3′-end fluorescence labelling of long RNAs using terminal deoxynucleotidyl transferase

Johannes Lidberg

Degree Project in Biophysical Chemistry, 2022
Department of Chemistry
Lund University
Sweden

BSc, 15 hp

LUND UNIVERSITY

CHALMERS
UNIVERSITY OF TECHNOLOGY

3′-end fluorescence labelling of long RNAs using terminal deoxynucleotidyl transferase

Johannes Lidberg

Degree Project in Biophysical Chemistry
2022
BSc, 15 hp

Supervisors:
Jesper R. Nilsson,
Pauline Pfeiffer,
L. Marcus Wilhelmsson,
Department of Chemistry and Biochemistry
Chalmers University of Technology,
Gothenburg

Examiner:
Emma Sparr,
Division of Physical Chemistry,
Lund University, Lund

Lund University
Department of Chemistry
Division of Physical Chemistry
P.O. Box 124
SE-221 00 Lund, Sweden

Popular science description

Every living cell making up the human body carries a copy of the genome. This is essentially the recipe for how to make a human. The recipe is written as long strands of the four letters A, C, G, and T, collectively known as nucleotides, spelling out genes as long strands of DNA. Genes get read by the cell and copies are made in the form of ribonucleic acids (RNAs) with the letter T replaced with U. The RNA can then be translated by the protein factory of the cell, the ribosome, into a functioning protein, able to perform the actions needed by the cell. This is the central dogma of molecular biology. A big field of research today is how to use RNA in new medical therapies, as was done for the vaccines used against the Covid-19 virus. However, delivering therapeutic RNA to cells remains a challenge. New techniques to study RNA in cells are therefore becoming increasingly important. One such technique is the use of fluorescence. Fluorophores are molecules that, when exposed to photons of the right energy level, emit new photons with lower energy which can be detected. The process where a fluorophore emits a photon is called fluorescence and allows for highly sensitive analysis. As the RNA building blocks are virtually non-fluorescent themselves, fluorescent labels have to be attached to the RNA to enable analysis.

This thesis explores the use of the enzyme terminal nucleotide transferase (TdT) to label the ends of long RNA strands using the fluorescent label tC^O. Conventional fluorescent labels are generally large and bulky, leading to possible adverse effects on the structure and function of the labelled RNA. tC^O is a fluorescent label that has been designed to replace and act as the natural C nucleotide within the RNA sequence, making it a less perturbing label for studying the native behaviour of RNA with fluorescence. To attach tC^O to RNA, the enzyme TdT is used. TdT is an enzyme that has the unique ability to extend nucleic acids, such as RNA, by taking the precursors to nucleotides, nucleoside triphosphates (tC^O-triphosphate in the case of tC^O), and adding them to one of the RNA ends. To work, TdT requires a metal ion present in the reaction called the cofactor. Reactions using different reaction conditions were performed to study their effect on the incorporation of tC^O.

TdT was successfully used in this thesis to label RNA using tC^O. The amount of incorporated tC^O was shown to be dependent on the amount of available tC^O triphosphate, the cofactor used, and the concentration of the cofactor. RNA sequence also seemed to affect the incorporation of tC^O. The number of tC^Os attached to the RNA strands was considerably less than what has previously been done with shorter RNA.

Abstract

Through recent years, RNA has rapidly gone from mainly being known as a passive information-carrying molecule within the central dogma of molecular biology to becoming the active ingredient in multiple therapeutics, such as in vaccines and potential drugs against cancer. However, one of the main challenges with RNA as a therapeutic is efficient delivery to cells. Understanding the mechanism through which RNA enters cells is therefore crucial to develop more effective RNA therapeutics. New methods to accurately study RNA in cells are becoming increasingly important to fully understand these processes. Fluorescence is one such method commonly used to analyse RNA *in cellulo*. Since natural nucleotides are virtually non-fluorescent, fluorescent labels have to be introduced to enable analysis. tC^O is a fluorescent cytosine analogue causing little perturbance to the RNA's structure and function. This thesis explores a novel way of enzymatically labelling RNA with tC^O using terminal deoxynucleotidyl transferase (TdT), a DNA polymerase capable of template-independent 3′-extension of nucleic acids.

RNA coding for enhanced green fluorescent protein (eGFP) was produced through *in vitro* transcription. These strands were then extended using the TdT enzyme to add a mixture of natural nucleotide triphosphates (NTPs) and tC^O-triphosphate (tC^OTP) to their 3′-ends. Reactions using varying tC^OTP /canonical NTPs-ratios, different cofactors, as well as different cofactor concentrations were performed to study their effect on the labelling performance. The RNA extension was evaluated using gel electrophoresis, as well as UV-vis absorption, excitation, and emission spectroscopy.

tC^O was successfully added to the 3′-ends of the RNA. The efficiency of the labelling seemed to be dependent on the available tC^OTP amount, the used cofactor, the cofactor concentration, and the RNA sequence. Altogether, TdT labelling of RNA was demonstrated as a feasible way of labelling long, functional RNAs with tC^O.

Keywords: Fluorescence, *in vitro* transcription, RNA, terminal deoxynucleotidyl transferase, nucleobase analogue.

Acknowledgement

I would like to express a big thank you to Pauline Pfeiffer and Jesper Nilsson for being great supervisors and helping me with all the laboratory work and with my writing. Another big thank you to Marcus Wilhelmsson for introducing me to the group and for all his help on this thesis. I would also like to thank Emma Sparr for being my examiner and for a great discussion at the presentation. One final thanks to my office mates for supporting me and always being there.

Table of contents

1 List of abbreviations

ATP	Adenosine triphosphate
CTP	Cytidine triphosphate
DNA	Deoxynucleic acid
dNTP	Deoxynucleoside triphosphate
eGFP	Enhance green fluorescent protein
FBA	Fluorescent base analogue
GTP	Guanosine triphosphate
IVT	*In vitro* transcription
mRNA	Messenger RNA
NTP	Nucleoside triphosphate
rNTP	Ribonucleoside triphosphate
RNA	Ribonucleic acid
TdT	Terminal deoxynucleotidyl transferase
tC^{O}	1,3-diaza-2-oxophenoxazine
$tC^{O}TP$	1,3-diaza-2-oxophenoxazine triphosphate
UTP	Uridine triphosphate

2 Introduction

2.1 RNA

Ribonucleic acid (RNA) is an essential molecule for life. Acting as the bridge between the cell's genetic information, stored as genes in its DNA, and those genes' expression as a protein, RNA was long thought of as just a conveyer of information within the central dogma of molecular biology. However, over the last few decades, the discovery of so-called non-coding RNAs has shone a light on RNA's many more varied functions within the cell. RNA is today at the centre of much research as RNA-based therapeutics are being developed as new cost-effective, flexible alternatives to conventional treatments [1].

RNA is a polymer made up of ribonucleotides consisting of one phosphate group, one ribose, and one nucleobase (Figure 1), most commonly being adenine, cytosine, guanine, or uracil, and is synthesized in cells through transcription. RNA polymerase utilizes the energy from cleaving off a pyrophosphate from a nucleoside triphosphate (NTP), together with a DNA template, to produce a polymer of nucleotides with a sequence complementary to the template sequence. Subsequent 3′-polyadenylation and 5′-capping is required to produce translatable messenger RNA (mRNA) in eukaryotic cells [2]. Using bacteriophage T7 RNA polymerase, transcription can be performed *in vitro* to produce RNA ranging from short oligonucleotides to kilobase long sequences. *In vitro* transcription (IVT) has made the production of long RNA simple and inexpensive, enabling the development of therapies based on longer, coding RNA such as the mRNA [3] recently used in vaccines against the Covid-19 virus [4], or its potential future use as a vaccine against cancer [5]. Still, delivery of RNA-based therapeutics *in vivo* remains a challenge[3]. To overcome this problem, techniques to study RNA *in cellulo*, such as using fluorescence microscopy, and flow cytometry, are becoming increasingly important.

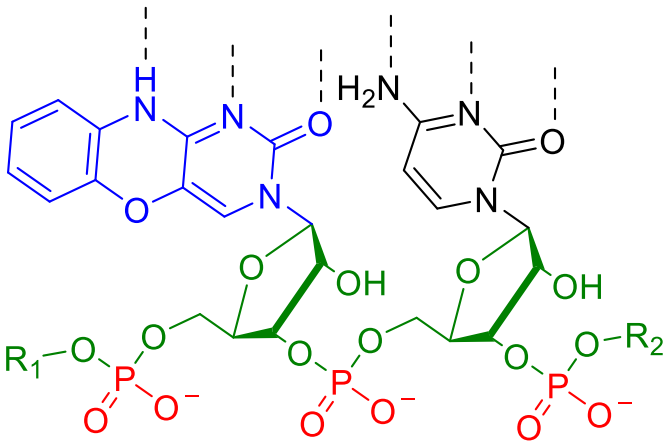


Figure 1. RNA structure with phosphate (red), ribose (green), cytosine (black), and tC^O (blue). Sites for base pairing are shown as dashed lines.

2.2 Fluorescence

After a molecule has absorbed a photon and moved from its ground state to an excited singlet state, the molecule can return to the ground state by either releasing the energy as a new photon or by releasing it through a non-radiative process. The process where the molecule releases a photon is called fluorescence, and the molecule is known as a fluorophore. Two aspects are characteristic of a fluorescence emission spectrum. Firstly, emitted photons are lower in energy, and therefore have a longer wavelength (red-shifted), compared to the exciting photon. Secondly, the emission spectrum is generally independent of the exciting photon's wavelength. Both of these phenomena are a consequence of Kasha's rule, which stipulates that a fluorophore, when excited, rapidly loses vibrational energy, causing emission to happen from the lowest vibrational level of the first excited singlet state. Unlike absorption spectroscopy, where the difference in light intensity between light passing through the sample and the reference is measured, fluorescence spectroscopy measures the light coming directly from the sample, without a comparison. Essentially, fluorescence is measured against a dark background compared to the bright background in absorption spectroscopy. This enables very high sensitivity for fluorescence spectroscopy making it suitable for use in biochemistry [6].

2.3 Fluorescent base analogues

Natural nucleotides are virtually non-fluorescent [6] making chemical modifications of nucleic acids necessary for analysis using fluorescence. One such modification is to covalently attach a fluorescent label. Many of these labels, however, are large and bulky and are attached either to the backbone of the RNA or to the nucleobase (so-called external modification), potentially altering the RNAs functions and properties, such as translational efficiency or protein binding. Fluorescent base analogous (FBAs) are fluorophores designed to substitute the nucleobases within a nucleic acid sequence (internal modification) and to closely resemble their canonical counterparts [7]. 1,3-diaza-2-oxophenoxazine (tC^O) is a tricyclic fluorescent cytosine analogue (Figure 1) that has been shown to act as a highly emissive label in RNA while causing little perturbation to the RNA's secondary structure [8]. Furthermore, tC^O-labelled mRNA has been shown to translate correctly *in cellulo* [9]. These properties make tC^O a good label for accurate study of RNA in living cells.

2.4 Terminal deoxynucleotidyl transferase

Terminal deoxynucleotidyl transferase (TdT) is a DNA polymerase expressed in lymphocytes where it plays an important role in the adaptability of the adaptive immune system [10]. TdT has the unique ability to catalyse 3′-end addition of up to thousands of nucleotides to DNA without the need for a template. Despite being a DNA polymerase, TdT shows low discrimination for deoxyribonucleoside triphosphates (dNTPs), over ribonucleoside triphosphates (rNTPs) and is, therefore, able to also extend the 3′-ends of RNA. Due to its high tolerance for nucleobase-modified NTPs, TdT is used to label 3′-ends with NTPs equipped with, for example, biotin moieties or fluorophores. This is commonly used in the terminal deoxynucleotidyl transferase dUTP nick end labelling (TUNEL) assay for detecting DNA fragmentation during apoptosis *in vivo*, and *in situ* [10-12].

TdT requires a divalent metal cation, such as Co^{2+}, Mg^{2+}, Mn^{2+}, or Zn^{2+}, as a cofactor to perform its catalysis. Depending on the cofactor that is used in the reaction, TdT shows different preferences for the different nucleobases. TdT shows a tenfold preference for purines over pyrimidines when Mg^{2+} is used. By using Co^{2+}, the enzyme instead prefers pyrimidines over purines. Adding sub-millimolar quantities of Zn^{2+} to a reaction using Mg^{2+} has been shown to increase tailing efficiency by 2-4 times when DNA is used as substrate [10, 13].

Recently, TdT was shown to label short RNA strands (≤61 bases) with >25 tC^Os [9]. In this thesis, the enzyme TdT is being explored as a tool to label 3′-ends of long, functioning *in vitro* transcribed RNA with the fluorescent base analogue tC^O. Different amounts of tC^OTP, different cofactors, and different cofactor concentrations are tested to study their effects on the reaction. Fluorescence spectroscopy, absorption spectroscopy, and gel electrophoresis are used to analyse the effectiveness of the labelling.

3 2. Materials and Methods

Unless stated otherwise, all reagents and materials were purchased from Thermo Scientific.

3.1 IVT reactions

To synthesize RNA for use in the TdT reactions, IVT reactions were performed using the sequence for enhanced green fluorescent protein (eGFP) bought from Addgene (Appendix I), or a slightly modified version of the same sequence still coding for eGFP but where stretches of 3 or more consecutive cytosines have been substituted for other bases, as DNA template.

The reactions were performed using materials from the TranscriptAid™ T7 High Yield Transcription Kit. Reactions were run in transcription buffer with added $MgCl_2$ to 4 mM. ATP, CTP, GTP, and UTP (all in tris-buffer) were all added each to 2 mM final concentration. DNA template was used in concentrations of 5, 10, or 20 ng/µL. The reaction had 1 U/µL RiboLock RNase inhibitor and 1.4 U/µL T7 RNA polymerase added to it. Lastly, RNase-free water was used to bring the reaction volume to 50 µL, and the reaction was incubated at 37 °C overnight. After the incubation, 0.1 U/µL Turbo DNase I was added followed by 15 min of incubation at 37 °C to digest the DNA template prior to clean-up.

3.2 TdT reactions

The enzyme, reaction buffer, and $CoCl_2$ were purchased as a Terminal Transferase kit (NEB). The protocol used was based on NEB's protocol for DNA tailing, with cofactor concentrations based on what Baladi *et al*. used in 2021 [9].

All TdT reactions were run as 50 µL reactions. The RNA was denatured by heating it to 65 °C for 5 min and then putting it directly on ice before adding it to the reaction to a final concentration of 200 ng/µL. The reaction ran in Terminal Transferase Buffer and RNase-free water with a final TdT concentration of 0.5 U/µL. All reactions used a mixture of ATP, CTP, GTP, and UTP in equal proportions, together with varying amounts of $tC^{O}TP$ (in-house synthesized dissolved in RNase-free water) for a total NTP concentration of 1 mM. Reactions were run with different fractions of $tC^{O}TP$, using different cofactors, and using different cofactor concentrations. The reactions were incubated for 4 h at 37 °C followed by immediate clean-up.

3.2.1 $tC^{O}TP$ dependency

Three reactions were set up as described in 3.2 with 1 mM total concentration of NTPs made up of 0%, 50%, or 75% $tC^{O}TP$, and the rest ATP, CTP, GTP, and UTP in equal amounts. 0.25 mM of Co^{2+} was used as cofactor by addition of $CoCl_2$.

3.2.2 Cofactor dependency

Different cofactors and cofactor combinations were tested by preparing five reactions as described in 3.2 using 75% $tC^{O}TP$. To the reactions either 0.25 mM $CoCl_2$, 0.25 mM $ZnCl_2$, 1 mM $MnCl_2$, 1 mM $MgCl_2$, or a combination of 1 mM $MgCl_2$ and 5 µM $ZnCl_2$ was added to the reaction before incubation.

3.2.3 Mg^{2+} concentration dependency

Different concentrations of Mg^{2+} as cofactor were studied by preparation of four reactions as described in 3.2 using 75% $tC^{O}TP$. The reactions were then run with 0.1 mM, 1 mM, 5 mM, or 10 mM $MgCl_2$ final concentration.

3.3 RNA purification

Following IVTs, and TdT reactions, the RNA was purified using EconoSpin® mini spin columns (Epoch Life Science). 350 µL of RLP buffer (Qiagen) and 250 µL of ethanol (99.7%) were added to the reaction tube and mixed. The solution was then transferred to the spin column. The column was centrifuged at 8 000 g for 3 min and the flow-through was discarded. The silica containing the RNA was then washed by adding 500 µL RPE buffer (10 mM Tris, 100 mM NaCl, 80 v-% EtOH) to the column, centrifuging the column at 8 000 g for 1 min, and then discarding the flow-through. This washing step was performed two times following an IVT reaction, but four times for TdT reactions to make sure that all unreacted $tC^{O}TP$ got washed away. The column was then centrifuged at 8 000 g for another 3 min to further rid the column of residual buffer. Finally, RNA was eluted twice by centrifugation at 8 000 g for 3 min into new collection tubes, each time using 30 µL of RNase-free water.

3.4 Concentration determination

The RNA concentrations of samples were determined using a NanoDrop 1000 (Thermo Fischer Scientific) and using RNase-free water as baseline. 1-2 µL of sample was loaded onto the NanoDrop and an absorption spectrum was recorded from 220 to 750 nm. The absorption value

of the sample at 260 nm (A_{260}) was then used together with the average molar absorptivity of the nucleobases in the RNA at 260 nm (ε_{260} = 9 750 $M^{-1}cm^{-1}$, based on a non-nearest neighbour base-composition method) in the Beer-Lambert law (Eq. 1) to calculate the RNA concentration. c is the concentration and l is the optical path length (1 mm for the NanoDrop).

$$A = \varepsilon \times c \times l \qquad \text{Eq. 1}$$

3.5 Agarose gel electrophoresis

Agarose gel electrophoresis was performed on all RNA samples to evaluate the length of the RNA. A mass-2% agarose/TBE buffer gel solution was prepared. 1x SYBR Safe was added as a nucleic acid stain.

The sample solutions were prepared by adding 0.2 µg of sample RNA to 5 µL of 2X RNA loading dye containing formamide and then adding RNase-free water until the solution volume reached 10 µL. The ladder was prepared by adding 4.95 µL of RiboRuler High Range RNA ladder to 4.95 µL of loading dye. The samples and the ladder were denatured by placing them on a heating block at 65 °C for 15 min before moving them directly onto ice. 10 µL of sample solution was loaded into each sample well and 4.5 µL of ladder solution was used in the ladder wells. The gel was run for 1h 40 min at 100 V in TBE buffer.

3.6 Spectroscopic analysis

UV-Vis absorption, along with fluorescence emission and excitation spectra were recorded for all purified products of TdT reactions, as well as for the tC^{O}TP monomer. All measurements were performed with 60 µL sample volume in a quartz cuvette with a 3 mm optical path length.

The TdT samples were prepared by pooling the two elutions from one reaction clean-up, estimating the sample volume using a pipette then diluting the sample to 70 µL by adding RNase-free water. For measurements of tCOTP, a solution of approximately 5 µM tCOTP in RNase-free water was used.

Absorption spectra were recorded using a Cary 50 Bio (Agilent Technologies) between 200 and 600 nm, at a scan rate of 600 nm/min, and a data interval of 1 nm. RNase-free water was used as baseline.

Fluorescence emission and excitation spectra were recorded on a Spex Fluorolog 3 (JY Horiba). Emission was measured between 370 and 700 nm (3 nm slit width, 1 nm increment, 0.1 s integration time), exciting the sample at 360 nm (3 nm slit width). The excitation spectrum was recorded by exciting the sample between 250 nm and 450 nm (3 nm slit width, 1 nm increment), measuring emission at 465 nm (3 nm slit width, 0.1 s integration time) [8].

The number of incorporated tC^Os per strand was approximated by first calculating the concentration of nucleotides in the sample. This was done using the average molar absorptivity of the strand at 260 nm (ε_{260} = 9 750 $M^{-1}cm^{-1}$), the absorption value at 260 nm and the Beer-Lambert law (Eq. 1). Similarly, the tC^O concentration was calculated the same way using the molar absorptivity of tC^O at 360 nm (ε_{360} = 10 000 $M^{-1}cm^{-1}$) and the absorption value at 360 nm. The ratio between these concentrations was then used to determine the number of nucleotides per tC^O. Assuming an approximate RNA sequence length of 800 bases, the number of tC^Os per strand was determined.

4 Results and discussions

4.1 *In vitro* transcription reactions

To produce RNA to be used in TdT reactions, IVT reactions were performed and the products were analysed using gel electrophoresis and absorption spectroscopy. For these reactions, DNA templates coding for eGFP was used. The absorption spectrum (Figure 2) shows a peak at 260 nm which is characteristic of nucleic acids [14]. This peak was then used to determine the RNA concentration using the Beer-Lambert law (Eq. 1). Initially, three different concentrations of DNA template were used. However, as they all yielded the same amount of RNA, the lowest concentration (5 ng/µL) was used for subsequent IVTs.

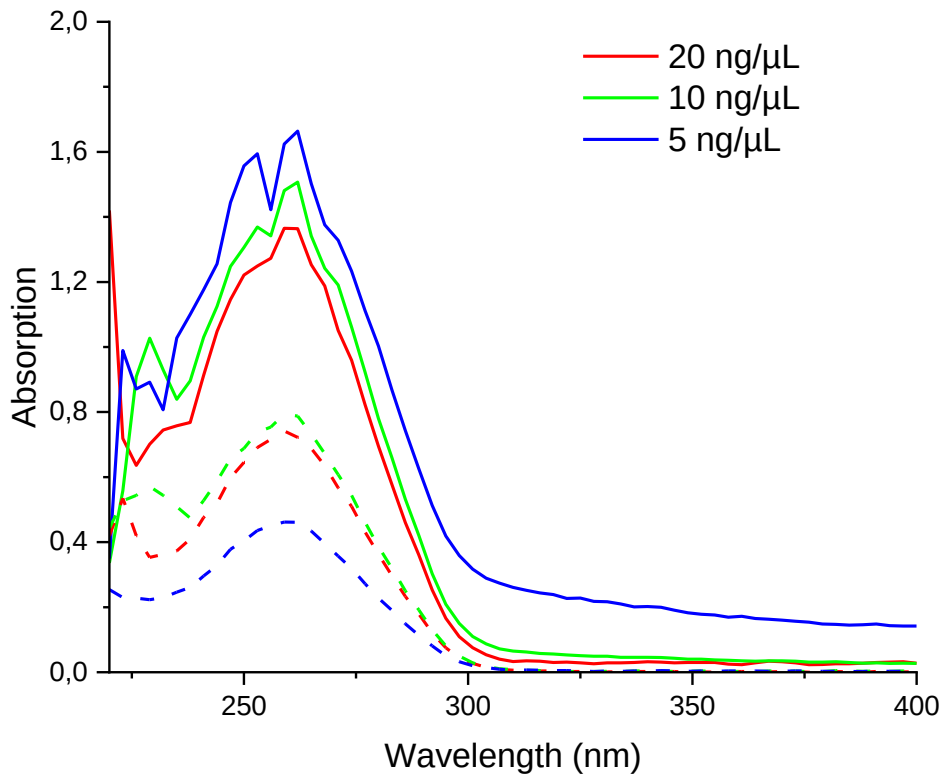


Figure 2. Absorption spectra of the first elution (solid line) and second elution (dotted line) of purified RNA from the second IVT reaction using 20 ng/µL (red), 10 ng/µL (green), or 5 ng/µL (blue) DNA template. The optical path length was 1 mm.

To determine the length of the RNA produced from IVTs, gel electrophoresis was performed. All gels of IVT samples showed one band at approximately 800 bases (Figure 3) for most

samples. This shows that the RNA produced in these reactions was of the expected length, and therefore, we assume, also the right sequence. Smears below 800 bases were observed for all samples with no apparent difference between most samples, except for a few reactions that showed a more intense smear, and a less intense band (Figure 3, 5 ng/µL (2) sample). This could be a sign of RNA degradation or, for some reason, aborted transcriptions. These were all omitted from future reactions.

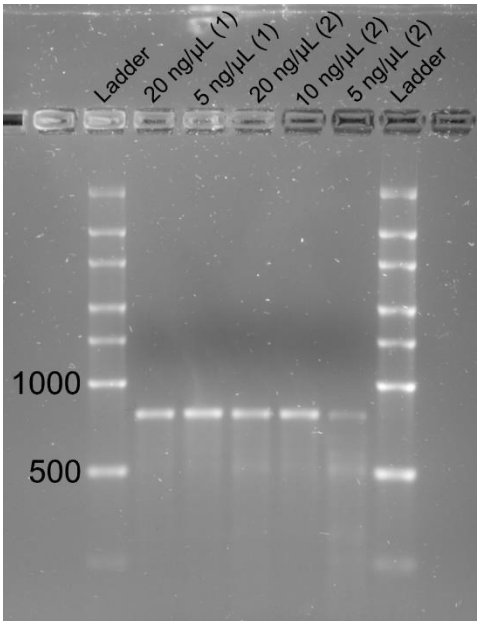


Figure 3. Agarose gel with 200 ng of RNA from IVT reactions using 5 ng/µL, 10 ng/µL, or 20 ng/µL DNA template from two separate reactions (parentheses) loaded per sample well. The ladder is RiboRuler High Range RNA ladder.

4.2 TdT reactions

To investigate whether tC^O labelling of RNA using TdT was at all possible, and to see whether it potentially would be affected by the fraction of available tC^OTP compared to canonical ATP, CTP, GTP, and UTP, three reactions were run; one reaction had 50% of all available NTPs being tC^OTP, one reaction had 75% tC^OTP, and one reaction with 0% tC^OTP was run as a control. A 100% tC^OTP reaction was omitted to avoid possible problems for the enzyme to consecutively add modified bases together. Another reason for opting against using 100% tC^OTP in the reactions is that self-quenching, as a result of too densely packed fluorophores, has been observed for tC^O in RNA [9]. The samples were analysed spectroscopically, and the resulting excitation and emission spectra were normalised to the corresponding absorption values at 260 nm to compensate for varying RNA concentrations.

The absorption spectra of the samples (Figure 4c) show the characteristic peak of nucleic acids at 260 nm. There also seems to be a shoulder at 360 nm, lining up with where tC^{O}TP has an absorption maximum (Figure 4a). However, the signal seems somewhat affected by scattering, showing a clear upward slope in absorption towards shorter wavelengths, making it hard to assess accurately. The excitation and emission spectra of the samples run with tC^{O}TP (Figure 4d) and for tC^{O}TP itself (Figure 4b) both show an excitation peak at 360 nm and an emission peak at 465 nm. The 0% reaction shows no such peaks. From previous experiments, we know that the four washing steps used in the RNA purification leave little to no unreacted tC^{O}TP in the product (data unpublished). This means that the emission recorded has to be attributed to incorporated tC^{O} and therefore shows that TdT successfully has extended the RNA and incorporated tC^{O} in the added tail. Furthermore, a larger fraction of tC^{O}TP in the reaction seems to give a more emissive product. Due to this, a reaction mixture with 75% tC^{O}TP was used for all following reactions.

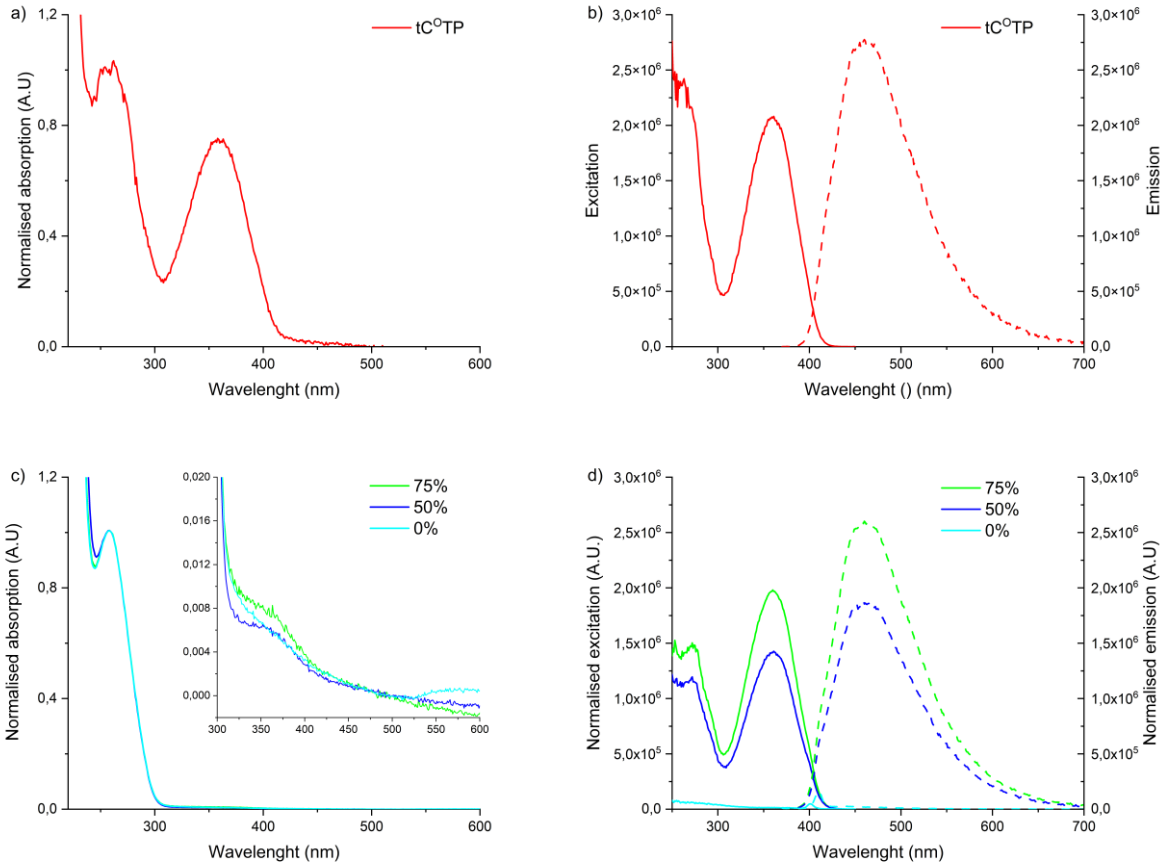


Figure 4. (a) Absorption spectrum (normalised to A_{260} = 1), and (b) excitation (solid line) and emission (dotted line) spectrum of tC^{O}TP. (c) Absorption spectra (normalised to A_{260} = 1), and

excitation (solid lines) and emission (dotted lines) spectra (normalised to A_{260}) of tC^O from TdT reactions run with 75% (red), 50% (green), or 0% (blue) tC^OTP.

To further investigate the tailed RNA's length, gel electrophoresis was used (Figure 5). Bands at ~800 bases were observed for all three reactions with no discernible difference between the labelled RNA and the unreacted RNA used as control. The gel also shows a similar smear below 800 bases as for the IVT products. The gels of all subsequent TdT reactions showed identical results (Appendix III). An important note is that this method uses the fluorescent dye SYBR Safe to visualize the RNA. In other words, the RNA does not have to be labelled with tC^O to be visible here.

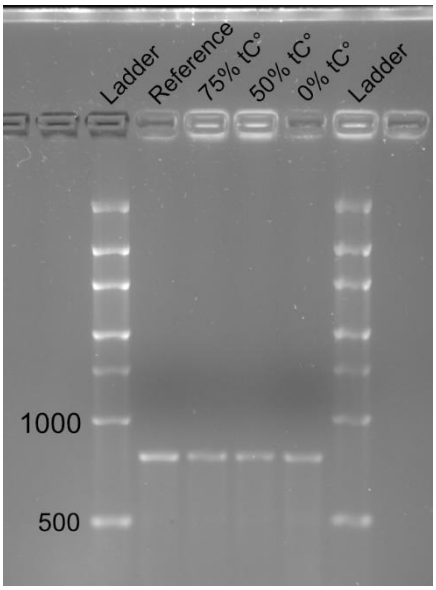


Figure 5. Agarose gel with 200 ng of RNA loaded per sample well from TdT reactions using 0%, 50%, or 75% tC^OTP. Reference is the starting RNA, and the ladder is RiboRuler High Range RNA ladder.

The fact that there is no apparent difference in length between the control RNA and the tailed RNA shows that the tail is short, *i.e.*, consists of only a few bases. Due to the resolution of the gel, it is not possible to distinguish between RNAs differing only by a couple of nucleotides, and it is therefore not possible to evaluate exactly how long the tail is. It is, however, considerably less than the thousands of dNTPs that have been shown to be added to DNA using the enzyme [10].

To investigate if the labelling of the RNA could be improved, reactions were run using different cofactors, as TdT exhibits different preferences for different nucleobases depending on the

cofactor that is present [10]. Reactions were run with either Co^{2+}, Zn^{2+}, Mn^{2+}, Mg^{2+}, or a combination of Mg^{2+} and Zn^{2+}. The concentration of each cofactor was based on what has previously been done by Baladi, et al. [9]. The concentration of Zn^{2+} in combination with Mg^{2+} was based on experiments using DNA [13]. The excitation and emission graph (Figure 6) shows that Mg^{2+} produces the most emissive product. Combining Mg^{2+} with a small amount of Zn^{2+} lowered the labelling efficiency. Absorption spectra for this reaction can be found in Appendix (IV).

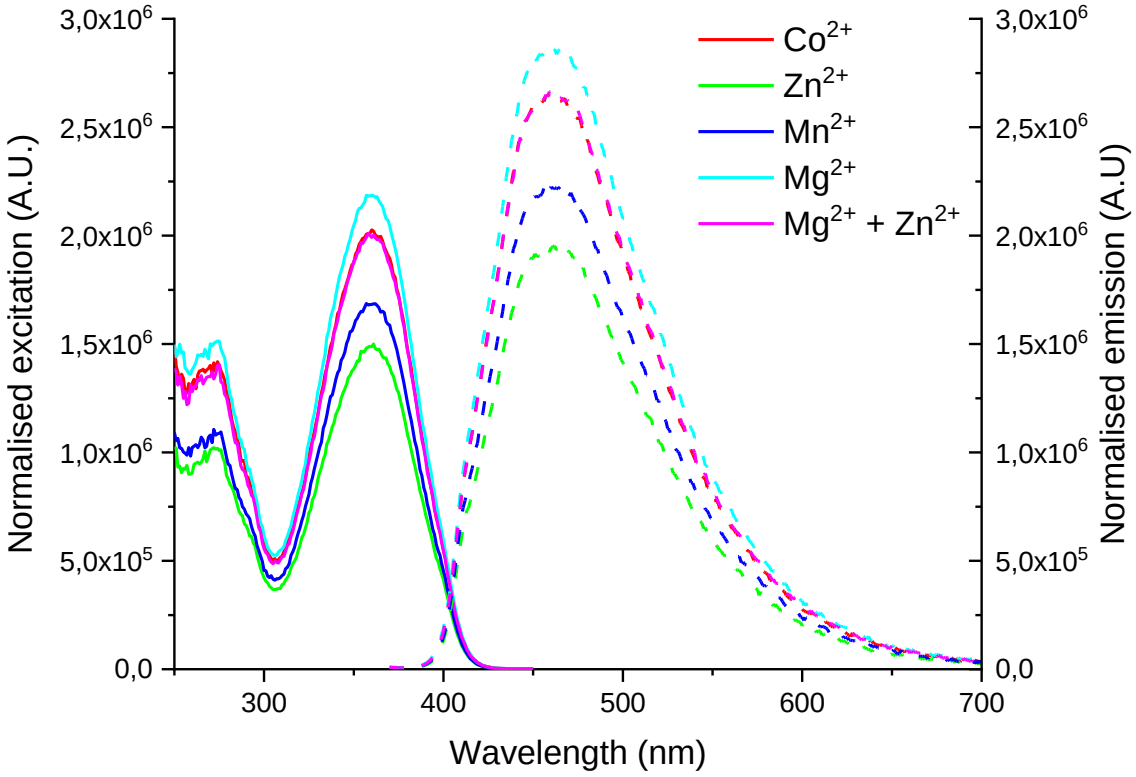


Fig. 6. Excitation (solid lines, normalised to A_{260}) and emission (dotted lines, normalised to A_{260}) spectra of tC^O from TdT reactions run with either Co^{2+} (red), Zn^{2+} (green), Mn^{2+} (blue), Mg^{2+} (cyan), or a combination of Mg^{2+} and Zn^{2+} (magenta).

There are two possible reasons for why Mg^{2+} increased the labelling of the RNA; one possibility is that Mg^{2+} increases TdT's affinity for tC^OTP relative to the other NTPs and in that way increased the frequency of which tC^O was incorporated into the tail. A second possibility is that the added tails were longer when Mg^{2+} was used as cofactor. This would result in more tC^Os per RNA even if the incorporation frequency stayed the same. The fact that Mg^{2+} helps

incorporate more tC^O is interesting as tC^O is an analogue of a pyrimidine. When dNTPs are used as substrate, TdT heavily favours purines over pyrimidines with Mg^{2+} as the cofactor [10]. This could indicate that the reason for the more emissive RNA is because of longer tails, rather than more frequent incorporation of tC^O. It could also be that the structure of tC^O is too different from purines or pyrimidines for this preference to be relevant. As previously discussed, one would need a higher-resolution gel to conclude whether the tail length changes or not with the different cofactors.

As Mg^{2+} seemingly helped produce more emissive products, four reactions were run to see how different Mg^{2+} concentrations would affect the reaction. The reactions were run with either 0.1 mM, 1 mM, 5 mM, or 10 mM Mg^{2+}. The concentrations were based on previous experiments with DNA [10]. The reactions all used the same mixture of two different RNA sequences coding for eGFP; one sequence was bought from Addgene, and the other is an in-house modified version of the same sequence where stretches of three or more consecutive Cs have been substituted for other bases.

The excitation and emission graph shows that 0.1 mM Mg^{2+} was most effective for labelling the RNA (Figure 7). This is not too surprising as lower Mg^{2+} concentrations have been suggested to increase tailing efficiency under certain conditions [10]. What is more surprising is the 5 mM concentration also seemed to produce more tC^O-labelled RNA than the 1 mM reaction, although it may not be a significant difference. Absorption spectra in Appendix IV.

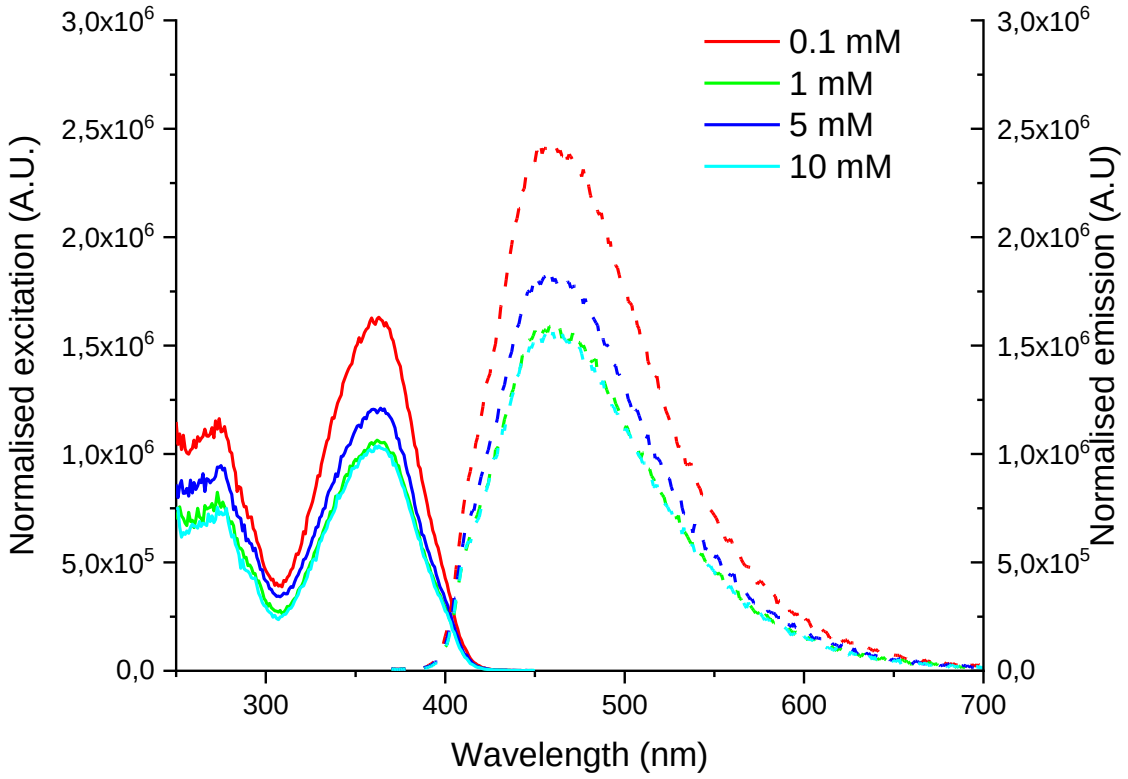


Figure 7. Excitation (solid lines, normalised to A_{260}) and emission (dotted lines, normalised to A_{260}) spectra of tC^O in products from TdT reactions run with Mg^{2+} concentrations of either 0.1 mM (red), 1 mM (green), 5 mM (blue), or 10 mM (cyan).

As the Mg^{2+} reactions from the cofactor comparison experiment used 1 mM Mg^{2+}, it is notable that the 1 mM in this reaction resulted in a considerably less emissive product. This is possibly due to the modified eGFP sequence that was used together with the Addgene sequence in this reaction and suggests that the labelling is sequence dependent. One possible cause for this could be steric hindrance, *i.e.*, if the modified sequence causes the RNA strands to fold in such a way that the 3′-end is blocked from the TdT. The enzyme would then have less access to the 3′-end resulting in a slower reaction and shorter tails. It could also be a result of the TdT showing less affinity for the modified sequence's 3′-end. Both of these possibilities would result in shorter tails.

Using the absorbance values at 260 and 360 nm of the reactions, it is possible to estimate the average amount of nucleotides per tC^O in the sample. For most reactions, this results in 1 tC^O per ~300 nucleotides. Since the gels show that the RNA strands haven't been extended

significantly, it is safe to assume that the RNA is still approximately 800 bases. This works out to be ~2-3 tC^Os per RNA strand. However, this is only an approximation since the 360 nm absorption is very low for all samples and is affected by scattering, therefore, it is hard to get an accurate value. Still, this is considerably less than the number of tC^Os that have been shown to be added to shorter RNA strands [9]. The difference may be that the RNA structure is unfavourable for the TdT reaction when it gets longer.

5 Conclusions

In this thesis, TdT was successfully utilized to enzymatically label the 3′-ends of RNA with the tC^O. The number of tC^Os added to the RNA was shown to depend on the amount of available tC^OTP, the cofactor in the reaction, the cofactor concentrations, and seemingly, the RNA sequence. 2-3 tC^Os were added to each strand.

6 Future aspects

For this thesis, all reactions were run only once. This makes it impossible to determine the statistical significance of the finding. It would therefore be recommended to continue this work by rerunning the same experiments to verify the results. Furthermore, trying more tC^{O}TP fractions and cofactor concentrations could be done to optimise this technique. Further studies could then include transforming the RNA to mRNA through 5′-capping and polyadenylation to see if the mRNA gets translated properly in cells and whether the mRNA is detectable. It could also be the case that polyadenylation is not possible due to the tail. If translation of the mRNA proves possible, this method could be used to study mRNA uptake and localisation *in cellulo* in the context of improving mRNA therapeutics.

7 References

1. Kim, Y.K., *RNA therapy: rich history, various applications and unlimited future prospects.* Exp Mol Med, 2022. **54**(4): p. 455-465.

2. Jeremy M Berg, L.S., John Tymoczko, Gregory Gatto, *Biochemistry*. Vol. 9. 2019, New York: W.H.Freeman & Co Ltd. 1208.

3. Sahin, U., K. Kariko, and O. Tureci, *mRNA-based therapeutics--developing a new class of drugs.* Nat Rev Drug Discov, 2014. **13**(10): p. 759-80.

4. Polack, F.P., et al., *Safety and Efficacy of the BNT162b2 mRNA Covid-19 Vaccine.* N Engl J Med, 2020. **383**(27): p. 2603-2615.

5. Damase, T.R., et al., *The Limitless Future of RNA Therapeutics.* Front Bioeng Biotechnol, 2021. **9**: p. 628137.

6. Lakowicz, J.R., *Principles of Fluorescence Spectroscopy*. Vol. 3. 2006, New York: Springer

7. Wilhelmsson, L.M., *Fluorescent nucleic acid base analogues.* Q Rev Biophys, 2010. **43**(2): p. 159-83.

8. Fuchtbauer, A.F., et al., *Fluorescent RNA cytosine analogue - an internal probe for detailed structure and dynamics investigations.* Sci Rep, 2017. **7**(1): p. 2393.

9. Baladi, T., et al., *Stealth Fluorescence Labeling for Live Microscopy Imaging of mRNA Delivery.* J Am Chem Soc, 2021. **143**(14): p. 5413-5424.

10. Sarac, I. and M. Hollenstein, *Terminal Deoxynucleotidyl Transferase in the Synthesis and Modification of Nucleic Acids.* Chembiochem, 2019. **20**(7): p. 860-871.

11. Majtnerova, P. and T. Rousar, *An overview of apoptosis assays detecting DNA fragmentation.* Mol Biol Rep, 2018. **45**(5): p. 1469-1478.

12. Motea, E.A. and A.J. Berdis, *Terminal deoxynucleotidyl transferase: the story of a misguided DNA polymerase.* Biochim Biophys Acta, 2010. **1804**(5): p. 1151-66.

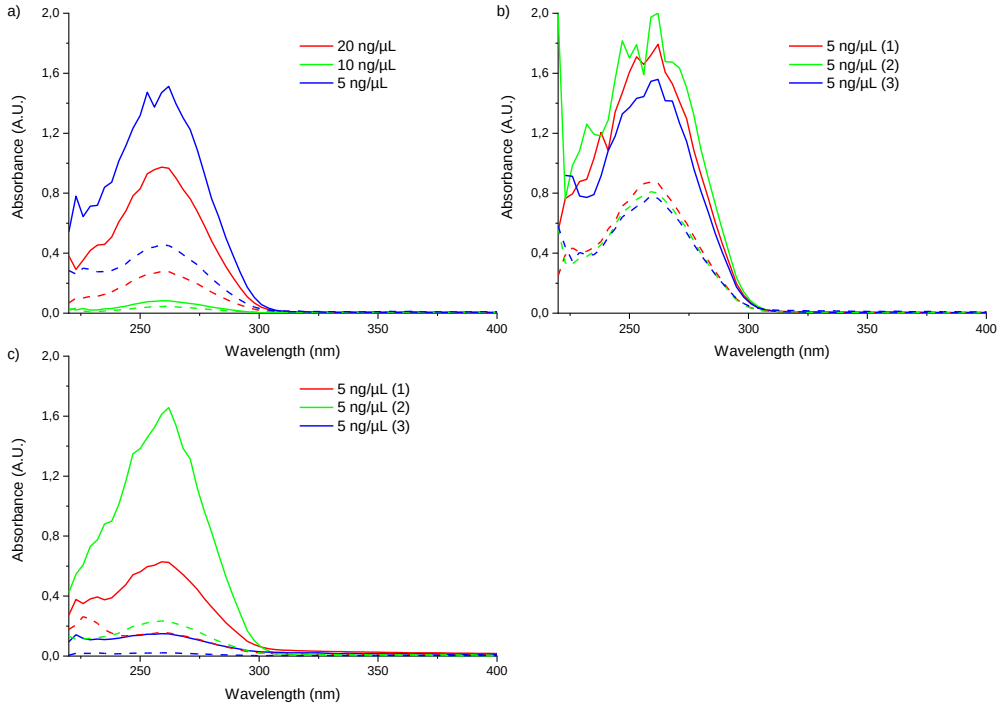
13. Chang, L.M. and F.J. Bollum, *Multiple roles of divalent cation in the terminal deoxynucleotidyltransferase reaction.* Journal of Biological Chemistry, 1990. **265**(29): p. 17436-17440.

14. Gallagher, S.R. and P.R. Desjardins, *Quantitation of DNA and RNA with absorption and fluorescence spectroscopy.* Curr Protoc Mol Biol, 2006. **Appendix 3**: p. Appendix 3D.

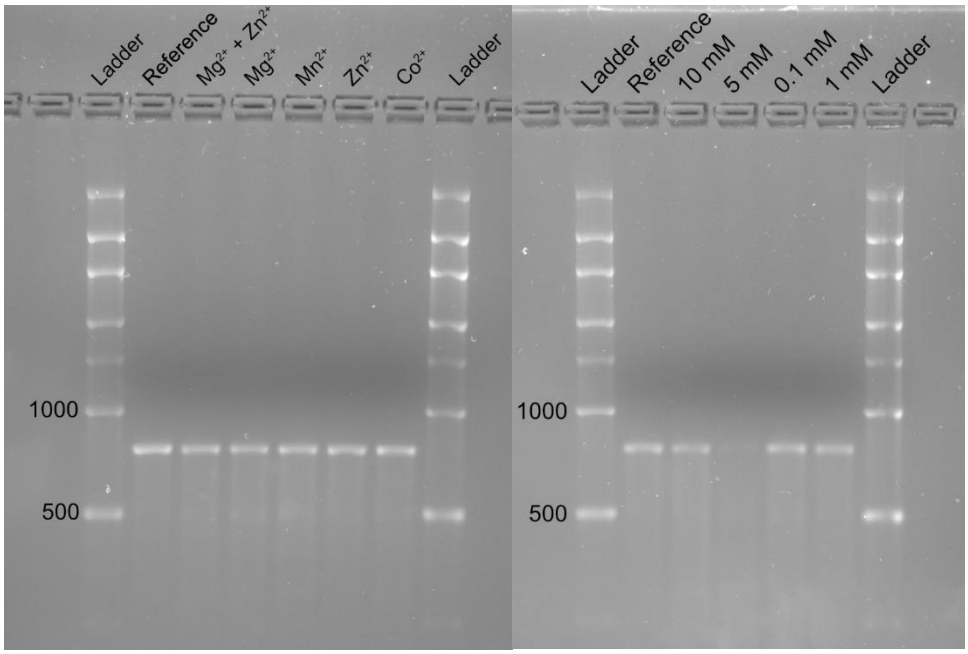
8 Appendix

AGCTTGATTTAGGTGACACTATAGATAATACGACTCACTATAGGATACATCGATTCGAATTCAAGGCCTCTCGAGCCTCT
AGAACTCGACGGTACCAAGGAGGCCACCATGGTGAGCAAGGGCGAGGAGCTGTTCACCGGGGTGGTGCCCATCCTGGTCG
AGCTGGACGGCGACGTAAACGGCCACAAGTTCAGCGTGTCCGGCGAGGGCGAGGGCGATGCCACCTACGGCAAGCTGACC
CTGAAGTTCATCTGCACCACCGGCAAGCTGCCCGTGCCCTGGCCCACCCTCGTGACCACCCTGACCTACGGCGTGCAGTG
CTTCAGCCGCTACCCCGACCACATGAAGCAGCACGACTTCTTCAAGTCCGCCATGCCCGAAGGCTACGTCCAGGAGCGCA
CCATCTTCTTCAAGGACGACGGCAACTACAAGACCCGCGCCGAGGTGAAGTTCGAGGGCGACACCCTGGTGAACCGCATC
GAGCTGAAGGGCATCGACTTCAAGGAGGACGGCAACATCCTGGGGCACAAGCTGGAGTACAACTACAACAGCCACAACGT
CTATATCATGGCCGACAAGCAGAAGAACGGCATCAAGGTGAACTTCAAGATCCGCCACAACATCGAGGACGGCAGCGTGC
AGCTCGCCGACCACTACCAGCAGAACACCCCCATCGGCGACGGCCCCGTGCTGCTGCCCGACAACCACTACCTGAGCACC
CAGTCCGCCCTGAGCAAAGACCCCAACGAGAAGCGCGATCACATGGTCCTGCTGGAGTTCGTGACCGCCGCCGGGATCAC
TCTCGGCATGGACGAGCTGTACAAGTAACTGACTGATAC

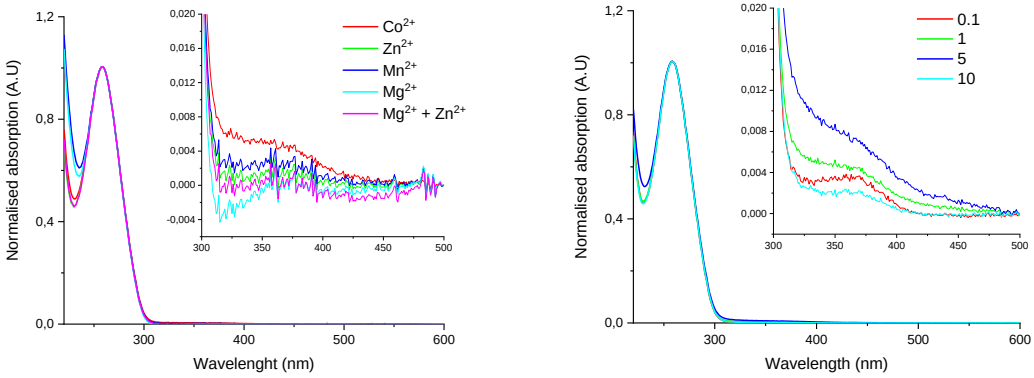
Appendix I. RNA sequence from Addgene coding for eGFP used in IVTs. T7 RNA polymerase promotor is highlighted in yellow, the coding sequence in green, and the stop codons in red.



Appendix II. Absorption spectra of first elution (solid line) and second elution (dotted line) of purified RNA from the first (a), third (b), and fourth (c) IVT reaction using 5 ng/µL, 10 ng/µL, or 20 ng/µL DNA template. The optical path length was 1 mm.



Appendix III. Agarose gel with 200 ng of RNA loaded per sample well from TdT reactions using different cofactors (left) or different Mg^{2+} concentrations (right). Reference is the starting RNA, and the ladder is RiboRuler High Range RNA ladder.



Appendix IV. (a) Absorption spectra of RNA from TdT reactions run with different cofactors, and (b) absorption spectra of RNA from TdT reactions run with different Mg^{2+} concentrations.