

Nanopore sequencing for N1-methylpseudouridine in RNA reveals sequence-dependent discrimination of the modified nucleotide triphosphate during transcription

Aaron M. Fleming* and Cynthia J. Burrows 

Dept. of Chemistry, University of Utah, Salt Lake City, UT 84112-0850, USA

Received August 19, 2022; Revised December 21, 2022; Editorial Decision January 15, 2023; Accepted January 16, 2023

ABSTRACT

Direct RNA sequencing with a commercial nanopore platform was used to sequence RNA containing uridine (U), pseudouridine (Ψ) or N1-methylpseudouridine ($m^1\Psi$) in >100 different 5-nucleotide contexts. The base calling data for Ψ or $m^1\Psi$ were similar but different from U allowing their detection. Understanding the nanopore signatures for Ψ and $m^1\Psi$ enabled a running start T7 RNA polymerase assay to study the selection of UTP versus Ψ TP or $m^1\Psi$ TP competing mixtures in all possible adjacent sequence contexts. A significant sequence context dependency was observed for T7 RNA polymerase with insertion yields for Ψ TP versus UTP spanning a range of 20–65%, and $m^1\Psi$ TP versus UTP producing variable yields that differ by 15–70%. Experiments with SP6 RNA polymerase, as well as chemically-modified triphosphates and DNA templates provide insight to explain the observations. The SP6 polymerase introduced $m^1\Psi$ TP when competed with UTP with a smaller window of yields (15–30%) across all sequence contexts studied. These results may aid in future efforts that employ RNA polymerases to make therapeutic mRNAs with sub-stoichiometric amounts of $m^1\Psi$.

INTRODUCTION

Native RNA across all phyla of life possesses >150 chemical modifications that include the addition of alkyl groups on the base and/or sugar, as well as isomerization, sulfurization, oxidation, or reduction of the nucleobases (1–4). These chemical changes are found in tRNA, rRNA, mRNA, small and large non-coding RNAs, and viral RNAs. The epitranscriptome refers to those modifications essential for the transcriptome's functional relevancy for cellular processes.

Chemically modified RNA has found its way into clinical applications where the successes of therapeutic siRNAs and mRNA vaccines are largely achieved due to the site-specific chemical decorations of the polymer (5,6). Identification and quantification of RNA modifications have been pursued by nuclease and phosphatase digestion of target strands to nucleosides for LC-MS/MS quantitative analysis, which also results in the complete loss of the sequence information (1). The development of nanopores as a third-generation sequencing platform has brought about many enabling advancements, one of which is the ability to directly sequence RNA, with the sequencing of chemical modifications as one of the many benefits of this method.

Nanopore sequencing on a commercial platform (Oxford Nanopore Technologies) is achieved by the use of a 3',5'-helicase as a motor protein to deliver the RNA into a protein nanopore at a rate that is ATP-dependent (Figure 1) (7). An electrophoretic force serves to guide the direction of the RNA 3' to 5' into the nanopore protein where a small central constriction zone exists. As the strand passes the constriction zone with a length of ~5-nt of RNA (i.e. the *k*-mer) and a diameter slightly larger than single-stranded nucleic acids, the ionic current changes as a function of sequence as it passes through the pore (8). In recent iterations of base-calling software, the current vs. time traces are deconvoluted via a recurrent neural network (9). Exciting developments using nanopores have showcased direct RNA sequencing for modifications in tRNA, rRNA, mRNA, small/large RNAs, and viral RNAs from biological sources (8,10–16). The native RNA modifications inspected with the greatest focus include pseudouridine (Ψ), *N*⁶-methyladenosine, and 2'-*O*-methyl nucleotides. Our work with Ψ demonstrated base calling error analysis combined with current and dwell time analysis minimizes the false discovery rate for modification detection in noisy nanopore data (13). Some challenges remain before this approach becomes a routine technique in the RNA researcher's toolbox.

*To whom correspondence should be addressed. Tel: +1 801 585 7290; Email: burrows@chem.utah.edu
Correspondence may also be addressed to Aaron M. Fleming. Email: afleming@chem.utah.edu

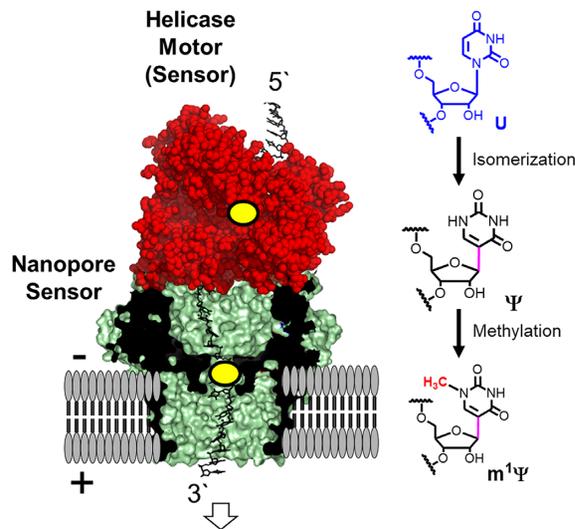


Figure 1. The Oxford Nanopore Technologies platform sequences U, Ψ and $m^1\Psi$ directly in RNA.

Chemically modified RNA has diverse clinical applications. At present, the best example is the SARS-CoV-2 mRNA vaccine that has the complete replacement of the U nucleotides with N1-methylpseudouridine ($m^1\Psi$) (Figure 1) (6). This modification results from isomerization and methylation of the uridine nucleotide and was first discovered in archaea tRNA by the McCloskey laboratory (17). Synthetic mRNA vaccines are produced by *in vitro* transcription (IVT), in which all U sites are converted to $m^1\Psi$ by feeding T7 RNA polymerase $m^1\Psi$ -nucleotide triphosphate ($m^1\Psi$ TP) instead of UTP (6). The first goal of the present study was to evaluate how the nanopore sequencing device responds to $m^1\Psi$ compared to Ψ and U in different k-mer sequence contexts.

The knowledge of how the nanopore sequencer responds during base calling to U, Ψ and $m^1\Psi$ from the first study then led us to apply direct RNA sequencing to monitor a running start RNA polymerase extension assay. More specifically, this enabled interrogation of T7 RNA polymerase NTP selection when UTP was mixed in a known ratio with either Ψ TP or $m^1\Psi$ TP. All immediate sequence contexts flanking a single competition site were evaluated ($5'$ -VXXV- $3'$ where V = A, C or G, and X = U/ Ψ or U/ $m^1\Psi$) and contexts with two adjacent competition sites ($5'$ -VXXV- $3'$). The analysis found that T7 RNA polymerase shows a sequence context bias for the selection of the competing NTP mixture, particularly for UTP vs. $m^1\Psi$ TP. Beyond demonstrating a new approach to monitoring a polymerase extension assay, the results suggest that any attempt to synthesize mRNA strands with sub-stoichiometric amounts of $m^1\Psi$ TP via T7-catalyzed IVT will generate strands with unbalanced levels of modifications across the sequence contexts. Additional experiments with modified template DNA strands for IVT identified possible polymerase template DNA interactions resulting in the sequence bias. Finally, this knowledge led us to study SP6 RNA polymerase for IVT generation of RNA to find a more balanced introduction of $m^1\Psi$ TP when competed with UTP. This observation inspired a fi-

nal set of goals of the present study that was to address why T7 RNA polymerase gave a sequence context impact when selecting competing NTPs. This knowledge led to a modified approach for mRNA synthesis via IVT to incorporate U and $m^1\Psi$ with similar stoichiometry in all immediate sequence contexts. The findings are discussed with their implications for analyzing RNA polymerase activity and the use of RNA polymerases for therapeutic mRNA synthesis.

MATERIALS AND METHODS

RNA synthesis by *in vitro* transcription

In vitro transcription was performed using the MEGAscript T7 transcription kit (Thermo Fisher) or HiScribe SP6 RNA synthesis kit (New England Biolabs) according to the manufacturer's instructions. The duplex DNA templates for the IVT reactions were synthesized via commercial sources to have a T7 or SP6 promoter for initiation of transcription and ended with a poly-A tail for sequencing library preparation (Supplementary Figure S1). The IVT reactions were incubated overnight at 37°C in a PCR thermocycler. After the overnight incubation, DNase I treatment was performed on all samples at 37°C, followed by purification using Quick Spin Columns for RNA purification (Sigma). To install Ψ or $m^1\Psi$, IVT was conducted in the presence of commercially available pseudouridine-5'-triphosphate (Ψ TP) or $m^1\Psi$ TP (Trilink Biotechnologies with purities > 99%) instead of UTP. Success in the synthesis of the RNA transcripts was verified by agarose gel electrophoresis by comparison to a ladder of known lengths. The synthetic RNA strands were synthesized by standard solid-phase synthesis protocols using commercially available phosphoramidites.

T7 or SP6 RNA polymerase studies with mixed and competing NTPs

Studies on NTP discrimination by T7 RNA polymerase were conducted using the MEGAscript T7 transcription kit or the SP6 RNA polymerase using the HiScribe SP6 RNA synthesis kit with some changes to the manufacturer's protocol as described. The NTP concentrations were 2 mM for ATP, GTP and CTP, while the UTP and Ψ TP or $m^1\Psi$ TP were 1 mM each to achieve a total concentration of U and its derivative of 2 mM. To ensure the UTP and Ψ TP or $m^1\Psi$ TP were at a 1:1 ratio, the stock solution concentrations were determined using UV-vis spectroscopy with established extinction coefficients (UTP: $\lambda_{262\text{ nm}} = 10\,000\text{ l mol}^{-1}\text{ cm}^{-1}$; Ψ TP: $\lambda_{262\text{ nm}} = 7550\text{ l mol}^{-1}\text{ cm}^{-1}$; $m^1\Psi$ TP: $\lambda_{271\text{ nm}} = 8870\text{ l mol}^{-1}\text{ cm}^{-1}$). The IVT reactions were allowed to progress for 2 h at 37°C before termination by the addition of DNase I following the manufacturer's protocol. Further studies on T7 RNA polymerase NTP selection were conducted with N1-ethylpseudouridine triphosphate ($e^1\Psi$ TP: $\lambda_{271\text{ nm}} = 7800\text{ l mol}^{-1}\text{ cm}^{-1}$) or N1-propylpseudouridine triphosphate ($p^1\Psi$ TP: $\lambda_{271\text{ nm}} = 8900\text{ l mol}^{-1}\text{ cm}^{-1}$; Trilink Biotechnologies) similarly to those described above. Replicate experiments were conducted to obtain errors.

Nanopore library preparation and sequencing

The poly-A tail containing RNAs generated by IVT or solid-phase synthesis were the input strands in the direct RNA sequencing kit (SQK-RNA002) from Oxford Nanopore Technologies (ONT). The protocol was followed without changes and the library-prepared samples (1–5 ng) were directly used for sequencing. The samples were applied to the ONT Flongle™ flow cell running the R9.4.1 chemistry following the manufacturer's protocol. The default settings were used with passed reads having a *Q* score >7.

Data analysis

The ionic current vs. time traces in fast5 file format passed by the sequencer were base called using guppy v.6.0.7 or v.6.3.2 to obtain the fastq sequencing read files used in the subsequent data analyses. The fastq files were aligned to the reference sequences using minimap2 with the command line '-ax map-ont -L' to generate aligned reads in bam file format (18). The bam file alignment statistics were determined with the flagstat function in Samtools and then the files were indexed with Samtools for visualization in Integrative Genomics Viewer (IGV) to obtain the base call information at the modification sites (19,20). Inspection of the sequencing reads to quantify the occupancy of U versus Ψ or $m^1\Psi$ was conducted using the ELIGOS2 or Nanopore-Psu tools (12). Calibration curves with known mixtures of U vs. Ψ or $m^1\Psi$ reads were obtained by first aligning the individual reads of the control RNAs to the reference genome. The number of aligned reads was determined using Samtools flagstat function. The pure U and Ψ or $m^1\Psi$ reads were then mixed in known ratios and submitted to ELIGOS2 or Nanopore-Psu for predicting the occupancy of the U isomer in each sequence context studied in the polymerase running start assay. This approach was important because the tools were designed to inspect data for Ψ occupancy, not $m^1\Psi$. A similar approach as described was used for the $e^1\Psi$ and $p^1\Psi$ studies. The data were plotted and analyzed in either python, Origin, or Excel for visualization.

RESULTS

Experimental setup

The RNA strands used for the study of base calling analysis at U, Ψ or $m^1\Psi$ were generated by IVT or solid-phase synthesis (Supplementary Figure S1). The use of commercially available $m^1\Psi$ TP or Ψ TP allowed the generation of RNAs by IVT with 100% incorporation of the modified nucleotides. The RNA strands sequenced had the U, Ψ or $m^1\Psi$ spaced >25 nts apart such that only one modification site interacted with the helicase-nanopore setup at a time. All k-mers of five nucleotides with a central U, Ψ , or $m^1\Psi$ with the other four nucleotides comprised of A, C or G were studied (5'-VVXVV-3'; 81 contexts). Additionally, 18 contexts were studied with two modifications adjacent to one another (5'-VVXXVV-3') with similar sequence context requirements as previously described. A limitation to IVT generation of RNA strands is sequence contexts with the parent nucleotide and modification cannot be generated; therefore, to study four k-mers that had the central U,

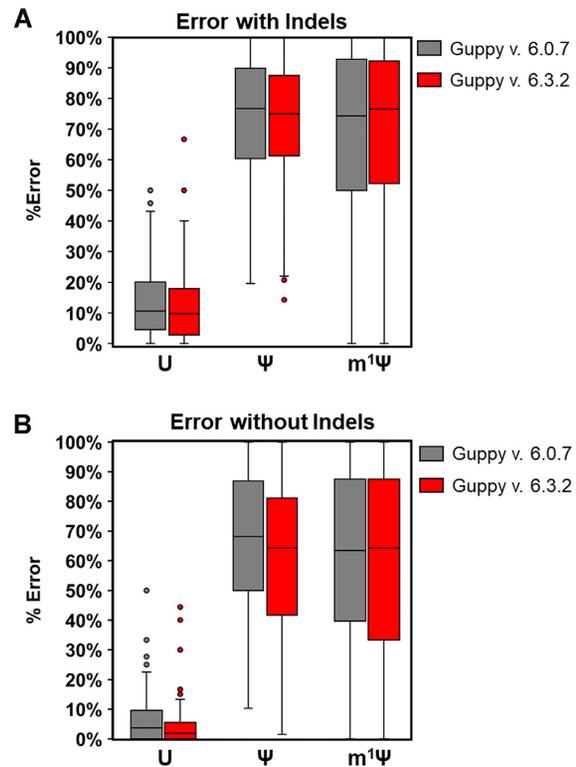


Figure 2. Analysis of base calling error from direct RNA nanopore sequencing data for strands with U, Ψ , or $m^1\Psi$ in 104 different 5-nt *k*-mer sequence contexts. The data for the plots are provided in Supplementary Figure S5. Guppy 6.0.7 (gray) and 6.3.2 (red) were used to base call the raw nanopore data. The values plotted had a sequencing depth of 10 or more.

Ψ , or $m^1\Psi$ in a context that included U, the RNA strands were made by solid-phase synthesis (*k*-mers = 5'-AUXAA-3', 5'-GAXUA-3', 5'-GUXGA-3' and 5'-AGXUG-3') and then 3' poly-A tailed enzymatically to enable library preparation. In total, 104 different 5-nt *k*-mer contexts in RNA were sequenced with U, Ψ or $m^1\Psi$ present (Supplementary Figure S1).

Base-calling error analysis

Nanopore sequencing was conducted on the RNA strands following standard protocols to generate current vs. time traces that were base called with guppy (v 6.0.7 or 6.3.2). The base-called data were aligned to the reference with minimap2 (Supplementary Figure S2), and visualization of the alignment was achieved with IGV (Supplementary Figure S3) (20). First, the base-calling errors identified by the presence of calls for C, A, G or insertion and deletions (indels) for the U, Ψ and $m^1\Psi$ RNAs were quantified. The data for *k*-mers that were sequenced at a depth of 10 or more are visualized in two different ways in Figure 2. The first displays the total error with indels included, and the second looked specifically at base calling errors without indels (Figure 2A and B; see Supplementary Figure S4 for data used in these plots). The U-containing RNA strands gave low percent errors in both analysis approaches, as expected, because the base calling algorithm was trained on RNA with this nucleotide (Figure 2A). The percent error for each of the

sequence contexts possessing Ψ ranged from $\sim 10\%$ to 100% when indels were considered (Figure 2A), and ranged from $\sim 20\%$ to 100% when indels were omitted (Figure 2B). The Guppy base caller version did not significantly impact the findings (Figure 2A–B and Supplementary Figure S4). In all cases, the average error decreased when indels were not considered (Figure 2A versus 2B). The percent base calling error for RNA with the $m^1\Psi$ modification also produced a similar range of errors from 10–100% when indels were considered (Figure 2A), and ranged from $\sim 25\%$ to 100% when indels were omitted (Figure 2B). Again, the Guppy base caller version did not alter the findings (Figure 2A–B and Supplementary Figure S4). The Ψ and $m^1\Psi$ base calling errors were not identical, but in general, the high error vs. low error sequences gave similar groupings (Supplementary Figure S5). The base call analysis identified modification of U to either $m^1\Psi$ or Ψ generated base-called data with much higher error.

The base calling data were then inspected more closely to determine the main modes of error in base calls that occur when Ψ or $m^1\Psi$ are directly sequenced with the nanopore system. Previous studies from our work and others found Ψ is miscalled as a C (11–13,15). These new data for Ψ support the previous observations, and now show that $m^1\Psi$ follows a similar error profile resulting in a higher frequency of C calls instead of U. The interesting observation that we re-confirmed for Ψ and now demonstrate for $m^1\Psi$ is that the error is highly dependent on the sequence context. A few additional observations are the percent error is generally lower when the sequence context has a greater nucleotide diversity around the site studied (Supplementary Figure S5). In contrast, k-mers with less sequence diversity generally give greater base calling error in the form of C miscalls and indels. For example, the high sequence diversity context 5'-CGXAC-3' produced the base call profile for Ψ of 92% U, 1% C, 1% A and 6% indels, and for $m^1\Psi$ the base calls were 85% U and 15% indels; in contrast, for the low diversity sequence context 5'-CCXCA-3', Ψ was called 75% C and 25% indels, and $m^1\Psi$ was called 100% as C (Supplementary Figure S5). Two adjacent modifications gave mixed results on the percent error that appears to follow the sequence diversity observations just discussed.

T7 RNA polymerase NTP selection for mixtures of UTP with Ψ TP or $m^1\Psi$ TP

The knowledge of nanopore sequencing signatures for Ψ or $m^1\Psi$ enabled exploration of NTP selection by T7 RNA polymerase during IVT. Currently, mRNA vaccines are produced by T7 RNA polymerase-mediated transcription, in which all the U nucleotides are completely replaced with $m^1\Psi$ (6). A prior study found the partial replacement of the canonical nucleotides with modified forms (m^5C and s^2U) in therapeutic mRNAs generated by IVT can be effective (21); however, the studies did not know whether there existed sequences that T7 RNA polymerase favored or disfavored for insertion of the non-canonical NTPs. Therefore, we addressed this question for UTP vs. Ψ TP or $m^1\Psi$ TP mixtures during T7 RNA polymerase synthesis of an mRNA using direct RNA nanopore sequencing as the readout for incorporation yields. The duplex DNA tem-

plates studied provided coding potential to interrogate all possible immediate sequences contexts in singly-modified and doubly-modified contexts (5'-VXXV-3' and 5'-VXXV-3'), which are all found in the BioNTech/Pfizer and Moderna vaccines (Supplementary Figure S6).

The RNA polymerase evaluation for NTP selection was first conducted with a 1:1 ratio of UTP to Ψ TP or UTP to $m^1\Psi$ TP. To ensure the competing NTPs were mixed in a 1:1 ratio, the stock solution concentrations were determined by UV-vis spectroscopy using extinction coefficients provided by the manufacturer of the nucleotides. The RNA strands were synthesized and then analyzed by gel electrophoresis to confirm the RNA strands synthesized with U, Ψ , or $m^1\Psi$ had the same length (Supplementary Figure S7). The confirmed RNA strands were prepared for sequencing and then sequenced with the nanopore using the default settings (pass = $Q > 7$). The obtained data were then used for quantification of Ψ or $m^1\Psi$ in each sequence context.

Quantification of RNA modifications from nanopore sequencing data can be approached by inspection of the current level data or the base calling data. The current levels for Ψ or $m^1\Psi$ relative to U were visualized using the Tombo tool that also allows quantification of modifications (Supplementary Figure S8). A challenge reconfirmed in the present studies is that when current level differences are observed, they generally do not occur when the modification is in the center of the 5-nt k-mer. The position of maximal current level difference between canonical and modified differs is dependent on the sequence context; moreover, some sequence contexts do not give current differences. As a result, using currents to monitor modification levels was not usable across all sequence contexts.

Inspection of the base calling data for identification of modifications by systematic errors introduced by the base caller when reading modified nucleotides is an alternative approach. Many tools are reported to conduct this analysis and two of them were used in the present study. The first is Nanopore-Psu, which was calibrated for the sequence space in which Ψ exists using rRNA from many species. The program predicts modification levels of unknown samples using the calibrated data. Another approach is provided by ELIGOS2 that compares base calling profiles between an RNA with modifications to a sequence-matched control void or diminished in modifications. The ELIGOS2 tool was selected because the RNA strands were judiciously designed and not of biological origin and are depleted in U nucleotides, which are critical differences compared to the rRNA for which the Nanopore-Psu tool was developed. Lastly, for each sequence context studied, calibration curves were developed by virtually mixing the pure U-containing RNAs with either the pure Ψ or $m^1\Psi$ RNAs in known ratios and submitting them to ELIGOS2 for analysis (Supplementary Figure S9). These calibration curves were used for quantification.

The U versus Ψ competition for the T7 polymerase active site in singly-modified contexts found Ψ was preferentially installed with a 40–70% yield (Figure 3A). A control was conducted with the same adjacent sequence context but a different 5-nt k-mer (GXC = 5'-AGXCA, and GXC-2 = 5'-CGXCG) to determine if this impacted the results (Figure 3A and C). The average measured Ψ TP

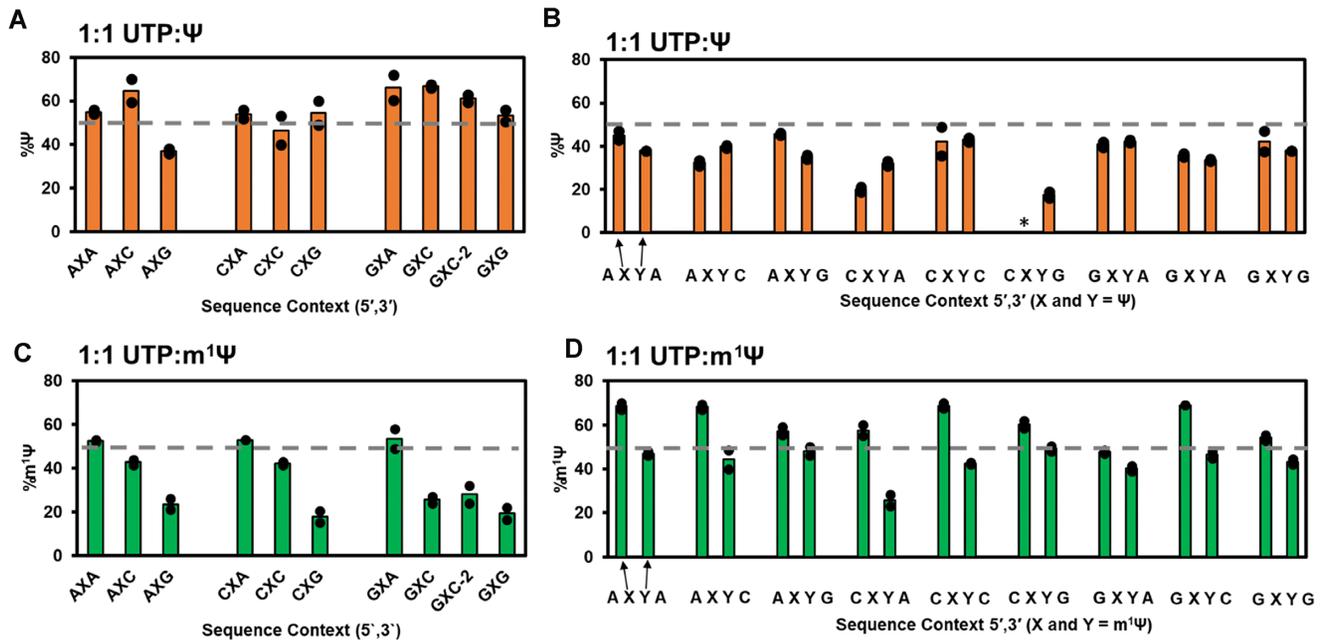


Figure 3. Yields of Ψ TP or $m^1\Psi$ TP incorporation when competed with an equimolar ratio of UTP for insertion and elongation by T7 RNA polymerase. Percent insertion yields for Ψ TP in (A) singly-modified or (B) doubly modified sequence contexts. Percent insertion yields for $m^1\Psi$ TP in (C) singly-modified or (D) doubly-modified sequence contexts. The expected yield assuming no sequence bias introduced by T7 RNA polymerase is shown by the dashed gray lines. The yields were determined via direct RNA sequencing with the commercial nanopore platform, and the base calling data were analyzed with the published tool ELIGOS2 (14). *A value for the insertion was not measured in the analysis.

levels inserted in these contexts were $GXC = 67\%$ and $GXC-2 = 61\%$, which were not significantly different ($P > 0.05$); however, this demonstrates there exists some variability in these measurements with the assumption T7 RNA polymerase selects NTPs with selectivity that only extends to the adjacent nucleotides at the competition site. The data were analyzed with Nanopore-Psu to give similar results in many sequence contexts (Supplementary Figures S10 and S11). In the same competition but for sites in which the template DNA strand codes for insertion of two adjacent U/ Ψ residues, the first insertion at the 5' site slightly favored Ψ TP ($>60\%$) insertion, while the second insertion at the 3' site gave a slight reduction in Ψ TP (Figure 3B). Exceptions were observed with the most noteworthy being the 5' site of the 5'-CXXG-3' context, in which no detectable Ψ TP insertion was measured (Figure 3B see '*'). For this sequence context the current level analysis by Tombo found Ψ did not alter the current strongly compared to U, which address the fact that there was no detectable signal in the unknown sample (Supplementary Figure S8). Nonetheless, this study identified T7 RNA polymerase selects competing UTP and Ψ TP with sequence context bias (Figure 3A–B).

Nanopore sequencing results for the UTP vs. $m^1\Psi$ TP competition for T7 RNA polymerase incorporation deviated from the UTP versus Ψ TP experiments. In sites where the template codes for a single modification, $m^1\Psi$ TP was installed with $<40\%$ yield with one exception; when the template DNA coded for ATP insertion 3' to the modified nucleotide, T7 RNA polymerase inserted UTP and $m^1\Psi$ TP with a yield similar to their solution concentrations (i.e. 1:1 or $\sim 50\%$ yield; Figure 3C). The sequence control experiments for two different k -mers with the same central se-

quence context (5'-GXC), were determined to have nearly identical $m^1\Psi$ TP insertion yields (Figure 3C). Sites that could incorporate two UTP or $m^1\Psi$ TP adjacent to one another led to a different result (Figure 3D). On the 5' site, $m^1\Psi$ TP was favorably inserted and the 3' site was disfavored ($\sim 65\%$ versus 50% ; Figure 3D). The ability to directly sequence RNA with nanopores and quantitatively call modifications has led to this discovery of the sequence-dependent bias that T7 RNA polymerase has for the selection of UTP versus $m^1\Psi$ TP.

T7 RNA polymerase can bias competing NTP selection based on the sequence context that was most profound for UTP vs. $m^1\Psi$ TP (Figure 3). Next, experiments to interrogate the RNA polymerase, duplex DNA template, and NTP identity were conducted to understand the bias (Figure 4A). Structural analysis of T7 RNA polymerase has found the active site features required for catalysis and proper NTP selection. The focus is on positions 639 and 644 of the polymerase (Figure 4B). Position 639 is a tyrosine residue required for proper NTP discrimination over dNTPs, while position 644 is phenylalanine that π stacks with the template DNA nucleotide of the still formed base pair 3' to the site directing insertion of the incoming NTP (22–26). There exists another phage-derived DNA-dependent RNA polymerase, SP6, with again Y639 while position 644 is a leucine that cannot π stack with the template nucleotide on the 3' base pair (Figure 4B). The UTP vs. $m^1\Psi$ TP competition assay in each sequence context was repeated with SP6 RNA polymerase. This new experiment found that when UTP and $m^1\Psi$ TP competed in a 1:1 ratio, the insertion yield for $m^1\Psi$ did not show a sequence context bias (Figure 4C). The $m^1\Psi$ yields within error across all sequence contexts were

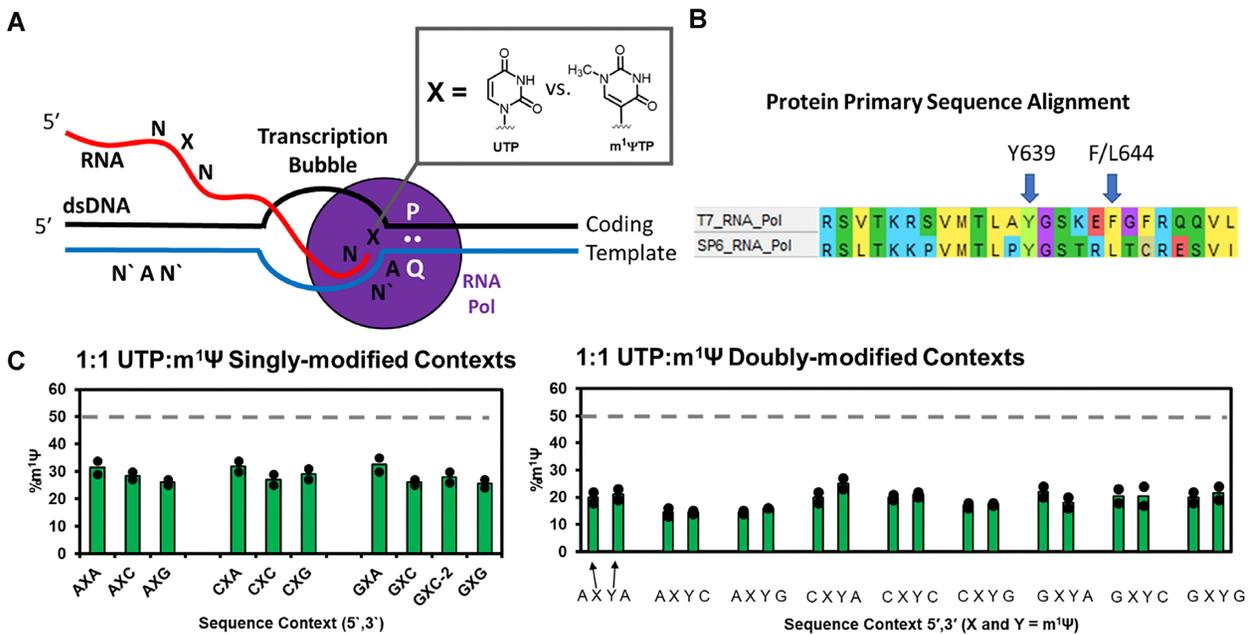


Figure 4. The transcription of mRNA when a mixture of UTP and m¹ΨTP is present to compete for the SP6 RNA polymerase active site shows minimal sequence context impact on the yield. (A) Model of a transcription bubble to illustrate the DNA-protein interactions involved in NTP selection. P:Q is the intact DNA base pair adjacent the insertion site. (B) Sequence alignment for the phage DNA-dependent RNA polymerases T7 and SP6. (C) The yields of m¹Ψ inserted and extended by SP6 RNA polymerase when competed in equal molar ratio with UTP. The expected yield assuming no sequence bias introduced by T7 RNA polymerase is shown by the dashed gray lines. The yields were determined via direct RNA sequencing with the commercial nanopore platform and the base calling data were analyzed with the published tool ELIGOS2 (14).

20–30% (Figure 4C), a range much reduced from that observed with T7 RNA polymerase (20–70%; Figure 3C and D). The SP6-catalyzed polymerization was repeated with a 3:1 m¹ΨTP:UTP ratio to identify the yield of m¹Ψ nearly doubled while maintaining the minimized sequence context bias for insertion yields (Supplementary Figure S12). The SP6 RNA polymerase study was also conducted with UTP vs. ΨTP to find differences compared to T7 RNA polymerase (Supplementary Figure S13). This study verifies that the RNA polymerase impacts NTP selection, as expected, and by changing the active site of the RNA polymerase, the NTP sequence context bias can be minimized for the UTP versus m¹ΨTP case.

The next focus was on the nucleic acids (i.e. dsDNA template and NTPs) and their role when two competing NTPs were allowed to gain access to the active site of the T7 RNA polymerase, particularly the unusual observation of having a 3' A in the RNA strand (or the corresponding dT in the DNA template strand) impacting the UTP versus m¹ΨTP selection. In the most favorable case for m¹ΨTP insertion, the template DNA strand has a dA nucleotide for the pyrimidine competition followed by a dT to direct insertion of ATP (Figure 4A, Q = dT). Structural analysis of T7 RNA polymerase found the DNA base pair 3' on the template strand (i.e. Figure 4A, P:Q) to the competition site is still formed with F644 π stacking with the dT nucleotide (22–26). This effect is not observed when the other pyrimidine dC (Figure 4A, Q = dC) is in this position to direct a GTP installment 3' to the competition site. The main difference in these base pairs is the number of hydrogen bonds.

To study this T7 RNA polymerase active site base pair interaction, a duplex DNA was designed, synthesized, and used for IVT that had a single DNA templating dA nucleotide for the UTP versus m¹ΨTP competition followed by a dC in position Q (Figure 4A). The other change was in the coding DNA strand in which the native dG that forms three hydrogen bonds with dC was replaced with 2'-deoxyinosine (dI; Figure 4A, P = dI) for pairing with two hydrogen bonds with the template dC (P:Q = dI:dC, Figure 4A). This situation would have a weaker base pair (two hydrogen bonds) but similar π stacking (dC) interaction with F644 in T7 RNA polymerase and direct the insertion of a G 3' to the U/m¹Ψ competition site (Figures 4B and 5A). The nanopore sequence analysis of the RNA transcript found in the weakened 3' base pair system (dI:dC) in comparison to the canonical system (dG:dC) the incorporation of m¹Ψ increased by >2-fold (Figure 5B). A control experiment was conducted with the dI in the template DNA strand and the results were compared with a dG in the same position to find minimal impact on the m¹Ψ insertion yields (Supplementary Figure S14). This observation supports the conclusion that base pair strength and F644 π stacking with the pyrimidine dT impact NTP selection.

The final study focused on the role of the non-canonical NTP to impact the selection when competing for T7 RNA polymerase insertion and extension with UTP. The alkyl group identity at N1 of ΨTP was varied from methyl, ethyl, to propyl (i.e. e¹ΨTP or p¹ΨTP), in which these Ψ derivatives have been proposed for use in therapeutic mRNAs (27). First, RNA was sequenced with pure N1-alkylpseudouridine derivatives and the base calling error

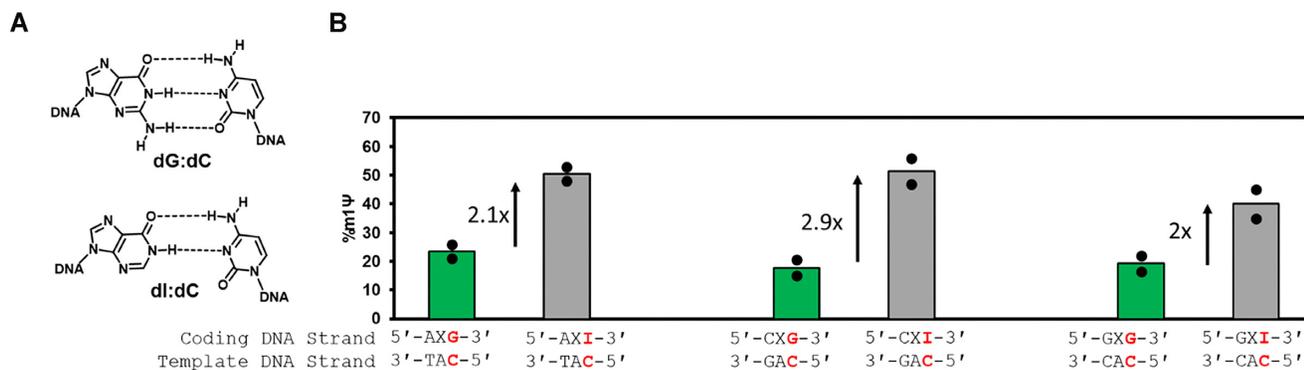


Figure 5. The base pair identity in the duplex DNA template on the 3' side of the templating nucleotide for an T7 RNA polymerase impacts the polymerization yield. (A) Structures for a dG:dC and dl:dC base pair illustrate they are similar except in the number of hydrogen bonds. (B) The yields of m¹Ψ incorporated when competed with UTP and there is a dG:dC versus dl:dC base pair in the duplex DNA 3' to the templating site for RNA polymerase. The yields were determined via direct RNA sequencing with the commercial nanopore platform and the base calling data were analyzed with the published tool ELIGOS2 (14).

analysis showed that the error profiles for e¹Ψ and p¹Ψ were similar to Ψ and m¹Ψ (Supplementary Figure S15). These alkyl derivatives were individually studied in equimolar ratio with UTP during IVT mRNA synthesis (Figure 6A). The RNAs made were then directly sequenced, and the yield of non-canonical NTP insertion was quantified via the base calling analysis using ELIGOS2 that was calibrated with each N1-alkylpseudouridine derivative (Supplementary Figure S9) (12). Only singly-modified sequence contexts were analyzed to find the range of insertion yields across the sequence contexts decreased with the larger N1-alkyl derivatives (Figure 6B). Specifically, the variability of yields for m¹Ψ was 42%, e¹Ψ was 13%, and p¹Ψ was 17%. This final dataset demonstrates all three components of the transcription process, polymerase, template DNA and NTP play a role when competing NTPs can be selected during mRNA synthesis.

DISCUSSION

Direct RNA sequencing with a nanopore for Ψ or m¹Ψ

The nanopore system can directly sequence RNA to enable detection and quantification of chemically modified nucleotides. The ONT nanopore sequencer utilizes the CsgG protein nanopore (v 9.4.1 flow cells) that has a sensing window or k-mer for RNA of ~5 nt (8) within which modifications can impact the sequencing signal. The RNA nucleotides U, Ψ and m¹Ψ in different 5-nt sequence contexts were sequenced, and the base calling errors were identified. The dominant and unique base calling signature for Ψ is to be called as a C (Figures S4 and S5), which was found herein and previously (11,13,15,16). The methylated base m¹Ψ also predominantly generated C miscalls (Supplementary Figures S4 and S5). The magnitude of the error for both Ψ and m¹Ψ was highly dependent on the sequence context (Figure 2, Supplementary Figures S4 and S5). Direct RNA nanopore sequencing for Ψ has been reported and tools developed to quantify the data to determine the extent of occupancy at modified sites (11–16). These tools can easily be applied to study m¹Ψ as was done in the present work.

A few noteworthy points arise regarding the sequence space of 5-nt k-mer sequences with a single RNA modification. The discussion that follows is the sequence space for differentiation of U from Ψ or from m¹Ψ, but not both. A k-mer with a single, central modification (5'-NNXNN-3'; where N = A, C, G or U and X = U or Ψ/m¹Ψ) would span 512 different sequence contexts. If we consider the modification can exist more than once in the k-mer, this would contain 3125 sequence contexts. Pseudouridine in rRNA and m¹Ψ in mRNA vaccines does exist with more than one modification in a five-nucleotide window (Supplementary Figure S6) (6,28); however, a single RNA modification in a k-mer is the most common presentation of these modifications in native mRNA (12,29). One approach employed to study this sequence space is to synthesize mRNA sequences by IVT that contain all possible k-mers, which has been done, with the limitation that k-mers with both the canonical base and its modified version present cannot be studied (11). Alternatively, mRNA strands have been synthesized by ligating a synthetic RNA with the canonical and modified nucleotide present into a longer construct for study (16). This approach provides a more realistic representation of modifications but requires laborious synthesis and low throughput. Thus, we took a middle approach to survey >100 representative k-mers using IVT or synthesis. As an aside, sequencing small oligomers with the nanopore is achievable, the data obtained fail at a very high rate compared to the longer sequences; nonetheless, the data do provide an indication of the base call error when canonical and non-canonical are present in the same k-mer (~50% error; Supplementary Figure S4).

The similarity in base calling features for Ψ and m¹Ψ point to a challenge for direct RNA sequencing for modifications with nanopores, and that is that different modifications can yield very similar signatures. After exploring the nanopore signatures for the >150 known modifications, the community might find, for example, that a particular U is modified in an unknown sample but the question of what the modification is will remain. Parallel biological and sequencing studies on native cells and those with suspected writer protein knockouts or knockdowns can aid in the

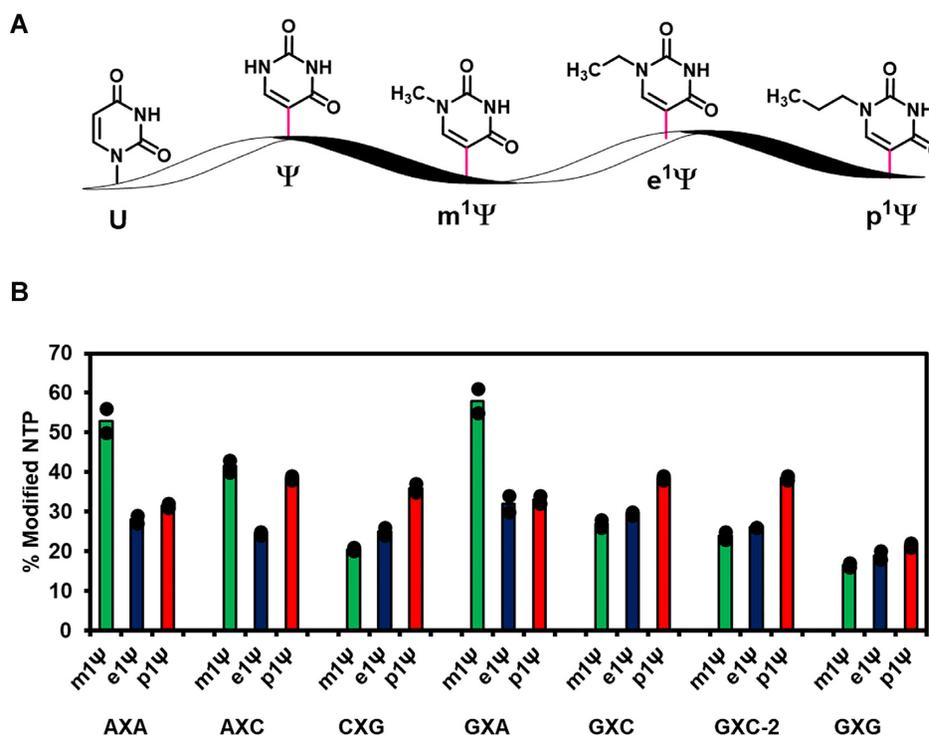


Figure 6. Comparison of the incorporation yields when UTP was allowed to compete with either m¹ΨTP (blue), e¹ΨTP (green) or p¹ΨTP (red) for RNA polymerization by T7 RNA polymerase. (A) The structures for U, Ψ and the N1-alkyl Ψ derivatives. (B) The incorporation yields for the NTP competition assays. The yields were determined via direct RNA sequencing with the commercial nanopore platform and the base calling data were analyzed with the published tool ELIGOS2 (14).

identification of the modification (8,11). Alternatively, the use of low-throughput and targeted assays such as SCAR-LET or mass spectrometry sequencing will be needed to identify the modification, and even these will be challenged by modification isomers (e.g. m⁵C, m⁴C and C_m) (30,31). Lastly, when natural U→C variations exist in the RNA they can masquerade as a Ψ or m¹Ψ when inspecting base calling features exclusively; we proposed a solution to this challenge by inspecting the raw ionic current vs. time traces for Ψ to minimize the false discovery rate (13). A similar approach would likely work for m¹Ψ, but this was not conducted in the present studies because the modification identity was always known.

T7 RNA polymerase selectivity for ΨTP when competing with UTP

Setting aside the challenges of sequencing native RNA for modifications, the goal here was to use the nanopore sequencer as a tool to monitor RNA polymerase selection of competing NTPs. The situation in which UTP versus ΨTP or m¹ΨTP competition would exist is during therapeutic RNA synthesis via IVT in which sub-stoichiometric amounts of modified NTP incorporation might be the goal. There is precedence for the successful use of therapeutic mRNAs with sub-stoichiometric amounts of modifications (m⁵C and s²U) (21). Soon there could be a growing need to synthesize mRNA with a mixture of the parent canonical and modified form of an NTP via T7 RNA polymerase. N1-Methylpseudouridine is the modification used in SARS-

CoV-2 mRNA vaccines to minimize immunogenicity of the foreign RNA (6); Ψ itself can also serve this purpose (32).

Previously reported details regarding T7 RNA polymerase will aid in understanding the results of the NTP competitions conducted. The phage RNA polymerase T7 is a single subunit DNA-dependent RNA polymerase, it does not have an exonuclease domain for proofreading, and with these limitations, the polymerase maintains high fidelity transcription (1 error in 10⁴ NTPs polymerized) (33). How T7 RNA polymerase maintains high fidelity RNA synthesis has been addressed by x-ray crystallography and computational studies (22–26). Experimental work identified that T7 RNA polymerase easily accepts modified NTPs (34), NTP elongation kinetics are influenced by steric factors (35), and NTP selection is impacted in solutions where the dielectric constant and water activity have been altered with PEG-200 (36). The key points in discrimination of the incoming NTP are to ensure the exclusion of dNTPs by checking the presence of the 2'-OH group via Y639, and that the NTP and templating DNA nucleotide form a viable base pair to maintain the fidelity of the mRNA. How does T7 RNA polymerase discriminate between UTP vs. ΨTP or m¹ΨTP that can all form viable base pairs with the dA nucleotide in the template DNA strand?

In the 1:1 UTP versus ΨTP competitions for the T7 RNA polymerase active site in singly modified sites, the yield of ΨTP incorporation was generally greater than the expected 50% yield (Figure 3A). We hypothesize the reason for the favorability of ΨTP incorporation is the greater freedom for *syn* and *anti*-glycosidic bond angles that favors the *syn* con-

formation (37). Both conformations of Ψ TP display a face to the templating dA nucleotide that can form two hydrogen bonds with similar base-pair shapes (Figure 7A). Thus, Ψ TP is favorably inserted because both base conformations, *syn* or *anti*, relative to the ribose yield viable dA base pairs, whereas UTP can only base pair with dA in the *anti* conformation (Figure 7B). This claim is supported by modeling work on T7 RNA polymerase that found phosphodiester bond formation during polymerization can only occur when the base pair has the right size and shape (26). When the template DNA strand had two adjacent dA nucleotides to direct insertion of the competing NTPs, Ψ TP insertion was reduced across all sequence contexts relative to the singly modified sites (Figure 3A and B). The most important observation is that Ψ TP insertion yields had a variability of nearly 50% from high to low across the sequence contexts demonstrating the nucleotide cannot be installed equally across all contexts (Figures 3A and B). Detailed structural studies would need to be conducted to address on the molecular scale the context dependency of Ψ incorporation by T7 RNA polymerase.

Considering $m^1\Psi$ TP vs. UTP for incorporation by T7 RNA polymerase, the findings differed compared to competitions with Ψ TP. In the sequence contexts that coded for a single U/ $m^1\Psi$, UTP was favorably selected, with the exceptions being those with a 3' A in the RNA that gave U and $m^1\Psi$ incorporation at a 1:1 ratio reflecting their relative concentrations in solution. There are two mysteries regarding these observations. (i) Why does UTP outcompete $m^1\Psi$ TP in all sequence contexts except one? (ii) Why does a 3' A in the RNA result in higher yields for the modified nucleotide triphosphate?

The Chow laboratory used NMR NOE measurements to report on the *syn* vs. *anti* conformational equilibrium for $m^1\Psi$ in the nucleoside context (37). They found $m^1\Psi$ had a greater preference for the *syn* conformation than Ψ . Unlike Ψ , $m^1\Psi$ *syn* cannot base pair with dA in the DNA template strand (Figure 7C); therefore, the dominant glycosidic bond conformation for this modification is not compatible with T7 RNA polymerase to catalyze phosphodiester bond formation. This glycosidic bond angle preference for $m^1\Psi$ results in UTP winning the competition for incorporation and elongation by this DNA-dependent RNA polymerase.

Regarding the 3' A effect where $m^1\Psi$ TP and UTP were selected based on their solution concentrations, the structures for the steps of NTP selection and polymerization by T7 RNA polymerase offer a clue that could be tested (22–26). The active site of T7 RNA polymerase is maintained in an open conformation by the O/O' helix during NTP selection that closes to generate the contacts needed for NTP discrimination and phosphodiester bond formation (22,25). Computational studies suggest closing of the O/O' helix is thermally regulated and that the open conformation provides space for NTPs to sample glycosidic bond conformations to yield viable base pairs for polymerization in the growing mRNA strand (24). In the open and closed conformations, the DNA base pair 3' to the templating nucleotide is still formed, and a π -stacking interaction exists between F644 of the O/O' helix, and the template DNA nucleotide (Figure 8A and B). In these 3'-A cases, the π stack would be with a dT on the template DNA strand to form

the weakest interactions possible, a π stack with a pyrimidine that has two hydrogen bonds with its complementary dA nucleotide. This may allow greater sampling of the open and closed states of the O/O' helix providing space and time for $m^1\Psi$ TP to find the *anti* conformation to base pair with the template dA and to result in successful phosphodiester bond formation (Figure 7C).

Support for these interactions resulting in UTP and $m^1\Psi$ TP equally competing when a 3' A is in the mRNA is derived from two studies. In the first, we switched the RNA polymerase to SP6, which maintains the Y639 NTP/dNTP proofreading ability but does not contain the F644 π stacking interaction and instead has a leucine at this position that cannot π stack (Figure 4B). When SP6 is the polymerase, $m^1\Psi$ TP is inserted at 20–30% yield in all adjacent sequence contexts, both singly and doubly modified. The reason why the yield is not \sim 50% may result from other differences in physical properties of UTP vs. $m^1\Psi$ TP that were not studied. Regardless, this observation means that to tailor the stoichiometry of canonical and modified nucleotides in a synthetic mRNA at all sites, the relative solution concentrations of the competing NTPs can be dialed in to the desired yields, which was found herein (Supplementary Figure S12).

A second experiment to support the F644 π stacking interaction with dT paired with dA on the 3' side of the competition is derived from a rationally designed study in which we altered the 3' A observation into a 3' G effect. In the coding DNA strand that would place a G on the 3' side of the U/ $m^1\Psi$ site, the native dG with three hydrogen bonds with dC was replaced with dI to generate a similar base pair with dC but with only two hydrogen bonds. This situation mimics the weaker A:T base pair and have a similar π stacking (dC) interaction with F644, which should direct the insertion of a G 3' to the U/ $m^1\Psi$ competition site (Figure 5A). The nanopore sequence analysis found in the weakened 3' base pair system in comparison to the dG:dC base pair that the incorporation of $m^1\Psi$ increased by $>$ 2-fold (Figure 5B). This observation provided more support for the DNA-protein interaction impacting selection of competing NTPs.

As a final study regarding the T7 RNA polymerase competition assay, UTP was mixed in equal ratios with either $e^1\Psi$ TP or $p^1\Psi$ TP during IVT (Figure 6A). The synthetic RNAs were then directly sequenced, and the modification inserted was quantified via the base calling analysis using Nanopore-Psu (12). Only singly-modified sequence contexts were analyzed. The analysis revealed that as the alkyl group length increased at N1 of pseudouridine, the incorporation yield ratio for the modified NTP increased in all contexts except those with a 3' A (Figure 6B). Prior work with long alkyl groups attached to the uridine nucleoside found them to adopt equal populations of *syn* and *anti* conformations (37). As the ratio of *anti* conformation increases in the methyl, ethyl, propyl N1-alkylpseudouridine series, the favorability of competing with *anti* UTP increases, as observed. The 3' A sites all had an \sim 50% yield for insertion of the modification as a result of the unique protein-DNA interactions described above.

Other methods for addressing RNA polymerase activity rely on gel-based (38), mass spectrometry-based (39), or

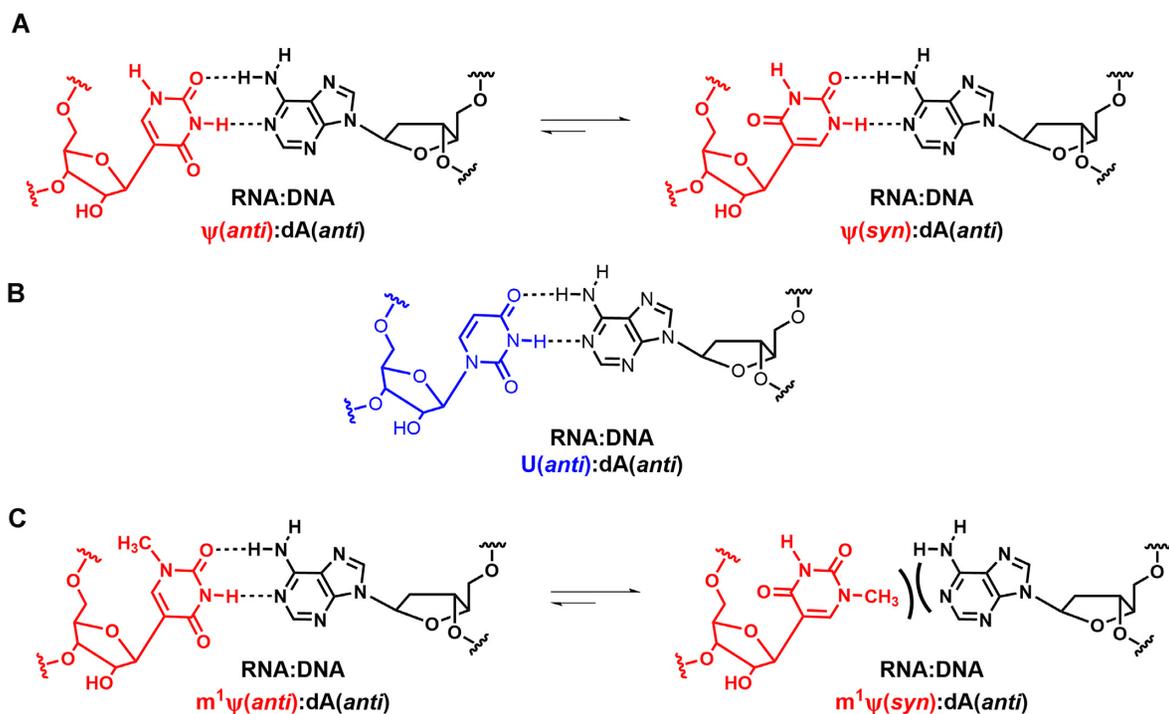


Figure 7. Base pairs formed by (A) Ψ , (B) U and (C) $m^1\Psi$ in RNA with a dA nucleotide in a template DNA strand.

sequencing-based assays (40); however, a drawback to these methods is seen when studying NTP selection during a running start assay in which the competition is between isocoding nucleotides such as UTP versus Ψ TP or $m^1\Psi$ TP. This is even more challenging for UTP versus Ψ TP because they are isomers and have the same mass. Direct RNA sequencing with the nanopore setup can provide a solution to these challenges in running start assays to enable quantification of U versus Ψ or $m^1\Psi$ in all adjacent sequence contexts. This study was possible because of the efforts of others to develop software to help interpret the data quantitatively (8,11,12,16). We used ELIGOS2 (14) and Nanopore-Psu (12) developed for the quantitative inspection of nanopore sequencing reads for Ψ , which was repurposed for $m^1\Psi$ quantification.

The work presented to monitor the running start polymerization of RNA nucleotides by T7 RNA polymerase via nanopore sequencing found a sequence context bias when UTP competed with $m^1\Psi$ TP. This observation identifies a challenge if the goal is to synthesize a therapeutic mRNA with sub-stoichiometric levels of $m^1\Psi$ with nearly equal representation in all sequence contexts. The studies identify DNA–protein interactions leading to this impact on sequence context. More importantly, by understanding the interactions, the SP6 RNA polymerase was identified to install $m^1\Psi$ TP with a similar yield across all sequence contexts (~25% when a 1:1 mixture of UTP and $m^1\Psi$ TP exists in solution). Based on the primary sequences for the BioNTech/Pfizer and Moderna vaccines, the present studies covered all sequence contexts except for three adjacent modifications (i.e. $(m^1\Psi)_3$; Supplementary Figure S6). The studies with two possible modification sites adjacent to one another allow us to predict the outcome when UTP and

$m^1\Psi$ TP compete. For T7 RNA polymerase, if a 3' A exists after the run of three, the last modification site will be lower in yield compared to the first two sites (Figure 3D); in contrast, SP6 RNA polymerase will install the three modifications with minimized sequence bias (Figure 4C). This provides one possible approach for the synthesis of mRNA vaccines with sub-stoichiometric modification levels that are similar in all sequence contexts. A minor challenge for using SP6 RNA polymerase is a slightly higher rate of mutations at A, C and G sites compared to T7 RNA polymerase when conducting IVT with $m^1\Psi$ TP present (40). Finally, a patent was filed for use of methyl, ethyl and propyl Ψ derivatives in therapeutic mRNAs (27). The nanopore sequencing of these modifications installed by T7 RNA polymerase catalyzed IVT may have relevance in the future.

CONCLUSIONS

Direct RNA sequencing with nanopores enables RNA modification analysis that has been conducted for nucleotides such as Ψ (11–14,16). In the present work, RNA containing either U, Ψ or $m^1\Psi$ in 83 different 5-nt sequence contexts that include four with canonical U present were nanopore sequenced. The base calling data obtained identified Ψ and $m^1\Psi$ behave similarly and, depending on the sequence context, are called as a C or generate indels (Figure 2). In the big picture, this analysis demonstrated different RNA modifications on the same nucleotide can yield similar signatures. This illustration should bring caution to RNA researchers using nanopore sequencing to conduct de novo analysis for chemical modifications. A site of modification can be found but the identity may remain questionable.

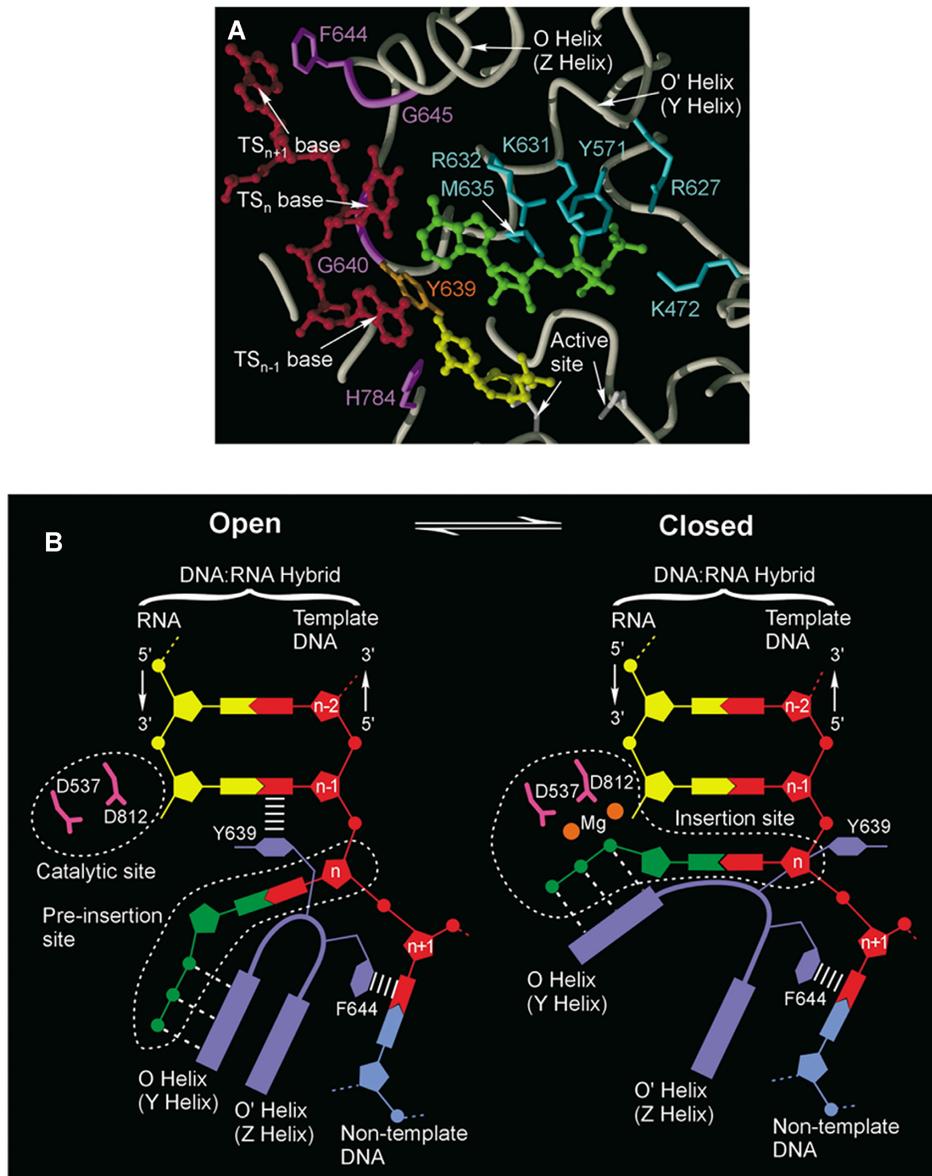


Figure 8. Previously reported structure for the T7 RNA polymerase active site in contact with the duplex DNA and incoming NTP (22). (A) Structure reported for key residues involved in NTP binding and template strand (TS) binding. (B) Schematic drawing of T7 RNA polymerase to show key contacts for substrate selection in the open conformation and catalysis in the closed conformation.

Understanding the nanopore sequencing data for Ψ or $m^1\Psi$ enabled a running start T7 RNA polymerase assay to be conducted to compete either Ψ TP or $m^1\Psi$ TP with UTP during IVT. The experiments found Ψ TP insertion yields were sequence-context dependent (Figure 3A and B). In contrast, UTP outcompeted $m^1\Psi$ TP in all contexts except when a 3' A occurred after the single modification site in the RNA, in which case the yields were the same as the NTP concentration ratio in the solution (Figure 3C). A model for the results is proposed based on the prior finding that $m^1\Psi$ strongly favors the *syn* conformation and cannot pair with dA in the *anti* conformation allowing UTP to outcompete the modification (37). The exception is when there is an A coded for in the RNA sequence 3' to the modification, which based on the solved structures for the poly-

merase (22,25), provides a more flexible active site to allow $m^1\Psi$ TP to adopt the *anti* conformation to pair with the templating dA. A change of the RNA polymerase to SP6 resulted in a more balanced incorporation of $m^1\Psi$ TP into the RNA in all sequence contexts when competing with UTP (Figure 4C). By changing the polymerase, a new solution for incorporating UTP and $m^1\Psi$ TP at a ratio with minimized sequence context dependency was found. The studies and information reported will be beneficial for the synthesis of therapeutic RNAs by IVT that have sub-stoichiometric levels of modifications present. Nanopore sequencing can aid in understanding many questions regarding DNA and RNA in cellulose, while the present studies demonstrate biochemical questions can also be addressed with this technology.

DATA AVAILABILITY

The data are available at Zenodo at public repository for data at doi: 10.5281/zenodo.7459451 searchable in the OpenAIRE explorer.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

The National Institutes of Health provided financial support for this project (R01 GM093099 and R35 GM145237). Oligonucleotide synthesis was provided by the University of Utah Health Sciences Core facilities that are supported in part by a National Cancer Institute Cancer Center Support grant (P30 CA042014).

FUNDING

National Institutes of Health [R01 GM093099, R35 GM145237]. Funding for open access charge: NIH.

Conflict of interest statement. A.M.F. and C.J.B. have a patent for nanopore sequencing licensed to Electronic BioSciences, and A.M.F. is a paid consultant at Electronic BioSciences advising on the chemistry of nucleic acids.

REFERENCES

- Jones, J.D., Monroe, J. and Koutmou, K.S. (2020) A molecular-level perspective on the frequency, distribution, and consequences of messenger RNA modifications. *Wiley Interdiscip. Rev. RNA*, **11**, e1586.
- Roundtree, I.A., Evans, M.E., Pan, T. and He, C. (2017) Dynamic RNA modifications in gene expression regulation. *Cell*, **169**, 1187–1200.
- Linder, B. and Jaffrey, S.R. (2019) Discovering and mapping the modified nucleotides that comprise the epitranscriptome of mRNA. *Cold Spring Harb Perspect. Biol.*, **11**, a032201.
- Netzband, R. and Pager, C.T. (2020) Epitranscriptomic marks: emerging modulators of RNA virus gene expression. *Wiley Interdiscip. Rev. RNA*, **11**, e1576.
- Hu, B., Zhong, L., Weng, Y., Peng, L., Huang, Y., Zhao, Y. and Liang, X.-J. (2020) Therapeutic siRNA: state of the art. *Sig. Transduct. Target Ther.*, **5**, 101.
- Nance, K.D. and Meier, J.L. (2021) Modifications in an emergency: the role of N1-methylpseudouridine in COVID-19 vaccines. *ACS Cent. Sci.*, **7**, 748–756.
- Branton, D. and Deamer, D. (2019) In: *Nanopore Sequencing An Introduction*. World Scientific Publishing Co. Pte. Ltd.
- Leger, A., Amaral, P.P., Pandolfini, L., Capitanich, C., Capraro, F., Miano, V., Migliori, V., Toolan-Kerr, P., Sideri, T., Enright, A.J. et al. (2021) RNA modifications detection by comparative nanopore direct RNA sequencing. *Nat. Commun.*, **12**, 7198.
- Furlan, M., Delgado-Tejedor, A., Mulrone, L., Pelizzola, M., Novoa, E.M. and Leonardi, T. (2021) Computational methods for RNA modification detection from nanopore direct RNA sequencing data. *RNA Biol.*, **18**, 31–40.
- Thomas, N.K., Poodari, V.C., Jain, M., Olsen, H.E., Akeson, M. and Abu-Shumays, R.L. (2021) Direct nanopore sequencing of individual full length tRNA strands. *ACS Nano*, **15**, 16642–16653.
- Begik, O., Lucas, M.C., Prysycz, L.P., Ramirez, J.M., Medina, R., Milenkovic, I., Cruciani, S., Liu, H., Vieira, H.G.S., Sas-Chen, A. et al. (2021) Quantitative profiling of pseudouridylation dynamics in native RNAs with nanopore sequencing. *Nat. Biotechnol.*, **39**, 1278–1291.
- Huang, S., Zhang, W., Katanski, C.D., Dersh, D., Dai, Q., Lolans, K., Yewdell, J., Eren, A.M. and Pan, T. (2021) Interferon inducible pseudouridine modification in human mRNA by quantitative nanopore profiling. *Genome Biol.*, **22**, 330.
- Fleming, A.M., Mathewson, N.J., Howpay Manage, S.A. and Burrows, C.J. (2021) Nanopore dwell time analysis permits sequencing and conformational assignment of pseudouridine in SARS-CoV-2. *ACS Cent. Sci.*, **7**, 1707–1717.
- Jenjaroenpun, P., Wongsurawat, T., Wadley, T.D., Wassenaar, Trudy M., Liu, J., Dai, Q., Wanchai, V., Akel, N.S., Jamshidi-Parsian, A., Franco, A.T. et al. (2020) Decoding the epitranscriptional landscape from native RNA sequences. *Nucleic Acids Res.*, **49**, e7.
- Smith, A.M., Jain, M., Mulrone, L., Garalde, D.R. and Akeson, M. (2019) Reading canonical and modified nucleobases in 16S ribosomal RNA using nanopore native RNA sequencing. *PLoS One*, **14**, e0216709.
- Makhamreh, A., Tavakoli, S., Gamper, H., Nabizadehmashhadroghi, M., Fallahi, A., Hou, Y.-M., Rouhanifard, S.H. and Wanunu, M. (2022) Messenger-RNA modification standards and machine learning models facilitate absolute site-specific pseudouridine quantification. bioRxiv doi: <https://doi.org/10.1101/2022.05.06.490948>, 06 May 2022, preprint: not peer reviewed.
- Pang, H., Ihara, M., Kuchino, Y., Nishimura, S., Gupta, R., Woese, C.R. and McCloskey, J.A. (1982) Structure of a modified nucleoside in archaeobacterial tRNA which replaces ribosylthymine. 1-Methylpseudouridine. *J. Biol. Chem.*, **257**, 3589–3592.
- Li, H. (2018) Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics*, **34**, 3094–3100.
- Danecek, P., Bonfield, J.K., Liddle, J., Marshall, J., Ohan, V., Pollard, M.O., Whitwham, A., Keane, T., McCarthy, S.A., Davies, R.M. et al. (2021) Twelve years of SAMtools and BCFtools. *GigaScience*, **10**, giab008.
- Robinson, J.T., Thorvaldsdóttir, H., Winckler, W., Guttman, M., Lander, E.S., Getz, G. and Mesirov, J.P. (2011) Integrative genomics viewer. *Nat. Biotechnol.*, **29**, 24–26.
- Kormann, M.S., Hasenpusch, G., Aneja, M.K., Nica, G., Lemmer, A.W., Herber-Jonat, S., Huppmann, M., Mays, L.E., Illenyi, M., Schams, A. et al. (2011) Expression of therapeutic proteins after delivery of chemically modified mRNA in mice. *Nat. Biotechnol.*, **29**, 154–157.
- Temiakov, D., Patlan, V., Anikin, M., McAllister, W.T., Yokoyama, S. and Vassilyev, D.G. (2004) Structural basis for substrate selection by T7 RNA polymerase. *Cell*, **116**, 381–391.
- Wang, B., Predeus, A.V., Burton, Z.F. and Feig, M. (2013) Energetic and structural details of the trigger-loop closing transition in RNA polymerase II. *Biophys. J.*, **105**, 767–775.
- Wu, S., Li, L. and Li, Q. (2017) Mechanism of NTP binding to the active site of T7 RNA polymerase revealed by free-energy simulation. *Biophys. J.*, **112**, 2253–2260.
- Yin, Y.W. and Steitz, T.A. (2004) The structural mechanism of translocation and helicase activity in T7 RNA polymerase. *Cell*, **116**, 393–404.
- Wu, S., Wang, J., Pu, X., Li, L. and Li, Q. (2018) T7 RNA polymerase discriminates correct and incorrect nucleoside triphosphates by free energy. *Biophys. J.*, **114**, 1755–1761.
- Schrum, J.P., Siddiqui, S. and Ejebe, K. (2010) In: *Modified Nucleosides, Nucleotides, and Uses Thereof*. USPTO US9334328B2, USA.
- Taoka, M., Nobe, Y., Yamaki, Y., Sato, K., Ishikawa, H., Izumikawa, K., Yamauchi, Y., Hirota, K., Nakayama, H., Takahashi, N. et al. (2018) Landscape of the complete RNA chemical modifications in the human 80S ribosome. *Nucleic Acids Res.*, **46**, 9289–9298.
- Carlile, T.M., Rojas-Duran, M.F., Zinshteyn, B., Shin, H., Bartoli, K.M. and Gilbert, W.V. (2014) Pseudouridine profiling reveals regulated mRNA pseudouridylation in yeast and human cells. *Nature*, **515**, 143–146.
- Liu, N., Parisien, M., Dai, Q., Zheng, G., He, C. and Pan, T. (2013) Probing N6-methyladenosine RNA modification status at single nucleotide resolution in mRNA and long noncoding RNA. *RNA*, **19**, 1848–1856.
- Su, D., Chan, C.T., Gu, C., Lim, K.S., Chionh, Y.H., McBee, M.E., Russell, B.S., Babu, I.R., Begley, T.J. and Dedon, P.C. (2014) Quantitative analysis of ribonucleoside modifications in tRNA by HPLC-coupled mass spectrometry. *Nat. Protoc.*, **9**, 828–841.
- Karikó, K., Buckstein, M., Ni, H. and Weissman, D. (2005) Suppression of RNA recognition by Toll-like receptors: the impact of nucleoside modification and the evolutionary origin of RNA. *Immunity*, **23**, 165–175.

33. Huang, J., Briebe, L.G. and Sousa, R. (2000) Misincorporation by wild-type and mutant T7 RNA polymerases: identification of interactions that reduce misincorporation rates by stabilizing the catalytically incompetent open conformation. *Biochemistry*, **39**, 11571–11580.
34. Milisavljević, N., Perliková, P., Pohl, R. and Hocek, M. (2018) Enzymatic synthesis of base-modified RNA by T7 RNA polymerase. A systematic study and comparison of 5-substituted pyrimidine and 7-substituted 7-deazapurine nucleoside triphosphates as substrates. *Org. Biomol. Chem.*, **16**, 5800–5807.
35. Ulrich, S. and Kool, E.T. (2011) Importance of steric effects on the efficiency and fidelity of transcription by T7 RNA polymerase. *Biochemistry*, **50**, 10343–10349.
36. Takahashi, S., Matsumoto, S., Chilka, P., Ghosh, S., Okura, H. and Sugimoto, N. (2022) Dielectricity of a molecularly crowded solution accelerates NTP misincorporation during RNA-dependent RNA polymerization by T7 RNA polymerase. *Sci. Rep.*, **12**, 1149.
37. Chang, Y.C., Herath, J., Wang, T.H. and Chow, C.S. (2008) Synthesis and solution conformation studies of 3-substituted uridine and pseudouridine derivatives. *Bioorg. Med. Chem.*, **16**, 2676–2686.
38. Oh, J., Fleming, A.M., Xu, J., Chong, J., Burrows, C.J. and Wang, D. (2020) RNA polymerase II stalls on oxidative DNA damage via a torsion-latch mechanism involving lone pair- π and CH- π interactions. *Proc. Nat. Acad. Sci. U.S.A.*, **117**, 9338–9348.
39. Tan, Y., You, C., Park, J., Kim, H.S., Guo, S., Schärer, O.D. and Wang, Y. (2022) Transcriptional perturbations of 2,6-diaminopurine and 2-aminopurine. *ACS Chem. Biol.*, **17**, 1672–1676.
40. Chen, T.-H., Potapov, V., Dai, N., Ong, J.L. and Roy, B. (2022) N1-methyl-pseudouridine is incorporated with higher fidelity than pseudouridine in synthetic RNAs. *Sci. Rep.*, **12**, 13017.