The Effect of Base Modification on RNA Polymerase and Reverse Transcriptase Fidelity

#707

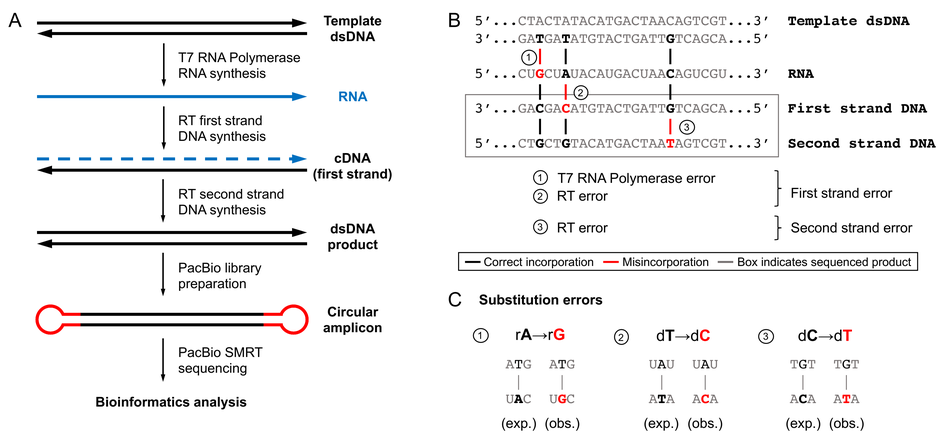
Jennifer L. Ong, Vladimir Potapov, Xiaoqing Fu, Nan Dai, Ivan R. Corrêa Jr., Nathan Tanner
New England Biolabs, Ipswich, MA 01938, USA

Introduction

Ribonucleic acid (RNA) is capable of hosting a variety of chemically diverse modifications. Post-transcriptional mRNA modifications can alter gene expression or mRNA stability, and can be conserved, regulated, and implicated in various cellular, developmental and disease processes. However, few studies have addressed how base modifications affect RNA polymerase and reverse transcriptase activity and fidelity, and hence, RNA sequencing data. Here, we describe the fidelity of RNA polymerization and reverse transcription of modified ribonucleotides using a fidelity assay based on Pacific Biosciences®' Single-Molecule Real-Time (SMRT®) sequencing. Several modified bases, including methylated (m^6A, m^5C and m^5U), hydroxymethylated (hm^5U) and isomeric bases (pseudouridine (Ψ)) were examined.

Methods/Results

Workflow

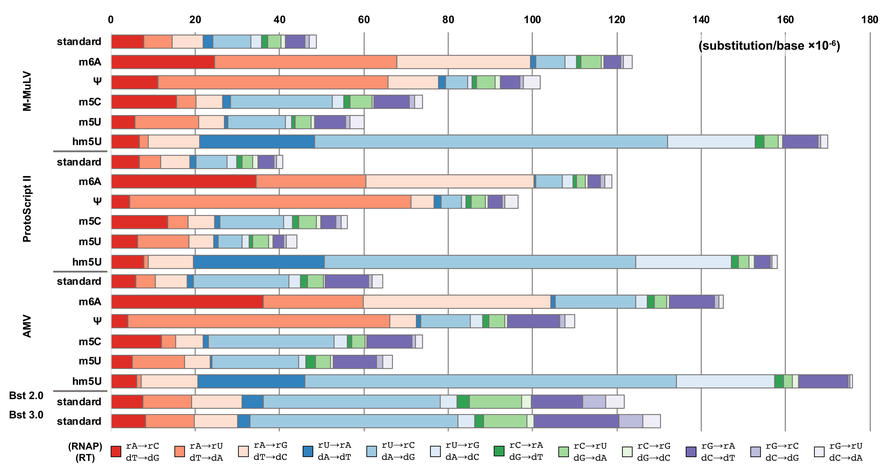


Measuring combined transcription and reverse transcription fidelity with PacBio® sequencing. (A) Workflow. DNA templates are transcribed by T7 RNA polymerase with standard and modified NTPs to produce RNA. RNA is replicated by a reverse transcriptase to produce cDNA, then the first strand is replicated by the same reverse transcriptase to produce double-stranded DNA, which is then prepared for PacBio sequencing by ligating SMRTbell™ adaptors. (B) Identical first strand errors can arise by misincorporation from either the RNA polymerase or the reverse transcriptase (error type 1 and 2 in the figure, respectively). Only first strand errors confirmed in the second strand are counted. Second strand errors produce a mismatch between the first and second strand and represent misincorporation by the reverse transcriptase on DNA templates (error type 3 in the figure). (C) Substitution errors arising from misincorporation.

Combined T7 RNA Polymerase and RT cDNA Errors

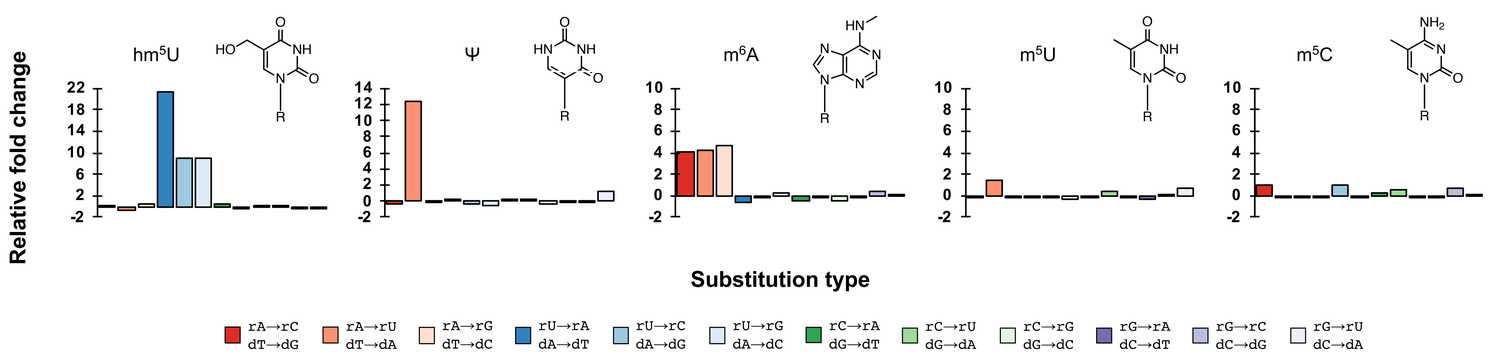
Table 1. Total error rates for cDNA strand synthesis of canonical and noncanonical RNA

Template	Total error rate errors/base ×10⁻⁶	Percentage of total errors: Substitution %	Deletion %	Insertion %
M-MuLV Reverse Transcriptase and T7 RNA Polymerase				
RNA	63 ± 12	78	11	11
m^6A	149 ± 21	86	9	5
Ψ	114 ± 23	89	6	6
m^5C	81 ± 18	86	9	5
m^5U	65 ± 12	87	9	4
hm^5U	185 ± 23	90	6	4
ProtoScript II Reverse Transcriptase and T7 RNA Polymerase				
RNA	56 ± 8	71	19	10
m^6A	152 ± 8	80	11	8
Ψ	101 ± 21	90	7	3
m^5C	70 ± 4	82	12	6
m^5U	54 ± 2	81	15	4
hm^5U	188 ± 24	87	9	5
AMV Reverse Transcriptase and T7 RNA Polymerase				
RNA	75 ± 11	87	5	8
m^6A	164 ± 11	89	5	6
Ψ	116 ± 22	94	4	3
m^5C	81 ± 2	92	3	5
m^5U	73 ± 5	91	5	3
hm^5U	192 ± 8	91	5	4
Bst 2.0 Reverse Transcriptase and T7 RNA Polymerase				
RNA	179 ± 105	78	16	6
Bst 3.0 Reverse Transcriptase and T7 RNA Polymerase				
RNA	181 ± 102	82	15	4



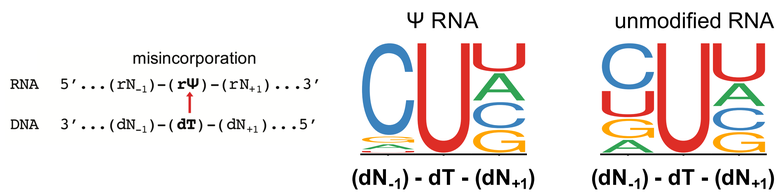
First strand (cDNA) synthesis error rates and error spectrum for standard and modified RNA. The RNA template is synthesized by T7 RNA polymerase, and then reverse transcribed by the reverse transcriptases shown in the workflow. For comparison, also shown are the first strand error rates of *Bst* 2.0 and 3.0 DNA polymerases, DNA polymerases which can be used to reverse transcribe RNA. Polymerase substitution errors are written as the equivalent RNA polymerase substitution (top substitution) or reverse transcriptase substitution (bottom substitution).

Base Modifications Can Alter Polymerase Fidelity



First strand error rates of modified RNA normalized to regular RNA (Protoscript® II reverse transcriptase). Relative substitution rates of each error type for each modification were normalized to unmodified RNA, for ProtoScript II reverse transcriptase (with T7 RNA polymerase). On the y-axis, E was calculated for each substitution type as (S - M) / S, where S is the substitution rate on unmodified RNA, and M is the substitution rate on RNA containing modified bases. An E value of 0 represents no change in fidelity compared to unmodified RNA, whereas the numerical values represent the fold-change relative to unmodified RNA. For each non-reference error identified during cDNA synthesis, the equivalent RNA polymerase error (top pair) and reverse transcriptase error (bottom pair) that could generate the corresponding first strand error are identified.

Sequence Context Analysis of Pseudouridine-induced Transcription Errors



Sequence context analysis of first strand errors. Sequence logos represent the identity of the bases surrounding each type of misincorporation, with respect to the reference RNA, for pseudouridine-containing and unmodified RNA. In each logo, bases are ordered most frequently (top) to least frequently (bottom) observed. In this example, T7 RNA polymerase was used to generate the RNA template, and ProtoScript II reverse transcriptase was used for reverse transcription.

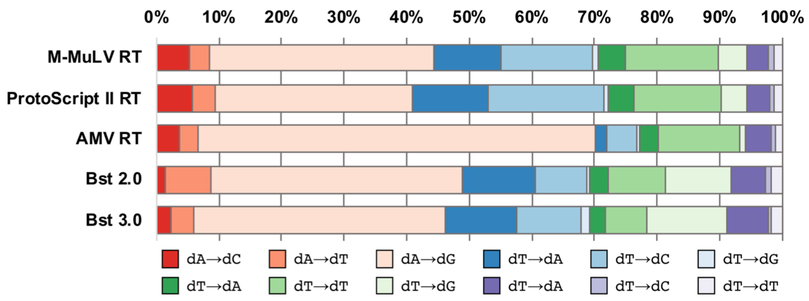
Reverse Transcriptase Fidelity

A

Second strand error rates

DNA Polymerase	Total error rate (errors/base ×10⁻⁶)	Substitution	Deletion	Insertion
M-MuLV RT	84 ± 19	92%	6%	3%
ProtoScript II RT	62 ± 9	91%	6%	3%
AMV RT	52 ± 4	93%	5%	2%
Bst 2.0	62 ± 5	92%	7%	1%
Bst 3.0	70 ± 23	89%	8%	3%

B



Second strand error rates, representing the error rate for reverse transcriptases or *Bst* DNA polymerases replicating DNA templates. (A) Total error rates and distribution of substitution, deletion and insertion errors. (B) Normalized mutational spectrum of second strand error rates. Polymerase substitutions are written as (expected base) → (observed base).

Summary

We developed an assay to measure the fidelity of cDNA synthesis on unmodified and modified RNA. cDNA (first strand) error rates are the combined error rates of T7 RNA polymerase and the reverse transcriptase. However, by normalizing base-specific error rates of the modified base to the equivalent standard RNA base, we were able to determine which modifications had an effect on either RNA polymerase or reverse transcriptase fidelity. 5-hydroxylmethyluracil and N6-methyladenosine both increased first strand error rates compared to the equivalent unmodified RNA base, whereas 5-methylcytosine and 5-methyluracil did not significantly affect first strand error rates. Pseudouridine (an isomer of uracil) was misincorporated across thymidine by T7 RNA polymerase at a greater frequency than uracil, and sequence context analysis revealed that a dT:dΨ mispair was also preferentially preceded by dG:rCTP incorporation. Reverse transcriptase-specific error rates were identified by an analysis of second strand errors, in which errors only arising during DNA templated DNA synthesis were counted.