mutant and the rendering of the figure were performed in PyMOL [84]. This figure was also published in [74].

2.3.1 Conjugate assembly using BP-23354

In the first step of the conjugate assembly, BP-23354 was coupled to a pa-dNTP to form a sulfhydryl-reactive linker-nucleotide (Figure 2.15). Next, trituration of the (linker-)dNTP pellet using ethylacetate was performed to remove excess linker. The linker-nucleotides pellet was resuspended, aliquoted and stored at −80 °C either after drying in a speed-vac or based on lyophilization. MTdT*c*302 was then incubated with the linker-nucleotides to perform the labeling reaction, and the protein was subsequently purified using amylose resin. While conjugates were previously stored in 50 % glycerol at −20 °C, we found that they could also be frozen in liquid nitrogen and stored at −80 °C without activity loss. In addition to the conjugate assembly procedure, Figure 2.15 shows the chemical details of the extension of a DNA molecule by TdT-dCTP.

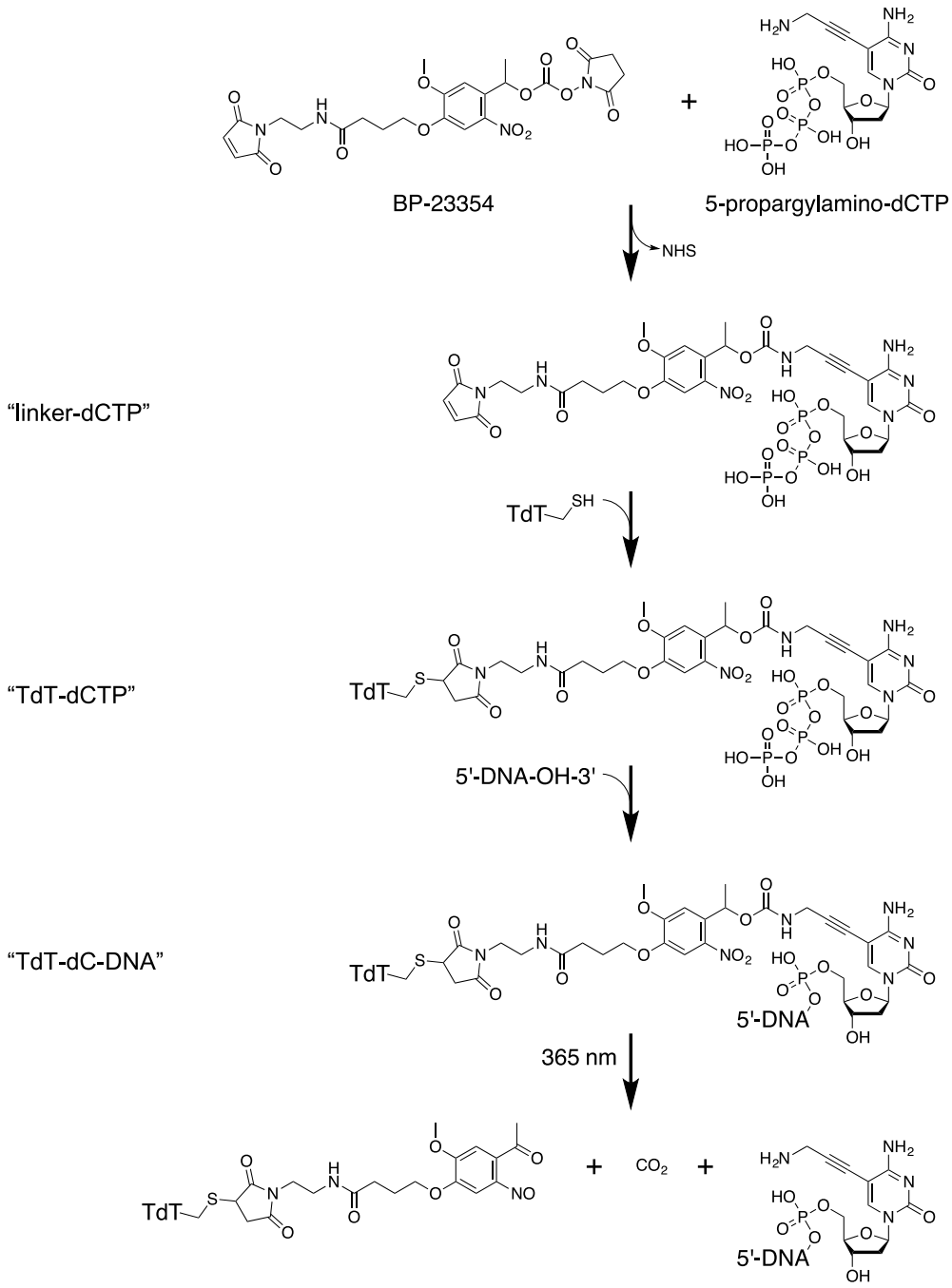
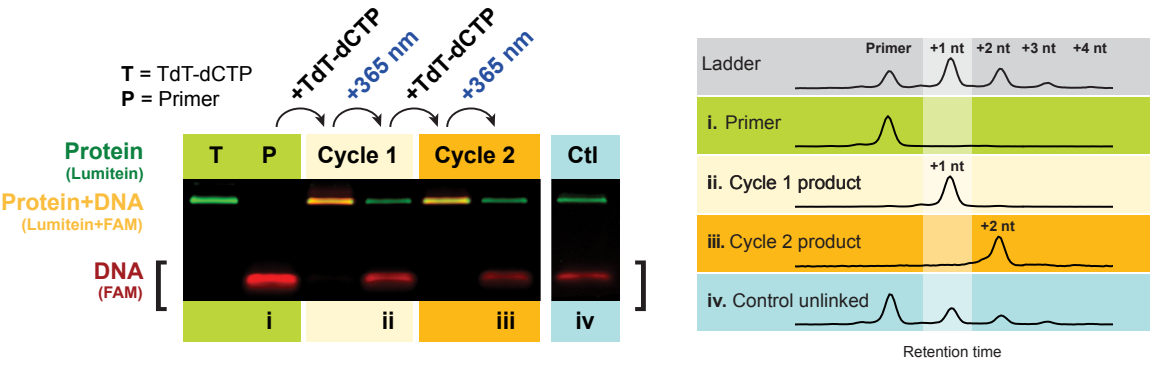


Figure 2.15: Chemical details of the preparation and usage of TdT-dNTP conjugates, shown for TdT-dCTP. First, pa-dCTP is coupled to the photocleavable crosslinker BP-23354 to form a cysteine reactive linker-dCTP. In the second step, the maleimide moiety is used to attach the linker-nucleotide to TdT to form TdT-dCTP. Upon incubation with a DNA primer, the conjugate incorporates its tethered dCTP, forming TdT-dC-DNA. The complex can be resolved by cleavage with 365 nm light. This figure was also published in [74].



(a) **(b)**

Figure 2.16: Analysis of two reaction cycles performed with TdT-dCTP conjugates. Both figures were also published in [74]. **(a)** Complex formation and dissolution between the protein (green channel) and the primer (i, red channel) during both cycles. Upon incubation with TdT-dCTP, tethering occurs and the primer migrates together with TdT, resulting in a yellow composite band. The complex is then cleaved with light of 365 nm (ii) before it is subjected to another round of conjugate-based extension and again dissociated (iii). In a control experiment where the conjugates are exposed to photocleavage conditions before incubation with the primer, no tethering occurs (iv). **(b)** Capillary electrophoresis data of all samples from panel (a) after photocleavage (indicated with brackets in panel (a)). The primer before extension (i), after the first (ii), and the second reaction cycle (iii) migrate together with the first three peaks of the ladder, indicating single nucleotide extensions. Incubation with conjugates previously subjected to photocleavage resulted in multiple extension products (control, iv). The ladder was generated by incorporation of free pa-dCTP using TdT.

2.3.2 Two cycles on SDS-PAGE and fragment analysis

Using the TdT-dNTP conjugates, single nucleotide incorporation and multiple steps of extension of a DNA molecule were tested. SDS-PAGE was used to detect tethering of a fluorescent primer to TdT, the protein was imaged based on the fluorescent stain Lumitein. In contrast to the scar left on PEG_4-SPDP upon cleavage, the small propargylamino-scar did not enable simple single-nucleotide resolution on PAGE. Therefore, capillary electrophoresis (CE) was used to analyze the reaction products in more detail.

We found that TdT-dCTP formed a high-molecular weight complex with the primer, indicated by a yellow band of composite color on the two-color fluorescent gel. No leftover primer was detected after 1 min of reaction time, indicating an excellent yield (Figure 2.16a). After irradiation of the complex with light of 365 nm, the distinct migration behavior of protein and primer could be regenerated. Using capillary electrophoresis, the reaction products after cleavage were analyzed and the vast majority of the product showed a running behavior corresponding to the +1 peak of the ladder (Figure 2.16b). The extended primer was then subjected to another round of extension, again using

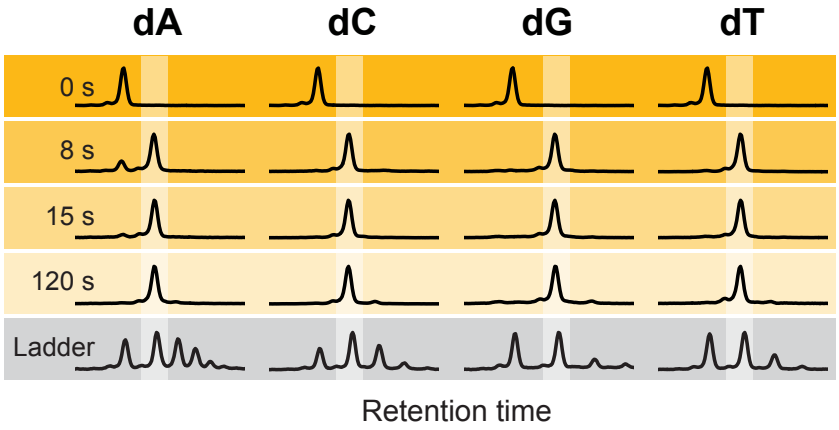


Figure 2.17: Timecourse showing fast turnover of 25 nM primer P2 by a high concentration of TdT-dNTP conjugates of all four different bases (16 µM). The ladder of product standards was generated by the incorporation of free propargylamino-dNTPs using TdT. The time stated corresponds to the incorporation time and does not include subsequent photocleavage of the linker. This figure was already published in [74].

TdT-dCTP. Again, primer and conjugate formed a high molecule weight complex visible on SDS-PAGE with a yellow composite color, and also this newly formed complex could be resolved by irradiation. On CE, as anticipated, the peak shifted towards the +2 position of the ladder, indicating successful addition of two pa-dCTP residues. In the control reaction, the conjugates were exposed to cleavage conditions prior to incubation with the oligo, and no tethering occurred during the incubation time. Instead, CE revealed a distribution of primer lengths ranging from the unextended primer up to the +3 extension product, which is in accordance with the incorporation of dNTPs that are no longer tethered to the polymerase.

2.3.3 Measuring incorporation speed

During experiments using the first two linkers, incubation times between 30 and 90 s typically resulted in quantitative turnover of a primer, which was a promising performance for establishing a practical DNA synthesis procedure. However, now that a system with minimal scars and the ability to incorporate all four bases had been developed, we sought to characterize the incorporation speed in more detail.

Fast turnover experiments were performed in a 37 °C room using a conjugate concentration of 16 µM and a primer concentration of 25 nM. For lack of a quench-flow instrument, we established a procedure where reagent addition, mixing and quenching were performed with the same pipette tip, enabling measurement of very short reaction times. After the incubation time of 8 / 15 s, the sample was quenched in Hi-Di Formamide. After cleavage of the linker (365 nm light), capillary electrophoresis analysis of the reaction products showed the full conversion of the primer into a product with the same running behavior as the +1 peak of the ladder in about 8 s for TdT-dCTP, -dGTP,

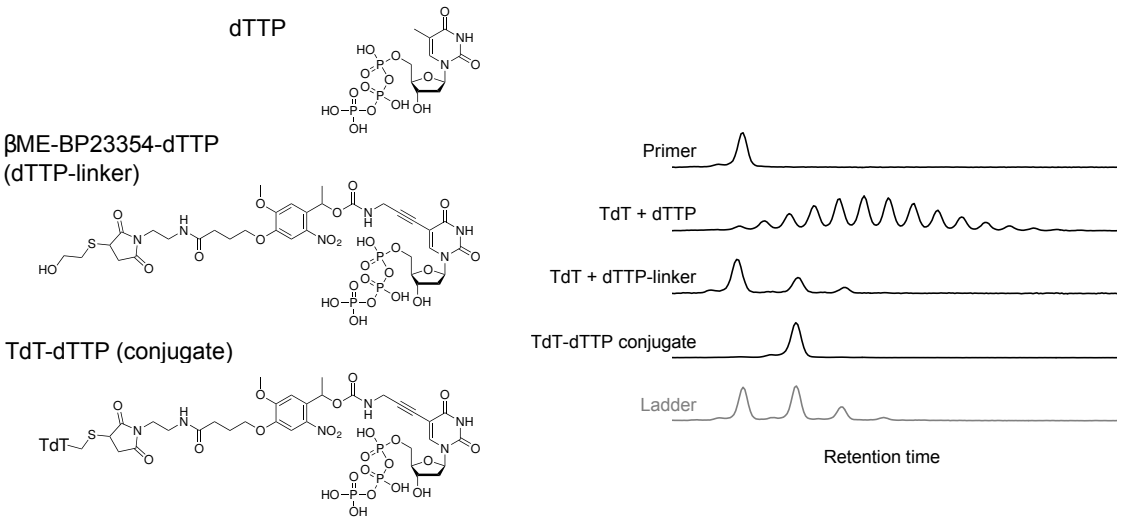


Figure 2.18: Analysis of the linker and the tethering effect on incorporation speed. In all experiments, the same concentration of TdT and an equimolar concentration of the respective (tethered) dNTP were used. Besides dTTP, the βME quenched BP23354 linker (free from solution), and the same linker attached to TdT were investigated. Both figures were also published in [74]. **(a)** The chemical structures of the three substrates tested are shown. **(b)** Capillary electrophoresis analysis of the primer and the three different reaction products. The ladder was generated using free pa-dUTP with TdT and therefore only works as reference for dTTP-linker and the TdT-dTTP conjugate.

-dTTP (Figure 2.17). For TdT-dATP, the reaction proceeds slower, with approximately 20 % of the starting material still present after 8 s. Almost all of the primer is converted by TdT-dATP after 15 s. In addition to the main reaction peak showing predominant single extension of the primer by all conjugates, a peak corresponding to the +2 peak of the ladder can be observed, indicating imperfect termination, i.e., some oligos that have been extended by two nucleotides, which is further discussed in Section 2.3.5.

The Master student Sebastian Barthel, who worked on the project under my and Daniel H Arlow's supervision, performed an experiment to analyze whether the difference in incorporation speed between the four bases is an intrinsic property of TdT with the four bases, or is induced by tethering. Sebastian found that also for the incorporation of free ddNTPs using MTdTc302, the adenosine nucleoside triphosphate is the slowest substrate, indicating that the different speed is an intrinsic property of the TdT variant used. The experiment is described in [74].

2.3.4 Effect of tethering on incorporation speed

Before we employed pa-dNTPs for the linker assembly, we performed a preliminary experiment to compare the incorporation speed of TdT with natural dTTP and pa-dUTP.

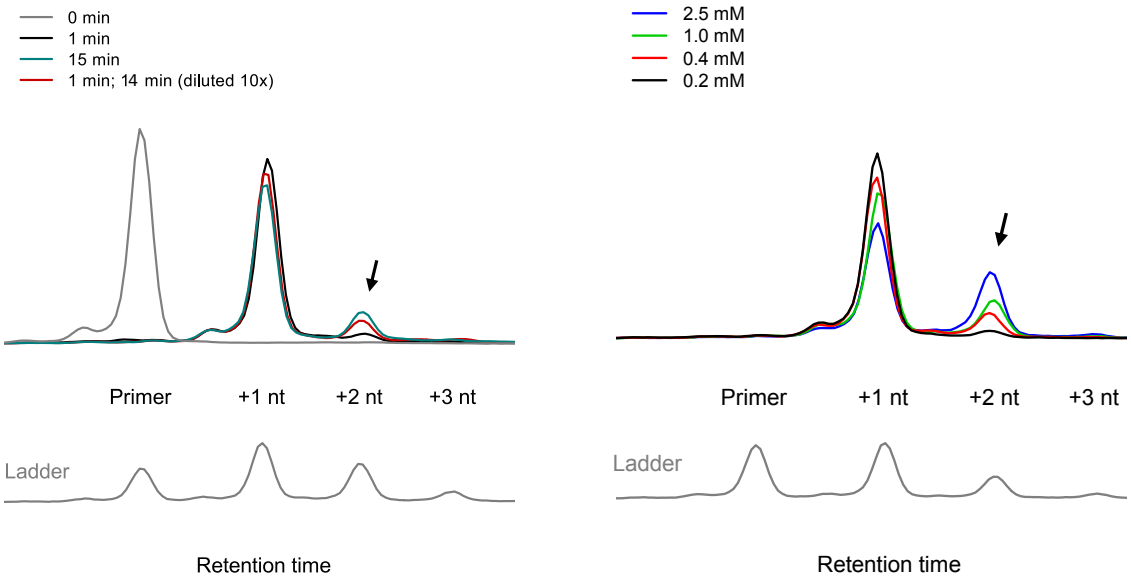
We found a very similar distribution of primer extension products for both nucleoside triphosphates and concluded that the propargylamino-modificiation itself is well tolerated by TdT, rendering it well suited for the attachment of the linker (data not shown). However, the complete linker structure attached to the base is much bigger and the photocleavable linker is relatively bulky. We therefore performed a similar experiment to test the influence of the complete linker structure (pa-dTTP plus attached BP-23354, with the maleimide moiety quenched with βME). In addition, we tested the effect of tethering the nucleotide-linker to TdT on the incorporation speed.

For the experiment, primer extension reactions were performed using the same concentration of MTdT*c*302 with 1) an equimolar amount of dTTP, 2) an equimolar amount of the linker-nucleotide BP23354-dTTP quenched with βME, and 3) the tethered linker-nucleotide BP23354-dTTP, i.e., the conjugate TdT-dTTP (Figure 2.18a). After incubation and photocleavage, the products were analyzed based on CE. While dTTP was incorporated very quickly, extending the primer by an average of 4 or 5 bases, attachment of the βME-quenched linker significantly slowed down the reaction (Figure 2.18). For the linker-nucleotide, some +1 and +2 product was observed, but the majority of primer was not turned over. In contrast, when the linker-nucleotide was attached to TdT, it was incorporated a lot faster than from solution. Complete turnover of the primer was found in this case, suggesting that tethering of the linker-nucleotide can speed up incorporation and compensate for some of the slow down caused by a bulky linker. This remarkable finding is further discussed in Section 3.1.1.

2.3.5 Conjugates with improved termination

We had previously observed imperfect termination with conjugates based on PEG_4-SPDP and as a result changed the purification method during conjugate preparation (see Section 2.1.7). However, in spite of the new purification strategy, non-termination was still observed. For example, Figure 2.17 shows the formation of a +2 product for all four bases after 120 s. While the amount of non-termination is small, even losses of 0.1 % are relevant for a DNA synthesis procedure, so that understanding and eliminating double extensions is of high importance.

Polymerase-tethered dNTPs may cause non-termination that can be classified in two different categories: The first category, **intra-molecular** non-termination, is based on multiple dNTPs that are tethered to the same TdT molecule, as observed with conjugates based on wildtype TdT in Section 2.1.1. Additional nucleotides can be tethered to the protein if a) multiple cysteines are accessible, or b) if the tethering chemistry is not specific, so that also other residues can be labeled. The second category, **inter-molecular** non-termination, is based on dNTPs that are tethered to one polymerase-molecule, but can still reach into another polymerase and be incorporated by it. Such non-termination can occur, e.g., if a long linker enables the dNTP to reach into another polymerase, but also includes special cases where an unfolded polymerase may have a structure that does not "shield" the nucleotide from other polymerase-molecules. For example, incomplete translation products, if not removed during protein-purification, could be labeled but not provide sufficient protection of the tethered dNTP.



(a) **(b)**

Figure 2.19: Generation of conjugates with improved termination. CE data is shown, the intensity axes were normalized to the peak area of the depicted region and the ladder was generated by incorporation of the free pa-dNTP using TdT. The peak corresponding to the +2 product is indicated with a black arrow in both panels. Both figures were also published in [74]. **(a)** Experiment to test if non-termination derives from a conjugate-concentration dependent or -independent process. TdT-dCTP prepared with 1 mM nucleotide in the labeling reaction was used to extend primer P2 (grey, 0 min) to the +1 product (black) in 1 min. Then, the primer was incubated for another 14 min with the same conjugate concentration (cyan) or at a ten-fold dilution (red). **(b)** Primer P2 turnover in 5 min using TdT-dTTP conjugates prepared with 0.2 mM to 2.5 mM nucleotide in the labeling reaction.

To differentiate between the two categories, we tested if the +2 formation arises from a conjugate-concentration-dependent or -independent process. After performing an initial primer-extension reaction with TdT-dCTP to form the +1 product (1 min reaction time), the extension reaction was split up into two different aliquots. One of the aliquots was diluted 10-fold with reaction buffer, while the other aliquot was kept as is, and both were incubated for another 14 min (Figure 2.19a). If the +2 formation was entirely based on a concentration-dependent process, the amount of non-termination observed in the diluted sample should be reduced significantly. However, instead of a 10-fold reduction, about 3/4 of the amount of non-termination from the undiluted reaction were found. These results indicated that the majority of +2 formation is concentration-independent, and therefore occurs based on two dNTPs tethered to the same TdT molecule.

To validate this observation and to reduce the amount of non-termination, we further tested various concentrations of (linker-)nucleotide in the labeling reaction and analyzed the effect on +2 formation (Figure 2.19b). We found that between a concentration of 2.5 mM to 0.2 mM (linker-)nucleotide, the amount of +2 formation of TdT-dTTP conjugates was reduced with lower labeling concentrations, which was in agreement with our understanding of the process. At the same time, the activity of the conjugates was maintained (data not shown). We later performed similar experiments titrating down the nucleotide labeling concentration of all four dNTPs, and found that 0.1 mM for TdT-dCTP and -dGTP and 0.2 mM for TdT-dATP and -dTTP lead to the lowest amount of non-termination while the maximal reaction speed was maintained.

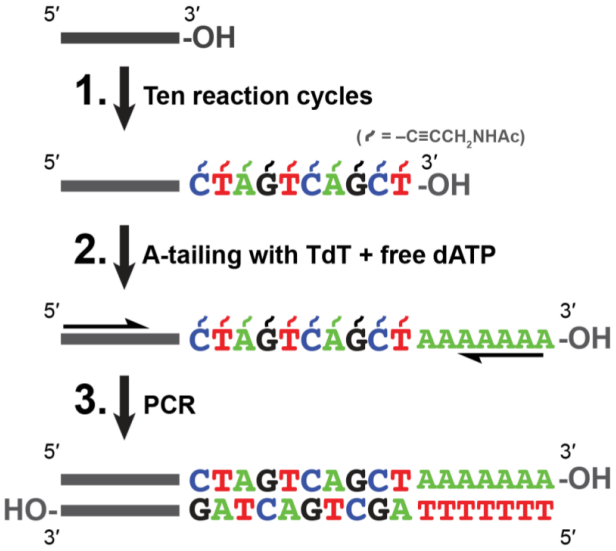


Figure 2.20: Scheme for the 10-mer synthesis and DNA amplification. First, the reaction cycles are performed, resulting in a DNA molecule extended by 10 scarred bases (1). In the next step, tailing of the reaction products using TdT and free nucleotides, in this case dATP, is performed, to create a reverse primer binding side (2). Finally, the reverse primer binding site is used to amplify the DNA molecule, generating a double stranded construct that contains the synthesized region as natural (unscarred) nucleotides (3). This figure was also published in [74].

2.4 10-mer synthesis using TdT-dNTP conjugates

2.4.1 Development of initial synthesis method

With the knowledge that DNA-containing propargylamino-scars can be amplified with reasonable fidelity and that TdT-dNTP conjugates can be used to perform at least two cycles, we next attempted to synthesize a defined oligonucleotide sequence. While there is ongoing effort to automate our system in a microfluidic device with the DNA attached to a solid support, the work I present here was based on experiments with the DNA in solution. In the long term, a DNA synthesis system requires a solid support for sufficient DNA recovery and to enable quick washing and addition of new reagents. However, *in solution* work appeared sufficient to validate that our system can achieve high yields over multiple synthesis cycles.

Synthesis scheme

The scheme for the 10-mer synthesis consisted of 1) extending an existing DNA molecule using the respective TdT-dNTP conjugate, 2) quenching the reaction and cleaving the linker, and 3) purification of the extended DNA to remove the reaction components and enable the next addition step. Once the desired sequence had been synthesized, a reverse primer binding site was added using TdT and free dNTPs (Figure 2.20). After the addi-

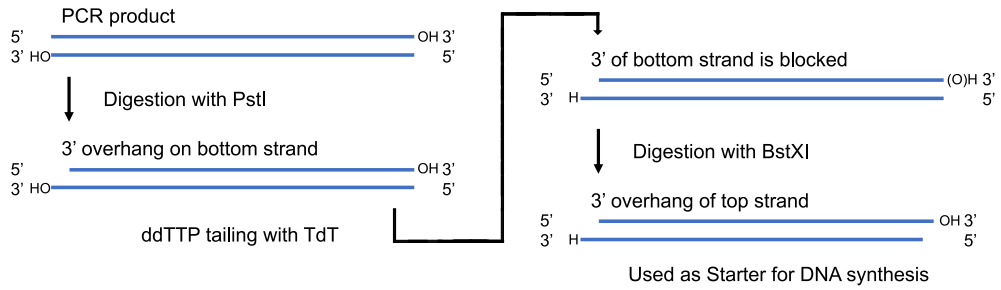


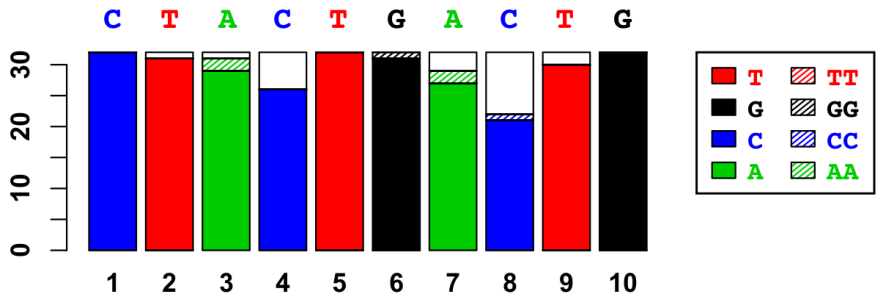
Figure 2.21: Scheme for generating the starter for the 10-mer synthesis. The PCR product derived from pET19 was first digested to generate a 3'-overhang on the bottom strand. This 3'-overhang enabled efficient ddTTP tailing of the bottom strand, that was performed in the next step. Last, another digestion was performed to generate a 3'-overhang and a free 3'-OH on the top strand that is later used for TdT-dNTP based extension. The complete starter was separated from other products on an agarose gel and purified. A similar figure was published in [74].

tion of a poly-A or poly-T tail, a complementary primer was used for the amplification of the DNA molecule.

DNA recovery

We first attempted synthesis of a defined sequence using a 90-mer oligonucleotide as the starter for TdT-based extension. Between the cycles, the DNA was purified using the Zymo OCC (Oligo Cleanup and Concentrator) kit, and a small aliquot was taken for tailing, amplification, and later analysis. To our surprise, when we analyzed those aliquots based on qPCR, we found much later amplification for reactions corresponding to ascending synthesis cycles, and the amplification of the cycle 10-product was comparable to the signal found for the no-template control. We investigated potential DNA loss during the OCC procedure based on qPCR, and indeed found that only 10 to 20 % of the DNA was recovered during the purification. Starting with an initial DNA amount of 1.6 pmol (16 µL with 100 nM oligo), such yields result in the recovery of only a few thousand molecules after cycle 10 ($1.6\,\text{pmol} * 0.15^{10} = 9.2\text{x}10^{-21}\,\text{mol}$). Small numbers of molecules can be amplified based on PCR, however, chances of amplification- and contamination bias increase, and larger DNA amounts would be highly favorable.

Based on the assumption that bigger DNA constructs can be easier purified than the single stranded oligonucleotide, we started using a double stranded PCR product as the starter. The PCR product was amplified from pET19 as a 359 bp long construct. To enable TdT-based extension on one end of the molecule, the PCR-product was cut in order to generate a 3'-overhang. Prior to that, the other end of the DNA was tailed with dideoxynucleotides to prevent further extensions (Figure 2.21). Using the double stranded starter together with Zymo Research DCC (DNA Clean and Concentrator) columns, purification yields of 50 % were achieved, enabling recovery of femtomoles of DNA after 10 synthesis steps, sufficient for the following PCR amplification.



32 clones sequenced x 10 nt = 320 attempted extensions

Correct extensions	Insertions	Deletions
291/320 (91%)	6/320 (2%)	23/320 (7%)

Figure 2.22: Analysis of the 10-mer synthesis performed using DCC purification columns. 32 clones containing the PCR-amplified synthesis product were sequenced, and the number of correct nucleotides (solid bars), double-extensions (hatched bars) and deletions (white bars) was analyzed for all positions. 13 clones (41 %) contained the complete 10-mer sequence.

2.4.2 10-mer synthesis using DCC purification columns

To synthesize the sequence 5'-CTACTGACTG-3', the double-stranded starter was subjected to ten extension cycles using the corresponding TdT-dNTP conjugates. All extension reactions were performed for 90 s at 37 °C, photocleavage of the linker was done for 20 min using a 365 nm LED. The conjugates used were generated with 0.4 mM of the respective (linker-)nucleotide in the labeling reaction, which corresponds to the conditions used before the linker-nucleotide concentration was changed to mitigate non-termination (see Section 2.3.5). After synthesis and tailing of the reaction products with dTTP, the product was PCR amplified and cloned into pUC 19 based on EcoRI and HindIII sites introduced by the PCR primers. The plasmids were transformed into *E.coli* DH10B cells and single colonies were grown overnight in LB medium, mini-prepped, and sequenced.

We sequenced the plasmids of 32 colonies that contained the PCR-amplified reaction product and found that 13 plasmids (41 %) contained the perfect 10-mer we intended to synthesize. To characterize the errors and estimate yields of the individual steps, all reads were aligned with the target sequence, and the base readout was assigned to the corresponding step (Figure 2.22). We found that some of the cycles had lead to the correct nucleotide addition in all of the 32 clones (steps 1,5 and 10). Others had a significantly lower yield, such as step 8, in which only 21 (65.6 %) of the sequences contained the right nucleotide. Out of the 320 attempted nucleotide additions (32 clones∗ 10 steps), 91 % were successful, 2 % resulted in non-termination (double extension) of the DNA with the respective nucleotide, and 7 % had failed.

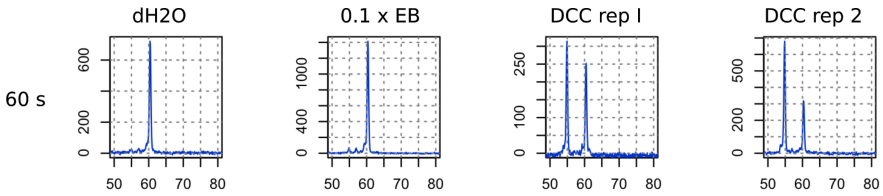


Figure 2.23: The inhibitory effect of Zymo Research DCC eluates on subsequent TdT reactions was tested. Conjugate-based primer extension reactions with the same amount of ddH_2O, 0.1 x Elution Buffer (EB), or DCC eluate (two replicates, rep) were performed. Reaction aliquots were quenched after 60 s and analyzed using CE. The chromatograms are scaled to the size of the highest peaks. The Y axis shows the fluorescence intensity, the X axis shows normalized sizing references for unmodified DNA oligonucleotides.

2.4.3 Development of improved synthesis method

While it was encouraging that many clones contained the complete sequence, the low yield observed for some of the steps was concerning. The variability between different synthesis cycles was not only found in the 10-mer presented, but also during other synthesis attempts, without an obvious explanation or detectable pattern. We therefore aimed to increase yields and reproducibility by investigating a number of possible modifications to the synthesis procedure, such as 1) optimization of TdT incorporation and linker cleavage conditions, 2) use of different conjugate labeling conditions (see Section 2.3.5), or 3) the effect of acetylation of the propargylamino scar between cycles. Some of those modifications are described below and were implemented in the next synthesis attempts to facilitate the procedure and to improve yields. However, they did not explain the large variability between stepwise yields of different synthesis cycles.

The first possible explanation for the variability of different steps we identified was loss of conjugate activity when the conjugates were stored in the presence of the cofactor cobalt. Activity loss of more than 50 % in 30 min was observed during storage at 4 °C, probably due to hydrolysis of the tethered dNTP in presence of the co-factor. We had often kept conjugates on ice in the presence of cobalt during the synthesis procedure, making it likely that corresponding activity loss had played a certain role. Without divalent ions, conjugates do not lose activity during 4 °C storage.

The second potential problem identified was related to the DNA purification between the cycles: We tested the DCC procedure for potential inhibitory effects on the next extension reaction. To do so, TdT-dTTP based primer elongation, either with ddH_2O or 0.1x EB as positive controls, or with the same amount of DCC eluate (column initially only loaded with ddH_2O, elution performed with ddH_2O) was performed. We found that conjugate reactions containing the eluate often performed worse, indicating inhibition of the reaction by some ingredient of the DCC elution (Figure 2.23). The eluates of different columns showed different extents of inhibition, potentially explaining why some of the 10-mer steps performed better than others. Modification of the DCC purification protocol with additional washing steps, spins, and incubation times to remove ethanol

were attempted, but we did not find conditions in which inhibition was consistently removed.

Modifications of the synthesis procedure

AMPure XP base purification: To eliminate possible inhibition arising from the DCC-based DNA cleanup as discussed above, AMPure XP beads were used for the DNA purification. Prior to the synthesis, we performed an experiment analogous to the one that revealed inhibitory effects of the DCC with the eluate of AMPure beads. If ethanol was sufficiently removed prior to the drying step, no slow-down of the TdT-dNTP reactions with the eluate was observed. The AMPure XP beads enabled DNA recovery above 50 % using the double stranded starter, sufficient for the synthesis.

Labeling reaction and storage: The conjugate labeling reaction was modified to reduce the amount of non-termination based on the findings documented in Section 2.3.5. For the initial 10-mer synthesis, all conjugates were produced with 0.4 mM (linker-)nucleotide in the labeling reaction. For the novel synthesis, TdT-dCTP and -dGTP were produced with 0.1 mM (linker-)nucleotide, and TdT-dATP and -dTTP were produced with 0.2 mM. To prevent activity loss during storage, we started buffer-exchanging the conjugates in TP8 (50 mM potassium acetate, 20 mM Tris-acetate, pH 7.9) instead of RBC (TP8 + 10 mM magnesium acetate, 0.25 mM cobalt chloride), after the labeling reaction.

Linker cleavage:. Photocleavage with the 365 nm LED used in the first synthesis (measured irradiance of 5 mW/cm^2) took about 20 min. To simplify the work-flow of the synthesis and also to demonstrate that the cleavage can be performed fast employing more powerful light sources, we replaced the LED with a 405 nm laser. The estimated irradiance of the novel cleavage system was 400 mW/cm^2, and enabled us to reduce the cleavage time to 1 min. As a precaution, sodium azide, a commonly used oxygen radical scavenger, which is used in buffers for photocleavage of reversible terminator dNTPs, such as in [61], was added to the quenching buffer.

Reaction conditions: While we initially performed all reactions for 90 s, the time for TdT-dATP was raised to 180 s for the new synthesis. TdT-dATP performs slower than conjugates of the other three nucleotides, so adjustment of the incubation times seemed reasonable.

Acetylation after photolysis: Upon cleavage of the linker, the resulting propargylamino scar of the dNTP contains an amine that is positively charged in our buffer conditions [63]. The charge can be removed, e.g., through a reaction of the amine with NHS-acetate (producing an N-acetyl-propargylamino scar) as also performed prior to CE analysis to improve running behavior. Assuming that multiple positively charged nucleobases might impede extension of an oligo by TdT, we performed a synthesis with

scar-neutralization using NHS-acetate after every step. The acetylation was not tested independently of other changes so we cannot conclusively say whether it is beneficial, however, the synthesis showed improved yields, so we continued using the procedure. To perform the acetylation, the quenching buffer was modified to contain 100 mM $NaHCO_3$, our buffer of choice for NHS reactions, and 40 mM NHS-acetate. The NHS-acetate was added immediately before the quenching was performed.

Phosphatase-based dNTP removal: In the first synthesis attempts using AMPure XP beads for DNA purification, we detected nucleotide insertions that did not follow the double-insertions scheme. The novel insertions instead followed a different pattern ("ABA-pattern") where the additional nucleotide was always of the same type as the pa-dNTP added one cycle ago. For example, we found the sequence CTAGTCAGC**G**T in 2.9 % of the reads of one synthesis, which was the desired sequence except for the G insertion highlighted in bold and underlined. We hypothesized that the insertions happened as the incorporation of a free pa-dNTP carried over from the previous cycle, after a DNA molecule had already been extended and was covalently bound to a TdT-conjugate from the next step. We tested whether the AMPure XP beads bind free pa-dNTPs and indeed found a substantial amount of nucleotide carryover during the purification procedure.

In order to reduce pa-dNTP carryover, additional washing and also variation of the ratio between water and ethanol during the washing procedure were investigated. However, the pa-dNTPs were binding to the beads even with modified washing conditions, so we decided to instead enzymatically digest the triphosphate after photocleavage. Alkaline phosphatase was used for the digestion, and the quenching buffer and the alkaline phosphatase buffer were modified accordingly to enable phosphatase activity. The EDTA concentration in the quenching buffer was reduced, to enable "over-buffering" of the EDTA with magnesium ions required for the phosphatase reaction. The phosphatase was added in Tris-HCl buffer so that the Tris would quench the NHS-acetate and prevent it from modifying the enzyme. The phosphatase treatment enabled a 10-fold reduction of ABA-pattern insertions (see Section 2.4.4).

2.4.4 10-mer synthesis using AMPure XP beads

Using magnetic bead DNA purification, we aimed to synthesize the 10-mer 5'-CTAGTCAGCT-3'. The original sequence attempted contained a tandem repeat of 5'-ACTG-3, which seemed disadvantageous. Taq can produce stutter products in which 4 bp repeats are deleted due to slippage [86, 96], and we found several synthesis products missing the last four bases during previous experiments. The new synthesis was performed with the modifications described in the previous Section, using extension reaction times of 90 s or 180 s for TdT-dATP and photoclavage time of 1 min with 405 nm light. After synthesis and tailing of the reaction products with dTTP, the product was PCR amplified, and the PCR products were sequenced.

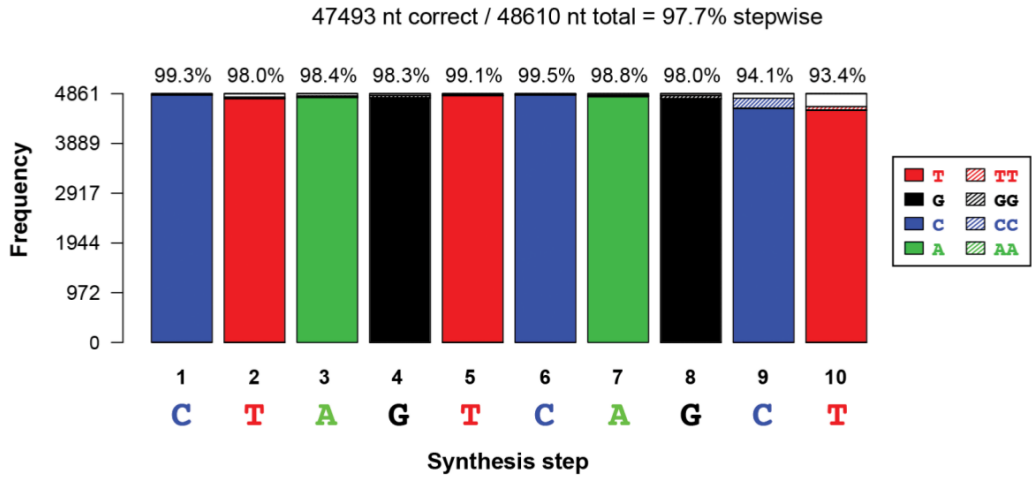


Figure 2.24: NGS-based analysis of the 10-mer synthesis using magnetic beads. The estimated yields of the steps are based on the analysis of 4,861 NGS reads (excluding singleton reads) that were aligned against the target sequence. The correct extensions are indicated with solid bars, white bars present deletions, and hatched bars indicate double-extensions. Substitutions were observed at a rate below 0.1 %. This figure was also published in [74].

NGS analysis of synthesis products

To enable a larger coverage than previously achieved with sanger-sequencing, the synthesis products were analyzed using NGS. The samples were submitted to the JBEI DiVA sequencing service, where they were barcoded and processed using the Nextera library preparation (Illumina) with some modifications [93]. Sequencing was performed on a MiSeq (Illumina), multiplexed together with other samples. The Illumina reads were filtered for sequences containing the 3'-end of the starter, up to 20 nucleotides, and a poly A tail of at least 6 bases, (5'-TCCAGATTT(N_{0-20})AAAAAA-3'). The selected reads were binned, and singleton reads were removed to avoid artifactual errors, e.g., based on index switching between different samples. About 1.5 % of the selected sequences were singleton reads.

The DiVA analysis resulted in 4,861 reads that contained the target region, out of which 3,913 (80.5 %) contained the perfect 10-mer that we intended to synthesize. We aligned the reads against the target sequence and identified ABA-pattern insertions in 0.7 % of the reads, that were then excluded from the analysis of stepwise yields. We then analyzed the number of correct extensions, deletions and double-extensions for every step. The average stepwise yield achieved was 97.7 %, with deletions occurring in 1.3 % of the steps, and insertions in 1.0 %. The individual steps of the synthesis had a yield between 99.5 % in step 6 and 93.4 % in step 10 (Figure 2.24, Table 2.1). Deletions ranged from 0.1 % in step 1 up to 5.2 % in step 10. Insertions were found in as little as 0.2 % (step 6) and up to 4.0 % (step 9) of the sequences. Interestingly, the step with the most as well as the one with the least insertions was the incorporation of TdT-dCTP.

Table 2.1: Stepwise deletions, insertions, and yield of correct incorporations based on the NGS analysis of the 10-mer synthesis using magnetic beads. The estimated yields are based on the analysis of 4,861 NGS reads (excluding singleton reads) that were aligned against the target sequence.

Step	1	2	3	4	5	6	7	8	9	10
Base	C	T	A	G	T	C	A	G	C	T
Del	0.1 %	1.4 %	1.0 %	0.8 %	0.4 %	0.3 %	0.7 %	0.8 %	1.9 %	5.2 %
Ins	0.5 %	0.6 %	0.6 %	0.9 %	0.4 %	0.2 %	0.4 %	1.3 %	4.0 %	1.4 %
Yield	99.4 %	98.0 %	98.4 %	98.3 %	99.2 %	99.5 %	98.8 %	98.0 %	94.1 %	93.4 %

Multiple sequence alignment

A multiple sequence alignment of all sequences with a frequency of more than 0.25 % is shown in Figure 2.25. The correct 10-mer was the most abundant sequence with a frequency of 80.5 %. The second most abundant sequence was a single deletion of position 10, which was found in 4.1 % of the reads, followed by a double insertion in position 9 (3.4 %). Single deletions and single insertions are alternating throughout the alignment, which is in accordance with the finding that both types of errors occur at similar rates (1 % deletions, 1.3 % insertions).

The most common errors are found towards the end of 10-mer. We suspected that this part of the sequence can be impacted by PCR bias, e.g., if the poly-T primer anneals into the sequence, forming a mismatch with the last base, and performed an experiment to test this hypothesis: The same 10-mer product was tailed with varying A-tail lengths, PCR amplified, and analyzed using NGS. Indeed, PCR products derived from shorter tails contained more "deletion" errors in the last position, however, we were not able to measure this effect consistently. A direct analysis of the synthesis products, e.g., based on HPLC or CE, would mitigate PCR bias, but would require a higher quantity of DNA and therefore a solid support.

The first sequence with two errors contains the combination of the two most common single-errors, which is in agreement with the expectancy assuming independent synthesis steps (frequency of 0.3 %). An ABA-pattern insertion (additional insertion of pa-dGTP from cycle 8 after step 9) was found with a frequency of 0.3 %. As previously discussed, reads with ABA-pattern insertions occurred at a frequency of up to 2.9 % before the phosphatase treatment was implemented in the synthesis procedure. This finding therefore indicates a reduction of the ABA-insertion rate by 10-fold through the additional step.

		5′- C	T	A	G	T	C	A	G	C	T -3′
		1	2	3	4	5	6	7	8	9	0
Count	**Freq.**	**Total: 4861 reads**									
3913	80.5%	C	T	A	G	T	C	A	G	C	T
201	4.1%	C	T	A	G	T	C	A	G	C	-
169	3.5%	C	T	A	G	T	C	A	G	CC	T
56	1.2%	C	T	A	G	T	C	A	G	C	TT
51	1.0%	C	T	A	G	T	C	A	G	-	T
46	0.9%	C	T	A	G	T	C	A	GG	C	T
45	0.9%	C	-	A	G	T	C	A	G	C	T
28	0.6%	C	T	A	GG	T	C	A	G	C	T
27	0.6%	C	T	-	G	T	C	A	G	C	T
21	0.4%	C	TT	A	G	T	C	A	G	C	T
19	0.4%	C	T	A	-	T	C	A	G	C	T
19	0.4%	CC	T	A	G	T	C	A	G	C	T
16	0.3%	C	T	AA	G	T	C	A	G	C	T
15	0.3%	C	T	A	G	T	C	A	G	CC	-
13	0.3%	C	T	A	G	T	C	A	G	CG	T

Figure 2.25: The multiple sequence alignment shows all NGS reads of the 10-mer synthesis using magnetic beads that occurred with a frequency of at least 0.25 %. An ABA-pattern insertion caused by dNTP carryover of the magnetic beads was found in 13 reads and is indicated by the grey “G”. This figure was also published in [74].

```
5'- C  C  C -3'

Count  Freq.   Total: 474 reads

  418  88.2%   C  C  C
   41   8.6%   C  C
   15   3.2%   C  C  C  C
```

Figure 2.26: All sequences of the 3-mer synthesis (5'-CCC-3') except for singleton reads are shown. This figure was also published in [74].

2.5 Synthesis of 5'-CCC-3' using TdT-dCTP conjugates

For the 10-mer sequence, repeats of the same base were separated by at least two other nucleotides to enable clear assignment of the sequencing readout to each step. However, direct repeats are an essential part of DNA sequences, so we performed an experiment to test if the conjugates were also able to repeatedly add the same base. The sequence 5'-CCC-3' was synthesized using polymerase-nucleotide conjugates following the same procedure and conditions employed for the magnetic bead synthesis of the 10-mer. The samples were then NGS-sequenced using the JBEI DiVA sequencing service.

All sequences observed during NGS of the synthesis products except for singleton reads are shown in Figure 2.26. Of 474 reads, 88.2 % contained the correct sequence, followed by 8.6 % readouts of two cytosine residues, presumably due to a single deletion. 3.2 % of the reads contained four cytosines, presumably due to a single double-insertion. Assuming three independent steps, the stepwise yield was calculated based on the amount of correct 3-mers to be $0.882^{1/3} = 95.9\,\%$. As in the 10-mer synthesis, deletion errors were found more often than insertions.

3 Discussion

3.1 Polymerase-nucleotide conjugates and reversible termination

Conventional reversible terminators are incorporated by a polymerase freely from solution. Upon extension of a DNA molecule, the RTdNTP modification prevents further nucleotide additions, while the primer has the ability to interact with other polymerases and dNTPs. The concept of polymerase-nucleotide conjugates follows a different strategy: By joining a polymerase and a nucleoside triphosphate, polymerase-nucleotide conjugates are generated that achieve termination because a polymerase shields the 3'-end of a primer towards other conjugates upon incorporation. The system only terminates towards other polymerase-nucleotide conjugates but not towards free dNTPs, which is one of the factors differentiating it from conventional reversible termination strategies. This thesis describes the development of polymerase-nucleotide conjugates from the initial validation of the idea up to the synthesis of a 10-mer sequence.

Validation of the basic concept

The idea that polymerase-nucleotide conjugates could be used to achieve single nucleotide extensions of a DNA molecule was initially tested using a PEG_4-SPDP linker to tether dNTPs to TdT. TdT is known for its high tolerance towards base-modifications [70], so it seemed likely that it could also incorporate tethered dNTPs with the linker attached to the base. Indeed, we found that multiple tethered nucleotides could be incorporated by wildtype TdT which has several attachment positions. Interestingly, while one would expected crowding of the catalytic site to occur during the first tethered incorporation, TdT was able to incorporate up to four tethered dNTPs, indicating a certain robustness of the tethered incorporation procedure. The encouraging results were further supported when, based on the same linker, four mutants with a single dNTP attachment positions were tested, and all showed the ability to incorporate their tethered dNTP. Importantly, all of the conjugates with a single attached dNTP converted a primer predominantly into the singly-extended product, validating the basic concept behind polymerase-nucleotide conjugates: A conjugate, upon incorporation of its tethered dNTP, shields the end of the primer and prevents further conjugate-based dNTP additions. Importantly, reaction times required for complete turnover were typically under 1 min, indicating that polymerase-nucleotide conjugates could fulfill the requirements for a DNA synthesis procedure in which a cycle should take $\ll 5$ min.

Photocleavable polymerase-nucleotide conjugates

The conjugates employing PEG_4-SPDP had limited usefulness due to a large scar left upon linker cleavage, so we developed a novel attachment strategy. A photolabile linker enabled cleavage back down to the propargylamino-group, leaving the minimal scar possible whilst using commercially available building blocks for conjugate assembly. DNA with propargylamino-modifications can be amplified by Taq polymerases with reasonable fidelity [74].

The photocleavable linker (based on PC-azido-NHS) that required an internal click reaction for assembly enabled quantitative incorporation of propargylamino-analogues of all four DNA bases in less than 1 min. These results suggested that we were close to a functional enzymatic DNA synthesis system, now having the ability to add all four bases with only small modifications. However, activity loss of the enzyme during the click reaction made us suspicious about potential side reactions during the copper-catalyzed click assembly. We therefore switched the linker chemistry once again, now using a linker with the same reactive moieties towards the dNTP and TdT, but without an internal assembly step (BP-23354).

Based on TdT-dNTP conjugates employing BP-23354, two synthesis cycles were demonstrated using PAGE and CE to monitor the individual steps (Figure 2.16). We found that repeating the synthesis cycle was possible, enabling the addition of multiple dNTPs. Tethering of the primer to conjugates looked very clean on SDS-PAGE with no detectable left-over starting material, and CE confirmed that the vast majority of primer was converted to the +1 and later into the +2 product. These findings therefore validated the suitability of the conjugates for multiple steps of DNA synthesis, indicating that the scars of previous iterations do not significantly hamper the next addition. Cleavage of the linker, analyzed by disassembly of the complex between the conjugate and primer on SDS-PAGE, also appeared to achieve a close to quantitative yield. For the cleavage step, a 365 nm light source with an irradiation of approximately 5 mW/cm^2 was employed, resulting in a cleavage time of 20 min, far away from a practical performance for DNA synthesis. However, this light-source can be replaced by more powerful instruments to speed up the cleavage procedure, as later shown based on the 405 nm laser that was used with an irradiation time of 1 min during the 10-mer synthesis.

3.1.1 Incorporation speed

Given sufficient process optimization, (enabling high accessibility of the DNA molecules, sufficient prevention of side reactions, etc.) the crucial factor determining synthesis length and quality is the speed at which the incorporation and deprotection reactions proceed. Characterizing primer extension speed was therefore attempted once we had conjugates suited for the synthesis of a defined sequence (Figure 2.17). Using 16 µM conjugates, a poly dT_{60}-mer was converted into the +1 product with quantitative yield in under 15 s for TdT-dCTP, -dGTP, and -dTTP, and in under 30 s for TdT-dATP. The reaction speed is below the maximal speed TdT can achieve using natural nucleotides, as discussed further below (k_{cat} of 180/min for dATP [10]). However, the turnover was in

the range required for a DNA synthesis system, and presented a good starting point for further improvements. For comparison, the incorporation cycle of an Illumina NovaSeq 6000, in which a highly engineered polymerase incorporates 3'-modified RTdNTPs, takes about 80 s according to the instrument log files.

Comparing conjugates of the four different bases, we consistently found a slower incorporation of TdT-dATP (> 2 x). Because free adenine-based nucleoside triphosphates were also incorporated slower from solution, we believe that the speed difference depends on the nucleobase and does not derive from a tethering- and therefore conjugate-specific effect. A speed comparison of different ddNTPs is described in [74]. Another factor of the incorporation speed could be the current terminal bases of the primer that closely interact with the incoming dNTP during the catalysis. Unfortunately, testing different primer sequences turns into a complex problem: Testing all four TdT-dNTP conjugates with all four different nucleotides in the last position results in $4^2 = 16$ reactions; if the effect of the last two bases shall be tested, $4^3 = 64$ combinations exist. With duplicates and multiple time points, those experiments quickly grow beyond the possible scope. Instead of doing a large screening experiment, we performed preliminary experiments that indicated comparable speed with different primers, and then decided to increase incorporation times during the 10-mer synthesis to compensate for possible slow downs during some steps (90 s and up to 180 s for TdT-dATP). In the future, incorporation times may be adapted based on the particular sequence that is synthesized if certain combinations require longer incubation.

Effect of nucleotide tethering

In the early phase of the project, we tested single primer extension using TdT with ddNTPs as an easy way to achieve single nucleotide incorporation with kinetics similar to natural nucleotides. The experiments showed that complete turnover of a primer within seconds is possible at the maximum speed of the enzyme. This turnover was not achieved using polymerase-nucleotide conjugates at 16 µM. However, we developed an experiment to test how the linker, and tethering the dNTP as opposed to incorporation from solution, impact the reaction speed.

In theory, tethering should lead to a high local concentration of the dNTP in the active site of the respective polymerase, resulting in fast incorporation by the conjugate. The tethering effect might therefore enable use of substrates that could barely be incorporated free from solution, and enable effective concentrations above the solubility limit of the tethered reagent. We indeed found that tethered linker-dNTP was incorporated much quicker than the free version at the same concentration in solution (Figure 2.18b). In comparison with the natural dNTP, free linker-dNTP was incorporated much slower. Despite of the high tolerance for base-modifications, the linker influences TdT incorporation speed, perhaps unsurprisingly given its bulk and size. Testing the incorporation of linkers with different flexibility and bulkyness, and how the reaction characteristics translate into the respective conjugate, could enable the generation of faster conjugates.

3.1.2 Non-termination

In the most simplified explanation, polymerase-nucleotide conjugates achieve termination when, upon dNTP incorporation, "the 3'-end of the primer remains covalently bound to TdT and is inaccessible to other TdTdNTP molecules". [74]. However, taking a closer look, termination turns out to be more complex. At least four requirements need to be fulfilled by the system in order to enable perfect termination:

1. Upon incorporation of its tethered dNTP, the polymerase moiety of a conjugate **hinders other polymerase-nucleotide conjugates from extending the 3'-OH** of the respective primer.

   This point is analogous to the short explanation given in the beginning of the section.

2. The dNTP tethered to one polymerase-nucleotide conjugate **may not be incorporated by a different polymerase molecule**.

   If dNTPs tethered to other polymerases can function as substrate for a polymerase that has already extended a primer and is now attached to the 3'-OH, non-termination occurs.

3. The conjugate itself may only **be labeled with a single dNTP** that it can incorporate into a primer.

   Our results show that multiply labeled conjugates can perform multiple dNTP incorporations, and we even found that unspecific labeling was one of the main sources for non-termination observed during our experiments.

4. All **free dNTPs must be removed** from the system.

To limit non-termination, a polymerase-nucleotide based DNA synthesis system therefore needs careful setup. To fulfill points 1 and 2, the length, flexibility, and attachment position of the linker, as well as the structure of the polymerase are expected to play a big role. Further, to fulfill point 2, the tertiary polymerase molecules should be intact, because an unfolded or partial polymerase moiety would likely offer less protection of a tethered dNTP than the intact protein. To fulfill point 3, measures need to be taken that prevent tethering of multiple dNTP molecules to the same polymerase. When we identified double-labeling to be the main cause for non-termination, labeling conditions were adjusted in order to reduce the effect (Figure 2.19). The non-termination could have occurred based on the additional two cysteines that are present in (M)TdTc302, or have derived from off-target reactions of the maleimide, e.g., with primary amines. In the long term, attachment strategies with higher specificity are desirable, such as the OPSS moiety employed in the PEG_4-SPDP linker. Additional cysteines could be removed, or unnatural amino acids could be used as attachment positions, e.g., in combination with copper-free click chemistry. To fulfill point 4, we generated the MBP-fusion of TdT that enabled improved purification of the conjugates after the labeling reaction, removing excess dNTPs reliably.

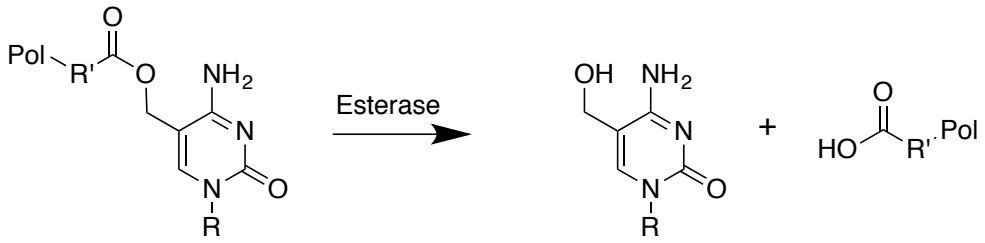


Figure 3.1: An ester-based linker attachment that can be cleaved by an esterase is shown. The nucleobase of a cytosine is depicted that would be attached to the sugar and triphosphate moiety at the **R** position. Further, **R'** includes the variable part of the linker that is then attached to the polymerase (**Pol**). Upon cleavage of the ester bond, only a small hydroxymethyl group is left on the nucleobase as a scar.

3.1.3 Linker strategies

The linker attachment position, length, and flexibility may impact dNTP incorporation speed and termination ability of the conjugates. However, there are two other aspects of the linker system that are not directly linked to the incorporation step, but are still essential. First, the scar that is left after cleavage can be of importance for follow up procedures. Second, cleavage of the linker to deprotect the extended DNA molecule needs to be extremely fast. The nucleotide addition is often considered the important step of a DNA synthesis system, however, the deprotection step has the same requirements with regard to yield and speed, and is therefore just as important.

Nucleobase attachment

Throughout this thesis, linker attachment was performed based on the C5 position of pyrimidines or the C7 position of 7-deazapurines, respectively. Modifications of these positions are known to be reasonably tolerated by polymerases (e.g., [36]) and the propargylamino attachment turned out to have little effect on Taq-based PCR amplification, which enabled the generation of natural DNA. However, it might be advantageous to further reduce the scar size in future implementations. One option for small nucleobase modifications that are found in nature are pyrimidine derivatives with a hydroxymethyl group in the C5 position [98]. The same group could be attached to the C7 position of 7-deazapurines, offering a simple solution for an attachment strategy that results in a small, uncharged modification upon cleavage. A scheme in which a linker is cleaved to leave a hydroxymethyl modification is depicted in Figure 3.1. Attachment of a linker to the exocyclic amines of adenine, cytosine and guanine [23], and attachment to an amine in the aromatic ring for thymine might enable scar-free cleavage.

Cleavage moieties

Two different cleavage moieties of the linker were used in this demonstration. The OPSS-based linker generated a disulfide bond that can be cleaved, e.g., using TCEP. Unfortunately, disulfide cleavage results in a sulfhydryl scar that can interfere with disulfide bonds in the next step by disulfide exchange. The photocleavable linker then enabled cleavage down to a propargylamino group and could be performed quickly with a sufficient light source. While we did not attempt optimization of the cleavage, similar nitrobenzole groups were cleaved quantitatively in 10 s with 365 nm light [90], which renders photochemistry very promising in the context of DNA synthesis. DNA damage induced by irradiation at this wavelength and high irradiation intensity has been studied in the context of RTdNTPs, and was not found to occur at significant levels [61]. Another option could be enzymatically cleavable linkers that promise fast reactions with high specificity. A strategy for a linker system employing an esterase for cleavage is shown in Figure 3.1. The esterase would in this case leave a hydroxymethyl modification on the nucleobase. To enable access of the esterase to the linker, the tethered polymerase could be digested using a protease, or denaturing reagents might be used to unfold TdT, rendering the linker accessible.

3.2 10-mer sequence

Synthesis of a defined sequence with a stepwise yield close to a practical DNA synthesis system was the proof of concept aimed for in this thesis. In the first 10-mer synthesis attempt, an average stepwise yield of 91 % was achieved. At this error rate, less than half of the DNA molecules are correct within 8 synthesis cycles. After the optimization of various parameters, the stepwise yield was increased to 97.7 % as measured by NGS. Phosphoramidite synthesis nowadays consistently achieves stepwise yields above 99 %, and synthesis of long oligos up to 150 requires yields exceeding 99.5 % [60]. Our demonstration does not achieve these numbers on average, and we further observed a considerable variability between different steps, with the worst yield being 93.4 %. However, the huge improvement achieved between the two synthesis demonstrations and various steps that exceeded a yield of 99 % are indicators that the system could be improved accordingly. Extrapolating from our current yield, synthesis of short primers for oligonucleotides should be possible ($.98^{40} = 45\%$), however, this would require a solid support in order to generate a sufficient amount of the primer.

While coupling times of 1.5 min (C, G, and T) and 3 min (A) were used in the final 10-mer demonstration, a realistic goal of an enzymatic DNA synthesis system is a cycle time below 1 minute. The capillary electrophoresis data in Figure 2.17 indicate that the conjugates can perform the incorporation reaction in 10–30 s, and reactions times under 10 s should be achievable given the k_{cat} of the enzyme. Deprotection, e.g., using photolabile groups is possible in a similar amount of time, as discussed in Section 3.1.3. Further, stepwise yields need to exceed those of phosphoramidite chemistry to enable the synthesis of longer and higher quality oligos. Approaches to reduce non-termination

have been discussed above. Ways to increase turnover further include automation to eliminate handling mistakes, adaption of buffer conditions or divalent ions used, or enzyme engineering to speed up the reaction. Variability between steps could also arise from secondary structure formation of the DNA that could render the 3'-end less accessible to TdT, slowing down the reaction. To prevent secondary structure formation, an increase of the reaction temperature or base-modifications that prevent secondary structure formation and are removed after synthesis [59] could be implemented.

Another very important step will be the implementation of a solid support for the growing DNA molecules, to enable automation and sufficient DNA recovery for analytical chemistry. NGS sequencing offers an easy way to estimating stepwise yields, and high coverage can be achieved. However, there are many additional steps between the synthesis and the NGS readout, in which bias can be introduced. One potential source of bias is the PCR, where annealing of the homopolymer primer into the synthesized region (producing a mismatch) could lead to an amplicon in which the last step of the synthesis allegedly appears to have many deletions. Methods such as HPLC will facilitate the analysis of reaction products, providing a more direct measure.

3.3 Costs of polymerase-nucleotide conjugates

It is important that the reagent costs of an enzymatic DNA synthesis procedure are not prohibitive for commercialization. One aspect to consider is that phosphoramidites are typically used at concentrations above 0.1 M to enable fast turnover, while much more efficient enzymatic reactions could work at 20 µM conjugates. Therefore, even if the production of the conjugates turns out to be much more expensive, they could potentially be used at a 5,000 fold lower concentration. In addition, during synthesis at a microscale level, e.g., for gene synthesis, reagent costs can be brought to an almost negligible level. A small number of DNA molecules can be PCR-amplified after the synthesis to generate the quantity required for follow up procedures. It seems likely that higher yields, increased reliability, and simplification or removal of DNA assembly procedures will be the cost-determining factors in this area.

Compared to an enzymatic DNA synthesis system using RTdNTPs, one could expect polymerase-nucleotide conjugates to have increased reagent costs, because they consume an amount of polymerase stoichiometric to the DNA. However, I think it is well possible, at least for microscale synthesis, that the enzyme concentration will be chosen in order to enable high reaction speed, and that the polymerase will be in an excess over the DNA molecules in both systems. It will be interesting to see how large-scale production can drive down costs for reagents for enzymatic DNA synthesis in the future, and to witness how larger scale DNA synthesis will be enabled based on the cost reduction.

3.4 Outlook

Before an enzymatic system will be competitive with conventional DNA synthesis, a large amount of process optimization is required. Phosphoramidite synthesis has achieved its high yields and reliability over decades, solving various scientific challenges such as the development of a suited solid support. However, I believe that various potential advantages of enzymatic systems will render it superior in the long term, as summarized in our publication: "1) the exquisite specificity of enzymes and mild conditions in which they function may reduce the formation of side products and DNA damage such as depurination, thereby enabling the direct synthesis of longer oligos; 2) reactions take place in aqueous conditions and need not generate hazardous waste; 3) synthesis could be initiated from natural DNA (i.e., DNA without protecting groups on the nucleophilic positions of the bases); and 4) enzyme engineering techniques such as high-throughput screens and selections can be employed to optimize the system in ways that are not possible using organic chemistry alone." [74] Synthesis in aqueous conditions further enables use of other proteins during the procedure, opening up the whole toolbox nature has created for handling DNA.

Besides our approach based on polymerase-nucleotide conjugates, several groups attempt to employ 3'-modified RTdNTPs with engineered polymerases for *de novo* DNA synthesis. The most prominent players in the field include DNA Script, Molecular Assemblies and Nuclera Nucleics. While certain benefits can result from the use of 3'-modified RTdNTPs, for example very reliable termination, I believe that the ability of polymerase-nucleotide conjugates to employ 3'-unblocked dNTPs is a key advantage for the synthesis of longer sequences. Despite large polymerase engineering efforts for 3'-O-modified RTdNTPs for sequencing applications [14], incorporation of 3-O-modified RTdNTPs still lags behind the incorporation kinetics of natural dNTPs for those systems [25]. The ability of polymerase-nucleotide conjugates to use dNTPs with an unmodified sugar moiety removes the need for enzyme engineering efforts, and enables fast incorporation rates that will translate into high yields and longer DNA sequences.

I hope that the presented work will be the starting point for the development of a practical enzymatic DNA synthesis technology. Such a system could have a huge impact on biotechnology, as summarized in a comment from Adam Clore [19]: "The work of Palluk et al. raises interesting questions about the future of DNA synthesis and synthetic biology. Facile production of long, high-fidelity DNA would likely galvanize the synthetic biology industry and could hasten the replacement of traditional chemical manufacturing by biological manufacturing." In my opinion, it is only a matter of time until enzymes will enable practical desktop oligonucleotide synthesizers and direct DNA production exceeding the synthesis length possible today.

4 Material and Methods

Large parts of this method section were already published in [74] and [4].

4.1 DNA construction

The TdT sequence used consists of the residues 132-510 from the short isoform (TdTs) of *Mus musculus* TdT (NCBI Accession number: NP_001036693.1) and lacks the N-terminal BRCT domain [30]. For the initial demonstration of tethered dNTP incorporation employing the PEG_4-SPDP crosslinker, TdT was expressed fused to an N-terminal His-Tag (TdT*wt*). The gene encoding TdT was ordered from Integrated DNA Technologies with codon-optimization for *E. coli* and inserted into pET19b using isothermal assembly [27]. To generate the TdTΔ5*cys* mutant, surface-exposed cysteine residues were identified in the crystal structure of TdT (PDB ID: 4I27) and the mutations Cys188Ala (PDB ID 4I27 numbering), Cys216Ser, Cys302Ala, Cys378Ala, and Cys438Ser were introduced using site-directed mutagenesis [34]. Based on this surface-cysteine free TdT variant, mutants with a single surface-exposed cysteine for linker attachment were then constructed by the re-insertion of cysteines into four positions near the catalytic site using site-directed mutagenesis, generating the mutants TdT*c*180 (Glu180Cys), TdT*c*188 (Ala188Cys), TdT*c*253 (THR253Cys), and TdT*c*302 (Ala302Cys). To generate the TdT variant that was used with the photocleavable linkers, maltose binding protein (MBP) was inserted between the His-Tag and the TdT domain of TdT*c*302 to yield MTdT*c*302. The sequence encoding MBP from *E. coli* was amplified from pMAL-c5X (NEB) and inserted into the pET19b plasmid harboring the TdT gene using isothermal assembly.

Sequences of the plasmids coding for all TdT variants can be downloaded from the JBEI Public registry (https://public-registry.jbei.org/folders/355). Respective cloning and expression strains harboring the plasmids were added to the JBEI strain archive and are available upon request (see Table 4.1).

Table 4.1: Plasmid and strain accession numbers. Sequences of the plasmids coding for all TdT variants can be downloaded from the JBEI Public Registry (https://public-registry.jbei.org/folders/355).

Construct	Plasmid	Cloning strain	Expression strain
pET19-TdT*wt*	JPUB_010259	JPUB_010258	JPUB_010276
pET19-TdTΔ5*cys*	JPUB_010261	JPUB_010260	JPUB_010275

pET19-TdT*c*180	JPUB_010267	JPUB_010266	JPUB_010274
pET19-TdT*c*188	JPUB_010265	JPUB_010264	JPUB_010273
pET19-TdT*c*253	JPUB_010263	JPUB_010262	JPUB_010272
pET19-TdT*c*302	JPUB_010257	JPUB_010256	JPUB_010271
pET19-MTdT*c*302	JPUB_008786	JPUB_010268	JPUB_008785

4.2 Section A: Conjugates based on PEG_4-SPDP

4.2.1 Protein expression and purification of His-tagged TdT

The expression of all TdT variants without MBP domain was performed in *E. coli* Rosetta-gami B(DE3)pLysS cells (Novagen) [55] that were grown in LB media containing carbenicillin (50 g/L), chloramphenicol (34 g/L), tetracycline (12.5 g/L) and kanamycin (15 g/L).

Starting cultures were grown for 24 h at 37 °C with shaking at 200 RPM and then used to inoculate expression cultures in shake flasks without baffles using a 1/20 dilution. Expression cultures were grown at 37 °C and 200 RPM until an OD_{600} of 0.6 was reached. IPTG was added to a final concentration of 0.5 mM and the expression was performed for 12 h at 30 °C with shaking at 200 RPM. Cells were harvested by centrifugation and resuspended in buffer A (20 mM Tris-HCl, 0.5 M NaCI, pH 8.3) + 5 mM imidazole. Protein purification steps were performed at 4 °C. Cell lysis was performed using sonication followed by centrifugation at 15,000 x g for 20 min. The supernatant was applied to a gravity column containing 1 mL of Ni-NTA agarose (Qiagen) equilibrated with 10 mL of buffer A + 5 mM imidazole [9]. The column was washed with 10 volumes of buffer A + 40 mM imidazole and elution was performed using 4 mL buffer A + 500 mM imidazole. The eluted protein was concentrated using Vivaspin 20 columns (MWCO 10 kDa, Sartorius) and then dialyzed against TdT pH 7.4 Storage Buffer (100 mM NaCl, 200 mM K_2HPO_4, pH 7.4) overnight at 4 °C using Pur-A-Lyzer Dialysis Kit Mini 12000 tubes (Sigma-Aldrich). The protein was stored at −20 °C after the addition of 50 % glycerol.

4.2.2 Preparation of TdT-PEG_4-dTTP conjugates

The scheme for the preparation of the polymerase-nucleotide conjugates is shown in Figure 2.2. The heterobifunctional amine-to-thiol crosslinker PEG_4-SPDP (no. 26128) and 5-aminoallyl-dUTP (aa-dUTP) were purchased from Thermo Fisher Scientific. First, aa-dUTP was reacted with PEG_4-SPDP to form the thiol-reactive linker-nucleotide OPSS-PEG_4-dTTP. The reaction contained 12.5 mM aa-dUTP, 3 mM PEG_4-SPDP and 125 mM sodium bicarbonate (pH 8.3) in a combined final volume of 8 µL. After an incubation of 1 h at RT, any potential unreacted PEG_4-SPDP was quenched by the addition of 1 µL of 100 mM glycine in PBS (Teknova; 137 mM sodium chloride, 2.7 mM potassium chloride, 4.3 mM disodium phosphate, 1.4 mM monopotassium phosphate, pH 7.4) and the reaction was incubated for an additional 10 min. To lower the pH for the subsequent OPSS-labeling reaction with the protein, 1 µL of 10 x TdT pH 7.4 Storage Buffer was added. The 10 µL reaction was then added to 70-100 µg TdT in 40 µL of 1x TdT pH 7.4 Storage Buffer to site-specifically label TdT at surface cysteine residues. After a 13 h incubation at RT, the TdT-PEG_4-dTTP conjugates were purified using the Capturem His-Tagged Purification Miniprep Kit (Clonetech) to remove free (i.e., unattached) linker-nucleotides. Elution of the protein from the Capturem columns was performed in 200 µL Capturem Elution Buffer, and the eluate was dialyzed against 100 mL TdT Reaction Buffer (50 mM potassium acetate, 20 mM tris-acetate, 10 mM magnesium acetate,

pH 7.9) for 4 h using Pur-A-Lyzer Dialysis Kit Mini 12000 tubes (Sigma-Aldrich). The conjugates were then directly used in oligo-extension reactions.

4.2.3 Generation of oligo P1 ladder with OPSS-PEG$_4$-dTTP

The ladder of oligo P1 (5-FAM-dT$_{35}$) OPSS-PEG$_4$-dTTP extension products for use as size standards in PAGE assays was generated by incorporating free OPSS-PEG$_4$-dTTP nucleotides using TdT. First, OPSS-PEG$_4$-dTTP was formed in a 20 µL reaction containing 15 mM aa-dUTP, 45 mM PEG$_4$-SPDP, and 250 mM sodium bicarbonate. The reaction was incubated at RT for 1 h and another 5 µL of 180 mM PEG$_4$-SPDP was added. After 1 h, the reaction was quenched by the addition of 5 µL 100 mM glycine in PBS and incubated for 10 min. The crude products containing the OPSS-PEG$_4$-dTTP nucleotides were then used directly in 6 TdT incorporation reactions with varying polymerase concentrations to achieve varying numbers of incorporations. The 15 µL incorporation reactions contained Reaction Buffer with Cobalt (RBC: 50 mM potassium acetate, 20 mM tris-acetate, 10 mM magnesium acetate, 0.25 mM cobalt chloride, pH 7.9), 1 µM oligo P1, 1 µL of the OPSS-PEG$_4$-dTTP crude products and 100, 50, 25, 12.5, 6.3, or 3.13 U of NEB TdT. Reactions were performed for 5 min at 37 °C and quenched with EDTA to a final concentration of 33 mM. Reactions were selectively combined to yield a ladder showing 0-4 OPSS-PEG$_4$-dTTP extensions. The quenched reactions were combined with an equal volume of 2x Novex Tris-Glycine SDS Sample buffer + 1 % β-mercaptoethanol (βME) and heated to 95 °C for 5 min before they were loaded into a polyacrylamide gel.

4.2.4 Investigation of TdT-PEG$_4$-dTTP activity by SDS-PAGE

To demonstrate the ability of TdT to incorporate tethered nucleotides, TdT-PEG$_4$-dTTP conjugates of TdT*wt*, TdTΔ5*cys*, TdT*c*180, TdT*c*188, TdT*c*253, and TdT*c*302 were used. Incorporation reactions shown in Figure 2.4 and Figure 2.6 contained 250 nM oligo P1 (See Table 4.5), RBC, and ~0.4 mg/mL of the respective TdT-PEG$_4$-dTTP conjugate. The reactions were performed for 20 s at 37 °C and were quenched by the addition of EDTA to a final concentration of 33 mM. The incorporation reactions shown in Figure 2.8 contained 500 nM of oligo P1, RBC, and ~0.1 mg/mL of the respective TdT-PEG$_4$-dTTP conjugate. Reactions were performed at 37 °C for 40 s and were quenched by the addition of EDTA to a concentration of 33 mM. To prepare the reactions for the PAA analysis, samples were mixed with an equivalent volume of 2x Novex Tris-Glycine SDS Sample Buffer (Thermo Fisher Scientific), or with 2x Novex Tris-Glycine buffer + 1 % βME to cleave the linkage between the protein and the nucleotides. All samples were heated to 95 °C for 5 min and run on a 4-20 % Mini-PROTEAN TGX SDS gel (Bio-Rad).

4.3 Section B: Conjugates based on PC-azido-NHS

4.3.1 Protein expression and purification of MBP-fused TdT

E. coli BL21(DE3) harboring pET19-MTdT*c*302, pET19-MTdT*wt* or pET19-M(BRCT)-TdT*wt* were grown in LB medium (Miller) with 100 µg/mL carbenicillin with shaking at 200 RPM throughout the expression procedure. Starting cultures were grown overnight at 37 °C and then used to inoculate expression cultures in shake flasks without baffles using a 1/60 dilution. Expression cultures were grown at 37 °C until an OD_{600} of $0.40-0.45$ was reached. The flasks were then cooled to room temperature (RT) for 45 min without shaking and then shaken at 15 °C for 45 min. Protein expression was induced with 1 mM IPTG and cells were grown overnight at 15 °C and harvested by centrifugation. All protein purification steps were performed at 4 °C. Cells were lysed in Buffer A + 5 mM imidazole using an Emulsiflex C3 homogenizer followed by centrifugation at 15,000 x g for 20 min. The supernatant was subjected to nickel affinity chromatography (HisTrap FF 5 mL, GE Healthcare) with an imidazole gradient (Buffer A + 5 mM imidazole to Buffer A + 500 mM imidazole). Fractions with sufficient purity were pooled, diluted 1:40 into 20 mM Tris-HCl pH 8.3, and subjected to anion-exchange chromatography (HiTrap Q HP 5 mL, GE Healthcare) in 20 mM Tris-HCl pH 8.3 using a gradient of 0 to 1 M NaCl. The protein eluted at 200 mM NaCl. The purest fractions were buffer exchanged into TdT pH 6.5 Storage Buffer (200 mM K_2PO_4, 100 mM NaCl) and concentrated to ~30 mg/mL using Vivaspin 20 columns (MWCO 10 kDa, Sartorius). Protein concentrations were estimated by absorbance spectrophotometry on a NanoDrop 2000 assuming an extinction coefficient of 108,750 $M^{-1}cm^{-1}$ at 280 nm. The protein was stored at −20 °C after the addition of 50 % glycerol. Subsequently, we found that MTdT*c*302 could be snap-frozen in liquid nitrogen and stored at −80 °C in TdT pH 6.5 Storage Buffer without loss of activity. The protein used in the experiments before Section 2.3.4 was stored at −20 °C in 50 % glycerol, whereas the protein used in all experiments in later sections was snap-frozen in aliquots using liquid nitrogen and stored at −80 °C.

4.3.2 Preparation of TdT-triazole-dNTP conjugates using MTdTc302

Synthesis of linker-dNTPs

The scheme for the preparation of TdT-triazole-dNTP conjugates is shown in Figure 2.11. The photocleavable PC-azido-NHS linker was purchased from Broadpharm (catalog number BP-22951), propargyl-maleimide was purchased from Jena Bioscience. The complete set of propargylamino-dNTPs (pa-dNTPs) was purchased from TriLink Biotechnologies. First, the pa-dNTP was coupled to PC-azido-NHS in a 6 µL coupling reaction containing 5 mM of the respective pa-dNTP, 4 mM PC-azido-NHS and 400 mM sodium bicarbonate buffer (pH 8.3) in ddH_2O. The reaction was incubated at RT for 1.5 h and then stored at 4 °C over night. Next, the click reaction with propargyl-maleimide was performed to build the complete linker-nucleotide. The following components were incubated for

1 h at RT: 6 µL of the initial coupling reaction, 3.75 µL of 20 mM propargyl-maleimide, 3.75 µL of a freshly prepared 1:5 solution of $CuSO_4$ (1 mM) and THPTA (5 mM; Tris((1-hydroxy-propyl-1H-1,2,3-triazol-4-yl)methyl)amine, Jena Bioscience), 7.5 µL of 25 mM freshly prepared sodium ascorbate in ddH_2O, 3.75 µL TP7 buffer (50 mM potassium acetate, 20 mM tris-acetate, pH 7) and 13.75 µL ddH_2O. Then 0.375 µL of 30 mM EDTA was added to quench the reaction.

To remove unreacted propargyl-maleimide, a liquid-liquid extraction with ethyl acetate was performed. 200 µL of ethyl acetate was added to the 40 µL reaction, followed by vortexing and spinning of the sample in a microcentrifuge at ~1000 RCF). The supernatant was removed and the extraction procedure was repeated three additional times. 15 µL of the aqueous phase were then taken and incubated in an open tube at 50 °C for 5 min to remove potential ethyl acetate contamination.

MTdTc302 labeling with linker-dNTPs and purification

The maleimide-labeling reaction of the protein was then performed by the addition of 6 µL 10 x TP7 buffer and 12 µL MTdT*c*302 at a concentration of 18 mg/µL to the 15 µL (linker-)nucleotide reaction. The reaction was incubated for 1.5 h at RT before the conjugates were purified using microcentrifuge-based amylose affinity chromatography. In detail, the labeling reaction was diluted into 100 µL MBP Column-Binding Buffer (NEB, 200 mM NaCl, 20 mM Tris-HCl, 1 mM and 1 mM DTT, pH 7.4) and loaded onto a 0.8 mL spin column (Pierce) containing 200 µL amylose resin (NEB) in MBP Column-Binding Buffer. All centrifugation steps were performed at 50 RCF, all reagents and buffers used throughout the procedure were precooled on ice. The column was incubated in a shaker block at 800 RPM for 10 min for binding. Next, the column was washed three times with TdT Reaction Buffer. Each washing step involved 1) addition of 500 µL buffer to the column, 2) incubation of the column for 1 min while shaking at 800 RPM, 3) centrifugation at 50 RCF for 1 min, and 4) removal of the flow-through. Elution of TdT-azido-dNTP conjugates was performed by 1) the addition of 150 µL TP8 Buffer + 10 mM maltose, 2) an incubation for 5 min while shaking at 800 RPM, and 3) centrifugation. The elution procedure was repeated twice. The eluates were then combined and concentrated using a 30 kDa MWCO column (Corning), diluted 1:10 with TdT Reaction Buffer to reduce the maltose concentration, and concentrated to ~4 µg/µL.

4.3.3 Investigation of TdT-triazole-dNTP activity by SDS-PAGE

Primer P3 (see Table 4.5) contains a 5'-amine that was labeled with the fluorescent dye LD650 (Lumidyne Technologies) in a reaction containing 50 µM of oligo P3 and 4.5 mM of the LD650 NHS-ester in sodium bicarbonate buffer, followed by OCC purification (Zymo Research). TdT-triazole-dNTP conjugates of all four pa-dNTPs were prepared. Reactions of all four conjugates as well as a negative control without the protein were performed and contained 75 nM labeled oligo P3, 1 x RBC, and ~1 mg/mL of the respective TdT-triazole-dNTP conjugate (except for the negative control). The reactions were performed for 1 min at 37 °C and quenched by the addition of EDTA to a final concen-

tration of 60 mM. To prepare the reactions for the gel, samples were mixed with an equivalent volume of 2x Novex Tris-Glycine SDS Sample Buffer + 1 % βME and heated to 95 °C for 5 min. The samples were then run on a 4-20 % Mini-PROTEAN TGX SDS gel (Bio-Rad) for 15 min.

4.4 Section C: Conjugates based on BP-23354

4.4.1 Preparation of TdT-dNTP conjugates using MTdTc302

The protein was expressed and purified as described in Section 4.3.1.

Synthesis of linker-dNTPs

The scheme for the preparation of TdT-dNTP conjugates is shown in Figure 2.15. The photocleavable NHS carbonate-maleimide crosslinker was purchased from Broadpharm (catalog number BP-23354). First, the pa-dNTP was coupled to BP-23354 to form a thiol-reactive linker-dNTP (maleimide) in a 30 µL reaction containing 1 µL of the respective 100 mM pa-dNTP (100 nmol), 1 µL of 10 x TdT pH 7.4 Storage Buffer, 26 µL of ddH_2O, and 2 µL of 100 mM BP-23354 dissolved in anhydrous DMSO (200 nmol) added last. The reaction was incubated at RT for 1 h with shaking. Initially, the linker concentration was above the solubility limit, but once the reaction progressed, enough (soluble) linker-dNTP product was formed that the remaining (unreacted) linker fully dissolved. The crude products were triturated with ethyl acetate (~2 mL) and centrifuged at 15,000 x g to pellet the linker-dNTPs. The supernatant was removed and the linker-dNTP-containing pellets were dried by speed-vac or lyophilization and stored at −80 °C.

MTdTc302 labeling with linker-dNTPs and purification

To site-specifically label TdT at surface cysteine residues with a linker-dNTP, a dried linker-dNTP pellet was resuspended in 1 x TdT pH 6.5 Storage Buffer and added to MTdT*c*302 (conc. 10-15 µg/µL by absorbance) in 1 x TdT pH 6.5 Storage Buffer. The (nominal) nucleotide concentration in the labeling reactions ranged from 0.1 mM to 2.5 mM, depending on the experiment. (The nominal nucleotide concentration was calculated based on the assumption that all (linker-)dNTPs precipitate quantitatively during trituration.) Unless indicated otherwise, all conjugates employing linker BP-23354 were prepared with nominal nucleotide concentrations of 0.1 mM (dGTP, dCTP) and 0.2 mM (dATP, dTTP). The labeling reaction was incubated for 1 h at RT, and TdT-dNTP conjugates were purified using amylose affinity chromatography to remove free (i.e., untethered) dNTPs: A spin column purification was performed using 0.8 mL spin columns (Pierce) that were filled with 250 µL amylose resin (NEB), and all centrifugation steps were performed at 50 RCF. All reagents and buffers used throughout the procedure were precooled on ice. Prior to binding, the amylose resin was washed twice with 500 µL of TdT pH 6.5 Storage Buffer. A typical 15 µL linker-dNTP labeling reaction containing ~200 µg of MTdT*c*302 was diluted into 200 µL TdT pH 6.5 Storage Buffer and loaded onto the spin column containing the amylose resin, which was then incubated in a shaker block at 800 RPM for 10 min for binding. Next, the column was washed twice with TdT pH 6.5 Storage Buffer. Each washing step involved 1) addition of 500 µL buffer to the column, 2) incubation of the column for 1 min while shaking at 800 RPM, 3) centrifugation at 50 RCF for 1 min, and 4) removal of the flow-through. Next, washing with TdT

Reaction Buffer or with TP8 Buffer (50 mM potassium acetate, 20 mM tris-acetate, pH 7.9), respectively, was performed twice. For the conjugates used in the initial 10-mer synthesis described in Section 2.4.1, TdT Reaction Buffer was used for the washing and the rest of the purification procedure, and the conjugates were then stored in 1 x RBC after the addition of 0.25 mM cobalt. For the 10-mer synthesis with improved methodology described in Section 2.4.3, TP8 buffer that does not contain divalent ions was used for the rest of the procedure and for the storage. Elution of TdT-dNTP conjugates was performed by 1) the addition of 150 µL TdT Reaction Buffer + 10 mM maltose or TP8 Buffer + 10 mM maltose, 2) an incubation for 5 min while shaking at 800 RPM, and 3) centrifugation. The elution procedure was repeated twice, and the eluates were combined and concentrated using a 30 kDa MWCO column (Corning). The reaction was then diluted 1:10 with TdT Reaction Buffer / TP8 Buffer to reduce the maltose concentration, and concentrated to ~2.5 µg/µL. The conjugates can be frozen in liquid nitrogen and stored at −80 °C. Notably, we observed a significant loss of activity when storing the conjugates on ice in the presence of cobalt ions.

4.4.2 Capillary electrophoresis (CE)

20 µL samples containing 0.5-1.5 nM 5'-FAM labeled oligonucleotides and ~0.3 µL GeneScan 600 LIZ dye size standard in 75 % Hi-Di formamide were submitted to the UC Berkeley Sequencing Facility for capillary electrophoresis (CE, also called Fragment Analysis). CE samples were run on an Applied Biosystems 3730xl DNA Analyzer with a 50 cm capillary array containing POP-7 Polymer, with 15 s of injection at 1.5 kV and a 41 min run at 15 kV, oven: 68 °C, buffer: 35 °C. Electropherogram data files were processed using custom software written in R (r-project.org) with comparable functionality to the Peak Scanner software from Applied Biosystems.

High ionic strength in a CE sample causes poor injection and distorted peaks, so DNA samples from extension reactions were either diluted 50-fold with 75 % formamide or desalted prior to CE as described in the respective method section.

It was observed that DNA containing multiple propargylamino groups had reduced injection yield and inconsistent migration in CE, likely due to the added positive charges. Therefore, all DNA samples containing propargylamino groups were derivatized using NHS-acetate prior to CE. Unless specified otherwise, acetylation reactions contained 20 mM NHS-acetate and 200 mM sodium bicarbonate.

4.4.3 Generation of oligo P2 ladder with pa-dNTPs for CE

Oligo P2 (5'-FAM-dT_{60}, see Table S1) extension products that were used as size standards (ladders) were generated by the incorporation of free pa-dNTPs using TdT. Reactions contained 100 nM oligo P2, 100 µM of one type of pa-dNTP, 1 x RBC, and either 0.05 U/µL or 0.03 U/µL NEB TdT. Reactions were performed at 37 °C and aliquots were quenched with EDTA to a final concentration of 33 mM after 2, 5 and 10 min. Quenched samples were then acetylated, desalted using OCC, and analyzed by CE. Samples with

detectable peaks for oligo P2 as well as the +1 and +2 pa-dNTP extension products were selected for use as ladders.

4.4.4 Two cycle demonstration using TdT-dCTP conjugates

The conjugates used in this experiment were generated using 1 mM nucleotide in the MTdT*c*302-labeling reaction. All extension reactions contained 50 nM of oligonucleotide P2, 0.25 mg/mL TdT-dCTP (or photolyzed TdT-dCTP, see below), and RBC. Reactions were performed at 37 °C and quenched after 2 min by the addition of an equal volume of 200 mM EDTA. Photolysis of the linker was performed using a Benchtop 2UV Transilluminator (UVP, LLC) on the 365 nm setting for 1 h on ice. The measured irradiance was ~5 mW/cm^2. Aliquots of all photolyzed samples were acetylated and desalted for CE using the Oligo Clean and Concentrator Kit (OCC; Zymo Research). Samples for PAGE were combined with 2x SDS loading buffer (Novex) + 1 % βME, and run on an 8-16 % PAA-gradient Mini-PROTEAN TGX gel (Bio-Rad). The gel was imaged on a MultiImager III (Alpha Innotech) for green fluorescence (5'FAM-labeled primer) and, after staining with Lumitein UV (Biotium), imaged for red fluorescence (total protein). Gel images were aligned and composited using Adobe Photoshop.

Two cycle experiment: A reaction containing TdT-dCTP conjugate and oligo P2 was performed and the reaction products were photolyzed. The DNA products were then purified by OCC and subjected to another extension reaction with TdT-dCTP, again followed by photolysis. Aliquots were taken after both extension reactions for PAGE and after both photocleavage reactions for PAGE and CE. **Control (pre-photolyzed conjugate) experiment:** TdT-dCTP conjugate was irradiated with 365 nm light for 1 h on ice to generate a stoichiometric mixture of unlinked MTdT*c*302(linker) + pa-dCTP. The photolysis products were then used in an extension reaction with oligo P2, and aliquots were taken for PAGE and CE.

4.4.5 Fast primer extension using all four TdT-dNTP conjugates

The conjugates used in this experiment were generated using 1 mM nucleotide in the MTdT*c*302-labeling reaction. Oligo P2 extension yield by 1.5 mg/mL (~16 µM) TdT-dNTP conjugates was measured at 8, 15, and 120 s. Reactions were performed in a 37 °C room by adding 4.5 µL of 2 mg/mL TdT-dNTP conjugate to 1.5 µL of 100 nM oligo P2 (final concentration: 25 nM), both in RBC. After rapid mixing, 4.5 µL of the reaction were quenched in 18 µL Quenching Solution (94 % Hi-Di Formamide with 10 mM EDTA) after 8 or 15 s. The remaining reaction volume was quenched with 6 µL Quenching Solution after 2 min. The samples were irradiated with 365 nm light on a Benchtop 2UV Transilluminator for 30 min. Photolysis products were acetylated using 100 mM NHS-acetate in 400 mM bicarbonate buffer. To desalt the samples for CE analysis, the oligos were captured onto DynaBeads M-280 StreptAvidin that were saturated with 5'-biotin-dA$_{60}$ oligo. Beads were washed with 1 x B&W buffer (5 mM Tris HCl pH 7.5, 0.5 mM EDTA, 1 M NaCl), 0.1 x B&W buffer, 0.01 x B&W buffer, and then eluted with 75 % formamide.

4.4.6 TdT-dNTP concentration-dependence of non-termination

The conjugates used in this experiment were generated using 1 mM nucleotide in the labeling reaction. An extension reaction was initiated using 50 nM oligo P2 with 0.25 mg/mL TdT-dCTP in RBC, and the reaction was split into three aliquots. After 1 min of incubation at 37 °C, sufficient time to form the +1 product complex with high yield, the first aliquot was quenched by the addition of EDTA to a final concentration of 100 mM, and the second aliquot was diluted 10-fold with RBC. The third aliquot was left as is. The latter two aliquots were quenched with EDTA after an additional 14 min of incubation. All three quenched reactions were photolyzed by 365 nm irradiation for 30 min using a Benchtop 2UV Transilluminator and then acetylated, purified by OCC, and analyzed by CE.

4.4.7 Effect of maleimide-labeling concentration on non-termination.

TdT-dTTP conjugates were generated using 0.2 mM, 0.4 mM, 1 mM and 2.5 mM nucleotide in the labeling reaction. Extension time courses of all conjugates at 0.53 mg/mL (~2.8 µM) were obtained from reactions in RBC using oligo P2 (100 nM) at 37 °C. 2 µL of the reactions were quenched in 8 µL Quenching Solution after 5 min. All samples were irradiated with 365 nm light on a Benchtop 2UV Transilluminator for 30 min. Photocleavage products were acetylated, purified by OCC, and analyzed by CE.

4.5 Section D: Synthesis of defined sequences using TdT-dNTP conjugates

4.5.1 Generation of the double stranded synthesis starter.

The double stranded DNA with a 3'-overhang used as initial substrate for the synthesis (starter) was prepared from a 359 bp PCR product derived from the pET19b plasmid. The PCR was performed using Phusion (Thermo Fisher Scientific) following the manufacturer's instructions and using primers C1 and C2 (see Table 4.5) (PCR program: 98 °C for 1 min, then 35 cycles of two step protocol: 98 °C for 10 s, 72 °C for 1 min). The PCR product was purified using the DNA Clean & Concentrator kit (DCC, Zymo Research) and digested with PstI, cutting the restriction site inserted by C1, to generate a 3'-overhang. The digested product was purified (DCC) and tailed with ddTTP to block the strand with the 3'-overhang from further incorporations (0.5 mM ddTTP, 1 U/µL TdT in RBC at 37 °C for 30 min). After tailing, the DNA was purified (DCC) and digested with BstXI to generate a 3'-overhang (5'-ATTT-3') for extensions by TdT-dNTP conjugates. The digestion product was separated from undigested DNA by 2 % TAE-agarose gel electrophoresis and gel-extracted using the Gel Recovery Kit (Zymo Research). A scheme for the preparation of the synthesis starter can be found in Figure 2.21.

4.5.2 Synthesis of 5'-CTACTGACTG-3'

Ten iterations of extension and deprotection of the starter were performed using TdT-dNTP conjugates prepared with 0.4 mM nucleotide in the labeling reaction. Extension reactions contained 1 mg/mL conjugate and were performed at 37 °C for 90 s in 1 x RBC. The reactions were then quenched by the addition of an equal volume of quenching buffer (250 mM EDTA, 500 mM NaCl). Photocleavage was performed for 20 min using a 365 nm LED (Thorlabs M365LP1-C1, operated by Thorlabs LEDD1B) filtered by a 365 nm bandpass filter with 10 nm FWHM (Chroma Technology ET365/10x, T387lp, E420lpv2). The measured irradiance was approximately 5 mW/cm^2. The first extension reaction contained ~40 nM of the initial substrate. After each cleavage step, the DNA products were purified using Zymo DCC columns according to the manufacturers protocol, and the recovered DNA was subjected to the next extension step. The following conjugates were used in the extension steps: 1) TdT-dCTP, 2) TdT-dTTP, 3) TdT-dATP, 4) TdT-dCTP, 5) TdT-dTTP, 6) TdT-dGTP, 7) TdT-dATP, 8) TdT-dCTP, 9) TdT-dTTP, 10) TdT-dGTP. The ten-cycle product was T-tailed using TdT and free dTTP + ddTTP at a ratio of 100:1 and acetylated using 20 mM NHS-acetat in bicarbonate buffer. The tailed product was then PCR-amplified using HotStart Taq (NEB) with primers C3 and C4 (see Table 4.5). PCR program: 98 °C for 2 min, 49 °C for 20 s, 68 °C for 12 min, then 35 cycles of: 98 °C for 30 s, 49 °C for 20 s, 68 °C for 30 s. The PCR product was ligated into pUC19 using EcoRI and HindIII sites that were introduced by the PCR primers. The plasmids were transformed into DH10B cells and single colonies were grown overnight in LB medium, miniprepped, and sequenced.

4.5.3 Testing of DCC-purification inhibition

The conjugates used in this experiment were generated using 1 mM nucleotide in the MTdT*c*302-labeling reaction. The DCC-eluate was generated by performing a DCC-purification following the manufacturers instructions and using 50 µL of ddH_2O as the input for the purification. Elution was performed in 8 µL ddH_2O. To test TdT inhibition, four reactions with a volume of 10 µL were set up, all containing 100 nM oligo P2, 1 x TP7 buffer (50 mM potassium acetate, 20 mM tris-acetate, pH 7), 5 mM $CoCl_2$ and 0.025 mg/mL TdT-dTTP. In addition, the 10 µL reactions contained either 4 µL of ddH_2O or 0.1x EB (positive controls), and two reactions contained 4 µL of the DCC-eluate. Reactions were performed at 37 °C and 2 µL were quenched in 8 µL Quenching Solution after 60 seconds. Next quenched samples were diluted 10-fold with 75 % formamide and analyzed by CE. The reaction buffer used in this experiment is different from RBC, but we believe that the findings are valid for both buffers, because we generally observed very similar behavior of conjugates in RBC and TP7+5 mM $CoCl_2$.

4.5.4 Synthesis of 5'-CTAGTCAGCT-3' and 5'-CCC-3'

Synthesis overview.

Nucleotide additions were performed using TdT-dNTP conjugates produced with a labeling concentration of 0.2 mM (dATP / dTTP) and 0.1 mM (dCTP, dGTP). The conjugates were used at 1 mg/mL in reactions in 1 x RBC at 37 °C. Extension reactions with TdT-dCTP, -dGTP and -dTTP were performed for 90 s, extensions with TdT-dATP for 180 s. Quenching of the reactions was performed by the addition of an equal volume of Quenching Buffer (100 mM $NaHCO_3$, 300 mM NaCl, 0.1 % TWEEN 20 (Sigma-Aldrich), 50 mM EDTA, 20 mM Sodium Azide, 20 mM NHS-acetate; NHS-acetate was added immediately before use). Photolysis was performed for 1 min using a 405 nm diode laser (500 mW, ToAuto M-33A405-500-G) filtered by a 405 nm bandpass filter with 10 nm FWHM (Thorlabs FBH405-10). After each cleavage step, the DNA products were purified using AMPure XP beads, and the recovered DNA was subjected to the next extension step. For the 10-mer synthesis, the following conjugates were used in the extension steps: 1) TdT-dCTP, 2) TdT-dTTP, 3) TdT-dATP, 4) TdT-dGTP, 5) TdT-dTTP, 6) TdT-dCTP, 7) TdT-dATP, 8) TdT-dGTP, 9) TdT-dCTP, 10) TdT-dTTP. To synthesize 5'-CCC-3', three cycles with TdT-dCTP were performed.

Detailed protocol of the extension cycles.

For the first step, 10 µL starter at ~30 nM was mixed with 2 µL Cofactor Mix (300 mM potassium acetate, 120 mM tris-acetate, 80 mM magnesium acetate and 2 mM cobalt chloride, pH 7.9). The 12 µL mixture was then added to 4 µL of TdT-dCTP at 4 mg/mL in TP8. The resulting 16 µL reaction in RBC was incubated for 90 s, before it was quenched by the addition of 16 µL Quenching Buffer. Afterwards, the reaction was photolyzed for 1 min using the 405 nm laser. A 3 min incubation at RT was performed to allow the acetylation reaction to proceed (NHS-acetate is a component of the Quenching

Buffer). Subsequently, 32 µL of FastAP Thermosensitive Alkaline Phosphatase (Thermo Fisher Scientific) at 0.32 U/µL in 40 mM Tris-HCl and 60 mM MgCl was added to the photolysis products to digest released dNTPs, and the reaction was incubated for 1 min at RT. The phosphatase treatment was performed because we found that there was a significant amount of dNTP carry-over during the AMPure XP cleanup, and that the dNTPs could be incorporated in the next reaction cycle, leading to insertions (see below, Amplification and next-generation sequencing analysis of synthesis products). Next, the DNA was purified using AMPure XP beads (Beckman Coulter): 115.2 µL AMPure XP beads were added to the 64 µL phosphatase reaction, and a binding step of 5 min was performed. The solution was then transferred into a well of a 96-well plate on a magnetic rack and incubated for 2 min for sedimentation of the beads. The liquid was removed and the beads were washed first with 400 µL of 70 % ethanol and then with 200 µL of 70 % ethanol. Subsequently, the beads were dried for 90 s, and the DNA was eluted with 10 µL of buffer EBT (1 mM Tris-HCl, 10 µM EDTA, 0.04 % TWEEN-20, pH 8.5). For the following cycles, the 10 µL purified synthesis product of the previous cycle were mixed with 2 µL cofactor mix, and the 12 µL mixture was added to 4 µL of the respective TdT-dNTP at 4 mg/mL in TP8. The resulting 16 µL reaction was then incubated for either 90 or 180 s, depending on the type of TdT-dNTP. The reaction was quenched, photolyzed, acetylated, phosphatase-treated and purified using AMPure beads in the same way as for the first synthesis step. The procedure was repeated until the complete sequence was synthesized.

Amplification and next-generation sequencing analysis of synthesis products.

10-cycle and 3-cycle synthesis products were A-tailed using 0.4 U/µL TdT (NEB) with 1 mM dATP in TdT Reaction Buffer for 30 min at 37 °C. The tailing products were purified by DCC and PCR-amplified using HotStart Taq (NEB) with primers C3 and C5 (See Table 4.5) according to the manufacturers instructions. PCR program: 98 °C for 2 min, 49 °C for 20 s, 68 °C for 12 min, then 35 cycles of: 98 °C for 30 s, 49 °C for 20 s, 68 °C for 30 s. Amplicons were purified by DCC and submitted to the JBEI DiVA DNA Sequencing Service for Nextera (Illumina) library preparation as described previously [93], multiplexed with other samples submitted by other users of the service. NGS was performed on a MiSeq (Illumina). Reads containing the sequence 5'-TCCAGATTT(N020)AAAAAA-3' were identified using a BioPython script, and non-singleton reads with a Q-score of at least 34 (error rate: ~1/2500 nt) for all bases in the target region were retained for analysis. Singleton reads accounted for 1.5 % of the data set and were excluded from analysis to avoid artefactual errors, e.g., due to index switching.

As mentioned in Section 2.4.3, early 10-mer synthesis attemts using AMPureXP beads showed that dNTPs can be retained in a manner that is resistent to washing. Therefore, some of the dNTPs that are released during photolysis of a quenched extension reaction are carried over into the next extension step, causing a characteristic type of (non-double) insertion error, (e.g., the G insertion in CTAGTCAGCGT observed in 0.27 % of reads, Figure 2.25). The effect was mitigated by a brief alkaline phosphatase treatment

following photolysis, but not completely eliminated. dNTP carryover-type insertions were definitively identified in 0.7 % of all reads total and were manually removed before estimation of stepwise yields.

Table 4.2: Overview of chemicals, laboratory materials and kits that were used during this thesis.

Chemicals and laboratory materials	Supplier
2-Mercaptoethanol	Sigma-Aldrich, St. Louis, MO, USA
2x Novex Tris-Glycine SDS Sample Buffer	Thermo Fisher Scientific, Waltham, MA, USA
5-aminoallyl-dUTP	Thermo Fisher Scientific, Waltham, MA, USA
Agarose	Merck Millipore, Billerica, MA, USA
AMPure XP	Beckman Coulter, Brea, CA, USA
Amicon, 30.000 kDa, 15 mL	Merck Millipore, Billerica, MA, USA
Amylose Resin	NEB, Ipswich, MA, USA
BP-22951	Broadpharm, San Diego, CA, USA
BP-23354	Broadpharm, San Diego, CA, USA
Carbenicillin disodium salt	Sigma-Aldrich, St. Louis, MO, USA
Cobalt(II) chloride ($CoCl_2$)	Sigma-Aldrich, St. Louis, MO, USA
Corning Spin-X UF, 30.000 kDa, 500 µL	Sigma-Aldrich, St. Louis, MO, USA
Dimethyl sulfoxide (DMSO)	Thermo Fisher Scientific, Waltham, MA, USA
DNA agarose gel recovery kit	Zymo Research, Irvine, CA, USA
DNA clean and concentrator kit (DCC)	Zymo Research, Irvine, CA, USA
DNA ladder, GeneRuler 1 kb	Thermo Fisher Scientific, Waltham, MA, USA
DNA loading dye, 6x	Thermo Fisher Scientific, Waltham, MA, USA
DNA size standard, GeneScan 600 LIZ	Thermo Fisher Scientific, Waltham, MA, USA
dNTPs	Thermo Fisher Scientific, Waltham, MA, USA
DynaBeads M-280 StreptAvidin	Thermo Fisher Scientific, Waltham, MA, USA
Ethanol, 99.5 %	VWR, Radnor, PA, USA

EDTA (Ethylenediamine-tetraacetic acid)	Sigma-Aldrich, St. Louis, MO, USA
GelRed	Biotium, Fremont, CA, USA
Glycerol	VWR, Radnor, PA, USA
Hi-Di Formamide	Thermo Fisher Scientific, Waltham, MA, USA
His Purification Miniprep Kit	Clonetech Laboratories, Mountain View, CA, USA
HisTrap FF	GE Healthcare, Chicago, Illinois, USA
HiTrap Q HP 5mL	GE Healthcare, Chicago, Illinois, USA
Hydrogen chloride (HCl), 37 %	Fisher Scientific, Hampton, NH, USA
Imidazole	Fisher Scientific, Hampton, NH, USA
Isopropyl-D-1-thioglactopyranoside (IPTG)	Sigma-Aldrich, St. Louis, MO, USA
LB agar plates with carbenicillin	Teknova, Hollister, CA, USA
LB broth, Miller	VWR, Radnor, PA, USA
LD650	Lumidyine Technologies, San Diego, CA, USA
Lumitine	Biotium, Fremont, CA, USA
Magnesium chloride ($MgCl_2$)	Thermo Fisher Scientific, Waltham, MA, USA
NHS-acetate	Broadpharm, San Diego, CA, USA
Ni-NTA agarose	Qiagen, Hilden, Germany
Nickel(II) chloride ($NiCl_2$)	Sigma-Aldrich, St. Louis, MO, USA
Oligo clean and concentrator kit (OCC)	Zymo Research, Irvine, CA, USA
PBS	Teknova, Hollister, CA, USA
PCR Clean-up Kit, DNA Clean	Concentrator-5 & Zymo Research, Irvine, CA, USA
PEG4 -SPDP (26128)	Thermo Fisher Scientific, Waltham, MA, USA
Pierce 0.8 mL centrifugation columns	Thermo Fisher Scientific, Waltham, MA, USA
Plasmid Miniprep kit, QiaPrep	Qiagen, Hilden, Germany
pMAL-c5X	NEB, Ipswich, MA, USA
Potassium acetate	Sigma-Aldrich, St. Louis, MO, USA

Potassium chloride (KCl)	Sigma-Aldrich, St. Louis, MO, USA
Potassium phosphate (KH_2PO_4)	VWR, Radnor, PA, USA
Protein ladder, PageRuler	Thermo Fisher Scientific, Waltham, MA, USA
Protein stain, GelCode Blue	Thermo Fisher Scientific, Waltham, MA, USA
Pur-A-Lyzer Dialysis Kit Mini 12000	Sigma-Aldrich, St. Louis, MO, USA
SDS loading dye, 2x Tris-Glycine SDS	Thermo Fisher Scientific, Waltham, MA, USA
SDS-PAGE, Mini-PROTEAN, 8-16 %	BioRad, Hercules, CA, USA
Set of ddNTPs	Thermo Fisher Scientific, Waltham, MA, USA
Set of pa-dNTPs	TriLink BioTechnologies, San Diego, CA, USA
Sodium chloride (NaCl)	VWR, Radnor, PA, USA
Sodium phosphate monobasic	VWR, Radnor, PA, USA
THPTA	Jena Bioscience, Jena, Germany
Tris acetate	Sigma-Aldrich, St. Louis, MO, USA
Tris acetate EDTA (TAE) buffer, 50x	BioRad, Hercules, CA, USA
Tris glycine SDS buffer, 10x	BioRad, Hercules, CA, USA
Tris(hydroxymethyl) aminomethane (Tris)	Sigma-Aldrich, St. Louis, MO, USA
Vivaspin 20, 10 kDa	Sartorius AG, Göttingen, Germany
Water, HPLC-grade	Honeywell, Morris Plains, NJ, USA

Table 4.3: Overview of enzymes and competent cells that were used during this thesis.

Enzymes and Cells	**Supplier**
E.coli DH10B, chemically competent	JBEI, homemade
DpnI	Thermo Fisher Scientific, Waltham, MA, USA
EcoRI	NEB, Ipswich, MA, USA
E.coli BL21 (DE3), chemically competent	NEB, Ipswich, MA, USA
FastAP Thermosensitive Alkaline Phosphatase	Thermo Fisher Scientific, Waltham, MA, USA
Gibson Assembly master mix	NEB, Ipswich, MA, USA
HindIII	NEB, Ipswich, MA, USA
Hot Start Taq DNA polymerase	NEB, Ipswich, MA, USA
Phusion Hot Start II DNA polymerase	Thermo Fisher Scientific, Waltham, MA, USA
E.coli Rosetta-gami B(DE3)pLysS chemically competent	MilliporeSigma (Novagen), Burlington, MA, USA
TdT (bovine)	NEB, Ipswich, MA, USA

Table 4.4: Overview of instruments that were used during this thesis.

Equipment	**Instruments**
Centrifuges	Eppendorf 5424 Beckman Coulter Allegra 25R Eppendorf 5810R
Electrophoresis systems	Thermo Scientific Owl EasyCast Bio-Rad Mini-PROTEAN Tetra
Gel imaging systems	UVP BioSpectrum Imaging System Alpha Innotech FluorChem Q MultiImage III UVP 2UV Transilluminator (LM-26)
Homogenizer	Emulsiflex C3 Avestin
Incubation shaker	Kuhner Shaker
pH meter	Thermo Scientific Orion 3-Star BenchTop pH Meter
Sonicator	QSonica Q700 Sonicator
Thermal cyclers	Applied Biosystems Veriti 96-well Thermal Cycler Bio-Rad CFX96 Touch Real-Time PCR Detection System
UV-Vis measurement systems	NanoDrop ND-1000 & 2000

Table 4.5: Oligonucleotide primers that were used during this thesis. All oligos were purchased from Integrated DNA Technologies (IDT). Primers P1, P2 and C1 were ordered with HPLC purification, and the remaining primers were ordered with standard desalting.

Name	Sequence (5' to 3')	Note
P1	/56-FAM/TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT	5' fluorescein dT_{35}
P2	/56-FAM/TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT	5' fluorescein dT_{60}
P3	/5AmMC6/CAACACACCACCCACCCAAC	5'-amine for NHS-labeling
C1	/5Phos/GCAGCCAACTCAGCTTCTGCAGGGGCTTTGTTAGCAGCCGGATCCTC	PstI site, 5'-phosphorylated
C2	AAACAAGCGCTCATGAGCCAGAAATCTGGAGCCCGATCTTCCCCATCGG	BstX1 site
C3	GTGCCGTGAGACCTGGCTCCTGACGAGGAtaagcttCTATAGTGAGTCGTATTAATTTCG	HindIII site
C4	AAAAgaattcAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA	EcoRI site
C5	TTTTgaattcTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT	EcoRI site

Sequences

TdT*wt* (His-TdT, wildtype)

```
  1 MGHHHHHHHH HHSSGHIDDD DKHMSPSPVP GSQNVPAPAV KKISQYACQR RTTLNNYNQL
 61 FTDALDILAE NDELRENEGS CLAFMRASSV LKSLPFPITS MKDTEGIPCL GDKVKSIIEG
121 IIEDGESSEA KAVLNDERYK SFKLFTSVFG VGLKTAEKWF RMGFRTLSKI QSDKSLRFTQ
181 MQKAGFLYYE DLVSCVNRPE AEAVSMLVKE AVVTFLPDAL VTMTGGFRRG KMTGHDVDFL
241 ITSPEATEDE EQQLLHKVTD FWKQQGLLLY CDILESTFEK FKQPSRKVDA LDHFQKCFLI
301 LKLDHGRVHS EKSGQQEGKG WKAIRVDLVM CPYDRRAFAL LGWTGSRQFE RDLRRYATHE
361 RKMMLDNHAL YDRTKRVFLE AESEEEIFAH LGLDYIEPWE RNA
```

TdTΔ5cys (His-TdT, surface cysteines removed)

```
  1 MGHHHHHHHH HHSSGHIDDD DKHMSPSPVP GSQNVPAPAV KKISQYACQR RTTLNNYNQL
 61 FTDALDILAE NDELRENEGS ALAFMRASSV LKSLPFPITS MKDTEGIPSL GDKVKSIIEG
121 IIEDGESSEA KAVLNDERYK SFKLFTSVFG VGLKTAEKWF RMGFRTLSKI QSDKSLRFTQ
181 MQKAGFLYYE DLVSAVNRPE AEAVSMLVKE AVVTFLPDAL VTMTGGFRRG KMTGHDVDFL
241 ITSPEATEDE EQQLLHKVTD FWKQQGLLLY ADILESTFEK FKQPSRKVDA LDHFQKCFLI
301 LKLDHGRVHS EKSGQQEGKG WKAIRVDLVM SPYDRRAFAL LGWTGSRQFE RDLRRYATHE
361 RKMMLDNHAL YDRTKRVFLE AESEEEIFAH LGLDYIEPWE RNA
```

TdT*c*180 (His-TdT, single attachment Cys 180)

```
  1 MGHHHHHHHH HHSSGHIDDD DKHMSPSPVP GSQNVPAPAV KKISQYACQR RTTLNNYNQL
 61 FTDALDILAE NDCLRENEGS ALAFMRASSV LKSLPFPITS MKDTEGIPSL GDKVKSIIEG
121 IIEDGESSEA KAVLNDERYK SFKLFTSVFG VGLKTAEKWF RMGFRTLSKI QSDKSLRFTQ
181 MQKAGFLYYE DLVSAVNRPE AEAVSMLVKE AVVTFLPDAL VTMTGGFRRG KMTGHDVDFL
241 ITSPEATEDE EQQLLHKVTD FWKQQGLLLY ADILESTFEK FKQPSRKVDA LDHFQKCFLI
301 LKLDHGRVHS EKSGQQEGKG WKAIRVDLVM SPYDRRAFAL LGWTGSRQFE RDLRRYATHE
361 RKMMLDNHAL YDRTKRVFLE AESEEEIFAH LGLDYIEPWE RNA
```

TdT*c*188 (His-TdT, single attachment Cys 188)

```
  1 MGHHHHHHHH HHSSGHIDDD DKHMSPSPVP GSQNVPAPAV KKISQYACQR RTTLNNYNQL
 61 FTDALDILAE NDELRENEGS CLAFMRASSV LKSLPFPITS MKDTEGIPSL GDKVKSIIEG
121 IIEDGESSEA KAVLNDERYK SFKLFTSVFG VGLKTAEKWF RMGFRTLSKI QSDKSLRFTQ
181 MQKAGFLYYE DLVSAVNRPE AEAVSMLVKE AVVTFLPDAL VTMTGGFRRG KMTGHDVDFL
241 ITSPEATEDE EQQLLHKVTD FWKQQGLLLY ADILESTFEK FKQPSRKVDA LDHFQKCFLI
301 LKLDHGRVHS EKSGQQEGKG WKAIRVDLVM SPYDRRAFAL LGWTGSRQFE RDLRRYATHE
361 RKMMLDNHAL YDRTKRVFLE AESEEEIFAH LGLDYIEPWE RNA
```

TdT*c*253 (His-TdT, single attachment Cys 253)

```
  1 MGHHHHHHHH HHSSGHIDDD DKHMSPSPVP GSQNVPAPAV KKISQYACQR RTTLNNYNQL
 61 FTDALDILAE NDELRENEGS ALAFMRASSV LKSLPFPITS MKDTEGIPSL GDKVKSIIEG
121 IIEDGESSEA KAVLNDERYK SFKLFCSVFG VGLKTAEKWF RMGFRTLSKI QSDKSLRFTQ
181 MQKAGFLYYE DLVSAVNRPE AEAVSMLVKE AVVTFLPDAL VTMTGGFRRG KMTGHDVDFL
241 ITSPEATEDE EQQLLHKVTD FWKQQGLLLY ADILESTFEK FKQPSRKVDA LDHFQKCFLI
301 LKLDHGRVHS EKSGQQEGKG WKAIRVDLVM SPYDRRAFAL LGWTGSRQFE RDLRRYATHE
361 RKMMLDNHAL YDRTKRVFLE AESEEEIFAH LGLDYIEPWE RNA
```

TdT*c*302 (His-TdT, single attachment Cys 302)

```
  1 MGHHHHHHHH HHSSGHIDDD DKHMSPSPVP GSQNVPAPAV KKISQYACQR RTTLNNYNQL
 61 FTDALDILAE NDELRENEGS ALAFMRASSV LKSLPFPITS MKDTEGIPSL GDKVKSIIEG
121 IIEDGESSEA KAVLNDERYK SFKLFTSVFG VGLKTAEKWF RMGFRTLSKI QSDKSLRFTQ
181 MQKAGFLYYE DLVSCVNRPE AEAVSMLVKE AVVTFLPDAL VTMTGGFRRG KMTGHDVDFL
241 ITSPEATEDE EQQLLHKVTD FWKQQGLLLY ADILESTFEK FKQPSRKVDA LDHFQKCFLI
301 LKLDHGRVHS EKSGQQEGKG WKAIRVDLVM SPYDRRAFAL LGWTGSRQFE RDLRRYATHE
361 RKMMLDNHAL YDRTKRVFLE AESEEEIFAH LGLDYIEPWE RNA
```

MTdT*c*302 (His-MPB-TdT, single attachment Cys 302)

```
  1 MGHHHHHHHH HHSSGHIDDD DKHMMKIEEG KLVIWINGDK GYNGLAEVGK KFEKDTGIKV
 61 TVEHPDKLEE KFPQVAATGD GPDIIFWAHD RFGGYAQSGL LAEITPDKAF QDKLYPFTWD
121 AVRYNGKLIA YPIAVEALSL IYNKDLLPNP PKTWEEIPAL DKELKAKGKS ALMFNLQEPY
181 FTWPLIAADG GYAFKYENGK YDIKDVGVDN AGAKAGLTFL VDLIKNKHMN ADTDYSIAEA
241 AFNKGETAMT INGPWAWSNI DTSKVNYGVT VLPTFKGQPS KPFVGVLSAG INAASPNKEL
301 AKEFLENYLL TDEGLEAVNK DKPLGAVALK SYEEELVKDP RIAATMENAQ KGEIMPNIPQ
361 MSAFWYAVRT AVINAASGRQ TVDEALKDAQ TNSSSNNNNN NNNNNLGIEG RISHMSMGGR
421 DIVDGSEFSP SPVPGSQNVP APAVKKISQY ACQRRTTLNN YNQLFTDALD ILAENDELRE
481 NEGSALAFMR ASSVLKSLPF PITSMKDTEG IPSLGDKVKS IIEGIIEDGE SSEAKAVLND
541 ERYKSFKLFT SVFGVGLKTA EKWFRMGFRT LSKIQSDKSL RFTQMQKAGF LYYEDLVSCV
601 NRPEAEAVSM LVKEAVVTFL PDALVTMTGG FRRGKMTGHD VDFLITSPEA TEDEEQQLLH
661 KVTDFWKQQG LLLYADILES TFEKFKQPSR KVDALDHFQK CFLILKLDHG RVHSEKSGQQ
721 EGKGWKAIRV DLVMSPYDRR AFALLGWTGS RQFERDLRRY ATHERKMMLD NHALYDRTKR
781 VFLEAESEEE IFAHLGLDYI EPWERNA
```

List of Abbrevations

aa-dUTP amino-allyl deoxyuridine triphosphate

bp base pairs

CE capillary electrophoresis

dATP deoxyadenosine triphosphate

dCTP deoxycytidine triphosphate

ddH_2O double-destilled H_2O

ddNTP dideoxynucleoside triphosphate

dGTP deoxyguanosine triphosphate

dNTP deoxynucleoside triphosphate

dTTP deoxythymidine triphosphate

EDTA ethylendiaminetetraacetic acid

FPLC fast protein liquid chromatography

IMAC immobilized metal ion chromatography

IPTG isopropyl β-D-1-thiogalactopyranoside

kbp kilo base pairs

kDa kilodalton

MSA multiple sequence alignment

NCBI National Center for Biotechnology Information

NGS next generation sequencing

NHS N-hydroxysuccinimide

nt nucleotide(s)

NTP nucleoside triphosphate

oligo oliconucleotide

OPSS orthopyridyl disulfide

pa-dNTP propargyl-amino deoxynucleoside triphosphate

PAGE polyacrylamide gel electrophoresis

PCA polymerase chain assembly

PCR polymerase chain reaction

RCF relative centrifugal force

RTdNTP reversible terminator deoxynucleoside triphosphate

SBS sequencing by synthesis

TdT terminal deoxynucleotidyl transferase

wt wild type

List of Figures

List of Tables

Bibliography

[1] Abel GR, Calabrese ZA, Ayco J, Hein JE, Ye T (**2016**) Measuring and Suppressing the Oxidative Damage to DNA During Cu(I)-Catalyzed AzideAlkyne Cycloaddition. *Bioconjugate Chemistry* 27(3):698–704

[2] Adessi C, Matton G, Ayala G, Turcatti G, Mermod JJ, Mayer P, Kawashima E (**2000**) Solid phase DNA amplification: characterisation of primer attachment and amplification mechanisms. *Nucleic Acids Research* 28(20):E87

[3] Agasti SS, Liong M, Peterson VM, Lee H, Weissleder R (**2012**) Photocleavable DNA BarcodeAntibody Conjugates Allow Sensitive and Multiplexed Protein Analysis in Single Cells. *Journal of the American Chemical Society* 134(45):18499–18502

[4] Arlow DH, Palluk S (**2017**), Nucleic acid synthesis and sequencing using tethered nucleoside triphosphates. US Patent Application WO2017223517A1. Available from: `https://patents.google.com/patent/WO2017223517A1/en`

[5] Barrangou R, Fremaux C, Deveau H, Richards M, Boyaval P, Moineau S, Romero DA, Horvath P (**2007**) CRISPR Provides Acquired Resistance Against Viruses in Prokaryotes. *Science* 315(5819):1709–1712

[6] Beaucage S, Caruthers M (**1981**) Deoxynucleoside phosphoramidites - A new class of key intermediates for deoxypolynucleotide synthesis. *Tetrahedron Letters* 22(20):1859–1862

[7] Bentley DR, Balasubramanian S, Swerdlow HP, Smith GP, Milton J, Brown CG, Hall KP, Evers DJ, Barnes CL, Bignell HR, Boutell JM, Bryant J, Carter RJ, Keira Cheetham R, Cox AJ, Ellis DJ, Flatbush MR, Gormley NA, Humphray SJ, Irving LJ, Karbelashvili MS, Kirk SM, Li H, Liu X, Maisinger KS, Murray LJ, Obradovic B, Ost T, Parkinson ML, Pratt MR, Rasolonjatovo IMJ, Reed MT, Rigatti R, Rodighiero C, Ross MT, Sabot A, Sankar SV, Scally A, Schroth GP, Smith ME, Smith VP, Spiridou A, Torrance PE, Tzonev SS, Vermaas EH, Walter K, Wu X, Zhang L, Alam MD, Anastasi C, Aniebo IC, Bailey DMD, Bancarz IR, Banerjee S, Barbour SG, Baybayan PA, Benoit VA, Benson KF, Bevis C, Black PJ, Boodhun A, Brennan JS, Bridgham JA, Brown RC, Brown AA, Buermann DH, Bundu AA, Burrows JC, Carter NP, Castillo N, Chiara E Catenazzi M, Chang S, Neil Cooley R, Crake NR, Dada OO, Diakoumakos KD, Dominguez-Fernandez B, Earnshaw DJ, Egbujor UC, Elmore DW, Etchin SS, Ewan MR, Fedurco M, Fraser LJ, Fuentes Fajardo KV, Scott Furey W, George D, Gietzen KJ, Goddard CP, Golda GS, Granieri PA, Green DE, Gustafson DL, Hansen NF, Harnish K,

Haudenschild CD, Heyer NI, Hims MM, Ho JT, Horgan AM, Hoschler K, Hurwitz S, Ivanov DV, Johnson MQ, James T, Huw Jones TA, Kang GD, Kerelska TH, Kersey AD, Khrebtukova I, Kindwall AP, Kingsbury Z, Kokko-Gonzales PI, Kumar A, Laurent MA, Lawley CT, Lee SE, Lee X, Liao AK, Loch JA, Lok M, Luo S, Mammen RM, Martin JW, McCauley PG, McNitt P, Mehta P, Moon KW, Mullens JW, Newington T, Ning Z, Ling Ng B, Novo SM, O'Neill MJ, Osborne MA, Osnowski A, Ostadan O, Paraschos LL, Pickering L, Pike AC, Pike AC, Chris Pinkard D, Pliskin DP, Podhasky J, Quijano VJ, Raczy C, Rae VH, Rawlings SR, Chiva Rodriguez A, Roe PM, Rogers J, Rogert Bacigalupo MC, Romanov N, Romieu A, Roth RK, Rourke NJ, Ruediger ST, Rusman E, Sanches-Kuiper RM, Schenker MR, Seoane JM, Shaw RJ, Shiver MK, Short SW, Sizto NL, Sluis JP, Smith MA, Ernest Sohna Sohna J, Spence EJ, Stevens K, Sutton N, Szajkowski L, Tregidgo CL, Turcatti G, VandeVondele S, Verhovsky Y, Virk SM, Wakelin S, Walcott GC, Wang J, Worsley GJ, Yan J, Yau L, Zuerlein M, Rogers J, Mullikin JC, Hurles ME, McCooke NJ, West JS, Oaks FL, Lundberg PL, Klenerman D, Durbin R, Smith AJ (**2008**) Accurate whole human genome sequencing using reversible terminator chemistry. *Nature* 456(7218):53–59

[8] Bolllum FJ (**1962**) Oligodeoxyribonucleotide-primed reactions catalyzed by calf thymus polymerase. *The Journal of Biological Chemistry* 237:1945–9

[9] Boulé JB, Johnson E, Rougeon F, Papanicolaou C (**1998**) High-level expression of murine terminal deoxynucleotidyl transferase in Escherichia coli grown at low temperature and overexpressing argU tRNA. *Molecular Biotechnology* 10(3):199–208

[10] Boulé JB, Rougeon F, Papanicolaou C (**2001**) Terminal Deoxynucleotidyl Transferase Indiscriminately Incorporates Ribonucleotides and Deoxyribonucleotides. *Journal of Biological Chemistry* 276(33):31388–31393

[11] Caruthers MH (**2011**) A brief review of DNA and RNA chemical synthesis. *Biochemical Society Transactions* 39(2):575–580

[12] Caruthers MH (**2013**) The Chemical Synthesis of DNA/RNA: Our Gift to Science. *Journal of Biological Chemistry* 288(2):1420–1427

[13] Chang LM, Bollum FJ (**1986**) Molecular biology of terminal transferase. *CRC Critical Reviews in Biochemistry* 21(1):27–52

[14] Chen CY (**2014**) DNA polymerases drive DNA sequencing-by-synthesis technologies: Both past and present. *Frontiers in Microbiology* 5(JUN):305

[15] Chen F, Dong M, Ge M, Zhu L, Ren L, Liu G, Mu R (**2013**) The History and Advances of Reversible Terminators Used in New Generations of Sequencing Technology. *Genomics, Proteomics & Bioinformatics* 11(1):34–40

[16] Chen F, Gaucher EA, Leal NA, Hutter D, Havemann SA, Govindarajan S, Ortlund EA, Benner SA (**2010**) Reconstructed evolutionary adaptive paths give polymerases accepting reversible terminators for sequencing and SNP detection. *Proceedings of the National Academy of Sciences* 107(5):1948–1953

[17] Church GM, Gao Y, Kosuri S (**2012**) Next-Generation Digital Information Storage in DNA. *Science* 337(6102):1628–1628

[18] Cline J, Braman JC, Hogrefe HH (**1996**) PCR fidelity of pfu DNA polymerase and other thermostable DNA polymerases. *Nucleic Acids Research* 24(18):3546–3551

[19] Clore A (**2018**) A new route to synthetic DNA. *Nature Biotechnology* 36(7):593–595

[20] Czar MJ, Anderson JC, Bader JS, Peccoud J (**2009**) Gene synthesis demystified. *Trends in Biotechnology* 27(2):63–72

[21] Deibel MR, Coleman MS (**1980**) Biochemical properties of purified human terminal deoxynucleotidyltransferase. *The Journal of Biological Chemistry* 255(9):4206–12

[22] Delarue M, Boulé JB, Lescar J, Expert-Bezançon N, Jourdan N, Sukumar N, Rougeon F, Papanicolaou C (**2002**) Crystal structures of a template-independent DNA polymerase: murine terminal deoxynucleotidyltransferase. *The EMBO Journal* 21(3):427–39

[23] Efcavitch WJ, Siddiqi S (**2013**), Methods and apparatus for synthesizing nucleic acids. US Patent Application US8808989B1. Available from: `https://patents.google.com/patent/US8808989B1/en`

[24] Fowler JD, Suo Z (**2006**) Biochemical, Structural, and Physiological Characterization of Terminal Deoxynucleotidyl Transferase. *Chemical Reviews* 106(6):2092–2110

[25] Gardner AF, Wang J, Wu W, Karouby J, Li H, Stupi BP, Jack WE, Hersh MN, Metzker ML (**2012**) Rapid incorporation kinetics and improved fidelity of a novel class of 3'-OH unblocked reversible terminators. *Nucleic Acids Research* 40(15):7404–7415

[26] Gibson DG, Glass JI, Lartigue C, Noskov VN, Chuang RY, Algire MA, Benders GA, Montague MG, Ma L, Moodie MM, Merryman C, Vashee S, Krishnakumar R, Assad-Garcia N, Andrews-Pfannkoch C, Denisova EA, Young L, Qi ZQ, Segall-Shapiro TH, Calvey CH, Parmar PP, Hutchison CA, Smith HO, Venter JC (**2010**) Creation of a Bacterial Cell Controlled by a Chemically Synthesized Genome. *Science* 329(5987):52–56

[27] Gibson DG, Young L, Chuang RY, Venter JC, Hutchison CA, Smith HO (**2009**) Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nature Methods* 6(5):343–345

[28] Gilfillan S, Benoist C, Mathis D (**1995**) Mice lacking terminal deoxynucleotidyl transferase: adult mice with a fetal antigen receptor repertoire. *Immunological reviews* 148:201–19

[29] Goldman N, Bertone P, Chen S, Dessimoz C, LeProust EM, Sipos B, Birney E (**2013**) Towards practical, high-capacity, low-maintenance information storage in synthesized DNA. *Nature* 494(7435):77–80

[30] Gouge J, Rosario S, Romain F, Beguin P, Delarue M (**2013**) Structures of intermediates along the catalytic cycle of terminal deoxynucleotidyltransferase: Dynamical aspects of the two-metal ion mechanism. *Journal of Molecular Biology* 425(22):4334–4352

[31] Hanwell MD, Curtis DE, Lonie DC, Vandermeersch T, Zurek E, Hutchison GR (**2012**) Avogadro: an advanced semantic chemical editor, visualization, and analysis platform. *Journal of Cheminformatics* 4(1):17

[32] He P, Yang W (**2018**) Template and primer requirements for DNA Pol θ-mediated end joining. *Proceedings of the National Academy of Sciences* 115(30):7747–7752

[33] Hein JE, Fokin VV (**2010**) Copper-catalyzed azide-alkyne cycloaddition (CuAAC) and beyond: new reactivity of copper(I) acetylides. *Chemical Society Reviews* 39(4):1302–15

[34] Ho SN, Hunt HD, Horton RM, Pullen JK, Pease LR (**1989**) Site-directed mutagenesis by overlap extension using the polymerase chain reaction. *Gene* 77(1):51–9

[35] Hogrefe RI, Midthune B, Lebedev A (**2013**) Current Challenges in Nucleic Acid Synthesis. *Israel Journal of Chemistry* 53(6-7):326–349

[36] Hottin A, Betz K, Diederichs K, Marx A (**2017**) Structural Basis for the KlenTaq DNA Polymerase Catalysed Incorporation of Alkene- versus Alkyne-Modified Nucleotides. *Chemistry - A European Journal* 23(9):2109–2118

[37] Humphrey W, Dalke A, Schulten K (**1996**) VMD: visual molecular dynamics. *Journal of Molecular Graphics* 14(1):33–8, 27–8

[38] Hutter D, Kim MJ, Karalkar N, Leal NA, Chen F, Guggenheim E, Visalakshi V, Olejnik J, Gordon S, Benner SA (**2010**) Labeled Nucleoside Triphosphates with Reversibly Terminating Aminoalkoxyl Groups. *Nucleosides, Nucleotides and Nucleic Acids* 29(11-12):879–895

[39] Jaynes J, Yang M, Espinoza N, Dodds J (**1986**) Plant protein improvement by genetic engineering: use of synthetic genes. *Trends in Biotechnology* 4(12):314–320

[40] Jensen MA, Davis RW (**2018**) Template-Independent Enzymatic Oligonucleotide Synthesis (TiEOS): Its History, Prospects, and Challenges. *Biochemistry* 57(12):1821–1832

[41] Johnson D, Morgan AR (**1976**) The isolation of a high molecular weight terminal deoxynucleotidyl transferase from calf thymus. *Biochemical and Biophysical Research Communications* 72(3):840–9

[42] Joyce CM, Steitz TA (**1994**) Function and Structure Relationships in DNA Polymerases. *Annual Review of Biochemistry* 63(1):777–822

[43] Juárez R, Ruiz JF, Nick McElhinny SA, Ramsden D, Blanco L (**2006**) A specific loop in human DNA polymerase mu allows switching between creative and DNA-instructed synthesis. *Nucleic acids research* 34(16):4572–82

[44] Keasling JD (**2010**) Manufacturing Molecules Through Metabolic Engineering. *Science* 330(6009):1355–1358

[45] Keasling JD (**2012**) Synthetic biology and the development of tools for metabolic engineering. *Metabolic Engineering* 14(3):189–195

[46] Kent T, Mateos-Gomez PA, Sfeir A, Pomerantz RT (**2016**) Polymerase θ is a robust terminal transferase that oscillates between three different mechanisms during end-joining. *eLife* 5

[47] Kepler TB, Borrero M, Rugerio B, McCray SK, Clarke SH (**1996**) Interdependence of N nucleotide addition and recombination site choice in V(D)J rearrangement. *Journal of Immunology (Baltimore, Md : 1950)* 157(10):4451–7

[48] Khorana H (**1968**) Nobel lecture. Nucleic acid synthesis in the study of the genetic code. *Nobel Lectures: Physiology or Medicine (1963-1970)* 350–356

[49] Khorana HG, Agarwal KL, Büchi H, Caruthers MH, Gupta NK, Kleppe K, Kumar A, Otsuka E, Raj Bhandary UL, Van de Sande JH, Sgaramella V, Terao T, Weber H, Yamada T (**1972**) Studies on polynucleotides. 103. Total synthesis of the structural gene for an alanine transfer ribonucleic acid from yeast. *Journal of Molecular Biology* 72(2):209–17

[50] Knapp DC, Serva S, D'Onofrio J, Keller A, Lubys A, Kurg A, Remm M, Engels JW (**2011**) Fluoride-Cleavable, Fluorescently Labelled Reversible Terminators: Synthesis and Use in Primer Extension. *Chemistry - A European Journal* 17(10):2903–2915

[51] Köster H, Biernat J, McManus J, Wolter A, Stumpe A, Narang C, Sinha N (**1984**) Polymer support oligonucleotide synthesis-XV: Synthesis of oligodeoxynucleotides on controlled pore glass (CPG) using phosphate and a new phosphite triester approach. *Tetrahedron* 40(1):103–112

[52] Kosuri S, Church GM (**2014**) Large-scale de novo DNA synthesis: technologies and applications. *Nature Methods* 11(5):499–507

[53] Krayevsky AA, Chernov DN (**1996**) Should the Asymmetry of Enzymatic Active Centers Always Correlate with the Asymmetry of their Substrates? *Journal of Biomolecular Structure and Dynamics* 14(2):225–230

[54] Krayevsky AA, Victorova LS, Arzumanov AA, Jasko MV (**2000**) Terminal deoxynucleotidyl transferase: catalysis of DNA (oligodeoxynucleotide) phosphorylation. *Pharmacology & Therapeutics* 85(3):165–173

[55] Kuan WL, Joy J, Mee NF, Perlyn KZ, Wen TS, Nguen T, James J, Chai E, Flotow H, Crasta S, Chua K, Peng NS, Hill J (**2010**) Generation of Active Bovine Terminal Deoxynucleotidyl Transferase (TdT) in E.coli. *Biochemistry Insights* 3:BCI.S5123

[56] Kyrylkova K, Kyryachenko S, Leid M, Kioussi C (**2012**) Detection of Apoptosis by TUNEL Assay. In *Methods in molecular biology (Clifton, N.J.)*, vol. 887, 41–47

[57] Lamers MH, Georgescu RE, Lee SG, O'Donnell M, Kuriyan J (**2006**) Crystal Structure of the Catalytic α Subunit of E. coli Replicative DNA Polymerase III. *Cell* 126(5):881–892

[58] Lebendiker M, Danieli T (**2017**) Purification of proteins fused to maltose-binding protein. In *Methods in Molecular Biology*, vol. 1485, 257–273

[59] Lefler CF, Bollum FJ (**1969**) Deoxynucleotide-polymerizing enzymes of calf thymus gland. 3. Preparation of poly N-acetyldeoxyguanylate and polydeoxyguanylate. *The Journal of Biological Chemistry* 244(4):594–601

[60] LeProust EM, Peck BJ, Spirin K, McCuen HB, Moore B, Namsaraev E, Caruthers MH (**2010**) Synthesis of high-quality libraries of long (150mer) oligonucleotides by a novel depurination controlled process. *Nucleic Acids Research* 38(8):2522–2540

[61] Litosh VA, Wu W, Stupi BP, Wang J, Morris SE, Hersh MN, Metzker ML (**2011**) Improved nucleotide selectivity and termination of 3-OH unblocked reversible terminators by molecular tuning of 2-nitrobenzyl alkylated HOMedU triphosphates. *Nucleic Acids Research* 39(6):e39–e39

[62] Ma S, Saaem I, Tian J (**2012**) Error correction in gene synthesis technology. *Trends in Biotechnology* 30(3):147–154

[63] Martínez-Martínez M, Coscolín C, Santiago G, Chow J, Stogios PJ, Bargiela R, Gertler C, Navarro-Fernández J, Bollinger A, Thies S, Méndez-García C, Popovic A, Brown G, Chernikova TN, García-Moyano A, Bjerga GE, Pérez-García P, Hai T, Del Pozo MV, Stokke R, Steen IH, Cui H, Xu X, Nocek BP, Alcaide M, Distaso M, Mesa V, Peláez AI, Sánchez J, Buchholz PC, Pleiss J, Fernández-Guerra A, Glöckner FO, Golyshina OV, Yakimov MM, Savchenko A, Jaeger KE, Yakunin AF,

Streit WR, Golyshin PN, Guallar V, Ferrer M (**2018**) Determinants and Prediction of Esterase Substrate Promiscuity Patterns. *ACS Chemical Biology* 13(1):225–234

[64] Mathews AS, Yang H, Montemagno C (**2016**) Photo-cleavable nucleotides for primer free enzyme mediated DNA synthesis. *Organic & Biomolecular Chemistry* 14(35):8278–8288

[65] Mathews AS, Yang H, Montemagno C (**2017**) 3-O-Caged 2-Deoxynucleoside Triphosphates for Light-Mediated, Enzyme-Catalyzed, Template-Independent DNA Synthesis. In *Current Protocols in Nucleic Acid Chemistry*, vol. 71, 13.17.1–13.17.38, John Wiley & Sons, Inc., Hoboken, NJ, USA

[66] Maxam AM, Gilbert W (**1977**) A new method for sequencing DNA. *Proceedings of the National Academy of Sciences of the United States of America* 74(2):560–4

[67] Metzker ML (**2010**) Sequencing technologies the next generation. *Nature Reviews Genetics* 11(1):31–46

[68] Meyer M, Kircher M (**2010**) Illumina sequencing library preparation for highly multiplexed target capture and sequencing. *Cold Spring Harbor Protocols* 2010(6):pdb.prot5448

[69] Michelson AM, Todd AR (**1955**) Nucleotides part XXXII. Synthesis of a dithymidine dinucleotide containing a 3: 5-internucleotidic linkage. *Journal of the Chemical Society* 0(0):2632–2638

[70] Motea EA, Berdis AJ (**2010**) Terminal deoxynucleotidyl transferase: The story of a misguided DNA polymerase. *Biochimica et Biophysica Acta (BBA) - Proteins and Proteomics* 1804(5):1151–1166

[71] Mullis K, Faloona F, Scharf S, Saiki R, Horn G, Erlich H (**1986**) Specific enzymatic amplification of DNA in vitro: the polymerase chain reaction. *Cold Spring Harbor Symposia on Quantitative Biology* 51 Pt 1:263–73

[72] Nick McElhinny SA, Ramsden DA (**2003**) Polymerase mu is a DNA-directed DNA/RNA polymerase. *Molecular and Cellular Biology* 23(7):2309–15

[73] Palluk S (**2015**) In silico Design of Murine Terminal Variants to Increase the Tolerance for 3'-O-Modified Nucleoside Triphosphates. *Masterthesis Biomolecular Engineering, TU Darmstadt*

[74] Palluk S, Arlow DH, de Rond T, Barthel S, Kang JS, Bector R, Baghdassarian HM, Truong AN, Kim PW, Singh AK, Hillson NJ, Keasling JD (**2018**) De novo DNA synthesis using polymerase-nucleotide conjugates. *Nature Biotechnology* 36(7):645–650

[75] Picard V, Ersdal-Badju E, Lu A, Bock SC (**1994**) A rapid and efficient one-tube PCR-based mutagenesis technique using Pfu DNA polymerase. *Nucleic Acids Research* 22(13):2587–2591

[76] Pretorius IS, Boeke JD (**2018**) Yeast 2.0 - connecting the dots in the construction of the world's first functional synthetic eukaryotic genome. *FEMS Yeast Research* 18(4)

[77] Rattray AJ, Strathern JN (**2003**) Error-Prone DNA Polymerases: When Making a Mistake is the Only Way to Get Ahead. *Annual Review of Genetics* 37(1):31–66

[78] Ravikumar VT, Andrade M, Wyrzykiewicz TK, Scozzari A, Cole DL (**1995**) Nucleosides, Nucleotides & Nucleic Acids Large-Scale Synthesis of Oligodeoxyribonucleotide Phosphorothioate Using Controlled-Pore Glass as Support. *Nucleosides, Nucleotides and Nucleic Acids* 0732–8311

[79] Reese CB (**2005**) Oligo- and poly-nucleotides: 50 years of chemical synthesis. *Organic & Biomolecular Chemistry* 3(21):3851–68

[80] Richardson SM, Mitchell LA, Stracquadanio G, Yang K, Dymond JS, DiCarlo JE, Lee D, Huang CLV, Chandrasegaran S, Cai Y, Boeke JD, Bader JS (**2017**) Design of a synthetic yeast genome. *Science* 355(6329):1040–1044

[81] Ruparel H, Bi L, Li Z, Bai X, Kim DH, Turro NJ, Ju J (**2005**) Design and synthesis of a 3'-O-allyl photocleavable fluorescent nucleotide as a reversible terminator for DNA sequencing by synthesis. *Proceedings of the National Academy of Sciences* 102(17):5932–5937

[82] Sander JD, Joung JK (**2014**) CRISPR-Cas systems for editing, regulating and targeting genomes. *Nature Biotechnology* 32(4):347–355

[83] Sanger F, Nicklen S, Coulson AR (**1977**) DNA sequencing with chain-terminating inhibitors. *Proceedings of the National Academy of Sciences of the United States of America* 74(12):5463–7

[84] Schrödinger, LLC (**2015**) The PyMOL molecular graphics system, version 1.8

[85] Seeman NC (**2007**) An Overview of Structural DNA Nanotechnology. *Molecular Biotechnology* 37(3):246–257

[86] Shinde D, Lai Y, Sun F, Arnheim N (**2003**) Taq DNA polymerase slippage mutation rates measured by PCR and quasi-likelihood analysis: (CA/GT)n and (A/T)n microsatellites. *Nucleic acids research* 31(3):974–80

[87] Sigel H (**1993**) Interactions of metal ions with nucleotides and nucleic acids and their constituents. *Chemical Society Reviews* 22(4):255

[88] Smith BT, Harvey MG, Faircloth BC, Glenn TC, Brumfield RT (**2014**) Target Capture and Massively Parallel Sequencing of Ultraconserved Elements for Comparative Studies at Shallow Evolutionary Time Scales. *Systematic Biology* 63(1):83–95

[89] Stemmer WP, Crameri A, Ha KD, Brennan TM, Heyneker HL (**1995**) Single-step assembly of a gene and entire plasmid from large numbers of oligodeoxyribonucleotides. *Gene* 164(1):49–53

[90] Stupi BP, Li H, Wang J, Wu W, Morris SE, Litosh VA, Muniz J, Hersh MN, Metzker ML (**2012**) Stereochemistry of Benzylic Carbon Substitution Coupled with Ring Modification of 2-Nitrobenzyl Groups as Key Determinants for Fast-Cleaving Reversible Terminators. *Angewandte Chemie International Edition* 51(7):1724–1727

[91] Tabatabaei Yazdi SMH, Yuan Y, Ma J, Zhao H, Milenkovic O (**2015**) A Rewritable, Random-Access DNA-Based Storage System. *Scientific Reports* 5(1):14138

[92] Tang L (**2018**) An enzymatic oligonucleotide synthesizer. *Nature Methods* 15(8):568–568

[93] Thompson MG, Sedaghatian N, Barajas JF, Wehrs M, Bailey CB, Kaplan N, Hillson NJ, Mukhopadhyay A, Keasling JD (**2018**) Isolation and characterization of novel mutations in the pSC101 origin that increase copy number. *Scientific Reports* 8(1):1590

[94] Tian J, Ma K, Saaem I (**2009**) Advancing high-throughput gene synthesis technology. *Molecular BioSystems* 5(7):714

[95] Ud-Dean SM (**2008**) A theoretical model for template-free synthesis of long DNA sequence. *Systems and Synthetic Biology* 2(3-4):67–73

[96] Walsh PS, Fildes NJ, Reynolds R (**1996**) Sequence analysis and characterization of stutter products at the tetranucleotide repeat locus vWA. *Nucleic acids research* 24(14):2807–12

[97] Wang W, Malcolm BA (**1999**) Two-Stage PCR Protocol Allowing Introduction of Multiple Mutations, Deletions and Insertions Using QuikChange Site-Directed Mutagenesis. *BioTechniques* 26(4):680–682

[98] Weigele P, Raleigh EA (**2016**) Biosynthesis and Function of Modified Bases in Bacteria and Their Viruses. *Chemical Reviews* 116(20):12655–12687

[99] Wu J, Zhang S, Meng Q, Cao H, Li Z, Li X, Shi S, Kim DH, Bi L, Turro NJ, Ju J (**2007**) 3'-O-modified nucleotides as reversible terminators for pyrosequencing. *Proceedings of the National Academy of Sciences* 104(42):16462–16467

[100] Xiong AS, Peng RH, Zhuang J, Gao F, Li Y, Cheng ZM, Yao QH (**2008**) Chemical gene synthesis: strategies, softwares, error corrections, and applications. *FEMS Microbiology Reviews* 32(3):522–540

[101] Xiong AS, Peng RH, Zhuang J, Liu JG, Gao F, Chen JM, Cheng ZM, Yao QH (**2008**) Non-polymerase-cycling-assembly-based chemical gene synthesis: Strategies, methods, and progress. *Biotechnology Advances* 26(2):121–134

[102] Yazdi SMHT, Gabrys R, Milenkovic O (**2017**) Portable and Error-Free DNA-Based Data Storage. *Scientific Reports* 7(1):5011

Ehrenwörtliche Erklärung

Ich erkläre hiermit ehrenwörtlich, dass ich die vorliegende Arbeit entsprechend den Regeln guter wissenschaftlicher Praxis selbstständig und ohne unzulässige Hilfe Dritter angefertigt habe.

Sämtliche aus fremden Quellen direkt oder indirekt übernommenen Gedanken sowie sämtliche von Anderen direkt oder indirekt übernommenen Daten, Techniken und Materialien sind als solche kenntlich gemacht. Die Arbeit wurde bisher bei keiner anderen Hochschule zu Prüfungszwecken eingereicht.

Berkeley, den 9. November 2018

Sebastian Palluk

Curriculum Vitae

Removed for electronic submission of thesis.

Acknowledgements

First of all, I would like to thank everyone on my thesis committee for their support and the consideration of my work: Prof. Beatrix Süß, Prof. Johannes Kabisch, Prof. Gerhard Thiel, and Prof. Heinz Köppl.

Developing a method for enzymatic DNA synthesis has played an important role in my life over the last years. My project began at TU Darmstadt, then continued at the Joint BioEnergy Institute (JBEI), and recently moved towards a new chapter after the foundation of Ansa Biotechnologies. There have been quite difficult times during this journey, and the project would have never happened without the great support and encouragement I received from many friends, mentors, co-workers, and my family. I am well aware of those contributions and want to express my deep gratitude to everyone who has helped me out on my way.

The person whose contributions I want to emphasize first is Daniel Lin-Arlow, my research partner at JBEI and now co-founder at Ansa Biotechnologies. Dan was the one who put the pieces together and came up with the idea of polymerase-nucleotide conjugates, which started an exciting research project that we pursued together. Dan has been an incredible colleague and friend, and I am very thankful for everything I learned from him over the last years. I am looking forward to the next steps and challenges we will face together. Further, I want to thank Dan's partner June Lin-Arlow much for being such a great support for Dan and for our project.

My first steps in research and working on enzymatic DNA synthesis happened at TU Darmstadt during iGEM, and I would like to thank everyone involved back then. In particular, I want to mention Sascha Hain, Andreas Schmidt, Thomas Dohmen, and Alex Schlauer. Further, two people were absolutely critical for this project: Bastian Wagner who set me up with the idea of enzymatic DNA synthesis in the first place, and Daniel Sachs who connected me with Dan Arlow and thereby enabled my time in Berkeley. Daniel later happened to be my bench neighbor in the US, helped me out on many occasions, and it was a huge pleasure having him around.

My work on TdT continued at the Computational Biology group in Darmstadt, where Prof. Kay Hamacher taught me a lot about critical thinking, science, and computational biology. In addition, Kay gave me the opportunity to work on my own research idea, and I am very grateful for his mentoring and support. Further, I would like to thank Sven Jager who supervised my Master's thesis on TdT and strongly encouraged me to continue with the project, Sabine Knorr, Frank Keul, Patrick Boba, Michael Schmidt, Benjamin Mayer, Julius Wiener, Martin Weinelt, and Caroline Pierre. At TU Darmstadt, I would also like to acknowledge Prof. Katja Schmitz who co-supervised my Master's thesis on TdT and provided great feedback on the project, Ina Rink, and Ulrich Göringer. From the Goethe-Universität in Frankfurt, I would like to thank Prof. Alexander Heckel and

Florian Schäfer who offered me the opportunity to perform chemical RTdNTP synthesis in their laboratory, and Prof. Joachim Engels. It was a great pleasure getting to know Prof. Engels, and his insight in early times of chemical DNA synthesis was absolutely fascinating.

Next, I moved to Berkeley and joined Jay Keasling's group at the Joint BioEnergy Institute which is part of the Lawrence Berkeley National Laboratory. I can hardly overstate how fortunate I was joining this extraordinary research location and having Jay as a mentor. Jay was incredibly motivating and positive, and always found time to help us out even though he is one of the busiest people I know. Our project was outside of the research focus of Jay's group, our idea was quite unconventional, but Jay still trusted us and strongly supported our work, for which I am very thankful. My other supervisor at JBEI was Nathan Hillson who met with our DNA synthesis group on a weekly basis, provided great feedback and always guided us in the right direction. Both, Jay and Nathan are involved in Ansa Biotechnologies, and I am very happy to continue working together.

Our DNA synthesis team further included our research associate Alisa Truong who taught me a lot during my first time at JBEI, and the undergraduates Matt Baghdassarian, Sanya Sehgal, Justine Kang and Rathin Bector. Working with and mentoring these aspiring researchers was a remarkable experience, and I am very thankful for all their help on the project. We also got amazing support from Tristan De Rond, whose critical thinking and guidance were crucial for the outcome of the project, and Sebastian Barthel from the TU Darmstadt who joined us to work on the project during his Master's thesis. Sebastian was an immense help and the best company for long nights in lab. Also, I would like to acknowledge the whole automation team including Peter Kim with whom it was a great pleasure to work with, and his team members including Akshay Tambe, Derek Wong, and Anup Singh.

Many other people supported me and the project over the last years in many ways. At JBEI, I would in particular like to thank Chris Eiben, Amin Zargar, Maggie Brown, Jesus Barajas, Mitch Thompson, Jacquelyn Blake-Hedges, Rachel Li, Heather Jensen, Andy de Giovanni, Nurgul Kaplan, Garima Goyal, and Jennifer Chiniquy. Further, Elizabeth Coyne and Susan Gardner saved our project multiple times with organizational help. I am also very thankful for the support of Maren Wehrs, Simon Harth, Michael Reiter, and Niklas Hummel. Further, I would like to acknowledge Karin Pless, Lars Jürgensen, Florian Kübler, Gerrit Rehs, Christian Immel, Stefan Illert, Julian Felger, Michael Sürder, Jonas Stumm, Kai Partenheimer, Dominik Krick, Matthias Knura, David Woog, Christina Antonini, Matthias Weißschuh, Tamara Schempp, Susann Weißheit, Vincent Klaer, Lukas Kaltschnee, Martin Weiß, Arthur Loubart, Tom Ruegg, Gavin Hu, Shaun Kerr, Peter Hargreaves, Izzy Ismailov, Lucas Frenzl, Maria José, Paulo Gonzales, Yi-Min Chun, Timo Schürg, Florian Groher, Thea Lotz, Sergej Franz, and everyone that I might have forgotten. Also, I want to thank Philipp Rosendahl, in particular for his help with submitting my thesis.

I further want to acknowledge four very close friends who hosted me during my recent stays in Germany and have closely followed and supported this project from the very beginning: Andre Pfeiffer and his partner Julia Gründel, and Philipp and Stephanie

Weißgraeber. A special thanks goes to Steffi for proof-reading this thesis, and to Philipp and her for hosting the first office of Ansa Biotechnologies.

Last, I want to thank my family, in particular my parents Ursula and Rainer Palluk, as well as my sister Juliane Lange for always being there, for all their help, and for believing in me. Further, I would like to acknowledge my grandparents Inge and Ulrich Weber, and Klara and Karl-Heinz Palluk for all their support and their excitement about my project. I also want to thank my sister's husband Robert Lange and his family. Finally, I want to express my deep gratitude to Glenn Reeder and Ken Pearce, who are like a second family here in the US.