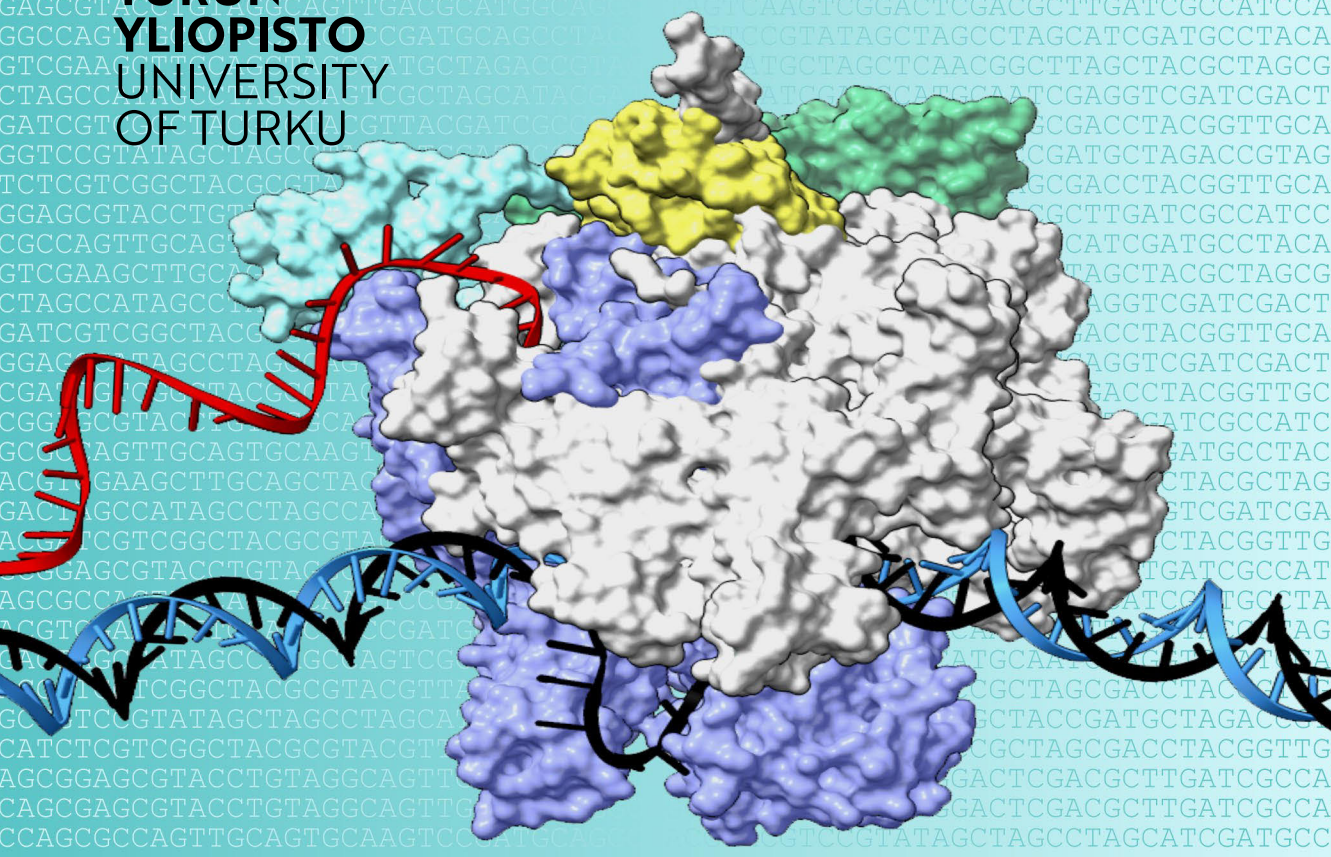
TURUN YLIOPISTO
UNIVERSITY OF TURKU



THE MECHANISM OF TRANSCRIPT ELONGATION

Substrate Selection and Proofreading by the Multi-Subunit RNA Polymerase

Janne J. Mäkinen

TURUN YLIOPISTON JULKAISUJA – ANNALES UNIVERSITATIS TURKUENSIS
SARJA – SER. AI OSA – TOM. 725 | ASTRONOMICA – CHEMICA – PHYSICA – MATHEMATICA | TURKU 2024

TURUN
YLIOPISTO
UNIVERSITY
OF TURKU

THE MECHANISM OF TRANSCRIPT ELONGATION

Substrate Selection and Proofreading by the Multi-Subunit RNA Polymerase

Janne J. Mäkinen

TURUN YLIOPISTON JULKAISUJA – ANNALES UNIVERSITATIS TURKUENSIS
SARJA – SER. AI OSA – TOM. 725 | ASTRONOMICA – CHEMICA – PHYSICA – MATHEMATICA | TURKU 2024

University of Turku

Faculty of Technology
Department of Life Technologies
Biochemistry
Doctoral programme in Technology

Supervised by

Associate professor, Ph.D. Georgiy A. Belogurov
Department of Life Technologies
University of Turku
Turku, Finland

Reviewed by

Professor Lari Lehtiö
Faculty of Biochemistry and Molecular Medicine
University of Oulu
Oulu, Finland

Assistant Professor, Ph.D. Tatiana V. Mishanina
Chemistry and Biochemistry,
University of California San Diego
California, United States

Opponent

Professor Craig T. Martin
Department of Chemistry
University of Massachusetts Amherst
Massachusetts, United States

Cover Image by Janne J. Mäkinen.

ISBN 978-951-29-9927-9 (PRINT)
ISBN 978-951-29-9928-6 (PDF)
ISSN 0082-7002 (Print)
ISSN 2343-3175 (Online)
Painosalama, Turku, Finland 2024

Today we still yearn to know why we are here and where we came from. Humanity's deepest desire for knowledge is justification enough for our continuing quest. And our goal is nothing less than a complete description of the universe we live in.

- Stephen Hawking, "A Brief History of Time"

UNIVERSITY OF TURKU
Faculty of Technology
Department of Life Technologies
Biochemistry
JANNE J. MÄKINEN: The Mechanism of Transcript Elongation; Substrate Selection and Proofreading by the Multi-Subunit RNA Polymerase
Doctoral Dissertation, 260 pp.
Doctoral Programme in Technology
November 2024

ABSTRACT

RNA polymerases (RNAPs) synthesize RNA using DNA or RNA as a template. Accurate RNA synthesis is essential for cellular functions and viral RNA replication, so RNAPs actively select the right nucleotides by probing for nucleobase and nucleosugar moieties. RNAPs transcribing cellular genomes are large multi-subunit enzymes, whereas mitochondrial genomes are transcribed by structurally distinct single-subunit RNAPs. Viral RNAPs from RNA viruses are distantly related to mitochondrial RNAPs, but they use RNA as a template. Nucleoside analogues that mimic the canonical ribonucleotide triphosphate substrates (rNTPs) can be used to inhibit the RNAPs of pathogens. The mechanism of substrate selection by all RNAPs needs to be studied in great mechanistic detail to optimize analogues for selective targeting. This thesis work elucidates the mechanisms of nucleosugar selection and transcriptional proofreading by multi-subunit RNAPs, and provides insights into the nucleobase selectivity mechanism by different RNAP structural families. First, we found that multi-subunit RNAPs differentiate nucleosugar in ribo- and deoxyribonucleoside triphosphates (2'dNTPs) by utilizing the invariant arginine residue. This residue promotes rNTP binding, but also disfavors 2'dNTP incorporation into the RNA by stabilizing the catalytically inert 2'-endo conformation of the nucleosugar. Second, we delineated the contributions of various regions of the active site for proofreading activity of multi-subunit RNAPs. Third, we evaluated the suitability of six nucleoside analogues as substrates for multi-subunit, mitochondrial and viral RNAPs. These RNAPs utilized the nucleoside analogues with different efficiencies and specificity. Several nucleoside analogues acted as dual coders, mimicking more than one canonical nucleobase. In particular, our data suggests that formycin A is a potent dual coder that may induce mutations during viral RNA synthesis. Overall, our results highlight the differences in substrate selection by cellular, mitochondrial and viral RNAPs, providing valuable information for the design of medically relevant transcription inhibitors.

KEYWORDS: RNA polymerase, transcription, substrate selection, proofreading, nucleoside analogues

TURUN YLIOPISTO
Teknillinen Tiedekunta
Bioteknologian Laitos
Biokemia
JANNE J. MÄKINEN: RNA-transkriptin pidentymisen mekanismi;
substraatin selektio ja oikoluku monialayksikköisessä RNA-polymeraasissa
Väitöskirja, 260 s.
Teknologian tohtoriohjelma
Marraskuu 2024

TIIVISTELMÄ

RNA-polymeraasit (RNAP:t) syntetisoivat RNA:ta käyttäen DNA:ta tai RNA:ta mallina. Tarkka RNA:n synteesi on välttämätöntä solun toiminnoille ja virusten RNA:n kahdentamiselle, joten RNAP:t valitsevat aktiivisesti oikeita nukleotidejä niiden emäs- ja sokeriosien perusteella. Solujen genomia transkriptoivat RNAP:t ovat suuria monialayksikköisiä entsyymejä, kun taas mitokondrioiden genomeja transkriptoivat rakenteellisesti erilaiset, pienemmät RNAP:t. RNA-virusten RNAP:t ovat kaukaista sukua mitokondrion RNAP:lle, mutta ne käyttävät RNA:ta mallina. Ribonukleosiditrifosfaatteja matkivat nukleosidianalogit voivat estää taudin-aiheuttajien RNAP:ien toimintaa. RNAP:ien substraatin valinnan mekanismia on tutkittava yksityiskohtaisesti, jotta analogit voidaan optimoida kohdistumaan taudinaiheuttajiin. Tämä väitöskirjatyö selventää substraatin valinnan ja transkription oikoluvun mekanismeja monialayksikköisissä RNAP:eissa, ja antaa uutta tietoa siitä, miten eri perheisiin kuuluvat RNAP:t tunnistavat nukleotidien emäsosia.

Ensiksi selvitimme, että monialayksikköiset RNAP:t erottavat ribo- ja deoksiribonukleosiditrifosfaattien sokeriosat toisistaan konservoituneen arginiinin avulla. Tämä aminohappotähde auttaa ribonukleosiditrifosfaattia sitoutumaan, mutta myös estää deoksiribonukleosiditrifosfaattien liittämistä RNA:han stabiloimalla sen sokeriosan katalyyttisesti inerttiä 2'-endo-konformaatiota. Toiseksi, tutkimme monialayksikköisen RNAP:n aktiivisen keskuksen eri alueiden osallistumista transkription oikolukuun. Kolmanneksi, arvioimme kuuden nukleosidianalogin soveltuvuutta monialayksikköisten, mitokondriaalisten ja virusten RNAP:ien substraatteina. Nämä RNAP:it käyttivät nukleosidianalogeja substraatteina eri tehokkuuksilla ja tarkkuudella. Useat nukleosidianalogit toimivat kaksois-koodaajina, matkien useampaa kuin yhtä ribonukleosiditrifosfaattia. Erityisesti formysiini A:n havaittiin olevan tehokas kaksoiskoodaaja, joka saattaa aiheuttaa mutaatioita viruksen RNA:n synteesin aikana. Tuloksemme korostavat, miten monialayksikköisten, mitokondriaalisen ja virusten RNAP:t eroavat toisistaan substraatin emäsosan tunnistuksessa, tarjoten arvokasta tietoa lääketieteellisesti merkittävien transkription estäjien suunnittelua varten.

ASIASANAT: RNA-polymeraasi, transkriptio, substraatin valinta, oikoluku, nukleosidianalogi

Table of Contents

Abbreviations

1-nt	1 nucleotide
2-nt	2 nucleotides
2’dNTP	2’deoxyribonucleoside triphosphate
6-MI	6-methylisoxanthopterin
8oA	8-oxoadenine
8oG	8-oxoguanine
8-oxo-ATP	8-oxoadenosine triphosphate
8-oxo-GTP	8-oxoguanosine triphosphate
A	adenine
asRNA	anti-sense RNA
ATP	adenosine triphosphate
BH	bridge helix
bp	base pair
C	cytosine
CTP	cytidine triphosphate
CVB3	coxsackievirus B3
dA	2’deoxyadenosine
dC	2’deoxycytosine
dG	2’deoxyguanosine
DNA	deoxyribonucleic acid
DNAP	DNA polymerase
dsDNA	double-stranded DNA
Eco	*Escherichia coli*
EDTA	ethylenediaminetetraacetic acid
EthS-Sdm	4-ethylthioshowdomycin
EthS-SdmTP	4-ethylthioshowdomycin triphosphate
For	formycin A
ForTP	formycin A triphosphate
G	guanosine
GTP	guanosine triphosphate
Hsa	*Homo sapiens*

LL	lid loop
MeS-Sdm	4-methylthioshowdomycin
MeSe-Sdm	4-methylseleniumshowdomycin
MD	molecular dynamics
miRNA	micro-RNA
mRNA	messenger RNA
MT	mitochondrial
NAC	nucleotide addition cycle
nt	nucleotide
ntDNA	non-template DNA
NMP	nucleoside monophosphate
NTP	nucleoside triphosphate
PAGE	polyacrylamide gel electrophoresis
PDB	protein databank (https://www.rcsb.org/)
PP_i	pyrophosphate
PPP	triphosphate
Pyr	pyrazofurin A
PyrTP	pyrazofurin A triphosphate
rA	riboadenosine
rC	ribocytosine
rG	riboguanosine
Rib	ribavirin
RibTP	ribavirin triphosphate
RNA	ribonucleic acid
RNAP	RNA polymerase
rNTP	ribonucleoside triphosphate
rRNA	ribosomal RNA
Sce	*Saccharomyces cerevisiae*
Sdm	showdomycin
siRNA	small interfering RNA
T	thymine
tDNA	template DNA
TEC	transcription elongation complex
TH	triple helix
TL	trigger loop
TLA	thiolacto
tRNA	transfer RNA
Tth	*Thermus thermophilus*
U	uracil
UTP	uridine triphosphate

Abbreviations of amino acid residues

A	Ala	alanine
C	Cys	cysteine
D	Asp	aspartic acid
E	Glu	glutamic acid
F	Phe	phenylalanine
G	Gly	glycine
H	His	histidine
I	Ile	isoleucine
K	Lys	lysine
L	Leu	leucine
M	Met	methionine
N	Asn	asparagine
P	Pro	proline
Q	Gln	glutamine
R	Arg	arginine
S	Ser	serine
T	Thr	threonine
V	Val	valine
W	Trp	tryptophan
Y	Tyr	tyrosine

List of Original Publications

This dissertation is based on the following original publications, which are referred to in the text by their Roman numerals:

I **Mäkinen JJ**, Shin Y, Vieras E, Virta P, Metsä-Ketelä M, Murakami KS, Belogurov GA (2021). The mechanism of the nucleo-sugar selection by multi-subunit RNA polymerases. *Nature Communications* 12(1):769.

II **Mäkinen JJ**, Belogurov GA. Mechanistic insights into RNA cleavage by bacterial RNA polymerase from a comprehensive mutational screen. (Manuscript, preprint at bioRxiv.org, DOI: 10.1101/2024.06.20.599782)

III Rosenqvist P, **Mäkinen JJ**, Palmu K, Jokinen J, Prajapati RK, Korhonen HJ, Virta P, Belogurov GA, Metsä-Ketelä M (2022). The role of the maleimide ring system on the structure-activity relationship of showdomycin. *Eur. J. Med. Chem*. 237:114342.

IV **Mäkinen JJ**, Rosenqvist P, Virta P, Metsä-Ketelä M, Belogurov GA (2024). Probing the nucleobase selectivity of RNA polymerases with dual-coding substrates. *Journal of Biological Chemistry* 9:107755.

List of Contemporaneous Publications

- Turtola M, **Mäkinen JJ**, Belogurov GA (2018). Active site closure stabilizes the backtracked state of RNA polymerase. *Nucleic Acids Res*. 46(20):10870–10887.
- Prajapati RK, Rosenqvist P, Palmu K, **Mäkinen JJ**, Malinen AM, Virta P, Metsä-Ketelä M, Belogurov GA (2019). Oxazinomycin arrests RNA polymerase at the polythymidine sequences. *Nucleic Acids Res*. 47(19):10296–10312.
- Huang Y, Trapp V, Puro O, **Mäkinen JJ**, Metsä-Ketelä M, Wahl MC, Belogurov GA (2022). Fluorogenic RNA aptamers to probe transcription initiation and co-transcriptional RNA folding by multi-subunit RNA polymerases. *Methods Enzymol*. 675:207–233.

1 Introduction

All organisms store their genetic information as DNA, which is transcribed into various types of RNA during gene expression. Transcribed RNAs contain instructions encoded in DNA and transfer them forward in the cell. RNAs have key roles in protein synthesis, where messenger-RNA (mRNA) delivers amino acid sequence information to ribosomes, which are made up of both protein and RNA (rRNA). Ribosomes, together with amino acyl transfer-RNAs (tRNA), decode mRNA and add the correct amino acids one by one to a growing amino acid chain, which makes up a protein. In addition to RNAs that participate in protein synthesis, various non-coding RNAs, like small interfering RNAs (siRNA), micro-RNAs (miRNA) and antisense RNAs (asRNA), actively participate in gene regulation. In other words, RNA works as a bridge between genetic information and the different biochemical functions happening in a cell.

RNA polymerase (RNAP) is the molecular machinery that transcribes the information encoded in DNA into RNA during transcription at the beginning of gene expression. All cellular organisms rely on at least one multi-subunit RNAP, while mitochondria and viruses express smaller RNAPs. RNAPs generate RNA transcripts by using ribonucleotidetriphosphates (NTPs) with adenine (A), uracil (U), guanine (G), or cytosine (C) nucleobases. DNA-polymerases (DNAPs) synthesize DNA utilizing 2'deoxyribonucleotidetriphosphates (2'dNTPs), which are synthesized from NTPs by ribonucleotide reductases. 2'dNTPs lack a 2'OH group, yet they also have adenine, guanine or cytosine, and thymine (T) in place of uracil. All known RNAPs and DNAPs utilize the same mechanism to catalyze the addition of nucleoside triphosphates (NTPs) to oligonucleotides (Steitz, 1998).

During gene expression RNAPs transcribe hundreds to thousands of DNA nucleotides, adding up to 100 NTPs to a growing RNA chain per second (Vogel and Jensen, 1994; Proshkin *et al.*, 2010; Singh and Padgett, 2009; Dundr *et al.*, 2002). While transcribing entire genes within seconds, RNAPs need to maintain a steady level of accuracy (i.e. fidelity) to ensure that the RNA transcripts have the correct structure and function. To achieve this, RNAP needs to select for a NTP that has (i) hydroxyl group (OH) at the 2' position and (ii) the correct nucleobase moiety that forms a Watson-Crick base pair with the template DNA nucleobase moiety. The right

type of nucleotide needs to be selected out of all available NTPs in the cell at each sequence position. If a multi-subunit RNAP adds a nucleotide that does not have matching nucleobase moiety to the RNA, it can proofread the RNA by excising the incorrect nucleotide from the RNA, after which it can retry to synthesize the RNA correctly. Together substrate selection and proofreading reduce errors in the final RNA transcripts down to one error in every 10^4-10^6 nucleotides of synthesized RNA (Lynch 2010; Gout *et al.*, 2013; Traverse and Ochman, 2016; Li and Lynch, 2020). While cells can tolerate transcriptional errors to some extent, inaccurate transcription has been linked to DNA damage (Dutta *et al.*, 2011) and proteotoxic stress (Vermulst *et al.*, 2015).

Bacterial and viral RNAPs are potential targets for drug development. While vaccines limit the spreading of bacterial and viral infections, there is high demand for new classes of antibacterial and antiviral drugs to treat the rising number of drug-resistant infections and new emerging pathogens. Nucleoside analogues that mimic natural NTP substrates of RNAPs could be modified into potential drugs to combat emerging viral infections, as many NTP analogues are clinically approved and effective against some DNA-viruses, RNA-viruses and retroviruses (reviewed in De Clercq and Li, 2016). Nucleoside analogues could prove valuable against antibiotic resistant bacterial infections as well, since so far only two classes of compounds that target bacterial transcription have been approved for clinical use: rifamycins and fidaxomicin (reviewed in Ma *et al*., 2016b). Because of the similarities of RNAP active sites, efficient inhibitors of bacterial and viral RNAPs can also inhibit eukaryotic multi-subunit RNAPs or mitochondrial single-subunit RNAPs thus exhibiting cytotoxic effects (De Clercq *et al.*, 1987; Feng *et al.*, 2016). Therefore, to selectively inhibit the target RNAP by exploiting its catalytic features, it is important to understand in detail the mechanisms of RNA synthesis by RNAPs of different organisms and viruses.

This thesis focuses on the two key mechanisms that multi-subunit RNAPs use to maintain accurate RNA synthesis: substrate selection and transcriptional proofreading (i.e. correcting RNA synthesis errors as they occur). While the mechanism of RNA synthesis is conserved in all known RNAPs, the mechanisms to support its accuracy are varied. In the literature review, I describe what is currently known about the mechanisms of substrate selection and proofreading in RNAPs, and how they affect the cellular functions of different organisms. The experimental part of this thesis expands upon current knowledge by introducing methods to study transcriptional fidelity by observing the kinetics of substrate binding and RNA proofreading. These methods are further utilized to study the nucleoside analogues of ribonucleotides, which could be modified to specifically inhibit bacterial or viral RNAPs.

2 Literature Review

2.1 Structure of RNAP and the transcription elongation complex

All known RNAPs can be divided into two structurally unrelated superfamilies: two-β-barrel RNAPs and right-hand RNAPs (Figure 1). RNAPs generate RNA transcripts as a reverse complement of DNA (DNA-dependent RNAPs), or of RNA (RNA-dependent RNAPs). Despite the structural diversity of RNAPs among the different domains of life, all transcribing and replicative nucleic acid polymerases utilize the same two-metal ion mechanism of nucleic acid synthesis (Steitz 1998).

2.1.1 Two-β-barrel RNA polymerases

In every organism there is at least one multi-subunit RNAP from the two-β-barrel superfamily that transcribes information from genomic DNA into RNA. The simplest model of a multi-subunit RNAP, bacterial RNAP, generally has five subunits: two α-subunits, β, β' and ω (often written as α2ββ'ω). This set of subunits forms the core of all multi-subunit RNAPs. The largest subunits, β and β' form together the active site of the polymerase and the clefts that encase the DNA strands and the RNA strand. β and β' together resemble a crab claw that clamps down on the nucleic acids. The two α-subunits form a homodimer, which brings together β and β'. The smallest of the subunits, ω, binds to the β' and assists in its folding.

Archaeal and eukaryotic multi-subunit RNAPs have their own variations of the bacterial subunits described above, but they also have additional subunits surrounding the core. These subunits increase the surface area of archaeal and eukaryotic RNAPs, allowing for more varied interactions with transcription factors and other regulators. While bacteria and archaea express only one RNAP that synthesizes all the cell's RNAs, the nuclei of eukaryotes have three multi-subunit RNAPs, each producing different RNAs: RNAPI (rRNA), RNAPII (mRNA) and RNAPIII (tRNA and other small RNAs). In plant chloroplasts, analogues of cyanobacterial RNAP transcribe a portion of chloroplast genes (reviewed in Börner *et al.*, 2015). Additionally, plants and fungi express small two-β-barrel RNAPs that

use RNA as a template (Makeyev and Bamford, 2002). These RNAi RNAPs synthesize short RNAs used in RNA interference.

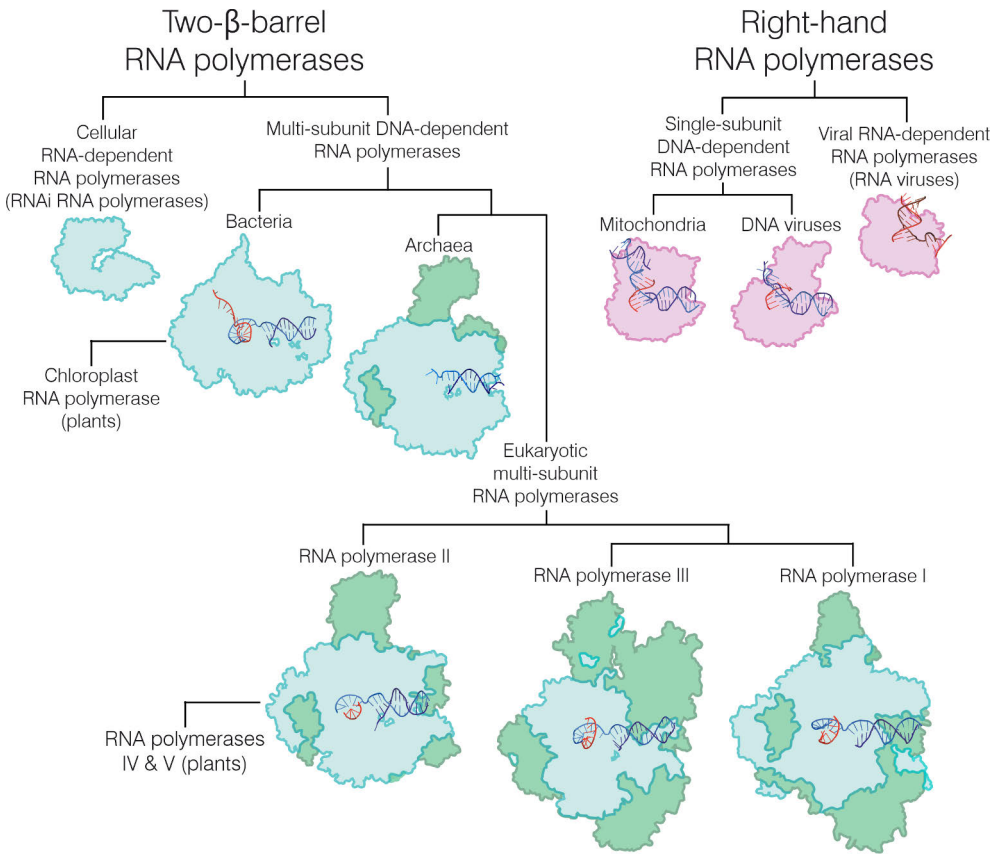


Figure 1. Structural diversity of the two distinct families of RNAPs. RNAPs are divided into two structurally distinct protein families: two-β-barrel RNAPs (cyan and green) and right-hand RNAPs (magenta). Displayed structures are silhouettes of structures uploaded from PDB database; cartoon representations of DNA (blue) and RNA (red) strands inside the polymerases are shown as reference (when available). Boundaries of individual subunits in multi-subunit RNAPs are omitted for clarity. Two-β-barrel RNAP superfamily consists of multi-subunit RNAPs from all domains of life: bacteria (PDB:2O5I), archaea (PDB:3HKZ) and eukaryotic RNAPII (PDB:3PO2), RNAPIII (PDB:5FJ8) and RNAPI (PDB:5M3F). The core of the two-β-barrel RNAPs (cyan) is surrounded by additional subunits (green) in archaeal and eukaryotic multi-subunit RNAPs. Chloroplast RNAP is structurally related to cyanobacterial RNAP and eukaryotic RNAPs IV and V present in plants are related to RNAPII. Plants and fungi express RNAi RNAPs (PDB:2J7N) from two-β-barrel RNAP superfamily that use RNA as template. Right-hand RNAP superfamily consists of DNA- and RNA-dependent viral RNAPs (PDB:1S77 and PDB:4K4S, respectively) and the mitochondrial RNAP (PDB:4BOC), which are structurally related to family I DNAPs and reverse transcriptases.

2.1.2 Right-hand RNA polymerases

While cellular organisms deploy large and globular RNAPs comprised of multiple subunits, mitochondria, DNA viruses and RNA viruses express smaller RNAPs. These RNAPs belong to the right-handed RNAP superfamily, which is not structurally related to the two-β-barrel RNAPs of cellular organisms. Instead, they are related to family I DNAPs and reverse transcriptases (Delarue *et al.*, 1990; Sousa *et al.*, 1993), sharing with them the highly conserved palm, fingers and thumb domains. These three domains shape the conserved right-hand motif that forms the active site of the polymerase while enclosing the nucleic acids in a partially clenched "hand". While right-hand RNAPs are not related to two-β-barrel RNAPs, they utilize the same mechanism of RNA synthesis (see chapter 2.2.2 for details).

In eukaryotic cells, mitochondria have a partially independent protein expression for which they require DNA-dependent RNAP to transcribe mitochondrial mRNAs. Mitochondrial RNAPs are related to the T7 phage DNA-dependent RNAP (Cermakian *et al.*, 1996). The active site of right-hand DNA-dependent RNAPs is formed by four structurally conserved motifs (A-D) that are located in the palm domain. RNA-viruses express RNA-dependent RNAPs, which replicate and transcribe viral RNA genomes inside a host cell. Picornaviruses and calciviruses express the smallest known RNA-dependent RNAPs, which contain the active site architecture found in larger viral RNAPs (Ferrero *et al.*, 2018). Some viral RNA-dependent RNAPs have N-terminal and C-terminal extensions, which include domains for the incorporation of an m7Gppp cap-1 structure on the 5′ end of RNA (flaviviruses) and the synthesis of poly-A tails for the 3' end of RNAs (picornaviruses and calciviruses), which provide additional chemical activities for viral RNAPs (reviewed in Ferrero *et al.*, 2018). In RNA-dependent RNAPs, there are seven conserved motifs that participate in NTP binding and RNA strand binding. Motifs A-E are in the palm domain, whereas F and G are located in the fingers domain. The distinct difference between viral RNA-dependent RNAPs and other right-hand polymerases is that the hand is closed instead of being partially open, due to the structure of the finger and thumb domains (Ferrer-Orta *et al*., 2006).

2.1.3 Formation of a transcription elongation complex

In order to transcribe DNA into RNA, a RNAP needs to first bind double stranded DNA (dsDNA) and open it so that the growing RNA chain can form a stable hybrid with the template DNA (tDNA). Initiation of transcription is the starting point of gene expression. RNAP (together with accessory proteins) binds to a promoter forming a closed complex, unwinds (with the help of accessory proteins) double stranded DNA, separating template DNA from non-template DNA and forming an open complex. Then RNA synthesis starts either *de novo* by binding single

nucleotides to the active site (multi-subunit RNAPs and DNA-dependent right-hand RNAPs) or by utilizing primers (some RNA-dependent RNAPs). Multi-subunit RNAPs require initiation factors to recognize the promoter sequence and aid in unwinding the DNA. Bacterial RNAPs utilize σ-factors for both promoter recognition and DNA unwinding (reviewed in Saecker *et al*., 2011), whereas eukaryotic RNAPs deploy a large number of transcription factors and other regulators to initiate transcription (reviewed in Haberle & Stark, 2018). Right-hand RNAPs utilize distinctly different methods for initiation. T7 phage RNAP can bind to a promoter and initiate transcription *de novo* without any additional factors, as it contains domains for promoter recognition (specificity loop and recognition loop) and opening dsDNA (intercalating hairpin) (Cheetham *et al.*, 1999). Human mitochondrial RNAP requires two factors for initiation: TFAM recruits the RNAP to a promoter and TFB2M assists the RNAP in opening the DNA to form the open complex (Hillen *et al.*, 2017). In contrast, viral RNA dependent RNAPs initiate RNA replication without accessory proteins either *de novo*, or by using a short primer (reviewed in Ferrero *et al.*, 2018).

Initially, RNAPs make short RNA transcripts in an attempt to form a stable complex of RNAP and nucleic acids. Once RNAP manages to form a long enough RNA transcript, the RNA will not dissociate from the active site, which results in a highly stable complex of RNAP and nucleic acids, referred to as a transcription elongation complex (TEC). From this point on this thesis mainly focuses on the multi-subunit RNAPs and is largely written from the perspective of bacterial RNAP, which is the simplest model of multi-subunit RNAPs. *Escherichia coli* numbering of amino acids is used throughout, unless otherwise stated. Eukaryotic RNAPs and right-hand RNAPs are referenced within the text where indicated.

2.1.4 Structure of the transcription elongation complex

In the case of multi-subunit RNAPs, the TEC contains about 35 nucleotides of DNA surrounded by amino acid residues of the RNAP (Figure 2). The clefts surrounding negatively charged nucleic acids contain multiple positively charged amino acids, which enable the RNAP to bind DNA in a sequence independent manner that still allows for movement of the nucleic acids during transcription (Cramer, 2001). The interactions of nucleic acids and RNAP make the TEC a highly stable complex in which the RNAP can remain bound to nucleic acids at high temperatures and at high salt concentrations (Landick, 2001). In multi-subunit RNAPs, dsDNA enters the main channel and is double stranded until it enters the active site. Since RNAP transcribes one strand of DNA to create a new strand of RNA, RNAP needs to separate the tDNA from ntDNA, to allow the growing RNA chain to form a hybrid with the tDNA. When dsDNA enters the RNAP, it is positioned so that its B-helical form is disrupted to

facilitate the separation of the tDNA from ntDNA. The RNAP separates the dsDNA arriving at the active site, guiding ntDNA away from the tDNA, while presenting the tDNA acceptor base at the active site to serve as the information for the RNA sequence. The unwound portion of dsDNA inside the RNAP is referred to as transcription bubble. Multiple RNAP domains facilitate the separation of nucleic acids. At the downstream edge of the transcription bubble, a domain called the bridge helix (BH) separates tDNA from ntDNA (Figure 2). A switch 2 domain attracts tDNA via positively charged amino acid residues, guiding it away from ntDNA, while a fork loop 2 domain prevents ntDNA from reannealing with the tDNA.

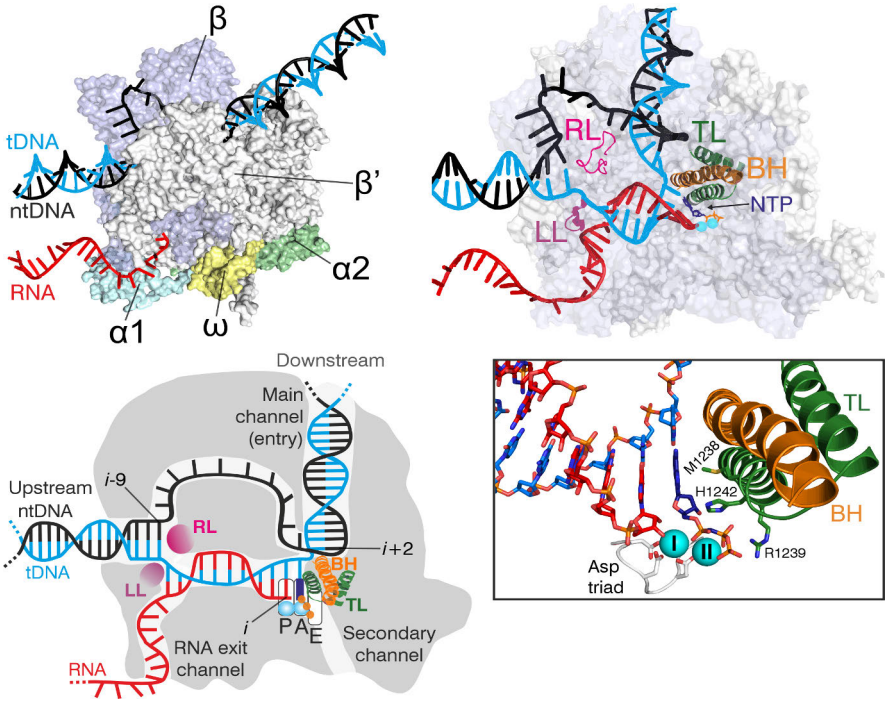


Figure 2. Structure of bacterial multi-subunit RNAP transcription elongation complex (TEC). Top left: structure of bacterial RNAP from *Thermus thermophilus* (*Tth*), originally presented in Turtola *et al.*, 2018. *Tth* RNAP consists of five subunits: α1 (light blue), α2 (light green), β (violet), β' (gray) and ω (yellow). In bacterial TEC there is 35 nucleotides of DNA (tDNA cyan, ntDNA black) and 16-17 nucleotides of RNA (red) surrounded by the RNAP. Top right: RNAP core comprised of β and β' subunits displayed as transparent surfaces (other subunits are omitted for clarity). Structure represents the insertion complex, where NTP (nucleoside dark blue, phosphates orange) has bound to A site in the closed active site. Catalytic Mg^{2+} ions are shown as cyan spheres. Key domains involved in stepwise RNA synthesis are shown as cartoon: TL (trigger loop, green), BH (bridge helix, orange), RL (rudder loop, magenta), LL (lid loop, purple). Bottom left: schematic of the transcription elongation complex, displaying the channels through which the nucleic acids move inside the RNAP. Bottom right: closeup of the NTP bound in the active site. Two catalytic Mg^{2+} ions (Mg-I and Mg-II) are shown as cyan spheres.

During active elongation, dsDNA is opened up for 10 nucleotides, while the nascent RNA chain forms a 9 to 10 base pairs (bp) long hybrid with the tDNA strand (Turtola and Belogurov, 2016; Kang *et al.*, 2017). RNAP adds one nucleotide to the RNA at a time, moving forward along the DNA by one nucleotide in between each nucleotide addition, which requires well-coordinated movement of all nucleic acid strands inside the RNAP. To maintain a stable transcription bubble the RNAP needs to rewind the dsDNA at the upstream side of the bubble, where the RNA strand is separated from the tDNA strand. A lid loop separates the RNA from the tDNA and, together with a rudder loop, directs the tDNA back together with the ntDNA so that they can reanneal, closing the transcription bubble (Vassylyev *et al.*, 2007a). Rewound dsDNA exits the RNAP through the more spacious opening of the main channel, while the growing RNA chain exits through the narrow RNA exit channel (Vassylyev *et al.*, 2007a). Separation and reannealing of tDNA and ntDNA, as well as separation of RNA and tDNA strands inside the RNAP is essential for the stepwise synthesis of RNA, which is discussed further in chapter 2.2.4.

2.2 Stepwise synthesis of RNA: the nucleotide addition cycle

Synthesis of RNA is a repeating sequence of events during which one nucleotide is added to the growing RNA chain and RNAP moves along the DNA to free the active site for the next nucleotide, all happening within milliseconds (summarized in Figure 3). This nucleotide addition cycle (NAC) requires structural rearrangements from both the nucleic acids and the RNAP, which are described in detail below.

2.2.1 Nucleoside triphosphate binding to the active site

At the beginning of the nucleotide addition cycle, the RNA forms a 9 bp long hybrid with DNA, leaving the A-site (*i*+1 site) vacant and allowing the next NTP to bind (i.e. RNAP is in a post-translocated state). In this state the RNA 3' end OH-group is positioned at the P-site (*i* site), next to the A-site. The incoming NTP to be added to the RNA arrives at the active site of RNAP through a narrow pore called the secondary channel. At this time, the mobile domain of the active site, the trigger loop (TL), is unfolded, keeping the active site open and allowing NTPs to diffuse in and out of it. NTP is first positioned at the E-site (entry site), where it is transiently bound by ionic interactions. From the E-site, the NTP twists towards the A-site, during which the nucleobase moiety of the incoming NTP is probed against the tDNA acceptor base, allowing the formation of hydrogen bonds between the two nucleobases (Westover *et al.*, 2004). At the same time, the invariant amino acid residues β'Arg425 and β'Asn458 can hydrogen bond with the OH-groups of the NTP

sugar moiety, which further stabilizes the binding of NTP. This is referred to as pre-insertion of the substrate. After NTP has bound to the A-site, its α-phosphate needs to be oriented for catalysis. At this stage the TL folds forming a "trigger helices" (TH) conformation with the bridge helix (BH), allowing TL residues to make multiple interactions with the bound NTP (Figure 3). In the TH conformation, TL folds over the substrate NTP, repositioning it from a pre-insertion pose to the insertion pose, which is needed to align the α-phosphate moiety in line for the nucleophilic attack of the 3'OH group during catalysis. The TL folding closes the active site and limits dissociation of the bound NTP from the A-site, while also preventing the binding of competing NTPs (Vassylyev *et al.*, 2007b). Interactions between the folded TL with the substrate create an induced fit that is estimated to increase the rate of catalysis by ~10 000 fold (Vassylyev *et al.*, 2007b; Wang *et al.*, 2006; Yuzenkova *et al.*, 2010; Windgassen *et al.*, 2014).

2.2.2 Catalysis of the nucleotidyl-transfer reaction

The new phopshodiester bond between the RNA primer and the bound NTP is formed through a nucleotidyl-transfer reaction. The reaction mechanism of the nucleotidyl-transfer, the S_N2 nucleophilic attack of nucleotide 3'OH to the α-phosphate of an incoming NTP, is believed to be the same in all replicative and transcribing DNA- and RNA-polymerases (Steitz, 1998). Catalysis of nucleotide addition requires (i) deprotonation of the nucleophilic OH group at the RNA 3' end, (ii) two catalytic metal ions for reactant coordination (Mg^{2+}), (iii) a 3'-endo sugar pucker conformation of the 3' end of the NMP, and possibly (iv) protonation of the leaving group (Figure 4). In the nucleotidyl-transfer reaction, the OH group of the RNA 3' end of NMP is deprotonated to O^-, which requires a catalytic Mg^{2+} ion (Mg-I) to lower the oxygen atom's affinity for the transferred hydrogen. The nucleophilic O^- at the P-site then attaches to the α-phosphate of the NTP bound to the A-site, forming the new phosphodiester bond between the nucleotides at the P- and A-sites. The β- and γ-phosphates of the triphosphate moiety form the leaving group; an inorganic pyrophosphate (PP_i) moiety bound by another catalytic Mg^{2+} ion (Mg-II). Interactions between the two catalytic Mg^{2+} ions with the reactants stabilize the pentacovalent transition state (Figure 4). In multi-subunit RNAPs, one of the catalytic metal ions, Mg-I, is strongly bound to the active site by an aspartate triad of the conserved NADFDGD motif: β'Asp460, β'Asp462 and β'Asp464 (Mustaev *et al.*, 1997; Sosunov *et al.*, 2005). Mg-I binds the 3'OH of the RNA 3' end of NMP, lowering its pK_a and making it a stronger nucleophile. The second catalytic metal ion, Mg-II, is weakly bound to the active site by the triphosphate moiety of the incoming NTP and is coordinated by one of the aspartates from the triad (β'Asp460) and a water molecule bound by βGlu813 (Sosunov *et al.*, 2003; Vassylyev *et al.*,

2007b). Mg-II stabilizes the PP_i group during catalysis, partially neutralizing it and making it a better leaving group. After catalysis, Mg-II leaves the active site with the PP_i by-product (Figure 3).

The catalytic mechanism of nucleotidyl-transfer in multi-subunit RNAPs is a subject of some controversy. For multi-subunit RNAPs, two alternative models for catalysis exist: general acid-base catalysis and positional (entropic) catalysis. The acid-base catalysis involving lysine, arginine, or histidine residues as general acids in the polymerase active sites has been proposed as the universal mechanism of nucleotidyl-transfer in all known DNAPs and RNAPs (Castro *et al.*, 2009). In the acid-base model of catalysis, two proton transfers occur: the deprotonation of the 3' end OH group by a general base (proton acceptor) and protonation of PP_i by a general acid (proton donor). The first protonation step has been observed in human DNAP η, where a water molecule is the acceptor of the 3'OH proton (Nakamura *et al.*, 2012). In the free-energy calculations performed with DNAP η structures by Roston *et al*. the deprotonation of 3'OH by active site water was found to be energetically the most plausible pathway of 3'OH deprotonation (Roston *et al*. 2019). It has been shown that in right-hand DNAPs and RNAPs, the protonation of the leaving PP_i group is a rate limiting step of the nucleotidyl-transfer reaction (Castro *et al.*, 2007) and substitutions of lysine residues to leucine in the active sites of these polymerases greatly affected reaction rates and pH dependency of the nucleotidyl-transfer (Castro *et al.*, 2009). In structurally distinct multi-subunit RNAPs, the protonation of PP_i may not be required, since the non-protonated, dianionic form of PP_i bound to Mg-II is already a strong enough leaving group. In addition, a suitable general acid donor for PP_i has not as yet been identified in multi-subunit RNAPs. The amino acid residues β'His936 and β'Arg933 of the TL were proposed to work as proton donors, since they are potential proton donors and based on structural data can interact with the triphosphate moiety of the NTP (Wang *et al.*, 2006; Vassylyev *et al.*, 2007b). However, substitutions of these basic residues with glutamine or alanine do not strongly affect RNA extension, indicating that they do not participate in catalysis as an acid or base (Mishanina *et al.*, 2017; Palo *et al.*, 2021 Zhang *et al.*, 2010; Kaplan *et al.*, 2008).

In the alternative model, positional catalysis, only one proton transfer (deprotonation of the RNA 3' end OH-group) occurs, while the TL works as the positional catalyst, which orients the NTP for the nucleotidyl-transfer reaction (Mishanina *et al.*, 2017; Palo *et al*. 2021). This view is further supported by findings of other enzymes that utilize NTPs, which use positional catalysis instead of acid-base catalysis (Lassila *et al.*, 2011). Positional catalysis provides a solution for the reaction mechanism without involving the protonation of PP_i, while active site water works as the base that deprotonates the 3'OH, which then attacks the α-phosphate (Figure 4).

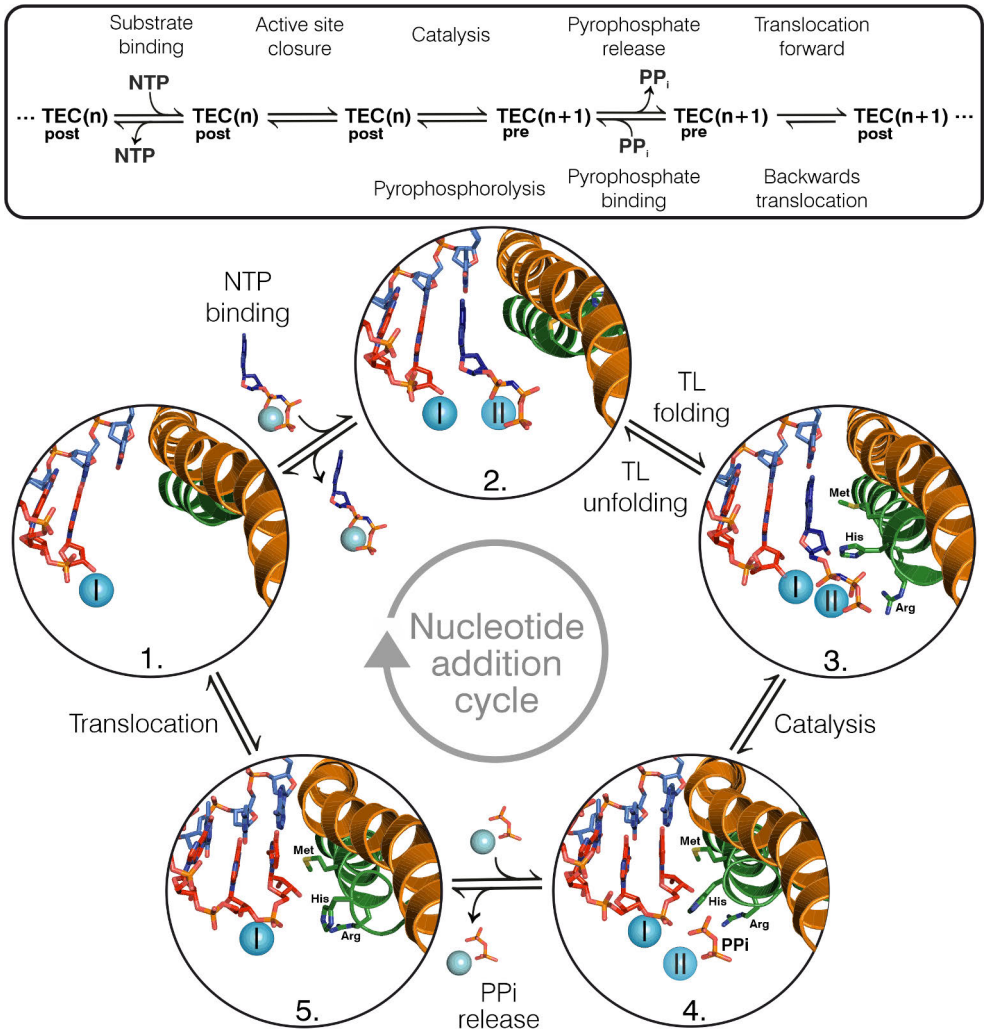


Figure 3. Overview of the nucleotide addition cycle. Structures of RNAPs are used to depict different stages of a single nucleotide addition (numbered from 1 to 5). RNA is shown as red sticks, tDNA light blue sticks, NTP dark blue sticks, Mg ions as cyan spheres, TL green cartoon and BH orange cartoon. TL residues β'Met932, β'Arg933 and β'His936 are shown as sticks when available. (1., PDB: 2O5I) Cycle starts from the post-translocated state. RNA-tDNA hybrid is 9 bp long and the A-site is unoccupied. (2., PDB: 2PPB) The incoming NTP binds initially to the E-site (not shown) from where it rotates to the A-site. (3., PDB: 2O5J) After NTP binds to the A-site the TL forms the trigger helices (TH) conformation together with BH and folds on the NTP closing the active site. TL folding stabilizes the bound NTP in the insertion complex accelerating catalysis, during which NMP is incorporated to the nascent RNA chain. (4., PDB: 5IPL) After NMP incorporation the PP_i byproduct is released from the active site and the TL unfolds. The original structure is modified so that the nucleotide at A-site is connected to the nascent RNA. (5., PDB: 5IPM) Pre-translocated state, in which the RNA-tDNA hybrid is 10 bp long and RNA 3'end occupies the A-site. RNAP translocates from the pre-translocated state to the post-translocated state shortening the hybrid to 9 bp and freeing the A-site, allowing the next NTP to bind (start of the next cycle).

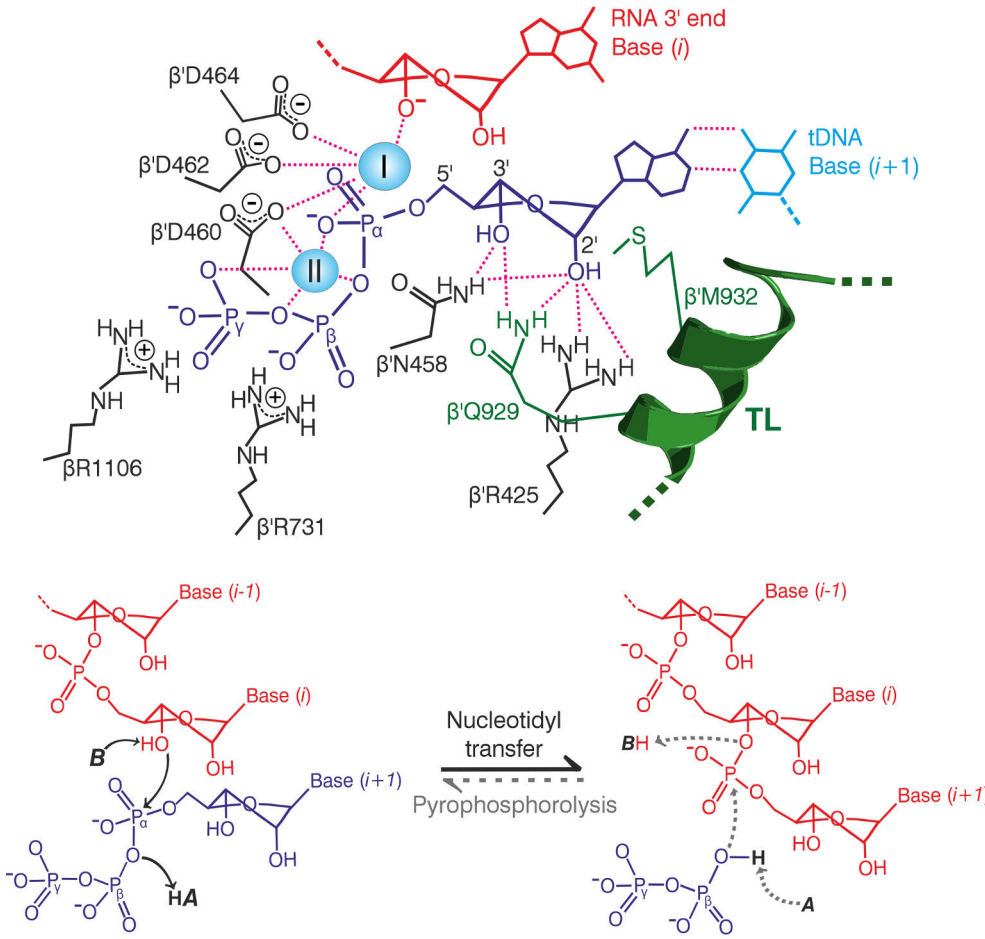


Figure 4. Two-metal catalysis of nucleotidyl transfer reaction in multi-subunit RNAP. Top: Interactions involved in substrate binding and the stabilization of reaction intermediate. Hydrogen bonds and polar interactions are shown as magenta dashed lines. Bound NTP at *i*+1 position (dark blue) is stabilized by Watson-Crick base pairing with template DNA nucleobase (cyan), base stacking with RNA 3' end NMP base (red), catalytic Mg ions I and II (cyan spheres), and multiple amino acid residues in RNAP active site (TL residues green, other residues gray). Bottom: Reaction mechanism of nucleotidyl transfer. Only the nucleotides are shown for clarity. Proton transfers in nucleotidyl transfer are shown with black arrows. The proton transfers in the reverse reaction, pyrophosphorolysis, are shown with dashed gray arrows. 3' OH at RNA 3' end is deprotonated by a general base (B). Resulting nucleophile (O-) attacks the α-phosphate of the bound NTP. Phosphodiester bond is formed between nucleotides at i and *i*+1 positions and PP_i is formed as a byproduct. NMP bound at A site of the RNAP becomes the new RNA 3' end nucleotide.

2.2.3 Pyrophosphate release

After NMP incorporation, the PP_i side-product rapidly dissociates from the active site. PP_i is released approximately 5-16 milliseconds after NMP incorporation (Malinen *et al.*, 2012). It is not certain if PP_i is released before or during translocation, but in bacterial RNAPs the TL residues β'His936 and β'Arg933 can interact with the bound PP_i, which stabilizes the folded state of TL (Liu *et al.*, 2016). When nucleotide addition reactions are supplemented with 1 mM of external PP_i, the translocation rate is reduced by 2-fold (Wang *et al.*, 1998), indicating that PP_i binding and release is connected to TL folding. It has been hypothesized that PP_i release could facilitate TL unfolding by disrupting the triple-helical conformation (Malinen *et al.*, 2012).

2.2.4 Translocation

RNAPs are motor proteins that move (i.e. translocate) along the DNA using a Brownian ratchet mechanism, where the thermal energy of the surrounding system provides energy for movement and the various domains of RNAP direct the movement of nucleic acids (Bar-Nahum *et al.*, 2005; Komissarova and Kashlev, 1997a; Abbondanzieri *et al.*, 2005). Before translocation, the newly added NMP at the RNA 3' end resides at the A-site and the RNA forms a 10 base pair long hybrid with the tDNA (Turtola and Belogurov, 2016; Kang *et al.*, 2017). This state is referred to as the pre-translocated state. To move the RNA 3' end from the A-site to the P-site, each RNA and DNA strand needs to move by one residue, which shortens the RNA-tDNA hybrid to 9 bp. During this one residue shift, one base pair between the tDNA and ntDNA unwinds at the downstream edge of the transcription bubble (at the i+2 position), tDNA reanneals with ntDNA by one base pair at the upstream edge of the bubble (at the i-9 position), and one base pair between the RNA and tDNA is unwound at the upstream edge (also at the i-9 position) (Vassylyev *et al.* 2007a). RNAP moves back and forth between the pre- and post-translocated states before incorporation of the next nucleotide (Hein *et al.*, 2011; Malinen *et al.*, 2012; Malinen *et al.*, 2014). Multiple biochemical studies conclude that the forward translocation rate is much faster than the backward translocation rate, thus favoring occupancy of the post-translocated state (Nedialkov *et al.*, 2012; Larson *et al.*, 2012; Malinen *et al.*, 2014). It has also been shown that the transcribed sequence itself affects the translocation dynamics of the TEC (Bai *et al.*, 2004; Hein *et al.*, 2011; Imashimizu *et al.*, 2015).

When the TEC is in the post-translocated state, the next NTP can bind to active site and start the next nucleotide addition cycle (Figure 3). Due to the effects of the transcribed sequence on translocation and the RNAPs varying affinities for different binding NTPs, the overall rate of the nucleotide addition cycle is not constant.

Additionally, secondary structures of the nascent RNA, transcription factors and external forces on the TEC (like trailing RNAPs upstream of the TEC and translating ribosomes moving along the growing RNA chain in bacteria) further contribute to the overall rate of transcription during the elongation phase (Epshtein and Nudler, 2003; Jin *et al.*, 2010; Proshkin *et al.*, 2010; Wee *et al.*, 2023). An important part of translocation is the TL unfolding, which frees space at the active site for the next incoming NTP. There is no clear consensus whether TL folding/unfolding is coupled to translocation. One view is that TL folding is independent of translocation, since RNAP can continue elongating RNA even when TL is locked in a folded state via cross-linking (Windgassen *et al.*, 2014). Yet, stabilizing the folded conformation of TL via cross-linking or site-directed mutagenesis reduces the RNA elongation rate and increases the pre-translocated fraction of TECs (Windgassen *et al.*, 2014; Nedialkov *et al.*, 2013; Malinen *et al.*, 2014), indicating that TL unfolding allows for more efficient forward translocation.

2.2.5 Off-pathway states of the nucleotide addition cycle; transcriptional pausing

During transcription, TEC can move out of the nucleotide addition cycle into different types of reversible paused states where NTP binding is blocked, rendering RNAP inactive. These paused states can be formed through asynchronous translocation, misincorporation of nucleotides, or because the transcribed sequence thermodynamically favors the pausing of TEC (Bai *et al.*, 2004; Larson *et al.*, 2014; Kang *et al.*, 2023). Transcriptional pausing has important roles during promoter-proximal pausing, transcriptional proofreading, cotranscriptional RNA folding, antitermination, mRNA splicing, and RNAP association with transcription factors (reviewed in Mayer *et al.*, 2017 and Chen *et al.*, 2018).

There are multiple transcribed sequences to which RNAPs respond by forming structurally distinct paused states, which have different rates of pause entry and escape (Gabizon *et al.*, 2018; Kang *et al.*, 2023). Genome sequencing methods have shown that pause sequences are widely distributed across genes of different organisms and sequence determinants of pause sites appear to be conserved in both prokaryotes and eukaryotes (Larson *et al.*, 2014; Imashimizu *et al.*, 2015; Gajos *et al.*, 2021). In many cases the transcript elongation rate slows down when TEC crosses a pause sequence, which provides more time for the off-pathway paused state to form (Gabizon *et al.*, 2018). At pause sequences, TECs can form multiple subpopulations that escape from the same pause at different rates (Kang *et al.*, 2023). Out of the known paused states the best characterized are the consensus pause and hairpin stabilized pause. The consensus pause sequence favors the formation of a stable pre-translocated state that pauses further RNA elongation (Larson *et al.*, 2014;

Kang *et al.*, 2023). During the hairpin stabilized pause, the synthesized RNA forms a hairpin at the RNA exit channel that alters the RNA structure inside the RNAP and results in a half-translocated state where RNA is post-translocated while the DNA remains pre-translocated (Kang *et al*., 2018; Kang *et al.*, 2023).

Because RNAP uses thermal energy to move along the DNA, it can translocate backwards from the pre-translocated state (i.e. backtrack), pushing the RNA 3' end past the A-site towards the secondary channel (Komissarova and Kashlev, 1997b; Nudler *et al.*, 1997). In this backtracked state the length of the RNA-tDNA hybrid is still 10 base pairs, but one or more unpaired nucleotides at the RNA 3' end are positioned towards the secondary channel (Wang *et al*., 2009; Sekine *et al*., 2015). RNAP can backtrack multiple times, pushing the RNA 3' end further away from the active site towards the secondary channel, eventually reaching outside of the RNAP. Backtracking can occur as a response to transcribed sequences that thermodynamically favor it (Nudler *et al.*, 1997; Bochkareva *et al.*, 2012), or because RNAP incorporates the wrong nucleotide at the RNA 3' end, forming a mismatch (Sydow *et al*., 2009; Imashimizu *et al.*, 2013). Backtracked nucleotides prevent NTPs from binding, so transcription is paused until RNAP (i) translocates forward enough times to free the active site, or (ii) removes the backtracked nucleotides that block the active site via endonucleolytic RNA cleavage. The latter is an important part of correcting transcriptional errors in two-β-barrel RNAPs (see chapter 2.4). Backtracked pauses last from seconds to minutes, thus having a substantial effect on the rate of transcription (Abbondanzieri *et al.*, 2005; Galburt *et al*., 2007).

2.3 Avoiding errors during transcription; substrate selection

While RNAP transcribes thousands of nucleotides of DNA, it adds up to 100 ribonucleotides to a growing RNA chain each second (Vogel and Jensen, 1994; Proshkin *et al.*, 2010; Singh and Padgett, 2009; Dundr *et al.*, 2002). During the rapid RNA elongation, RNAP needs to select the correct substrate at each sequence position. RNAP utilizes four ribonucleotides that have different nucleobase moieties: two with purine bases (guanine and adenine) and two with pyrimidine bases (cytosine and uracil). Out of these substrates only one is correct at each sequence position, depending on the tDNA acceptor base presented at the A-site. There are also 2'deoxyribonucleotides in a cell that have the same guanine, cytosine and adenine bases as ribonucleotides, in addition to the pyrimidine base thymine. NTPs are more abundant than 2'dNTPs, having approximately a 10-fold higher concentration in the cell (Traut 1994). The 2'deoxyribonucleotides make up DNA, so RNAP needs to distinguish them from the ribonucleotides that make up RNA. This chapter focuses on the mechanisms with which multi-subunit RNAPs recognize

the nucleobase and sugar moieties of NTPs to select the correct substrate (Figure 5). To compare selection mechanisms between nucleic acid synthesizing proteins, substrate selection by DNAPs and right-hand RNAPs are also briefly discussed.

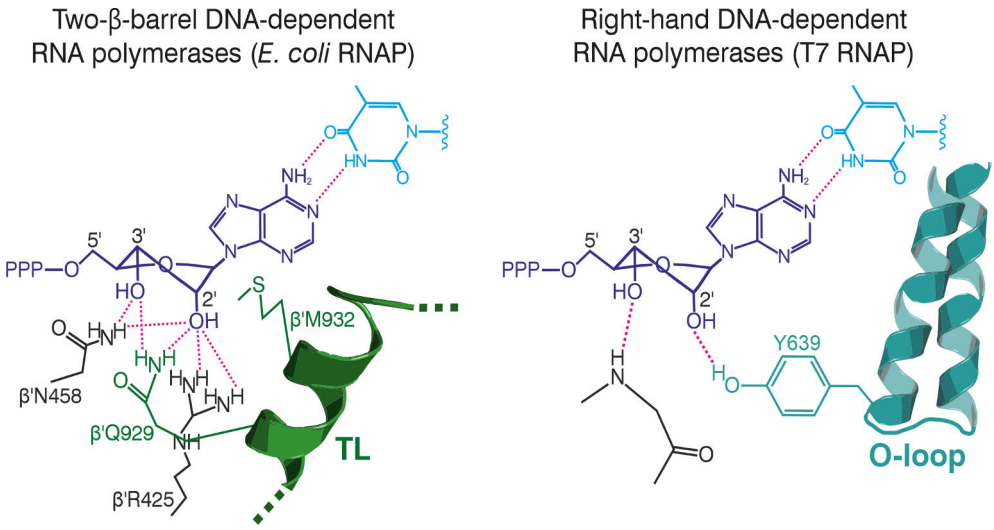


Figure 5. Substrate selection in two-β-barrel multi-subunit RNAPs and right-hand DNA-dependent RNAPs. Template DNA base and residues involved in probing for the correct sugar and base are shown. Hydrogen bonds and polar interactions are shown as magenta dashed lines. Trigger loop (TL, green) in multi-subunit RNAPs and O-loop (teal) in right-hand RNAPs close the active site creating an induced fit for the bound substrate.

2.3.1 Accurate transcription depends on substrate nucleobase recognition

Since RNA is synthesized as a reverse complement of tDNA, RNAP needs to bind nucleotides with the nucleobase moiety that corresponds to the tDNA acceptor base. Like in most enzymes that bind double stranded nucleic acids, the nucleic acid bases in a double stranded hybrid form Watson-Crick base pairs in RNAPs. NTP bases pair with 2'dNTP bases via different numbers of hydrogen bonds as follows: guanine with cytosine (3 bonds), uracil with adenine (2 bonds), and adenine with thymidine (2 bonds). When NTP enters the open active site, hydrogen bonding between the binding NTP and template DNA base, together with base stacking interactions, stabilizes the binding of the correct NTP and position it optimally for catalysis (Temiakov *et al.*, 2004; Vassylyev *et al.*, 2007b).

Since all nucleobase moieties of NTPs can form one or more hydrogen bonds with a DNA acceptor base, how do RNAPs and DNAPs account for the binding of nucleotides with incorrect base moieties (i.e. noncognate NTPs)? Numerous studies

with DNAPs, right-hand RNAPs and two-β-barrel RNAPs show that the major determinant of cognate substrate binding is in fact the optimal geometry that results from the Watson-Crick base pair. Hydrogen bonds are strong and directional, meaning that during pre-insertion positioning of the bound NTP at the A-site will be largely dominated by the hydrogen bonds between the NTP base and tDNA acceptor base. In multi-subunit RNAPs, when TL folds over the bound NTP during insertion, initial positioning of the NTP can either favor catalysis by placing the α-phosphate at an optimal distance from the RNA 3' end (cognate NTP), or disfavor catalysis due to α-phosphate being too far from the nascent RNA 3' end (noncognate NTP) (Vassylyev *et al*., 2007b, Yuzenkova *et al*., 2010). Active sites of DNAPs, right-hand RNAPs and two-β-barrel RNAPs contain residues that interfere with noncognate NTP binding thus contributing to base recognition. For instance, in crystal structures of T7 DNAP and *Bacillus stearothermophilus* DNAP I all four Watson-Crick base pairs place the bound NTPs so that the protein side chains in the active site interact with the hydrogen bond acceptor bases (Doublié *et al*., 1998; Kiefer *et al*., 1998). T7 RNAP can recognize base moieties of bound NTP analogues that have been modified with non-polar atoms, which removes hydrogen bonding between the NTP base and tDNA base (Hirao *et al*., 2006; Seo *et al*., 2009; Ulrich and Kool, 2011). Loss of hydrogen bonding between a NTP and tDNA acceptor base reduces affinity for the substrate, but since the substrate does not clash with active side residues, the substrate is recognized and incorporated. In multi-subunit RNAPs there are multiple amino acid residues around the NTP-dNTP pair that could participate in cognate substrate recognition. β'Pro427 at the active site cleft and the BH residue β'Thr790 can interact with the bound NTP base moiety (Vassylyev *et al*., 2007b; Maffioli *et al*., 2017). TL residues β'Met1238 in *T. thermophilus* RNAP and *S. cerevisiae* (*Sce*) RNAPII Rbp-Leu1081 (which replaces methionine in eukaryotic RNAPII TL) can form stacking interactions with a bound substrate base moiety and could therefore participate in substrate base recognition (Vassylyev *et al*., 2007b; Wang *et al*., 2006; Basu *et al*., 2014). A study where *Thermus aquaticus* β'Met1238 was targeted for mutagenesis showed that mutation from methionine to alanine reduced selectivity by 3-fold and methionine to valine reduced selectivity by 0.6-fold, while mutation from methionine to leucine (the corresponding residue in eukaryotic RNAPs) retained similar selectivity as the wild type (Yuzenkova *et al*., 2010).

Another major determinant of substrate recognition by two-β-barrel and right-hand RNAPs is the closure of the active site, which creates an induced fit with the bound NTP. Transcriptional fidelity is linked to active site closure: a more stable closed active site promotes catalysis of both cognate and noncognate NTPs, which increases the overall rate of RNA synthesis, but at the same time promotes misincorporation. In T7 RNAP the recognition of a NTP base moiety is facilitated

by the induced fit during active site closure, since mutations that stabilize the closed conformation of the T7 RNAP active site increase the misincorporation rate (Huang *et al*., 2000). In polioviral RNA-dependent RNAP, multiple mutations increase the fidelity of RNA replication while decreasing the elongation rate by altering the palm domain folding that closes the active site (Campagnola *et al*. 2015). In two-β-barrel multi-subunit RNAPs, the TL, which closes the active site by folding over the NTP, favors cognate substrate recognition kinetically by selectively increasing the catalysis rate of cognate NTPs (Kireeva *et al*., 2008; Larson *et al*., 2012). The *E. coli* RNAP variants β'G1136S and β'F773V, as well as *Sce* RNAPII variant Rbp1-E1103G, which stabilize TL folding, display faster kinetics of NTP incorporation, but also have higher misincorporation rates than wild type enzymes (Bar-Nahum *et al*., 2005; Nedialkov *et al*., 2013; Kireeva *et al*., 2008; Kaplan *et al*., 2008; Larson *et al*., 2012). The *E. coli* β'I1134V variant that destabilizes TL folding has decreased catalysis rates for both cognate and noncognate NTPs, but it favors incorporation of the cognate NTPs over noncognate ones more than wild type, thus increasing transcriptional fidelity (Bar-Nahum *et al*., 2005). Interestingly, when the entire TL is removed, RNAP discriminates noncognate base moieties poorly (Yuzenkova *et al*., 2010), indicating that some degree of TL folding is required to maintain fidelity, possibly through contacts which β'Met932 or other TL residues make with the NTP nucleobase moiety.

Noncognate NTPs are frequently misincorporated into RNA, which results in a mismatch between the RNA 3' end NMP and tDNA acceptor base. Certain mismatches are more probable than others (Kashinka *et al*., 2006; Sydow *et al*., 2009; Yuzenkova *et al*., 2010; Gout *et al*., 2013; Imashimizu *et al*., 2013; Imashimizu *et al*., 2015; Gout *et al*., 2017; Chung *et al*., 2023) and occur more frequently at specific transcribed positions (Imashimizu *et al*., 2015; James *et al*., 2017; Cheung *et al*., 2020). The probability of mismatches is also dependent on the varying concentrations of available NTPs: if the concentration of one NTP greatly exceeds the others, it can be more easily misincorporated in place of the correct NTP (Larson *et al*., 2012). When intracellular concentrations of NTPs are in balance, RNAPs can retain accurate transcription even when NTP concentrations are limited due to growth conditions (Traverse and Ochman, 2016). A misincorporated nucleotide at the RNA 3' end does not properly form a hydrogen bond with the tDNA acceptor base, which results in fraying of the ultimate 3' end nucleotide and backtracking of the TEC. After misincorporation, RNAP either (i) translocates forward and continues transcribing the DNA leaving the error in the RNA transcript, or (ii) removes the misincorporated nucleotide from the RNA 3' end and tries to transcribe the same position again correctly (transcriptional proofreading is discussed further in chapter 2.4).

2.3.2 Differentiation of NTPs from 2'dNTPs

RNAPs synthesize RNA using NTPs and DNAPs synthesize DNA using 2'dNTPs. Both enzymes need to actively separate NTPs from 2'dNTPs by probing the 2' position of the nucleo-sugar moiety, which either has an OH group (NTP) or not (2'dNTP). The atoms at the 2' position affect the helical form a nucleic acid assumes in a hybrid (A-helical form of RNA and B-helical form of DNA), so incorporation of the nucleotides with the right nucleo-sugar moiety is essential for both DNAPs and RNAPs to ensure the correct structure of the synthesized oligonucleotide. 2'dNTPs are synthesized from NTPs by ribonucleotide reductases that remove the 2'OH groups from NTPs.

DNAPs utilize a steric gate that clashes with the 2'OH group of NTP, while allowing the smaller 2'dNTP to bind efficiently (reviewed in Brown and Suo, 2011). For RNAPs the challenge of nucleo-sugar selection arises from the fact that 2'dNTPs are smaller than NTPs, due to the missing 2'OH group, so RNAPs need to disfavor binding of a noncognate substrate that fits in the active site better than the cognate one. Right-hand RNAPs and two-β-barrel RNAPs use distinctively different mechanisms to select NTPs with the right sugar moiety. Despite being closely related to family A DNAPs, the right-hand RNAPs do not apply a steric gate, but instead utilize a conserved tyrosine in the O-loop (Tyr639 in T7 RNAP), which forms a hydrogen bond with the 2'OH group of an incoming NTP serving as the major determinant of nucleo-sugar selection (Figure 5) (Sousa and Padilla, 1995; Temiakov *et al*. 2004; Huang *et al*., 1997; Yin and Steitz, 2004).

While the mechanism of nucleo-sugar selection by right-hand RNAPs has been largely determined, the mechanism of nucleo-sugar selection by the two-β-barrel RNAPs remains poorly understood. Biochemical data shows that RNAPs from various organisms bind 2'dNTPs very poorly, exhibiting 500- to 5000-fold differences in selectivity between NTPs and 2'dNTPs (Svetlov *et al*., 2004; Wang *et al*., 2006; Kaplan *et al*., 2008; Zhang *et al*., 2010; Yuzenkova *et al*., 2010), which means that multi-subunit RNAPs have an efficient mechanism to reduce affinity for the smaller substrate. There are multiple residues in a multi-subunit RNAP that come in close proximity to the 2' position of NTPs bound to the A-site: β'Arg425, β'Asn458, β'Gln929 and β'Met932 (Figure 5) (Vassylyev *et al*., 2007b; Wang *et al*., 2006; Maffioli *et al*., 2017). Since β'Gln929 and β'Met932 are in the TL, they come into close contact with the bound substrate only upon TL folding. The *T. aquaticus* RNAP variant with a TL deletion was found to be unselective for 2'dNTPs (Yuzenkova *et al*., 2010). However, studies on the *Sce* RNAPII and *E. coli* RNAP suggested that TL would contribute only 5- to 10-fold to an overall 500- to 5000-fold selectivity of NTPs over 2'dNTPs, which is supported by findings that in the absence of a TL, *E. coli* RNAP can still efficiently discriminate 2'dNTPs (Kaplan *et al*. 2008; Zhang *et al*., 2010). Therefore, while TL does have an effect, other contacts

with the sugar moiety and RNAP are likely the major determinants for selection of NTPs over 2'dNTPs. Mutations of the invariant β'Asn458 residue affected selectivity of NTPs over 2'dNTPs only by about 5-fold (Svetlov *et al*., 2004; Wang *et al*., 2006). The invariant residue β'Arg425 has been suggested to participate in sugar moiety recognition due to its positioning in the substrate loading structures (Vassylyev *et al*., 2007b), and MD simulations by Roßbach and Ochsenfeld suggested that β'Arg425 would be the major determinant of nucleo-sugar selection (Roßbach and Ochsenfeld, 2017). We recently determined the role of β'Arg425 in 2'dNTP discrimination and showed that it favors binding of NTPs over 2'dNTPs by more than 100-fold (Mäkinen *et al*., 2021).

2.4 Correcting errors during transcription; proofreading

While substrate selection works as a major step in ensuring the accuracy of synthesized oligonucleotides, both RNAPs and DNAPs periodically incorporate the wrong nucleotides into a growing nucleic acid chain. Multi-subunit RNAPs and replicative DNAPs have evolved to resolve such errors by removing the misincorporated nucleotides from the 3' end of the nucleic acid chain (Scheuermann *et al*., 1983; Orlova *et al*., 1995). This activity allows the proofreading of nucleic acids as they are synthesized, which further improves the accuracy of nucleic acid synthesis. There is a major difference in how replicative DNAPs and multi-subunit RNAPs perform the mismatch removal. Replicative high-fidelity DNAPs have a separate 3'-5' exonuclease active site, located far away from the polymerase active site, to which the mismatched DNA 3' end is directed for mismatch removal via exonuclease reaction (Wang *et al*., 1997; Shamoo and Steitz, 1999; Fernandez-Leiro *et al*., 2017). In contrast, in multi-subunit RNAPs the same active site that catalyzes incorporation of NTPs into RNA also catalyzes the hydrolysis of nucleotides from the RNA, meaning that the multi-subunit RNAP active site has evolved to catalyze two opposite reactions (Orlova *et al*., 1995; Sosunov *et al*., 2003). In multi-subunit RNAPs, mismatch removal occurs via endonucleolytic RNA cleavage, meaning that two or more nucleotides are removed from the RNA 3' end. RNAPs from the right-handed family do not have intrinsic endonucleolytic cleavage activity in the same peptide with the polymerase activity, so they rely on substrate selection to maintain accurate RNA synthesis. However, some RNA viruses, such as coronaviruses, utilize additional viral proteins that can provide proofreading activity outside the viral RNAP core to improve fidelity of viral genome replication (reviewed in Robson *et al*., 2020). In the following chapter, the endonucleolytic RNA cleavage activity of multi-subunit RNAPs is discussed.

2.4.1 The two-metal ion mechanism of endonucleolytic RNA cleavage by multi-subunit RNAPs

Both nucleotidyl-transfer and endonucleolytic RNA cleavage reactions are dependent on the same catalytic Mg-ions and active site domains. As described above, the role of endonucleolytic RNA cleavage in multi-subunit RNAPs is to provide a mechanism for correcting transcriptional errors during RNA elongation, but it also allows the RNAP to escape from backtracking dependent pauses. After mismatch incorporation the RNAP is inclined to backtrack by one residue, because the mismatched NMP at the RNA 3' end does not base pair properly with the tDNA acceptor base, destabilizing the RNA-tDNA hybrid (Sydow *et al*., 2009; Imashimizu *et al*., 2013). In a backtracked state the RNAP catalyzes the cleavage of the phosphodiester bond at the same position where it is formed during the nucleotidyl-transfer reaction: between *i* and *i*+1 sites (Figure 6). In the backtracked state, two or more nucleotides are removed from the RNA depending on the number of steps the RNAP backtracks (Zhang *et al*., 2010; Yuzenkova and Zenkin, 2010; Sosunova *et al*., 2013; Turtola *et al*., 2018).

It has been proposed that switching between the RNA extension and RNA cleavage functions of the RNAP active site is achieved by altering the position of the second catalytic Mg ion, Mg-II (Sosunov *et al*., 2003; Sosunova *et al*., 2013), which is supported by findings that mutations in residues binding Mg-II impair the RNA cleavage reaction (Sosunov *et al*., 2003). Sosunova *et al*. compared the structure of backtracked *Sce* RNAPII (PDB: 3GTG) to a pre-translocated elongation complex of *Sce* RNAPII (PDB: 1I6H) and modelled a possible reorientation of Mg-II that would favor RNA cleavage over RNA extension (Sosunova *et al*., 2013; Wang *et al*., 2009; Gnatt *et al*., 2001). Coordination of Mg-II allows the multi-subunit RNAP to switch between its activities so that the RNA is not cleaved during active elongation. While the altered coordination of Mg-II provides an elegant solution for the central dilemma of how RNAPs can utilize the same active site for two different reactions, it is unlikely that the positioning of Mg-II is the only factor regulating tuning between the RNA extension and cleavage reactions. Otherwise, RNAP would rapidly cleave RNA in a pre-translocated state making RNA extension and RNA cleavage competing reactions, which would slow down transcription and needlessly deplete NTP pools. In the case of mismatch removal, backtracking is the consequence of a mismatched NMP losing contact with the tDNA acceptor and fraying away from the A-site, which provides a way for RNAP to probe for RNA synthesis errors. Therefore, it makes sense that Mg-II tuning to RNA cleavage mode occurs in response to RNAP backtracking, which alters the conformation of nucleic acids in the active site.

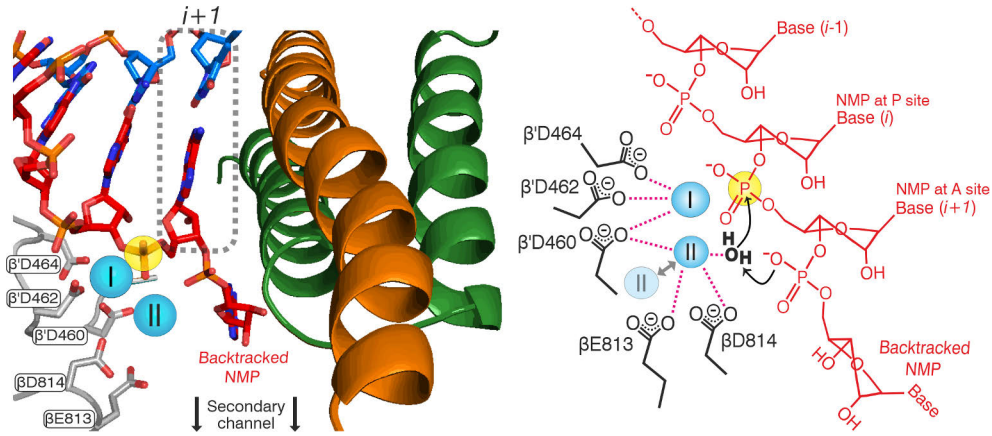


Figure 6. Endonucleolytic RNA cleavage of backtracked nucleotides by the multi-subunit RNAP. (Left) The model of *T. thermophilus* RNAP in one nucleotide backtracked state with TL (green) folded on the backtracked NMP (originally presented in Turtola *et al.*, 2018). Residues coordinating the catalytic Mg^{2+} ions (I and II, cyan spheres) are shown as sticks and labeled accordingly. Numbering of the residues (black text) follows *E. coli* nomenclature. The *i + 1* (substrate binding) site is highlighted with a dashed grey line. The phosphodiester bond that is cleaved in the reaction is highlighted with a yellow sphere. (Right) Mechanism of transcript mediated endonucleolytic RNA cleavage (as described in Ka Man Tse *et al.*, 2019). The position of Mg-II changes when active site is tuned for RNA cleavage (shown with a grey arrow). Polar interactions are shown as magenta dashed lines.

2.4.2 General acid and base in the endonucleolytic RNA cleavage reaction

Due to the strong pH dependency of the RNA cleavage reaction, the mechanism of endonucleolytic RNA cleavage has been considered to occur via acid-base catalysis, but the general acid and base involved have not been identified to date. TL has been suggested to participate in the acid-base mechanism, because removal of the entire TL or inhibition of TL folding by salinamide (in bacterial RNAP) greatly decreases the rate of RNA cleavage (Yuzenkova and Zenkin, 2010; Mosaei and Zenkin, 2021). Additionally, multiple point mutations directed at different TL residues or to residues that affect TL folding have varying positive or negative effects on RNA cleavage rates (Yuzenkova and Zenkin, 2010; Turtola *et al*., 2018; Riaz-Bradley *et al*., 2020). It was initially proposed that TL could provide catalytic residues for the cleavage reaction, explaining the strong effect the TL folding has on the reaction (Yuzenkova and Zenkin, 2010). However, the only TL residues that could work as an acid or base in reactions catalyzed by bacterial RNAP (β'R933 and β'H936), are not required for the nucleotidyl-transfer or cleavage reaction (Mishanina *et al*., 2017). Similarly, Rbp1-His1085 substitution to leucine in the *Sce* RNAPII TL had no measurable effect on the rate of intrinsic RNA hydrolysis (Palo *et al*., 2021). Endonucleolytic RNA cleavage can also proceed without TL folding (Fouqueau *et al*., 2013; Esyunina

et al., 2016; Miropolskaya *et al*., 2017; Mosaei & Zenkin, 2021), further challenging the view that TL residues could work as an acid or base. It was suggested that TL works as a positional catalyst instead, directing the RNA 3' end NMP in the cleavage reaction into an induced fit mechanism (Mishanina *et al*., 2017). This view is supported by the finding that TL folding stabilizes the one nucleotide backtracked state, which in turn increases the RNA cleavage rate (Turtola *et al*., 2018).

The general acid or base could be in the nascent RNA chain itself, since multiple classes of ribozymes (RNA structures that catalyze chemical reactions) catalyze endonucleolytic cleavage (Perrault *et al*., 2011; Webb and Lupták, 2011; Roth *et al*., 2014). The NMP at the RNA 3' end has multiple functional groups that could work as a general acid or base and their roles in RNA cleavage have been extensively studied. Different mismatched NMPs are cleaved at different rates (Zenkin *et al*., 2006; Sydow *et al*., 2009), meaning that either (i) the different 3' end NMPs provide different functional groups for RNA cleavage from their base moieties, or (ii) the differences in RNA cleavage rates are dependent on the structure of the RNA-tDNA hybrid, which affects backtracking and the different conformations RNA adopts in the backtracked state. A study by Zenkin *et al*. suggested that different NMP bases at the RNA 3' end provide different reaction pathways for RNA cleavage, which could explain the differences in sequence dependency of RNA cleavage (Zenkin *et al*., 2006).

Later, Sosunova *et al*. proposed that the oxygen at the ultimate 3' end phosphodiester bond in the RNA transcript could work as the general base in the cleavage reaction and showed that alterations to the oxygen group greatly reduced the RNA cleavage rate (Sosunova *et al*., 2013). Detailed QM/MM analysis on *Sce* RNAPII provided further insight into the possibility that non-bridging phosphate oxygen in the phosphodiester bond between the ultimate and penultimate NMPs could work as the general base (Ka Man Tse *et al*., 2019). In the proposed reaction model by Ka Man Tse *et al*. the non-bridging phosphate oxygen deprotonates water that is coordinated by Mg-II, after which the water attaches to the next phosphodiester bond upstream, resulting in the formation of a dinucleotidemonophosphate and a new RNA 3' end NMP at *i*-1 (Figure 6). This reaction model is not dependent on the nucleobases at the RNA 3' end and it provides a logical solution for how reactants are regenerated. Ka Man Tse *et al*. also showed that altering the position of the phosphodiester bond from the 5'-3' to 5'-2' position slows down the RNA cleavage rate *in vitro*. The theory that RNA participates in its own cleavage concurs with the view that the role of TL is to direct the RNA 3' NMP for cleavage, while the Mg-ions coordinate water in the endonucleolytic cleavage reaction. Further studies are still needed to clarify the identity of the general acid and base in the cleavage reaction.

2.4.3 Roles of the RNAP domains in the endonucleolytic cleavage reactions

While the cleavage reaction mechanism is believed to be conserved due to the high similarity of RNAP active sites, the multi-subunit RNAPs of different organisms cleave RNA at very different efficiencies (Miropolskaya *et al*., 2017). Amino acid residues coordinating the catalytic Mg-ions are conserved, meaning that other species-specific differences around the active site affect RNA cleavage. The E-site could be one of these areas, since it can bind nucleotides (Wang *et al*., 2006; Zhang *et al*. 2014) and point mutations in the E-site residues β'Gln504 and β'Lys598 greatly affect the RNA cleavage reaction without affecting stability of the backtracked state (Turtola *et al*., 2018). In multiple structures of RNAPs in 1 nucleotide long backtracked states, the backtracked nucleotide is positioned close to the P-site (Wang *et al*., 2009, Sekine *et al*., 2015, Wee *et al*., 2023), indicating that it might have a role in RNA cleavage. While the role of the TL in RNA cleavage is disputed, the species-specific differences in RNA cleavage are largely contributed to TL folding dynamics, as removal of the TL in *E. coli*, *D. radiodurans* and *T. aquaticus* RNAPs erases the large differences between their RNA cleavage efficiencies (Miropolskaya *et al*., 2017). Also, many amino acid residues that affect TL folding dynamics have been shown to affect backtracking and RNA cleavage. For example, the *E. coli* RNAP variants β'F773V and β'I937T, in which the *E. coli* residues are substituted with corresponding cyanobacterial residues, display greatly increased RNA cleavage rates (Riaz-Bradley *et al*., 2020; Turtola *et al*., 2018). Understanding these differences is important to determine how RNAPs from different organisms regulate RNA cleavage, which affects two important events of transcription: removal of mismatches and pause release.

2.4.4 Cleavage factors improve the rate of RNA cleavage

Multiple organisms express cleavage factors, which accelerate the RNA cleavage rate by coordinating the catalytic Mg-II: GreA and GreB in bacteria (Borukhov et al., 1993), TFS in archaea (Lange and Hausner, 2004), and TFIIS in eukaryotes (Reines, 1992). Cleavage factors are inserted through the secondary channel into the active site of the RNAP, where they provide additional functional amino acid groups for the RNA cleavage reaction. GreA assists with the cleavage of RNA dinucleotides. GreB accelerates the cleavage of longer stretches of RNA with multiple nucleotides at a time, which the RNAP has difficulties cleaving independently due to TL clashing with the multiple backtracked nucleotides (Cheung and Cramer 2011; Abdelkareem *et al*., 2019). Gre factors are not uniformly distributed among bacteria, since bacteria have genes for both factors, one, or none. For example, *E. coli* expresses both Gre factors, while cyanobacteria do not express

any Gre factors, possibly because they can intrinsically cleave RNA at high efficiency without them, retaining a similar accuracy for RNA synthesis as *E. coli* (Riaz-Bradley *et al*., 2020). In all domains of life the cleavage factors majorly contribute to the overall fidelity of transcription by increasing the proofreading activity of RNAP (Imashimizu *et al*., 2015; Lange and Hausner, 2004; Gout *et al*., 2017; James *et al*., 2017).

2.5 Effects of transcriptional fidelity at the cellular level

While the primary structures of DNA and RNA are similar, the accuracy with which they are synthesized is drastically different, reflecting their respective roles in the cell. Genetic information encoded in the DNA is transferred to every future generation of the cell, so DNA needs to stay intact through the entire life cycle of the cell. Cells tolerate increasing amounts of DNA damage and errors poorly because uncorrected errors keep affecting everything the DNA encodes. Cells spend a lot of energy on minimizing DNA damage and errors, though sometimes errors remain unrepaired driving genetic variation in organisms in the form of mutations. It is estimated that high-fidelity DNAPs make errors once in every 10^8–10^{10} nucleotides of replicated DNA (Lang and Murray, 2008; Lynch, 2010). In contrast, RNAPs make errors once in every 10^4–10^6 nucleotides of transcribed RNA (Lynch, 2010; Gou *et al*., 2013; Imashimizu *et al*., 2013; Traverse and Ochman, 2016; Li and Lynch, 2020, Gout *et al*., 2017; Chung *et al*., 2023), which is a multiple orders of magnitude higher error rate compared to DNAPs.

RNAs are synthesized and degraded actively, so the effect of the errors in RNA transcripts is dependent on the lifetime of the transcripts. The strength of the effect a single transcriptional error has on the cell depends on how many mRNA transcripts are produced for one gene and how many proteins are produced from a single mRNA. In microbes and embryonic stem cells, the average number of specific mRNA molecules can be close to one (Pelechano *et al*., 2010; Islam *et al*., 2011) and the number of proteins that are expressed from one mRNA molecule can vary from 2000 to 4000 proteins (Schwanhäusser *et al*., 2011), meaning that transcriptional errors can affect a large number of translated proteins that are expressed from one gene. Because of the transient nature of RNA, cells have better tolerance for transcriptional errors than for DNA replication errors, yet inaccurate transcription slows down RNA synthesis due to mismatch induced pausing (James *et al*., 2017), needlessly depletes NTP pools, and can lead to the production of dysfunctional proteins (Vermulst *et al*., 2015).

The overall error rate of transcription is largely dependent on the inherently error-prone activity of RNAP, while some errors are induced by DNA damage.

Misincorporated nucleotides left in the final RNA transcripts are one source of errors, while another source is transcriptional slippage in which RNAP either skips a template position(s) without adding an NTP to the nascent RNA chain, or adds extra nucleotides against the same template position resulting in heterogeneous RNA transcripts (reviewed in Anikin *et al*., 2010). Codon frame shifts in RNA transcripts, which can be caused by transcriptional slippage, generate nonsense proteins causing proteotoxic stress to cells, which has been linked to neurodegenerative diseases such as Alzheimer's disease (van Leeuwen *et al*., 1998). RNAPs can transcribe past DNA lesions resulting in erroneous RNA transcripts, which is referred to as transcriptional mutagenesis (Viswanathan and Doetsch, 1998; Marietta and Brooks, 2007; Wang *et al*. 2018). In eukaryotes, errors made by RNAP can lead to splicing defects, such as intron retention, in the final mRNA transcript (Carey, 2015).

Studying the transcriptional fidelity at a transcriptome level is difficult due to multiple variables that are related to gene expression, including cell cycle phases, availability of nutrients, loci dependent differences in the half-lives of mRNAs (Wilusz *et al*., 2001), and the effects of mutations in used cell strains. Since RNAs are reverse-transcribed to cDNA prior to sequencing, the inherent error rate of reverse transcriptase introduces errors during RNA sequencing (Reid-Bayliss and Loeb, 2017). Additionally, cytosine deamination to uracil after an RNA transcript has been generated introduces errors in the sequence analysis (Traverse and Ochman, 2016). Next-generation RNA sequencing techniques (Gout *et al*., 2013; Carey, 2015) and reporter genes that produce functional biomarkers only after transcriptional errors occur (Strathern *et al*., 2013; Vermulst *et al*., 2015) have provided ways to study transcriptional fidelity *in vivo*. While the detected error rates can vary by orders of magnitude depending on the studied organism and used sequencing methods (Bradley *et al*., 2019), eukaryotes appear to have an overall higher transcriptional fidelity than bacteria (Li and Lynch, 2020). This difference is partly due to the eukaryotic mRNA decay pathways that detect and degrade incorrectly synthesized mRNAs, which reduces the number of erroneous mRNA transcripts (reviewed in Isken and Maquat, 2007). In eukaryotes the accuracy of rRNA and mRNA synthesis is higher than the accuracy of tRNA and mitochondrial RNA synthesis (Reid-Bayliss and Loeb, 2017; Gout *et al*., 2017; Chung *et al*., 2023), indicating that eukaryotic RNAPs display different levels of transcriptional fidelity.

As described in chapter 2.4.4, most prokaryotes, archaea and eukaryotes depend on cleavage factors for efficient proofreading and pause release, so deletions of these factors, or mutations that impair their function can provide insight into how large an effect transcriptional proofreading has on cells. *E. coli* cells with one or both Gre factors deleted have a significantly higher number of mismatches in the final RNA transcripts than the wild type strain (Imashimizu *et al*., 2015; Bubunenko *et al*., 2017; James *et al*., 2017), emphasizing the impact that proofreading has on fidelity of

transcription. Deletion of TFIIS from yeast cells also results in significantly increased error rates in mRNA transcripts (Gout *et al*., 2017; James *et al*., 2017).

Two mutant *S. cerevisiae* strains that have reduced transcriptional fidelity have been extensively studied: rpb1-E1103G and Δrpb9. The former has a point mutation in the eukaryotic RNAPII TL that impairs the function of cleavage factor TFIIS and increases the misincorporation of NTPs (Malagon *et al*., 2006, Kireeva *et al*., 2008). The latter has a deletion of a nonessential subunit of RNAPII, Rpb9, which results in a more stable unfolded state of TL, thus increasing misincorporation of NTPs (Walmacq *et al*., 2009). These mutations result in a 3- to 9-fold higher transcription error rate than the wild type strain (Kireeva *et al*., 2008; Walmacq *et al*., 2009; Reid-Bayliss and Loeb, 2017; Gout *et al*., 2017). A study by Vermulst *et al*. showed that rpb1-E1103G and Δrpb9 mutant strains had elevated levels of chaperones that assist in the refolding of misfolded proteins (Ydj1) and prevent protein aggregation in cells (Ssa1). Deletions of these chaperones resulted in a dramatic reduction of cell growth rate for the mutant strains, but not for the wild type strain, which indicates that the increased error rate in RNA transcripts affected the mutant strain cells by inducing proteotoxic stress (Vermulst *et al*., 2015).

Conflicts between transcribing RNAPs and replisomes can result in DNA damage. Stalling RNAPs can meet with replisomes moving in the opposite direction leading to head-on collisions between the two, which causes DNA damage (Pomerantz and O'Donnell, 2008; Pomerantz and O'Donnell 2010; Nudler, 2012). To avoid these collisions, most highly expressed genes in all domains of life are oriented so that the transcribing RNAPs do not meet with replisomes head-on (Srivatsan *et al*., 2010; Petryk *et al*., 2016). Conflicts between RNAPs and replisomes are more likely to occur when there are more molecules moving along DNA, which often occurs in cancer cells where both transcription and replication are upregulated (Kotsantis *et al*., 2016; Gaillard *et al*., 2015). In the absence of cleavage factors, RNAPs dwell in the paused states for longer, increasing the probability of conflicts between replisomes and RNAPs, thus promoting DNA damage (Dutta *et al*., 2011). Mismatches, which are a major source of RNAP pausing *in vivo* (James *et al*., 2017), increase the probability of RNAP-replisome collision. Translocases (Mfd) and macromolecules (trailing ribosomes in bacteria) that can force RNAP forward past a mismatch can safeguard cells from DNA damage at the expense of transcriptional fidelity (Le *et al*., 2018; Wee *et al*., 2023).

2.6 Nucleoside analogues as RNAP inhibitors

All RNAPs synthesize RNA from the same NTP precursors. The mechanisms for substrate selection and proofreading are different for the two-β-barrel RNAPs and right-hand RNAPs, which opens up possibilities for creating selective inhibitors for

viral RNAPs and bacterial RNAPs that mimic natural NTP substrates. Indeed, multiple clinically approved antivirals are nucleoside analogues (reviewed in De Clercq and Li, 2016). Due to the increasing prevalence of antibiotic resistant bacteria, new sources of potential antibacterial compounds, such as nucleoside analogues, should be explored. At the time this thesis was written there are no clinically approved antibacterial compounds that are nucleoside analogues, although pseudouridimycin shows great potential as a bacterial RNAP inhibitor (Maffioli *et al*., 2017). Currently available inhibitors of bacterial transcription are allosteric inhibitors that alter the structure of different RNAP domains, compounds that sterically block parts of the active site thus preventing NTP binding or RNA elongation, and compounds that disrupt protein-protein interactions between RNAP and transcription factors (Table 1). Out of all these compounds, only rifamycins and fidaxomicin have received approval for clinical use.

2.6.1 Structure and function of nucleoside analogues

The base, sugar and triphosphate moieties of nucleotides can be chemically modified to alter their selectivity and binding efficiency. A wide variety of naturally occurring nucleobase analogues have been discovered (Figure 7) and chemically modified to create drug compounds. Sugar moieties can be modified by adding additional chemical groups to different positions around the ribose, or replacing existing ones (3' and 2' hydroxyl groups). Since NTP analogues are substrate derivatives of NTPs, they are subjected to the same catalytic and regulatory steps as NTPs: substrate binding (involving recognition of nucleobase and sugar moieties), catalysis, translocation and proofreading.

Because modified base analogues interact with the tDNA acceptor base differently, they can be susceptible to transcriptional proofreading in two-β-barrel RNAPs, if the nucleoside analogue increases the fraction of backtracked TECs after incorporation. Pseudouridine and oxazinomycin closely resemble uridine, which is why they are incorporated into RNA efficiently by *E. coli* RNAP, *Sce* RNAPII and human mitochondrial RNAP. RNAPs proceed rapidly with elongation after incorporating pseudouridine and oxazinomycin, but in the presence of oxazinomycin RNAPs pause strongly at polythymidine sequences, where multiple oxazinomycins are incorporated into the RNA (Prajapati *et al*., 2019). Cleavage factors can remove nucleoside inhibitors from the RNA limiting the inhibition of multi-subunit RNAPs (Arnold *et al*., 2012; Prajapati *et al*., 2019), while the lack of proofreading activity makes single-subunit RNAPs especially vulnerable to nucleoside inhibitors.

Table 1. Compounds that inhibit bacterial multi-subunit RNAPs.

Mode of action	Compounds	References
Allosterically inhibits TL folding by destabilizing contacts between BH and TL.	CBR703, CBR9393 and CBR9379	Malinen *et al.*, 2014
	D-AAP1	Lin *et al.*, 2017
	Salinamides	Degen *et al.*, 2014
	Streptolydigin	Temiakov *et al.*, 2005
Occupies the same area as PP_i in the active site, stabilizes the folded state of TL.	Tagetitoxin	Malinen *et al.*, 2012
Binds in between the switch domains allosterically locking the clamp domains in closed state, which inhibits DNA loading to the active site cleft during initiation.	Fidaxomicin*	Tupin *et al.*, 2010
	Myxopyronine	Belogurov *et al.*, 2009
	Corallopyronine	Mukhopadhyay *et al.*, 2008
	Ripostatin	Mukhopadhyay *et al.*, 2008
	Squaramides	Molodtsov *et al.*, 2015
Binds to the active site blocking NTPs from binding.	Pseudouridimycin	Maffioli *et al.*, 2017
	GE20377	Zhang *et al.*, 2014
	Microcin J25	Mukhopadhyay *et al.*, 2004
Binds to a cleft in primary channel near the active site sterically blocking RNA chain elongation limiting it to 2-3 nucleotides during initiation phase.	Rifamycins*	Campbell *et al.*, 2001
	Sorangicin	Campbell *et al.*, 2005
	Kanglemycin A	Mosaei *et al.*, 2018
Destabilizes RNAP contacts with σ-factors, which disrupts holoenzyme formation and inhibits transcription initiation.	GKL003	Ma *et al.*, 2013
	DSHS00507	Ma *et al.*, 2016a
	SB11 and SB15	André *et al.*, 2004
Inhibits ATP hydrolysis of termination factor Rho at termination phase.	Bicyclomycin	Carrano *et al.*, 1998

* Approved for clinical use.

Nucleoside analogues can target viral or bacterial RNAPs by (i) inhibiting further RNA elongation after incorporation (chain terminators), (ii) occupying the active site preventing substrate binding (non-hydrolyzable analogues), (iii) inducing pausing after incorporation, or (iv) altering the structure of the final RNA transcripts thus leading to dysfunctional RNA structures that are misread either during RNA replication (viruses), or during translation (i.e. error catastrophe). As described in chapter 2.5, altering the accuracy of transcription can cause adverse effects in cells, such as DNA damage and proteotoxic stress, so incorporated nucleoside analogues that alter RNA structure and/or ribosome reading during translation could be utilized

to limit the growth and proliferation of pathogenic bacteria. In the case of viral RNAPs, nucleoside analogues are used to inhibit both translation of viral RNA by the host and replication of the viral RNA genome (Figure 8). Nucleoside analogues that are incorporated into RNA in the target organism could work as "Trojan horses" that inhibit transcription or downstream events that follow RNA synthesis by altering the structure of RNA.

A nucleobase moiety can rotate along the nucleosidic bond forming interconvertible conformers where the nucleobase is facing the acceptor tDNA nucleobase differently, altering hydrogen bonding. The *anti*-conformation is the canonical conformation for cognate nucleotides that form Watson-Crick base pairs, whereas in the *syn*-conformation the base moiety (or a functional group in it) has rotated so that the potential hydrogen donors and acceptors in the base moiety face in a different direction. This results in an effect called "dual coding", which allows one nucleotide to be incorporated in place of multiple nucleotides. The best-known dual coders are the oxidized derivatives of adenine and guanine, called 8-oxoadenine (8oA) and 8-oxoguanine (8oG). In *anti*-conformation, 8oA and 8oG pair with uridine and cytidine, respectively, but they can also adopt a *syn*-conformation allowing them to pair with guanine and adenine, respectively (Koag *et al*., 2019; Kamiya *et al*., 1998; Batra *et al*., 2014; Brieba *et al*., 2004). Ribavirin, an inhibitor of viral RNAPs, is a dual coder that can pose as both adenine and guanine because of rotation of the carboxamide group in the base moiety (Crotty *et al*., 2000). Ribavirin incorporation into RNA causes increasing amounts of mutagenesis during consecutive rounds of RNA replication by viral RNAPs leading to an "error catastrophe", gradually inhibiting production of viral proteins in the host cell due to translational errors (Crotty *et al*., 2002; Ortega-Prieto *et al*., 2013).

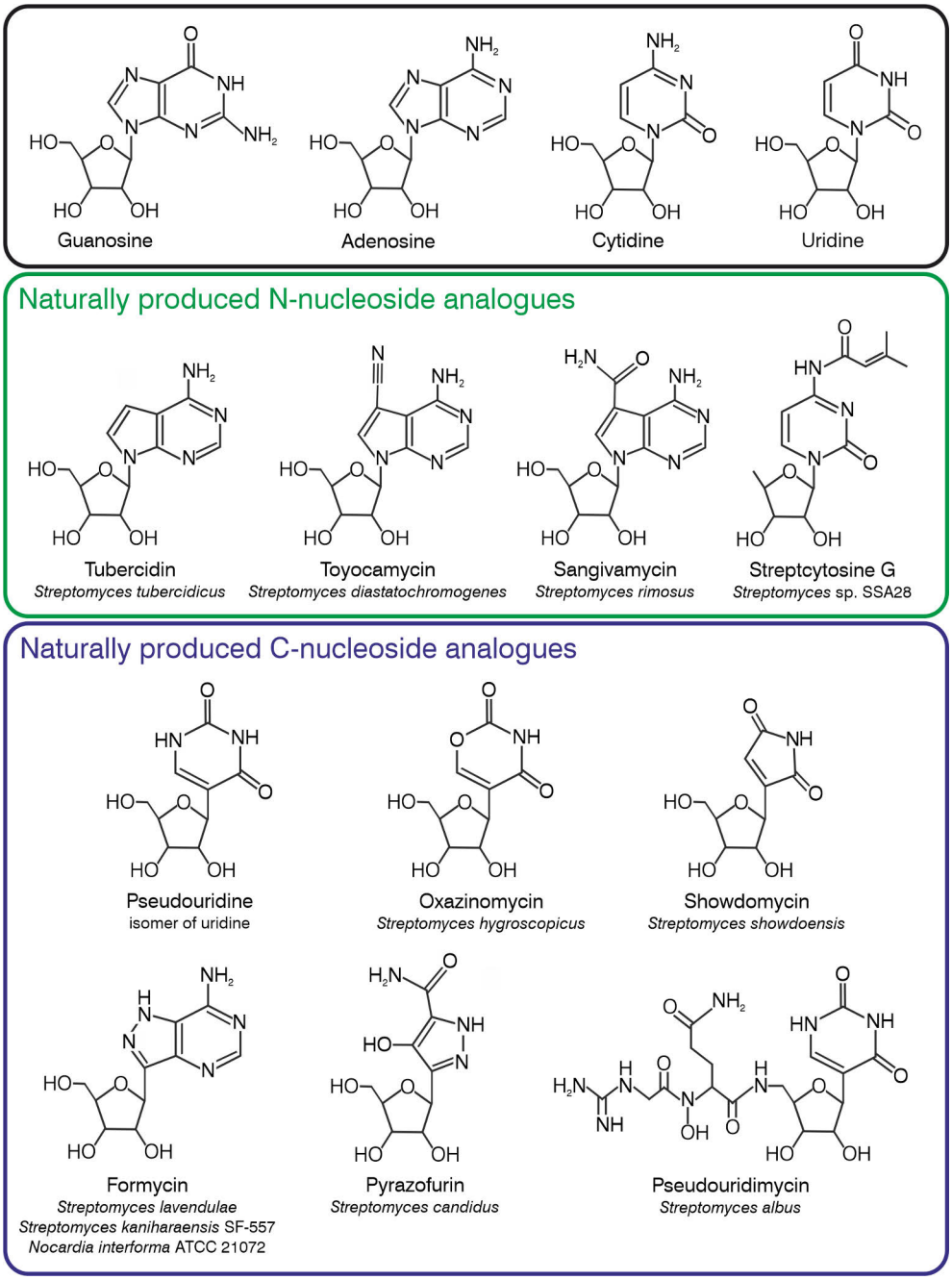


Figure 7. Examples of naturally occurring nucleoside analogues. On the top panel are the cognate ribonucleotides presented as nucleosides. Naturally produced nucleoside analogues of ribonucleotides are presented in two separate panels (N-nucleosides in the middle panel, C-nucleosides in the bottom panel). The origin of the analogue is written below the name.

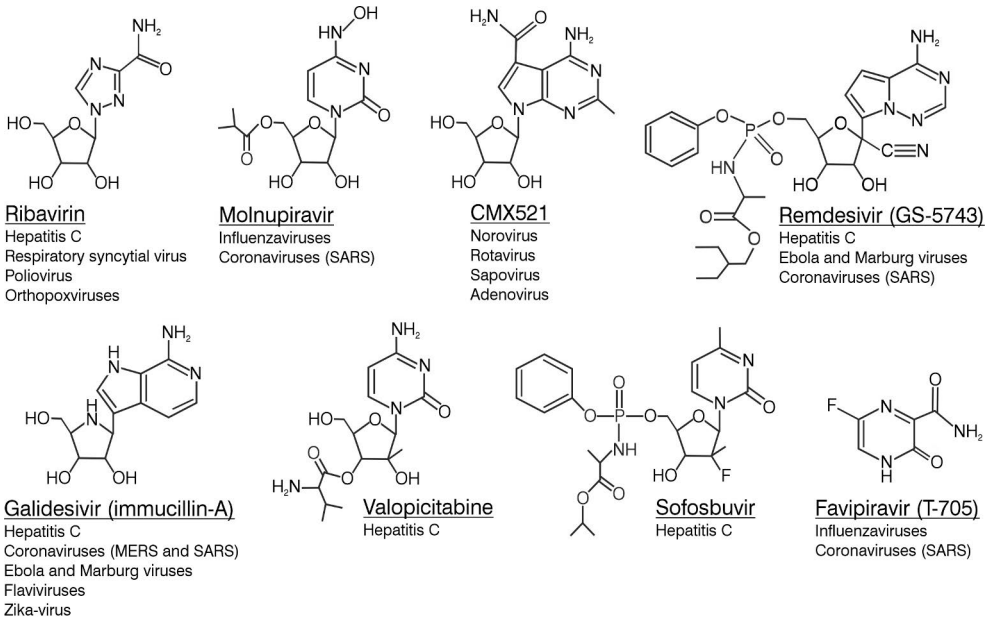


Figure 8. Examples of nucleoside analogues that inhibit viral RNA-dependent RNAPs. The targeted viruses (i.e. RNAPs from these viruses are inhibited by the nucleoside analogue *in vitro*) are shown below the name of the compound. The compounds are named according to the commercial names. Ribavirin, CMX521 and galidesivir are presented as nucleosides, other compounds are presented as prodrugs (forms in which they are administered to the patients).

The ribonucleotides utilized in RNA synthesis are all N-nucleosides, meaning that the glycosidic bond between the nucleo-sugar and base moiety forms between a carbon atom (at the 1' position of the nucleo-sugar) and a nitrogen atom (in the base moiety). Actinobacteria produce diverse natural products as secondary metabolites, which include N-nucleoside and C-nucleoside analogues of ribonucleotides. C-nucleosides have the same ribonucleo-sugar moiety as NTPs, but different base moieties. These C-nucleosides have a glycosidic bond between two carbons atoms, meaning that unlike N-nucleosides, they are not susceptible to enzymatic degradation by phosphorylases, which improves their half-life inside cells. C-nucleosides have been studied for decades for their antiviral and anticancer properties (reviewed in De Clercq, 2016).

To be effective as an antibacterial or antiviral drug, a nucleoside analogue needs to compete with natural NTP substrates. At the same time, the nucleoside analogue must be selective for the intended target, as inhibition of eukaryotic and mitochondrial RNAPs results in cytotoxic effects. Some nucleoside inhibitors of hepatitis C virus are incorporated into RNA by eukaryotic mitochondrial RNAP, inhibiting mitochondrial RNA synthesis (Arnold *et al*., 2012), which limits the synthesis of mitochondrial proteins leading to a reduction in cellular respiration

(Feng *et al*., 2016). Some nucleoside analogues with noncognate base moieties, like the adenosine analogues tubercidin, toyocamycin, sangivamycin and formycin A, display cytotoxic effects in mammalian cells (De Clercq *et al*., 1987). In these cases, the cytotoxicity is likely caused due to eukaryotic and mitochondrial RNAPs utilizing the analogues in place of native substrates, but could also be linked to inhibition of other enzymes that utilize nucleotides in chemical reactions (Van Rompay *et al*., 2003). The cytotoxic and antiviral effects of nucleoside analogues can be altered by modifying the sugar and/or base moieties (De Clercq *et al*., 1987; Gupta *et al*., 1989). However, while chemical modifications to the nucleoside can reduce cytotoxicity, they can also reduce the compound's inhibitory potency (De Clercq *et al*., 1987), meaning that finding the optimal chemical composition for a nucleoside inhibitor is not a straightforward task.

2.6.2 Nucleoside analogue delivery into cells

A major challenge in the development of nucleoside inhibitors is transporting the compounds into the target cells (bacterial cells, or eukaryotic cells infected by a virus). In bacteria the outer cell membranes are coated with negatively charged molecules (lipopolysaccharides in gram-negative bacteria and lipoteichoic acids in gram-positive bacteria), which repel the negatively charged triphosphate moieties of NTPs preventing their diffusion through the cell membrane (reviewed in Auer and Weibel, 2017). Only a few bacteria have been reported to express NTP transporters at cell membranes. The intracellular pathogens *Rickettsia prowazekii* and *Chlamydia trachomatis*, as well as *Protochlamydia amoebophila* (endosymbiont of Acantamoeba) utilize NTP transporters to import NTPs from their surroundings due to the lack of self-sustaining nucleotide synthesis pathways (Winkler and Neuhaus, 1999; Audia and Winkler, 2006; Tjaden *et al*., 1999; Haferkamp *et al*., 2006). A negatively charged outer membrane and lack of NTP transporters prevent the delivery of active (triphosphorylated) nucleoside analogues into bacterial cells and eukaryotic cells.

A triphosphate moiety is needed for nucleoside analogue incorporation into RNA by RNAPs, but at the same time it limits transport into the target cells. To circumvent this problem a compound can be synthesized as a prodrug that can pass through the cell membrane efficiently and then be metabolized into an active form in the cytoplasm (Figure 9). The ProTide Technology refers to methods where the hydroxyl of the monophosphate group masked with an amino group and aromatic group is added to the 5' position of nucleotides, which are removed enzymatically when the nucleoside analogue is transferred into the cell (reviewed in Mehellou *et al*., 2018).

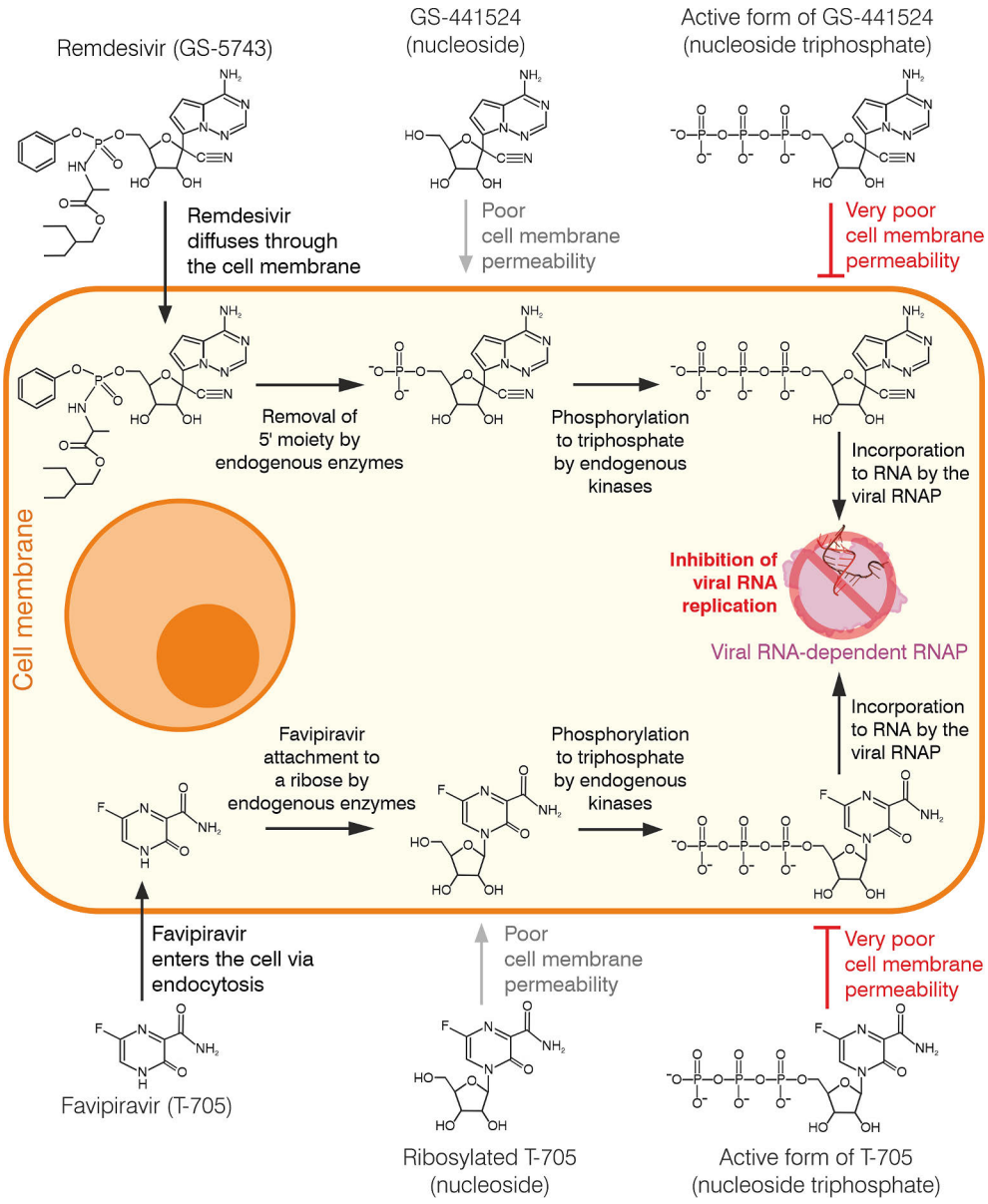


Figure 9. Strategies of nucleoside analogue delivery into a cell infected by a virus. Triphosphorylated (active) forms of nucleosides cannot efficiently pass through the cell membrane, so they are administered as prodrugs. When remdesivir (prodrug of GS-441524) enters the targeted cell (mammalian cell infected by virus), the endogenous enzymes inside the target cell remove the aryloxy phosphoramidate moiety producing the monophosphate of the GS-441524. Favipiravir is administered as a nucleobase that is attached to a ribose forming a nucleoside analogue after it enters the cell. Inside the cell the nucleosides are phosphorylated by the endogenous kinases producing the biologically active nucleoside triphosphates, which are utilized by the viral RNA-dependent RNAP inhibiting viral RNA replication.

Remdesivir (a prodrug of GS-441524) is synthesized with an aryloxy phosphoramidate moiety at the 5' position that improves nucleoside transport into cells. After entry the 5' moiety is enzymatically cleaved from the nucleotide by enzymes in the target cell, after which the nucleotide is phosphorylated by kinases in the cell (Eastman *et al*., 2020). Favipiravir is administrated as a nucleobase that is taken in by the cell. Inside the target cell the nucleobase is added to a ribose and the resulting ribonucleotide is phosphorylated into ribonucleotide triphosphate by endogenous enzymes (Smee *et al*. 2009). Nucleoside analogues need to undergo multiple phosphorylations in the cell to become an active nucleoside triphosphate. Phosphorylation is limited by the selectivity of kinases, which affects the activity of nucleoside analogues in host cells (Balzarini *et al*., 1998; Van Rompay *et al*., 2003). Prodrugs that maintain a triphosphate moiety can be used to bypass the phosphorylation steps. For instance, addition of lipophilic acyl moieties to γ-phosphate of the thymine analogue 3'-deoxy-2',3'-didehydrothymidine improved cell membrane permeability allowing efficient delivery of the nucleoside analogue into human CD4+ T-lymphocyte cells, without the need to remove the triphosphate moiety (Gollnest *et al*., 2015).

Design of any drug involves careful balancing of potency and selectivity. To achieve both of these requirements for nucleoside analogues, the mechanism of substrate selectivity and proofreading need to be understood in great detail for the targets of the drug (bacterial and viral RNAPs), as well as for host off-targets (eukaryotic and mitochondrial RNAPs). Therefore, studying the effects of modified nucleoside compounds on different RNAPs *in vitro* provides valuable information on their specificity and inhibitory mechanisms. Various cell culture models are also needed to evaluate the compounds' inhibitory effects and cytotoxicity *in vivo*, while computational methods can be utilized for chemical optimization of compounds. A combination of these methods creates a pipeline for researching drugs that target the RNA synthesis of pathogenic bacteria and viruses.

3 Aims of the Study

The synthesis of RNA from DNA is a vital step in gene expression in all cellular organisms. Transcription needs to be carried out accurately by RNAP so that the many types of RNA carry the information as it is encoded in the genomic DNA. The structure and many functions of multi-subunit RNAPs have been described in great detail, yet some of the key functions of RNAP that mediate transcriptional fidelity remain uncertain, or are strongly debated. For instance, it is well established that RNAP incorporates 2'dNTPs into RNA at a very low efficiency compared to NTPs, thus favoring the synthesis of RNA over DNA. However, it is unclear how RNAP achieves this, since 2'dNTPs are substructures of NTPs and cannot be excluded through steric blocks. The second step of transcriptional fidelity, proofreading, has a major effect on the number of transcriptional errors in the final RNA transcripts, but the way the multi-subunit RNAP regulates RNA cleavage activity remains largely unsolved. While the selectivity of nucleobase moieties is well understood, there are multiple noncognate nucleobases that haven't been extensively studied in the different classes of RNAPs. We developed methods to elucidate the mechanisms of substrate selection and transcriptional proofreading to study transcriptional fidelity. Our methods can be used to study the function of possible nucleoside inhibitors in RNAPs from different organisms and viral RNAPs.

Specific aims of the study:

1) Resolving the mechanism of nucleo-sugar selection by multi-subunit RNAPs.
2) Investigating the mechanism of transcriptional proofreading in multi-subunit RNAPs.
3) Studying the selectivity and transcriptional effects of nucleobase analogues in multi-subunit RNAPs, mitochondrial RNAPs and viral RNAPs.

4 Materials and Methods

The experimental setups and data analysis have been described in detail in the original publications (I-IV). In this section the used materials and methods are briefly discussed.

4.1 Protein expression vectors

E. coli RNAP wild type and variants used in this study were expressed using plasmids that contained all five RNAP subunits controlled by the T7-promoter, except for the RNAP variants βE813A and β'N458S, for which the ω-subunit was expressed with a separate plasmid pIA839 controlled by the araB-promoter. Expressed *E. coli* RNAPs, where mutations were directed to rpoB gene, contained N-terminal His6 tag in the β-subunit. When mutations were targeted to rpoC, a C-terminal His6 or His10 tag was included in the β'-subunit. The expression plasmid for human mitochondrial RNAP (mtRNAP) contained an N-terminal His6-tag. The expression plasmid for coxsackievirus B3 (CVB3) RNAP contained an N-terminal His6-tag attached to GB1-solubility tag that was cleavable by TEV protease. Site-directed mutagenesis was carried out by inserting a linear double stranded DNA fragment bearing the mutation into the linearized plasmid (vector) using either restriction enzymes and DNA ligase (enzymes provided by Thermo Fisher Scientific, USA) or Gibson Assembly Master Mix (New England Biolabs, Ipswich, MA, USA). Sequences of the newly constructed plasmids were verified via DNA sequencing of the inserted region. Expression vectors are listed in Table 2.

Table 2. Protein expression vectors used in this study.

Expressed protein	Plasmid description	Source	Studies
Eco RNAP wild type	pVS10 (T7p-α-β-β'_His_6-T7p-ω)	Belogurov *et al.*, 2007	I, II, III, IV
Eco RNAP βE813A	pIA624 (T7p-α-His_6_β[E813A]-β')+pIA839 (araBp-ω)	Vassylyev *et al.*, 2005	II
Eco RNAP βE813Q	pEV1 (T7p-α-His_6_β[E813Q]-β'-T7p-ω)	This work	II
Eco RNAP βD814A	pKS1 (T7p-α- His_6_β[D814A]-β'-T7p-ω)	This work	II
Eco RNAP βD814N	pEV2 (T7p-α-His_6_β[D814N]-β'-T7p-ω)	This work	II
Eco RNAP β'R425K	pAM18 (T7p-α-β-β'[R425K]_TEV_His_{10}-T7p-ω)	This work	I, II
Eco RNAP β'R425L	pAM17 (T7p-α-β-β'[R425L]_TEV_His_{10}-T7p-ω)	This work	I, II
Eco RNAP β'Y457F	pKS3 (T7p-α-β-β'[Y457F]_TEV_His_{10}-T7p-ω)	This work	II
Eco RNAP β'N458S	pIA528 (T7p-α-β-β'[N458S]_His_6)+pIA839 (araBp-ω)	Svetlov *et al.*, 2004	I, II
Eco RNAP β'A459T	pMV13 (T7p-α-β-β'[A459T]_TEV_His_{10}-ω)	This work	II
Eco RNAP β'Q504R	pJM14 (T7p-α-β-β'[Q504R]_TEV_His_{10}-ω)	Turtola *et al.*, 2018	II
Eco RNAP β'K598W	pJM16 (T7p-α-β-β'[K598W]_TEV_His_{10}-ω)	Turtola *et al.*, 2018	II
Eco RNAP β'P750L	pTG12 (T7p-α-β-β'[P750L]_His_{10}-T7p-ω)	Malinen *et al.*, 2014	II
Eco RNAP β'F773V	pVS48 (T7p-α-β-β'[F773V]_His_6-T7p-ω)	Svetlov *et al.*, 2007	II
Eco RNAP β'Q929A	pAM13 (T7p-α-β-β'[Q929A]_TEV_His_{10}-T7p-ω)	Turtola *et al.*, 2018	II
Eco RNAP β'Q929M	pJM17 (T7p-α-β-β'[Q929M]_TEV_His_{10}-ω)	This work	I, II
Eco RNAP β'Q929R	pMV18 (T7p-α-β-β'[Q929R]_TEV_His_{10}-ω)	This work	II
Eco RNAP β'M932A	pAM12 (T7p-α-β-β'[M932A]_TEV_His_{10}-T7p-ω)	Malinen *et al.*, 2012	I, II
Eco RNAP β'M932L	pAM14 (T7p-α-β-β'[M932L]_TEV_His_{10}-T7p-ω)	Turtola *et al.*, 2018	II
Eco RNAP β'R933A	pIA846 (T7p-α-β-β'[R933A]_His_6-T7p-ω)	Artsimovitch *et al.*, 2011	II
Eco RNAP β'R933N	pRP10 (T7p-α-β-β'[R993N]_TEV_His_{10}-T7p-ω)	This work	II
Eco RNAP β'H936A	pGB130 (T7p-α-β-β'[H936A]_TEV_His_{10}-T7p-ω)	Malinen *et al.*, 2012	II
Eco RNAP β'H936Q	pJM2 (T7p-α-β-β'[H936Q]_TEV_His_{10}-ω)	Turtola *et al.*, 2018	II
Eco RNAP β'I937T	pJM18 (T7p-α-β-β'[I937T]_TEV_His_{10}-T7p-ω)	This work	II
Eco RNAP β'T1135V	pJM9 (T7p-α-β-β'[T1135V]_TEV_His_{10}-ω)	This work	II
Eco RNAP β'G1136S	pJM1 (T7p-α-β-β'[G1136S]_TEV_His_{10}-ω)	Turtola *et al.*, 2018	II
Eco RNAP β'ΔSI3	pVS10-ΔSI3(b) (T7p-α-β-β'[Δ945-1132]_His_6-T7p-ω)	Esyunina *et al.*, 2016	II
Eco GreA	pIA578 (T7p-GreA-His_6)	Furman *et al.*, 2013	IV
Eco σ^{70}	pET28-His_6-σ^{70} (T7p-σ^{70}_His_6)	Marr & Roberts, 1997	IV
Hsa MT RNAP wild type	pRP9 (T7p-His_6_TEV_RpoM)	This work	III, IV
CVB3 RNAP	pGB161 (T7p-His_6_GB1 solubility tag_TEV_3DPol)	This work	IV
T7 RNAP P266L	T7p-His_6_T7 RNAP[P266L]	Guillerez *et al.*, 2005	IV
Yeast inorganic PPase	pYWT	Heikinheimo *et al.*, 1996	IV

Eco = *E. coli*, *Hsa* = *H. sapiens*, CVB3 = coxsackievirus B3

4.2 Protein production and purification

Expression plasmids for *E. coli* RNAP and variants were transformed into *E. coli* Xjb(DE3) strain (Zymo Research, Irvine, CA, USA) or T7 Express lysY/Iq strain (New England Biolabs, Ipswich, MA, USA). Cells were grown in LB medium containing selective antibiotic (100 µg/ml carbenicillin, 30 µg/ml kanamycin or 30 µg/ml chloramphenicol), shaking at 30 °C until OD280 > 1.5. Protein expression under T7-promoter was induced with 1 mM IPTG and expression under araB-promoter was induced with 0.1 % arabinose. Cells were grown for 8-10 hours after induction at 22 °C and the cells were collected by centrifugation (7000 x g, 10 minutes, 4 °C) and stored at -80 °C. Cells were suspended in *E. coli* lysis buffer (50 mM Tris-HCl, pH 6.9, 0.5 M NaCl, 5 % glycerol, 1 mM β-mercaptoethanol) supplemented with 0.1 % Tween-20, 0.1 mM Na_2-EDTA, 1 protease inhibitor tablet (Roche Diagnostics, Mannheim, Germany) and 0.2 mg/ml lysozyme. After 1 hour incubation on ice the cells were lysed by sonication. Lysate was clarified by centrifugation at 57 000 x g for 45 minutes at 4 °C. RNAPs bearing His6 or His10 tag were captured from the cleared lysate with Ni-sepharose matrix (GE Healthcare) and eluted with lysis buffer supplemented with 300 mM imidazole. *E. coli* RNAPs that eluted from Ni-sepharose were initially purified with affinity chromatography using heparin resin (5 ml HiTrap heparin column, GE Healthcare), after which the *E. coli* RNAPs were purified further with anion exchange chromatography using Q-sepharose resin (6 ml Resource Q column, GE Healthcare). During these steps the proteins were eluted with increasing NaCl gradient by mixing buffer A (50 mM Tris-HCl, pH 6.9, 5 % glycerol, 1 mM β-mercaptoethanol, 0.1 mM Na_2-EDTA) and buffer B (buffer A supplemented with 1.5 M NaCl). Fractions collected after anion exchange step were pooled, concentrated and dialyzed against *E. coli* RNAP Storage buffer (20 mM Tris-HCl, pH 7.9, 50 % glycerol, 0.1 mM Na_2-EDTA, 150 mM NaCl, 0.1 mM DTT). Purity of the *E. coli* RNAP batches were determined by SDS-PAGE and stored at -20 °C.

E. coli Gre factors (GreA and GreB) were purified similarly to *E. coli* RNAPs, except that all the buffers contained 1 M NaCl to prevent protein precipitation and the Ni-sepharose eluates were purified using size exclusion chromatography (HiPrep™ 16/60 Sephacryl™ S-200 HR column, Ge Healthcare). Eluted Gre factors were concentrated and dialyzed against Gre Storage buffer (20 mM Tris-HCl, pH 7.9, 50 % glycerol, 0.1 mM Na_2-EDTA, 1 M NaCl, 0.1 mM DTT). Purity of the Gre factor batches were determined by SDS-PAGE and stored at -20 °C.

H. sapiens MT RNAP was purified similarly to *E. coli* RNAPs, except that cation exchange with S-sepharose resin (6 ml ResourceS column, GE Healthcare) was used instead of anion exchange due to positive charge of the protein at pH 6.9. Fractions collected after cation exchange step were pooled, concentrated and dialyzed against

Storage buffer. Purity of the *H. sapiens* MT RNAP batches were determined by SDS-PAGE and stored at -80 °C in aliquots.

S. cerevisiae RNA polymerase II (*Sce* RNAPII) was expressed using genetically modified *S. cerevisiae* cell line SHy808 (provided by Mikhail Kashlevs' research group, NIH, National Cancer Institute, Frederick, MD, USA) in which a N-terminal His6 tag has been added to the Rpb3 subunit encoded in the chromosome IX. Yeast cells were grown in modified YPD medium (10 g/l yeast extract, 20 g/l tryptone, 3% (w/v) glucose, 0.002 % adenine), shaking at 30 °C until OD280 > 5.0. Cells were collected by centrifugation, suspended in 3 x yeast lysis buffer (150 mM Tris-HCl (pH 7.9), 1.5 mM Na_2-EDTA, 30 µM $ZnSO_4$, 30 % (v:v) glycerol, 3 % (v:v) DMSO, 1.5 mM TCEP) and stored at -80 °C. *Sce* RNAPII was purified from SHy808 cells by lysing the suspended yeast cells by bead beating in yeast lysis buffer (50 mM Tris-HCl (pH 7.9), 0.5 mM Na_2-EDTA, 10 µM $ZnSO_4$, 10 % (v:v) glycerol, 1 % (v:v) DMSO, 0.5 mM TCEP) and clearing the lysate by centrifugation at 16 000 x g for 45 minutes at 4 °C. *Sce* RNAPII bearing His8 tag in Rpb3 subunit was captured by binding it to Ni-sepharose matrix (Cytiva) to separate *Sce* RNAPII bearing His8 tag from RNAPI and RNAPIII. *Sce* RNAPII was eluted from Ni-sepharose matrix with 300 mM imidazole and purified further with using heparin resin (GE Healthcare) and eluted with steep gradient. *Sce* RNAPII that eluted from heparin was purified further with anion exchange chromatography using Q-sepharose resin (6 ml Resource Q column, GE Healthcare). During heparin and anion exchange steps the proteins were eluted with increasing K-acetate gradient by mixing RNAPII buffer A (50 mM Tris-HCl, pH 7.9, 5 % glycerol, 1 mM β-mercaptoethanol, 10 µM $ZnSO_4$) and RNAPII buffer B (RNAPII buffer A supplemented with 1.5 M K-acetate). Fractions collected after anion exchange step were pooled, concentrated and dialyzed against RNAPII Storage buffer (20 mM Tris-HCl, pH 7.9, 50 % glycerol, 10 µM $ZnSO_4$, 0.1 mM Na_2-EDTA, 200 mM K-acetate, 0.5 mM TCEP). Purity of *Sce* RNAPII was determined by SDS-PAGE and the protein was stored to -20 °C.

CVB3 RNAP was expressed in T7 Express lysY/Iq. The cells were grown in LB medium supplemented with 30 µg/ml kanamycin at 30 °C until OD 1.0, after which the culture was transferred to 22 °C and protein expression was induced for 10 h by the addition of 1.0 mM IPTG. Cells were harvested by centrifugation at 7000 × g, 4 °C for 10 min and stored at -80 °C. Dry pellet was resuspended in lysis buffer (50 mM Tris-HCl pH 7.9, 500 mM NaCl, 5% glycerol) supplemented with 1 mM β-mercaptoethanol, a tablet of EDTA-free protease inhibitors (Roche Applied Science, Mannheim, Germany) and 1 mg/ml lysozyme. Cells were incubated on ice for 45 min and disrupted by sonication. The lysate was cleared by centrifugation at 57 000 × g for 50 min at 4 °C. The supernatant was supplemented with 10 mM imidazole and loaded onto a Ni-sepharose (GE Healthcare, Chicago, IL, USA) column pre-

equilibrated with lysis buffer. Protein was eluted using a step gradient (30, 60, 300 mM) of imidazole in lysis buffer. The 300 mM imidazole fraction containing RNAP was treated with TEV protease for 8 h at 6 °C to remove the N-terminal His6 tag and solubility tag. TEV treated protein was purified using 6 ml Resource-Q column (GE Healthcare, Chicago, IL, USA) in virus buffer A (50 mM Tris-HCl pH 7.9, 5% glycerol, 1 mM β-mercaptoethanol, 0.1 mM Na_2-EDTA) and virus buffer B (virus buffer A supplemented with 1.5 M NaCl). CVB3 eluted from Resource-Q column at > 20% B. CVB3 was purified further by gel filtration using HiPrep™ 16/60 Sephacryl™ S-200 HR column running at constant ratio of 33% B (~500 mM NaCl). CVB3 eluted from gel filtration column at ~55-60 ml after loading the sample to the column. The fractions containing the purified protein were concentrated using Amicon Ultra-4 centrifugal filter (Merck Milipore, Burlington, MA, USA), diluted by adding two parts of storage buffer (75% glycerol, 30 mM Tris-HCl pH 7.9, 225 mM NaCl, 0.15 mM Na_2-EDTA, 0.75 mM DTT) to one part of protein sample and stored at -80 °C.

T7 RNAP variant P266L (provided by Thomas Steitz, Yale University, New Haven, Connecticut, United States) was expressed in T7 Express lysY/Iq. The cells were grown in LB medium supplemented with 100 µg/ml carbenicillin at 30 °C until OD 1.0, after which the culture was transferred to 22 °C and protein expression was induced for 10 h by the addition of 1.0 mM IPTG. Cells were harvested by centrifugation at 7000 × g, 4 °C for 10 min and stored at -80 °C. Dry pellet was resuspended in lysis buffer (50 mM Tris-HCl pH 7.9, 500 mM NaCl, 5% glycerol) supplemented with 1 mM β-mercaptoethanol, a tablet of EDTA-free protease inhibitors (Roche Applied Science, Mannheim, Germany) and 1 mg/ml lysozyme. Cells were incubated on ice for 45 min and disrupted by sonication. The lysate was cleared by centrifugation at 57 000 × g for 50 min at 4 °C. The supernatant was supplemented with 10 mM imidazole and loaded onto a Ni-sepharose column pre-equilibrated with lysis buffer. Protein was eluted using a step gradient (30, 60, 300 mM) of imidazole in lysis buffer. T7 RNAP eluted from Ni-sepharose was purified with 5 ml Heparin column (GE Healthcare, Chicago, IL, USA) using buffer A and buffer B. T7 RNAP eluted from Heparin at > 25% B. T7 RNAP eluted from Heparin was purified using 6 ml Resource-Q column in buffer A and buffer B. T7 RNAP eluted from Resource-Q column at > 12% B. The fractions containing the purified protein were concentrated using Amicon Ultra-4 centrifugal filter (Merck Milipore, Burlington, MA, USA), diluted by adding two parts of storage buffer to one part of protein sample and stored at -80 °C.

4.3 DNA and RNA oligonucleotides

PAGE and HPLC grade DNA and RNA oligonucleotides were synthesized by IBA GmbH (Gottigen, Germany), Fidelity systems (Gaithersburg, MD, USA) and Eurofins MWG Operon (Edensburg, Germany). The oligonucleotides were designed so that the RNA and the corresponding tDNA had 9 to 11 matching bases between them allowing the primer RNA to form a stable hybrid with the tDNA. The ntDNAs were designed to be fully complementary with the tDNAs in the used experimental setups.

Long RNA oligos for processive RNA elongation tests with CVB3 RNAP were synthesized using T7 RNAP variant P266L in reaction with 0.2 µM dsDNA template with T7 promoter, 2.3 mM of each NTP, 2.3 mM triphosphorylated formycin A in place of ATP (where indicated), 10 mM $MgCl_2$, 5 mM DTT, 0.2 µM yeast inorganic PPase and 1.0 µM T7 RNAP in TB buffer (40 mM HEPES-KOH pH 7.5, 10 mM $MgCl_2$, 80 mM KCl, 5% glycerol, 0.1 mM Na_2-EDTA, and 0.1 mM DTT). Reaction was incubated for 3 hours at 37 °C after which the reaction was supplemented with 0.02 units/µl of RNase free DNase I (Invitrogen, Thermo Fisher Scientific) and incubated for additional 20 minutes at 25 °C. Reaction was loaded to 1 ml CaptoQ HR anion exchange column (Cytiva, Marlborough, Maine, U.S.) and purified using buffer NA (50 mM Tris-HCl pH 7.9, 5% glycerol, 1 mM β-mercaptoethanol, 1.0 mM Na_2-EDTA) and buffer NB (buffer NA supplemented with 1.5 mM NaCl). RNA products eluted > 35% of B. Purity of RNA containing fractions was checked on agarose gel and fractions containing RNA were mixed with 0.6-fold volume of 100% isopropanol and incubated for 20 minutes at 22 °C. Precipitated RNA was spined down at 21 000 x g, 4 °C for 10 minutes. RNA pellets were washed with 1 ml of 70 % (v:v) ethanol and centrifuged at 15 000 x g, 22 °C for 5 minutes. Pellets were dried and suspended in scaffold buffer (10 mM HEPES-KOH pH 7.5, 0.1 mM Na_2-EDTA).

4.4 Substrates

Commercially available NTPs, 2'dATP and 3'dGTP were purchased from Jena Bioscience (Jena, Germany) and Thermo Fisher Scientific (Waltham, Massachusetts, U.S.). 2'dGTP, 2'dUTP, and 2'dCTP were from Bioline Reagents (London, UK). Showdomycin derivatives were produced, modified and triphosphorylated by Petja Rosenqvist at University of Turku (Department of Chemistry). Petja also triphosphorylated formycin A and pyrazofurin A nucleosides, which were purchased from Merck (Darmstadt, Germany). 8-oxo-ATP, 8-oxo-GTP and ribavirin triphosphate were purchased from Jena Bioscience.

4.5 TEC assembly

Assembly of TEC was carried out *in vitro* using a procedure described in Komissarova *et al.*, 2002. RNA primer (10 µM) was annealed with tDNA oligonucleotide (14 µM) at 70 °C for 5 minutes. 1.0 µM of the annealed RNA-tDNA hybrid was incubated with RNAP (1.5 µM) in TB buffer supplemented with 0-10 mM $MgCl_2$ depending on the experimental setup, for 10 minutes at 25 °C. After initial incubation the ntDNA (2.0 µM) was added and the mixture was incubated for additional 20 minutes at 25 °C to create the artificial transcription bubble. When *Sce* RNAPII was used the second incubation step was shortened to 10 minutes. After TEC assembly the TECs were diluted with TB buffer to the concentration needed for the experiment (0.1–0.5 µM). In case of TECs with CVB3 RNAP, a duplex of template RNA annealed with RNA primer (1.0 µM) was incubated with CVB3 RNAP for 20 minutes at 25 °C in VTB buffer (40 mM HEPES–KOH pH 7.5, 5% glycerol, 0.1 mM Na_2-EDTA, and 5 mM DTT).

When the RNA needed to be extended by nucleotides before the experiment, the diluted TECs were mixed with a subset of nucleotides and incubated for 2 minutes at 25 °C in TB buffer supplemented with 1 mM $MgCl_2$. When the pre-extended RNA would be rapidly cleaved by the RNAP during pre-extension, additional NTPs were added to allow regeneration of the primer RNA. After incubation the TECs were washed from excess substrates and $MgCl_2$ with ZebaTM Spin (40K molecular-weight cut-off) desalting columns (ThermoFisher Scientific, USA) equilibrated with TB buffer.

4.6 Single nucleotide addition assays

Transcription reactions were initiated by the addition of 10 µl of NTPs and/or triphosphorylated nucleoside analogues in TB buffer supplemented with 10 mM $MgCl_2$ to 10 µl of the assembled TEC in TB buffer also supplemented with 10 mM $MgCl_2$. The final concentrations of the TEC, NTPs and triphosphorylated nucleoside analogues were 0.1 mM, 10–20 µM and 100 µM, respectively. The reactions were incubated for 1 min at 25 °C and quenched with 30 µl of gel loading buffer (94% formamide, 20 mM Li_2-EDTA and 0.2% Orange G). RNAs were analyzed as described in section 4.10.

4.7 Time-resolved single nucleotide addition experiments with rapid kinetic instruments

Time-resolved measurements of nucleotide addition were performed in an RQF 3 quench-flow instrument (KinTek Corporation, Austin, TX). 14 µl of TEC (0.2 µM) was rapidly mixed with 14 µl substrate and then quenched with 86 µl of either 0.5

M HCl or 0.45 M Na_2-EDTA in TB buffer. When quenched with 0.5 M HCl, the reaction samples were mixed with 171 µl of neutralizing sample buffer (94 % formamide, 290 mM Tris base, 13 mM Li_2-EDTA, 0.2 % Orange G) and stored at -20 °C. Samples quenched with 0.45 M EDTA were mixed with non-neutralizing sample buffer (97 % formamide, 0.2 % Orange G) and stored at -20 °C. Extended RNAs were analyzed as described in section 4.10.

Time-resolved fluorescence-based measurements of translocation after nucleotide addition were observed using SX.18MV stopped-flow instrument (Applied Photophysics, Leatherhead, UK). The 6-MI fluorophores in the tDNA were excited at 340 nm and the emitted light was collected through a 400-nm long pass filter. Nucleotide addition reactions were initiated by mixing 64 µl TEC (0.2 µM) with 64 µl of substrate mix and measuring the fluorescence signal in real time. A minimum of three traces were averaged in each individual experiment.

4.8 Processive transcript elongation experiments

Processive transcript elongation was carried out by mixing 10 µl of TEC with 10 µl of substrate mix containing four (or more) substrates that allow transcription of all sequence positions of the tDNA downstream from the RNA 3' end. When comparing 2'dNTPs to NTPs, one of the NTPs was replaced with 2'dNTP bearing the corresponding base moiety. When studying competition between UTP and triphosphorylated showdomycin derivatives, all four NTPs (100 µM each) were added to the reaction together with 2 mM showdomycin derivatives. Reactions were incubated for 5 minutes in 25 °C and quenched with gel loading buffer (94% formamide, 20 mM Li_2-EDTA, and 0.2% Orange G). Extended RNAs were analyzed as described in section 4.10.

4.9 Intrinsic and Gre assisted RNA cleavage

For the RNA cleavage experiments the TEC was assembled in TB buffer without $MgCl_2$ to inhibit RNA cleavage before the reaction, unless the TECs were pre-extended before the cleavage reaction, in which case the TECs were assembled in TB buffer supplemented with 1 mM $MgCl_2$, pre-extended and washed from $MgCl_2$ as described in section 4.5. During intrinsic cleavage reactions TECs (0.1 µM) were mixed with 2 mM $MgCl_2$ to start the reaction and incubated for up to 4 hours at 25 °C while taking samples at specific time points. During Gre factor mediated cleavage reactions TECs (0.2 µM) were mixed with 10 mM $MgCl_2$ and 10 µM GreA or GreB to start the reaction and incubated up to 16 minutes at 25 °C while taking samples at specific time points. Samples were quenched with STOP-buffer and the shortened RNAs were analyzed as described in section 4.10.

4.10 Electrophoresis and scanning of $Atto_{680}$-labeled RNA

Reaction samples containing Atto680-labeled RNA in neutralizing sample buffer, non-neutralizing sample buffer or STOP buffer were heated at 95 °C for 2 minutes and loaded onto 16 % denaturating urea-polyacrylamide gel. Different length RNAs in the samples were separated by electrophoresis in TBE buffer (100 mM Tris, pH 8.4, 90 mM boric acid, 1 mM EDTA). Gels were scanned with Odyssey Infrared Imager (Li-Cor Biosciences, Lincoln, NE, USA) using the 700 nm channel to detect the Atto680-label. The RNA band intensities were quantified from the 16-bit TIF files using ImageJ software (Abràmoff *et al.*, 2004).

4.11 Monitoring RNA elongation over long template using fluorescent light-up aptamer

Linear dsDNA template used in the experiments was prepared by digesting plasmid pOP004 with restriction enzyme XhoI in Red Buffer (Thermo Fisher Scientific) for 6 hours. Digested plasmid was purified with 1 ml CaptoQ HR anion exchange column (Cytiva) in buffer NA and buffer NB. Linear plasmid DNA eluted at > 55 % B. Fractions containing plasmid DNA were mixed with 0.6-fold volume of 100% isopropanol and incubated for 20 minutes at 22 °C. Precipitated DNA was spined down at 21 000 x g, 4 °C for 10 minutes. DNA pellets were washed with 1 ml of 70 % (v:v) ethanol and centrifuged at 15 000 x g, 22 °C for 5 minutes. Pellets were dried and suspended in water.

The transcription reactions were initiated by mixing 30 µl of holoenzyme mix (2 µM *Eco* RNAP, 8 µM *Eco* σ^{70}, 0.1 µM linearized plasmid, and 0.1 µM *S. cerevisiae* inorganic PPase in TB buffer) with 30 µl of NTP mix (2 mM NTPs, 0–10 mM ForTP or EthS-SdmTP, 10–20 mM $MgCl_2$, and 20 µM DFHBI-1T in TB buffer), both pre-heated to 37 °C prior to mixing. Mixed reactants were transferred to a cuvette (pre-heated to 37 °C) and after 30 seconds from starting the reaction the fluorescence was measured at 507 nm for 600 seconds using LS-55 Spectrofluorometer (Perkin Elmer) at 37 °C. After 700 seconds from mixing the reactants the reaction was quenched in 120 µl gel loading buffer. RNAs were separated on 6 % TBE-Urea Gel (Thermo Fisher Scientific), which were stained post-run with SYBR® Gold (Thermo Fisher Scientific). Stained RNA was visualized with Azure Sapphire™ Biomolecular Imager (Azure Biosystems Inc., Dublin, CA, U.S.), band intensities were quantified using the ImageJ software.

4.12 Data analysis

The kinetic data were analyzed with different methods depending on the case. KinTek Explorer software (KinTek Corporation, Austin, TX) was used to simultaneously fit the translocation time traces (stopped flow) and the time-resolved GMP incorporation data (HCl and EDTA quenched reactions) to a three-step model (model 1) using the numerical integration capabilities of the software (Johnson, 2009) as described in Prajapati *et al*., 2019. In model 1, initial TEC16 reversibly binds the GTP substrate, undergoes the irreversible transition to TEC17 upon incorporation of the nucleotide into RNA, and then translocates. The EDTA quenched reactions were modeled using the pulse-chase routine of the Kin-Tek Explorer software. Time-resolved 2′dGMP incorporation concentration series (translocation time traces) were globally fitted to a stretched exponential function (equation 1) using Origin 2015 software (OriginLab, Northampton, MA, USA). In equation 1, the exponent followed a hyperbolic dependence on the 2′dGTP concentration. Rate constant k, Michaelis constant K_m and the stretching parameter β were shared by all curves in the dataset. Stretched exponential function was used for fitting the rates of intrinsic RNA cleavage of the nascent RNA as the slow reactions were poorly described by single and double exponential functions. The median reaction times were calculated as in Turtola & Belogurov, 2016. The time courses of nucleotide incorporation by TECs pre-extended with nucleoside analogues were fitted to a sum of exponential (models the fast phase) and stretched exponential (models the slow phase) functions. GreA cleavage of C-nucleoside analogues from the nascent RNA were fitted with stretched exponential function.

5 Results and Discussion

5.1 Discrimination of 2'dNTPs by multi-subunit RNAPs (Study I)

The fundamental property of RNAPs is the synthesis of RNA, which requires the selection of NTP over 2'dNTPs by probing for the nucleo-sugar moiety bearing the 2'OH group. DNAPs, which utilize 2'dNTPs in the synthesis of DNA, sterically block the binding of NTPs with residues that occupy the space that 2'OH would fill in the active site (reviewed in Brown and Suo 2011). Single-subunit RNAPs from the right-hand RNAP superfamily differentiate between NTPs and 2'dNTPs with a multifunctional conserved Tyr residue that stabilizes the binding of NTPs and inhibits the binding of 2'dNTPs (Huang *et al*., 1997). In the case of multi-subunit RNAPs from the two-β-barrel superfamily, the selection of NTPs over 2'dNTPs occurs by a different mechanism, since the active site of multi-subunit RNAPs is spacious enough to accommodate both NTPs and the smaller 2'dNTPs. While the cellular concentrations of NTPs exceed 2'dNTPs by approximately 10-fold (Traut 1994), that alone does not explain the observed 500- to 5000-fold differences in incorporation rates between NTPs and 2'dNTPs by multi-subunit RNAPs (Kaplan *et al*. 2008; Zhang *et al*., 2010). Three conserved amino acid residues in bacterial multi-subunit RNAPs can interact with the 2'OH moiety: β'Arg425, β'Asn458 and β'Gln929 (Vassylyev *et al*., 2007b; Wang *et al*., 2006). Residues β'Gln929 and β'Met932 (which is replaced by Leu in eukaryotic RNAPs) are close enough to interact with 2'OH upon TL folding into TH. While TL has been shown to affect the discrimination of 2'dNTPs (Zhang *et al*., 2010), it contributes by 5- to 10-fold to selectivity, which is only a fraction of the total effect. Substitutions to the residue β'Asn458 have only minor effects on the differentiation of nucleo-sugar moieties (Svetlov *et al*., 2004; Wang *et al*., 2006). Substitutions to the residue β'Arg425 have not been experimentally studied, even though multiple X-ray crystal structures place the residue near the NTP 2'OH at the different steps of substrate binding (Vassylyev *et al*., 2007b). Computational modelling of *Sce* RNAPII suggests that the Arg residue coordinates bound 2'dATP differently compared to rATP, increasing the distance between the 3'OH and α-phosphate of 2'dATP by 1 Å, which would inhibit phosphodiester bond formation (Roßbach and Ochsenfeld 2017). Therefore, we

investigated how mutations in residues β'Arg425, β'Asn458, β'Gln929 and β'Met932 would affect the selectivity of NTPs and 2'dNTPs in multi-subunit RNAPs.

5.1.1 Comparing the binding of NTPs, 2'dNTPs and 3'dNTPs in wild type RNAP

We first determined how *Eco* RNAP wild type utilizes substrates with different nucleo-sugar moieties. We used a stopped-flow instrument to measure concentration series for GTP and 2'dGTP to estimate the catalytic constant (k_{cat}) and dissociation constant (K_D) for these substrates (Figure 10). We used a HCl quench method to determine the incorporation rates for GMP, 2'dGMP and 3'dGMP. EDTA quenching was used to estimate the dissociation of bound GTP and 2'dGTP. EDTA chelates all Mg^{2+} ions in the reaction except for the ones bound in the active site with the substrate, stopping incorporation of substrates that bind after EDTA addition, while allowing incorporation of substrates that have bound to the active site prior to the addition of EDTA (Gohara *et al*., 2004; Kireeva *et al*., 2008). The faster EDTA quenched reactions are, the higher the affinity for the substrate, as substrates are incorporated into RNA instead of dissociating. Kinetic parameters inferred from these experiments are presented in Table 3.

Table 3. Kinetic parameters for the substrate utilization by *E. coli* RNAP wild type. Parameters were fit to model 1 (see Figure 10). Reaction products were modeled as sums of independent contributions by fast and slow fractions of RNAP; contributions of each fraction were modeled as model 1. Upper and lower bounds were calculated at a 10% increase in Chi2 by the FitSpace routine of KinTek Explorer software. Table is adapted from Mäkinen *et al*., 2021 (study I).

Substrate	k_{on} $\mu M^{-1}s^{-1}$	k_{off} s^{-1}	k_{cat} (fast) s^{-1}	k_{cat} (slow) s^{-1}	Fast fraction %	K_D µM
GMP	2.5 (2.2–3.0)	7.7 (4.0–9.6)	31 (28–42)	2.6 (0.8–7.6)	88 (77–93)	3.1 (2.3–4.0)
2'dGMP	>0.24	>45	1.4 (1.1–2.3)	0.35 (0.22–0.49)	52 (30–70)	190 (160–220)
3'dGMP	2.7 (2.2–4.4)	87 (70–140)	12 (11–13)	0.74 (0.5–1.4)	80 (75–85)	32 (28–38)

From the concentration series and EDTA quenching data we estimated that wild type RNAP has a 60-fold higher affinity for GTP over 2'dGTP. At saturating concentrations GMP was incorporated 50-fold faster than 2'dGMP (Figure 11A), while GMP was incorporated only 1.5-fold faster than 3'dGMP, which lacks the

3'OH group. This indicates that the lack of 2'OH is more detrimental for the incorporation of the substrate than the lack of 3'OH.

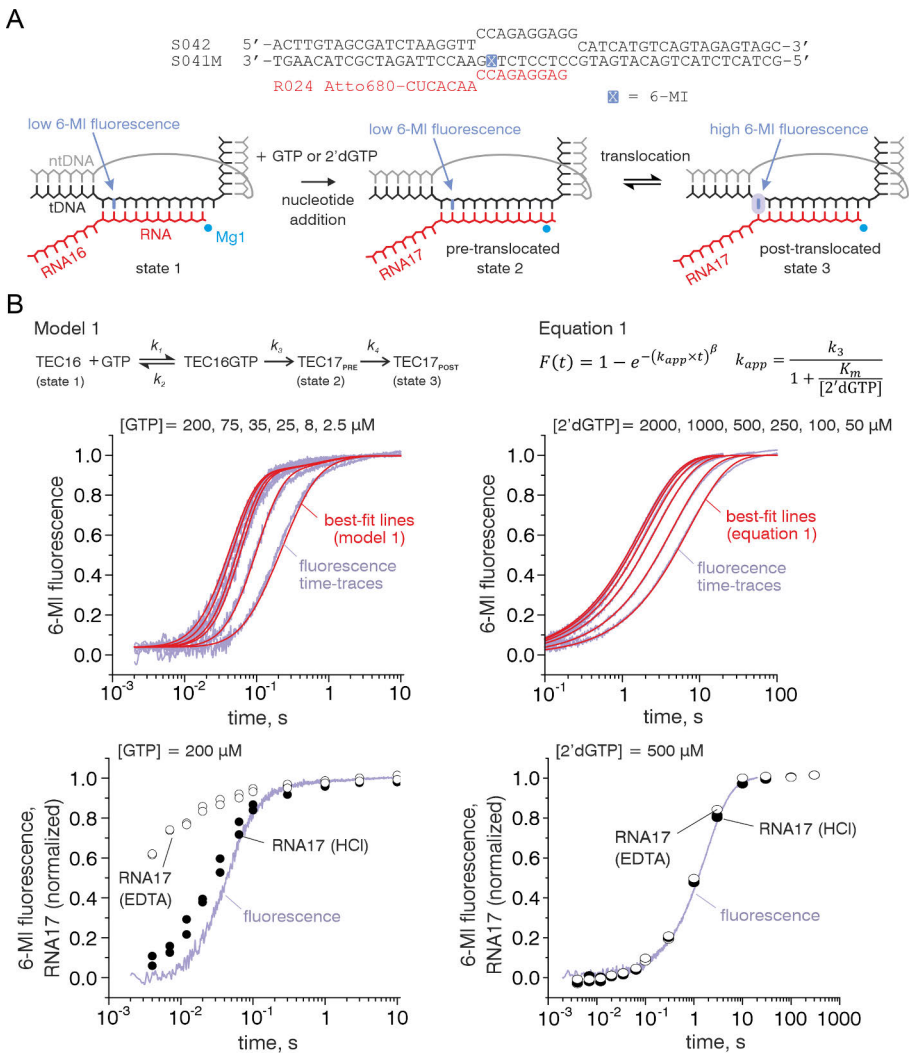


Figure 10. Time-resolved measurements of GTP and 2′dGTP utilization by the wild type *E. coli* RNAP. (A) The nucleic acid scaffold employed in the assays. The fluorescence of guanine analogue 6-MI (cyan) was quenched by neighbouring base pairs in the initial TEC (state 1) and pre-translocated TEC that formed following the nucleotide incorporation (state 2), but increased when the 6-MI relocated to the edge of the RNA-DNA hybrid upon translocation (state 3). Template DNA, non-template DNA, RNA, and the catalytic Mg^{2+} ions are colored black, gray, red, and cyan, respectively. (B) Kinetic data of GTP and 2'dGTP utilization by *E. coli* RNAP wild type. GTP concentration series was fit to model 1 and 2′dGTP concentration series was fit to equation 1. The best-fit lines and fluorescence time-traces are colored red and purple, respectively. The HCl and EDTA quenched data points are shown as closed and opened circles, respectively. Figure was adapted from Mäkinen *et al.*, 2021 (study I).

5.1.2 Substitutions to the conserved residue β'Arg425 reduce the discriminating effect on 2'dNTPs

We then investigated the effects of site-directed mutagenesis on the β'Arg425, β'Asn458, β'Gln929 and β'Met932 residues to learn how they affect the selectivity of NTPs over 2'dNTPs in *E. coli* RNAP. We studied *E. coli* RNAP variants in which the mutations altered the properties of the target residues without compromising catalytic activity: β'R425K, β'N458S, β'Q929M and β'M932A. We compared the binding and catalytic efficiency of GTP and 2'dGTP with these variants by measuring the rate of translocation at different concentrations of substrates, the same as above (Figure 11A).

When binding selectivity (difference in K_D between GTP and 2'dGTP) was considered, the variants β'R425K, β'N458S and β'Q929M had significantly reduced selectivity compared to the wild type (17-, 4- and 30-fold, respectively). β'M932A displayed a 2.5-fold increase in binding selectivity compared to the wild type. In the case of incorporation selectivity (difference in incorporation rates between GMP and 2'dGMP), β'N458S and β'Q929M had a higher incorporation selectivity compared to the wild type enzyme (1.6- and 4-fold, respectively). β'M932A had a 3-fold reduced incorporation selectivity compared to the wild type, whereas β'R425K significantly increased the incorporation rate of 2'dGMP while reducing the incorporation rate of GMP, resulting in a 25-fold reduction in incorporation selectivity compared to the wild type. β'R425K incorporated 2'dGMP at a rate of 9.7 s^{-1}, which is 10-fold faster than β'M932A and 15-fold faster than the wild type.

While the single nucleotide addition experiments provide detailed information on the binding properties and catalytic efficiency of NTPs and 2'dNTPs, these experiments are limited to a single easily transcribed sequence position, where the TEC starts at a predetermined conformation (post-translocated state). Therefore, additional proof is needed to show that the kinetic properties of the 2'dNTPs observed in the RNAP variants are not limited to the scaffold used in the single nucleotide addition experiments. We utilized a processive transcript elongation method to provide additional information on transcript elongation by 2'dNTPs in multiple different sequence positions (Figure 11B). While this method provides only semi-quantitative data on transcript elongation, the major advantage is that transcription through multiple sequence positions can be observed in a single experiment.

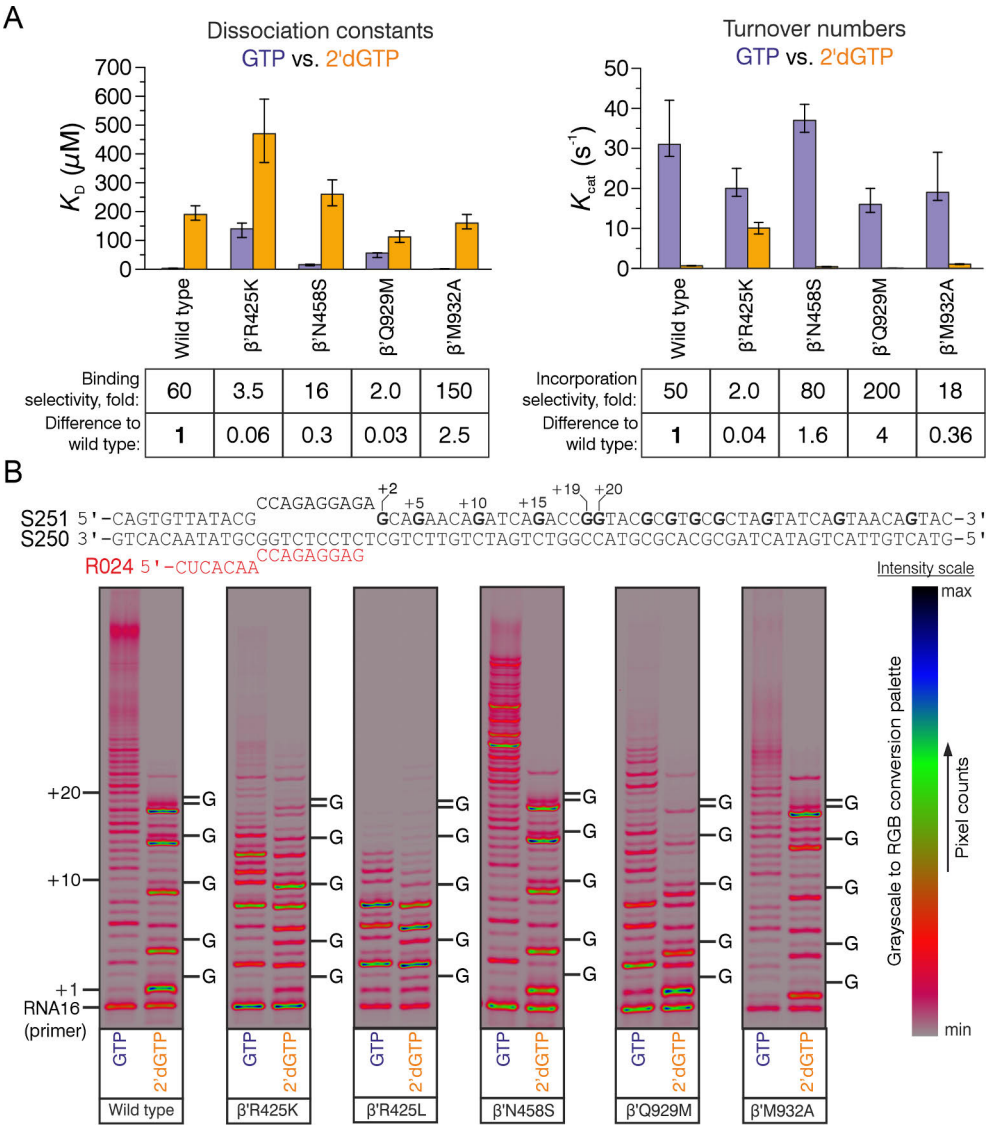


Figure 11. Nucleo-sugar selectivity of *E. coli* RNAP wild type and variants. (A) Kinetic parameters of GTP and 2'dGTP binding and incorporation by *Eco* RNAP wild type and variants. Equilibrium dissociation constants (K_D) for the reversible binding of GTP and 2′dGTP and turnover numbers (k_{cat}) for incorporated GMP and 2'dGMP are shown as bar graphs. (B) Processive transcript elongation by RNAPs with 2'dGTP in place of GTP. TECs were assembled using the scaffold shown above the gel panels and chased with 50 μM ATP, CTP, UTP, and GTP or 2′dGTP for 2 min at 25 °C. The positions of GMPs in resolved stretches of the transcribed sequence are marked along the right edge of gel panels; 16-bit grayscale scans were normalized using max pixel counts within each gel panel and pseudo-colored using inverted Rainbow RGB color palette of ImageJ. Figure was adapted from Mäkinen *et al.*, 2021 (study I).

To see how the effects of mutations on 2'dNTP binding and incorporation affect processive transcript elongation, we assembled TECs on scaffolds where the RNAP can transcribe along the DNA for 49 nucleotides and chased them with 50 μM ATP, CTP and UTP, and with 50 μM GTP or 2'dGTP for 2 minutes (Figure 11B). When GTP was substituted with 2'dGTP, the wild type enzyme arrested strongly before each guanine adding position, showing that incorporation of 2'dNTP was rate limiting for processive transcription elongation. Variants β'N458S, β'Q929M and β'M932A had similar properties as the wild type. β'N458S ran further from the primer position with GTP than the wild type, yet it had a very similar pausing pattern when GTP was substituted with 2'dGTP. β'M932A did not pause as strongly before guanine adding positions as the wild type. β'Q929M paused at early sequence positions much more than other variants in the presence of 2'dGTP, limiting RNA extension largely within 9 nucleotides from the primer. The amount of pausing observed with 2'dGTP on these variants is consistent with their catalytic efficiencies.

In contrast to the aforementioned variants, β'R425K was more defective in processive transcript elongation (as expected due to its reduced catalytic efficiency and reduced affinity for NTPs), yet it did not arrest strongly before any of the guanine incorporating positions it was able to reach within the time frame of the experiment. In fact, β'R425K paused mostly at the same positions during both GTP and 2'dGTP chases, indicating that 2'dGTP had a much smaller effect on the overall transcription rate of β'R425K than what was observed for other variants and the wild type. Additionally, we tested another β'Arg425 variant, β'R425L, to see if the observed reduction in nucleo-sugar selectivity is dependent on the target residue or the type of amino acid substitution. Despite being more defective than β'R425K in processive elongation, β'R425L also did not accumulate TECs before the guanine incorporating positions when GTP was substituted with 2'dGTP. These experiments were performed for each different NTP on different scaffolds to verify that the observed effects were not dependent on 2'dGTP. The same effects were observed with 2'dATP, 2'dUTP and 2'dCTP: β'R425K and β'R425L utilized 2'dNTPs more readily than the wild type and other variants. As such, mutations targeted at the invariant residue β'Arg425 altered the nucleo-sugar selection properties of the multi-subunit RNAP. Since Arg-to-Lys and Arg-to-Leu mutations had similar effects, the reduction in 2'dNTP discrimination is likely a result of the loss of contact between the β'Arg425 residue and binding NTP.

5.1.3 β'Arg425 discriminates against 2'dNTPs by interacting with their 3'OH moieties and altering their sugar pucker conformation

While it is clear that mutations to β'Arg425 reduce selectivity for the nucleo-sugar, the reason behind this effect is not apparent from transcription reactions alone. Substituting β'Arg425 greatly reduces affinity for both GTP and 2'dGTP indicating a loss in a key interaction with both substrates, but at the same time an increase in catalysis rate for 2'dGTP is observed (while the catalysis rate of GTP is reduced) compared to the wild type, suggesting that the contacts which β'Arg425 makes with 2'dNTPs are unfavorable for catalysis in the wild type enzyme. These findings can be explained by comparing X-ray crystal structures of NTP and 2'dNTP bound at the active site of the *T. thermophilus* RNAP (Figure 12).

The X-ray crystal structures with bound 2'dCTP (PDB: 6WOX) and 3'dCTP (PDB: 6WOY) at the active site of the *T. thermophilus* RNAP were resolved by Katsuhiko Murakami's research group (Department of Biochemistry and Molecular Biology, The Center for RNA Molecular Biology, Pennsylvania State University, University Park, PA, USA). The structure with bound 2'dCTP shows that β'Arg425 interacts with 3'OH when the 2'OH group is absent. 2'dCTP was observed in a 2'-endo conformation, which brings the 3'OH within hydrogen bonding distance with β'Arg425. In this orientation the 3'OH largely occupies the same space as the 2'OH of CMPCPP in structure 4Q4Z (Basu *et al*., 2014). The TL was observed in an unfolded state, likely due to the loss of interaction between β'Gln929 and the 3'OH group that is sequestered by β'Arg425. 2'dNTPs bind to the active site of multi-subunit RNAP in a 2'-endo conformation, which leads to β'Arg425 making the initial contact with the 3'OH group, as 2'OH is absent. This contact stabilizes the 2'-endo conformation of the 2'dNTP and sequesters the 3'OH group, which prevents contact between β'Gln929 and the 3'OH group. In contrast, 3'dCTP was in a similar conformation as observed in CMPCPP, as the substrate was in a 3'-endo conformation while β'Arg425 and β'Asn458 form hydrogen bonds with the 2'OH group. In this structure the TL was observed in an unfolded state, again likely due to loss of interaction between β'Gln929 and the absent 3'OH group. In our biochemical assays, site-directed mutagenesis in residue β'Arg425 reduces the binding affinity for both NTPs and 2'dNTPs. However, substituting β'Arg425 also increases the incorporation rate of 2'dNTPs as the missing arginine residue will not sequester 3'OH, allowing for efficient interaction between β'Gln929 and 3'OH promoting TL folding and catalysis of 2'dNTPs. These findings are further supported by the X-ray crystal structures, which show how the 3'OH of 2'dCTP is sequestered by β'Arg425. TL folding provides additional effects through the β'Gln929 and β'Met932 residues. β'Gln929 interacts with the 3'OH group promoting catalysis of substrates with a cognate nucleo-sugar moiety, while β'Met932 likely probes for substrate orientation

during TL folding. In conclusion, the conserved β'Arg425 residue at the active site is the major determinant of nucleo-sugar selectivity in multi-subunit RNAPs, where it promotes NTP binding by hydrogen bonding with the 2'OH group, while also selectively reducing the incorporation of 2'dNTPs by interacting with the 3'OH group instead, sequestering it from the folding TL.

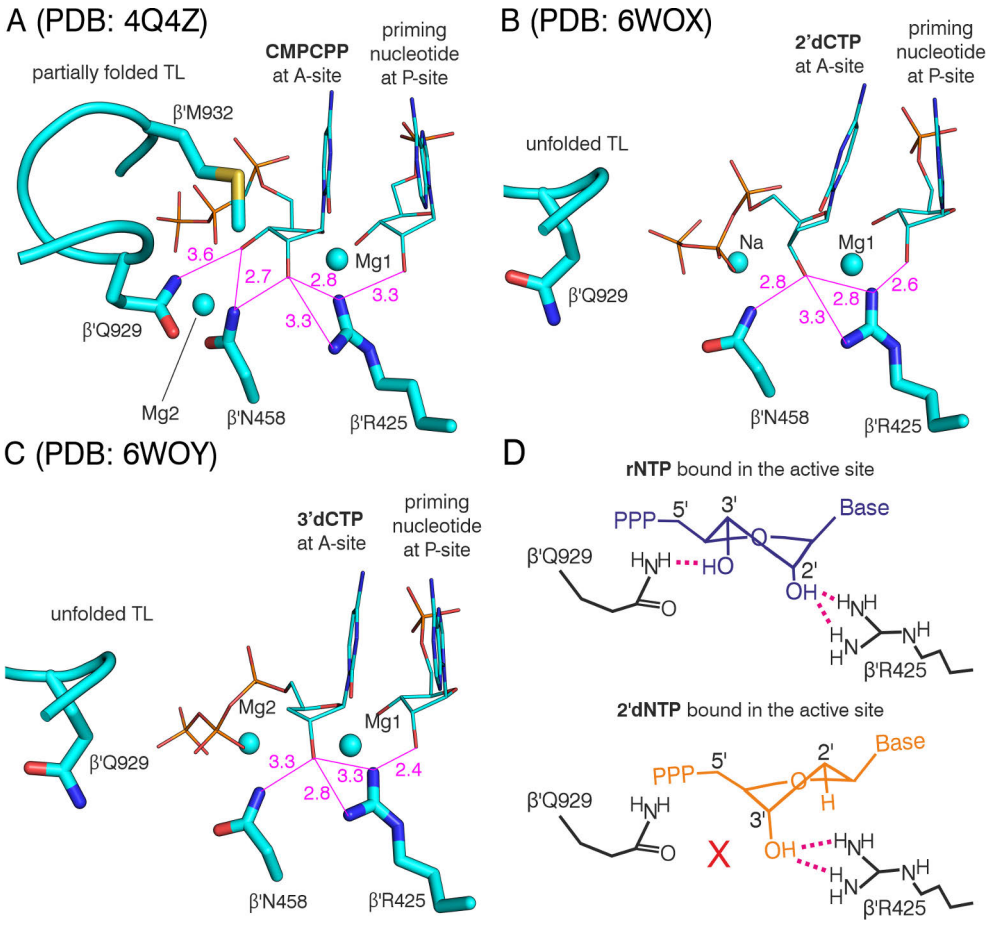


Figure 12. X-ray crystal structures of *T. thermophilus* RNAP with CMPCPP, 2'dCTP and 3'dCTP substrates. Figure shows binding poses of CMPCPP (Panel A, PDB:4Q4Z), 2′dCTP (Panel B, PDB:6WOX), and 3′dCTP (Panel C, PDB:6WOY) in the active site of *T. thermophilus* RNAP. Numbering of amino acid residues follows *E. coli* nomenclature. Magenta numbers are interatomic distances in Å. Pre-catalytic complexes in panels B and C were trapped due to low reactivity of deoxyribonucleoside substrates and the slow catalysis by RNAP in crystallo. Panel D illustrates how β'Arg425 binds rNTPs and 2'dNTPs differently. β'Arg425 interacts with 2'OH of rNTPs allowing interaction between 3'OH and β'Gln929. In the absence of 2'OH (2'dNTPs) β'Arg425 interacts with the 3'OH instead, which prevents its interaction with β'Gln929 limiting TL folding and catalysis. Figure was adapted from Mäkinen *et al.*, 2021 (study I).

5.2 Mechanism of intrinsic RNA cleavage by multi-subunit RNAPs (Study II)

Multi-subunit RNAPs exhibit endonuclease activity allowing them to remove nucleotides from the backtracked RNA 3' end, which it utilizes during the pause release of stalled TECs and transcriptional proofreading (i.e. removal of mismatched NMPs). The proofreading mechanism has remained elusive, as the reaction occurs slowly (in the absence of cleavage factors) in an off-pathway state. While seemingly most multi-subunit RNAPs have this activity, the efficiency of RNA cleavage differs drastically between RNAPs from different organisms, making it difficult to find a unifying theory on mismatch removal. Bacterial RNAPs from *Deinococcus radiodurans*, *Thermus aquaticus* and cyanobacteria can rapidly cleave the backtracked RNA, whereas *E. coli* RNAP cleaves RNA slowly, requiring Gre factors to assist with RNA cleavage. While the species-specific differences in RNA cleavage activity have been mapped to TL and regions that modulate its folding (Esyunina *et al*., 2016; Riaz-Bradley *et al*., 2020), it has been shown that TL does not donate any catalytical residues to the cleavage reaction (Mishanina *et al*., 2017; Palo *et al*., 2021). Instead, it is more likely that TL catalyzes the cleavage reaction by stabilizing the backtracked state when folding into TH and by positioning the RNA in the cleavage reaction (Turtola *et al*., 2018; Mishanina *et al*., 2017). While the role of TL has been extensively studied, the roles of other regions in the active site have been overlooked. Therefore, the exact mechanism of how RNAP directs misincorporated RNA for endonucleolytic cleavage remains unsolved.

RNA cleavage has been previously studied by many research groups, but experiments were often performed at a high concentration of Mg^{2+} (> 10 mM) and high pH to speed up the reaction (Fouqueau *et al*., 2013; Mishanina *et al*., 2017; Yuzenkova *et al*., 2010; Mosaei & Zenkin, 2021). We reasoned that under such conditions the reaction may possibly follow physiologically irrelevant mechanisms because it is being assisted by Mg^{2+} and OH^- ions bound at sites that are not occupied under physiological conditions. In bacterial cells the physiological concentration of Mg^{2+}, which is mostly bound by NTPs and amino acids, is about 54 mM (Outten & O'Halloran, 2001), with the concentration of unbound hydrated Mg^{2+} ions only 1.5–3 mM (Yamagami *et al*., 2018). We thus aimed to investigate the cleavage reaction at low Mg^{2+} (2 mM) and neutral pH (7.5). We selected two scaffold systems from a large (200+) library of scaffolds available in our laboratory to assemble TECs that are cleaved by the wild-type RNAP and most variants within 4 h at 25 °C at neutral pH and low Mg^{2+} concentration. This allowed us to perform cleavage assays in a reasonable time frame so that RNAP remains stable in the reaction mix. One TEC contained a single mismatched base at the RNA 3' end (1BKT), so we expected it to occupy a 1-nt backtracked state where the penultimate RNA nucleotide occupies the substrate site and the 3' terminal RNA nucleotide is extruded into the secondary

channel. The other TEC contained two mismatched RNA bases at the RNA 3' end (2BKT), so we expected it to occupy a 2-nt backtracked state where the third nucleotide from the RNA 3' terminus occupies the substrate site and two 3' terminal RNA nucleotides are extruded into the secondary channel.

Previous biochemical studies involving structural modelling suggested that one backtracked nucleotide can be accommodated within the closed RNAP active site and that TL folding can stabilize the TEC in a 1-nt backtracked state (Turtola *et al*. 2018). In a 1-nt backtracked state, TL can assist cleavage as a positional catalyst by stabilizing the state and positioning the backtracked nucleotide at the active site (Mishanina *et al*., 2017). At the same time, two backtracked nucleotides cannot conceivably fit into the closed active site (Abdelkareem *et al*., 2019), so TL should not affect the stability of the 2-nt backtracked state and, if assisting the cleavage reaction, should do so in the open or partially open conformation. The RNA cleavage reaction may follow distinct mechanisms in a 1-nt backtracked TECs and 2-nt backtracked TECs (or longer backtracked TECs). Our experimental setups were chosen to investigate the reaction mechanisms in both scenarios.

We proceeded to measure the rate of RNA cleavage in 1BKT and 2BKT using a large collection of *Eco* RNAP variants available in the laboratory. We chose a total of 28 variants in which the type of amino acid substitutions varied greatly. In some cases, both alanine and more conservative substitutions of the active site residues were available and tested (e.g. βE813A and βE813Q). In other cases, mutations mimicked the variations at the active site structure between RNAPs from different species (e.g. β'Q504R, β'I937T and β'ΔSI3). Another subset of substitutions located at the periphery of the active site were chosen because they are known to strongly stimulate TL folding (β'P750L, β'F773V). Finally, some substitutions such as β'K598W had complex reasonings behind their design, but were found to strongly affect RNA cleavage activity in our previous studies (Turtola *et al*., 2018). The effects of amino acid residue substitutions on RNA cleavage rate are summarized in figures 13 and 14.

5.2.1 TL folding improves proofreading of single mismatched nucleotides, but interferes with RNA cleavage at a 2-nt backtracked state

TL is the focus of many studies involving RNA cleavage, so we sought to test how various mutations targeting TL or domains involved in TL folding affect RNA cleavage in our scaffolds (Figure 13). Substitutions that promote TL folding stimulated cleavage in 1BKT and inhibited cleavage in 2BKT, reinforcing our idea that folded TL stimulates cleavage in the former but not the latter. Most substitutions in TL had neutral or positive effects on the cleavage in 1BKT, except for β'Q929M

and β'R933N, which decreased the rate of RNA cleavage several fold. Substitutions with positive effects in BKT1 (β'F773V, β'P750L, β'I937T, β'G1136S, β'ΔSI3) likely act by promoting TL folding. Most substitutions in TL had neutral or negative effects on the cleavage in 2BKT, except for β'R933A and β'Q929R, which increased the rate of cleavage 4.7- and 14-fold, respectively. Substitutions with negative effects can either promote TL folding in 2BKT thereby inhibiting cleavage activity, or they inhibit the cleavage activity by other means.

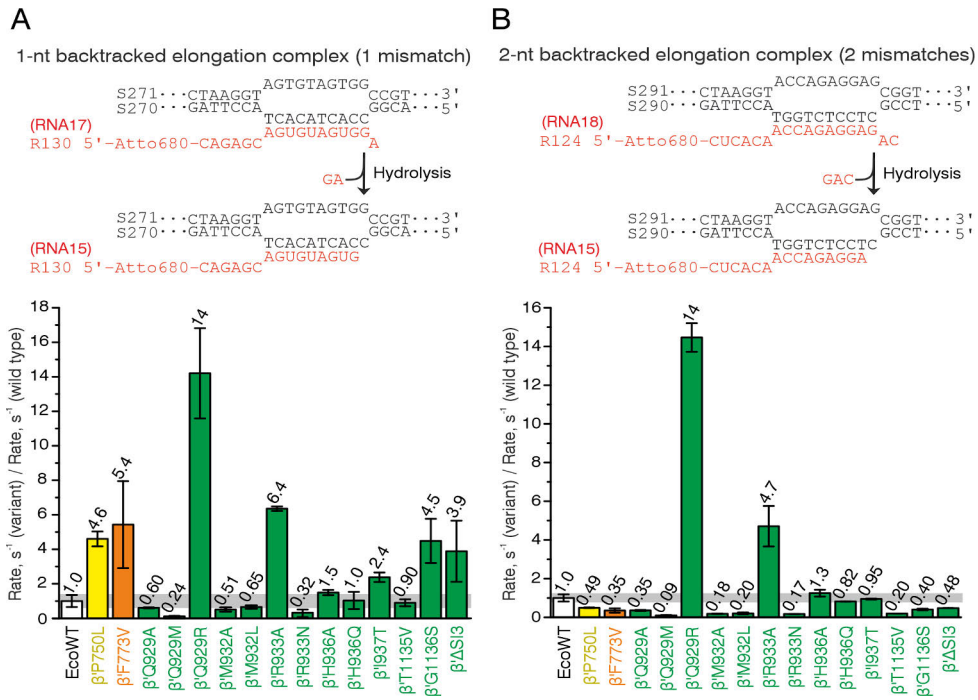


Figure 13. Rates of intrinsic RNA cleavage at 1-nt and 2-nt backtracked states by TL variants and by variants that affect TL folding. The structures of the nucleic acid scaffolds used in the experiments are shown above the column plots. The column plot values are RNA cleavage rates relative to the RNA cleavage rate of wild type enzyme (i.e. how many times faster the variant is compared to the wild type). The average values are shown above the columns. Error bars are ranges of duplicate measurements. Grey horizontal boxes depict the standard deviation in the reactions with *E. coli* wild type enzyme (EcoWT). Wild type is shown as white column, F-loop variant β'P750L yellow, BH variant β'F773V orange, and TL variants green. (A) RNA cleavage in 1-nt backtracked scaffold with rA-dG mismatch at RNA 3'end. (B) RNA cleavage in 2-nt backtracked state with rA-dG and rC-dC mismatches at RNA 3'end.

Noteworthy, the TL substitutions β'Q929M and β'R933N had negative effects on cleavage in both 1BKT and 2BKT, whereas β'Q929R and β'R933A had positive effects on the cleavage rate in both 1BKT and 2BKT. β'Gln929 and β'Arg933 are located within the first helical turn of the N-terminal helix of TL and can reach into

the active site upon partial folding of TL, with the former perhaps not even requiring the folding of TL to do so. These residues are thus likely to modulate the RNA cleavage rate by direct contact with backtracked RNA. Our data suggests that β'Gln929 facilitates RNA cleavage and β'Arg929 is even more stimulatory. In contrast, β'Arg933 likely inhibits RNA cleavage and β'Asn933 is even more inhibitory.

5.2.2 Mutations in regions other than the TL

In addition to the TL, we sought to investigate the various regions in an active site that could potentially interact with backtracked RNA and thus be involved in the RNA cleavage mechanism (Figure 14). Within the static core of the active site (regions other than the mobile TL), substitutions of βAsp814 (βD814A and βD814N) had the most pronounced negative effect. βAsp814 is one of the five conserved residues forming an acidic patch in the RNAP active site that binds the catalytic Mg^{2+} ions. From the two residues contributed to the patch by the β subunit, βGlu813 has a relatively small impact on RNA cleavage activity considering the margins of experimental uncertainty. In contrast, substitutions of βAsp814 reduced the cleavage rate several fold in both 1BKT and 2BKT. These effects are consistent with the hypothesis that βAsp814 participates in coordination of the Mg^{2+} ion responsible for activating the water molecule for nucleophilic attachment during the cleavage reaction (Sosunov *et al*., 2003; Sosunova *et al*., 2013). Noteworthy, βAsp814 is not critical for the nucleotide addition activity of RNAP, as βD814V RNAP is also viable (Holmes *et al*., 2006).

Among the cleavage promoting substitutions within the static core of the active site, β'Q504R and β'K598W are located at the E-site that was previously proposed to accommodate backtracked nucleotides (Wang *et al*., 2006; Sosunov *et al*., 2003). Our data indicates that the E-site structure affects cleavage of both 1BKT and 2BKT, as both β'Q504R and β'K598W increased RNA cleavage rates in both TECs. One of the most interesting observations was the strong stimulation of the cleavage rate by β'N458S (in both 1BKT and 2BKT) and β'R425K (in 2BKT) substitutions. These two residues are responsible for positioning the RNA nucleotide (or NTP) at the substrate site. In fact, if we aggregate all observations, substitutions in the majority of residues that may come into direct contact with the RNA nucleotide at the substrate site (β'R425, β'N458, β'Q929, β'M932, β'R933) had a strong positive or negative effect on the RNA cleavage rate. While the structural consequences of substitutions like β'Q929R are difficult to predict, substitutions of β'R425 and β'N458 very likely relax the position of the RNA nucleotide at the substrate site and allow it to adopt altered conformations compared to that of a substrate NTP aligned to the nucleotide addition reaction (Sydow *et al*., 2009; Sosunova *et al*., 2013). These

observations lead us to infer that the optimal cleavage proficient geometry is likely achieved when the RNA nucleotide at the substrate site (*i*+1) forms canonical base pairing with the DNA template, but its sugar-phosphate moieties are at least partially displaced from a canonical location characteristic of the substrate NTP or RNA 3'NMP at the pre-translocated state. Sosunova *et al*. 2013 reached similar conclusions based on purely structural considerations. They compared structures of backtracked (PDB: 3GTG) and pre-translocated (PDB: 1I6H) *Sce* RNAPII elongation complexes and proposed that the backtracked NMP at *i*+2 forces the NMP at *i*+1 to shift (compared to RNA 3'NMP at the pre-translocated state) without losing the canonical base pairing with the template DNA, resulting in a possible cleavage proficient geometry of RNA (Sosunova *et al*., 2013).

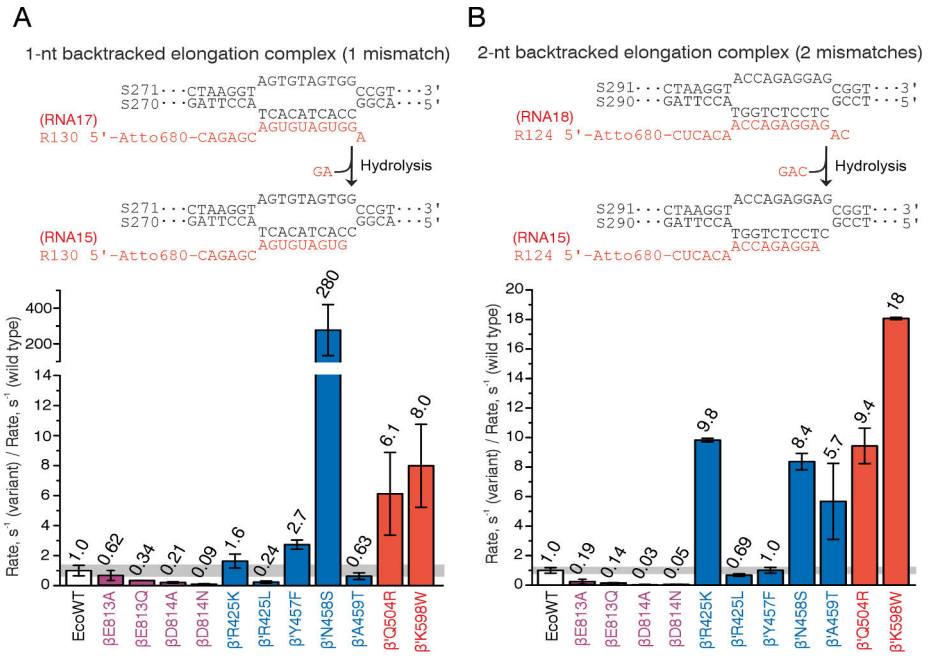


Figure 14. Rates of intrinsic RNA cleavage at 1-nt and 2-nt backtracked states by variants in which mutations were directed to MgII-binding residues, A-site and E-site. The structures of the nucleic acid scaffolds used in the experiments are shown above the column plots. The column plot values are RNA cleavage rates relative to the RNA cleavage rate of wild type enzyme (i.e. how many times faster the variant is compared to the wild type). The average values are shown above the columns. Error bars are ranges of duplicate measurements. Grey horizontal boxes depict the standard deviation in the reactions with *E. coli* wild type enzyme (EcoWT). Wild type is shown as white column, MgII-binding variants purple, A-site variants blue, and E-site variants red. (A) RNA cleavage in 1-nt backtracked scaffold with rA-dG mismatch at RNA 3'end. (B) RNA cleavage in 2-nt backtracked state with rA-dG and rC-dC mismatches at RNA 3'end.

5.2.3 Roles of the RNAP active site domains and clefts in endonucleolytic RNA cleavage

In previous studies on intrinsic RNA cleavage, the focus has been mostly on either the residues that coordinate MgII (Sosunov *et al*., 2003; Sosunova *et al*., 2013), or the role of the TL (Zhang *et al*., 2010; Yuzenkova and Zenkin, 2010; Mishanina *et al*., 2017). In our study, mutations directed at the A-site and E-site greatly affected RNA cleavage, which indicates that these clefts at the active site are involved in the RNA cleavage mechanism alongside TL. However, the reasons why some mutations at the active site (especially at the E-site) have such high impacts on the cleavage reaction in our assays are not apparent from our RNA cleavage experiments alone. Our tests with RNAP variants are limited to two scaffolds, meaning that we cannot observe the sequence dependent effects on the cleavage reaction, which very likely affect the positioning of the backtracked RNA 3' end nucleotides at the active site. Indeed, some mismatches are proofread by RNAPs much more efficiently than others (Zenkin *et al*., 2006; Sydow *et al*., 2009), and in some cases the RNA cleavage reaction can proceed in largely different ways depending on the structure of nucleic acid scaffold used (Mosaei & Zenkin, 2021), meaning that a more detailed examination of variants in different scaffolds is needed to ascertain the roles of active site residues in the proofreading reaction. In addition to the variants tested here, there are other residues at the active site that could participate in the reaction that we did not test, like β'Arg1106, which could potentially form a salt bridge with βAsp814 affecting the coordination of MgII during RNA cleavage (Sosunov *et al*., 2003).

Our data suggests the following insights into the mechanisms of intrinsic RNA cleavage by multi-subunit RNAPs: (i) folding TL assists cleavage in 1BKT but not in 2BKT, (ii) the first helical turn of the N-terminal helix of TL (up to β'R933) modulates the cleavage rate in both 1BKT and 2BKT, (iii) βAsp814 always plays an important role in cleavage, likely by coordinating the nucleophile-generating Mg^{2+}, (iv) the E-site structure modulates the cleavage rate in both 1BKT and 2BKT, and (v) the RNA nucleotide at the substrate site must be base-paired to template DNA, but is likely shifted to some extent from its canonical location during the RNA cleavage reaction.

5.3 Nucleobase analogues as substrates for different RNAPs (Studies III and IV)

In addition to NTPs and 2'dNTPs, RNAPs can incorporate other substrates that have triphosphate-, nucleosugar- and nucleobase moieties. A wide variety of nucleoside analogues have been developed to target viral nucleic acid polymerases and other viral proteins. Many species of actinobacteria produce C-nucleosides (see chapter 2.6.1) as secondary metabolites, which have been studied as potential antiviral and

anticancer drugs (De Clercq, 2016). C-nucleosides have the ribonucleosugar moiety and varying base moieties that can form hydrogen bonds with tDNA acceptor base moieties. Nucleoside analogues with 5'-triphosphate moieties can be incorporated into RNA instead of the cognate NTPs, which provides the basis for developing RNAP inhibitors for bacterial and viral RNAPs. In studies III and IV we focused on the recognition of different non-cognate nucleobase moieties.

Showdomycin, a C-nucleoside isolated from *Streptomyces showdoensis*, has wide spectrum antibacterial activity, as well as antitumor activity (Nishimura *et al*., 1964; Matsuura *et al*., 1964). Addition of showdomycin to growth media greatly reduces the biosynthesis of both nucleic acids and proteins in the *E. coli* K-12 strain in minimal media, but the inhibitory effect is much less potent in organic culture media that is rich in nucleotides, or when additional nucleotides are added to minimal media (Komatsu and Tanaka, 1968; Nishimura and Komatsu, 1968). These findings imply that showdomycin competes with nucleotides in the biosynthesis of nucleic acids. Formycin A is a C-nucleoside isolated from *Nocardia interforma* ATCC 21072, *Streptomyces kaniharaensis* SF-557 (ATCC 21070) and *Streptomyces lavendulae* (Hori *et al*., 1964; Aizawa *et al*., 1965). Formycin A is shown to have antibacterial, antiviral and antitumor activity (Ishizuka *et al*., 1968). Pyrazofurin A, which has antiviral and antitumor activity, is a C-nucleoside produced by *Streptomyces candidus* (Gutowski *et al*., 1975). In addition to C-nucleosides we also wanted to study known dual coders, which can be incorporated in place of more than one nucleotide due to interconvertible *anti* and *syn* conformers. 8-oxoadenine (8oA) pairs with uridine in an *anti*-conformation and with guanine in a *syn*-conformation (Kamiya *et al*., 1998; Koag *et al*., 2019). 8-oxoguanine (8oG) pairs with cytosine in an *anti*-conformation and with adenine in a *syn*-conformation (Brieba *et al*., 2004; Batra *et al*., 2012). Ribavirin, an inhibitor of viral RNAPs, can pair with uridine and cytosine due to rotation of the carboxamide moiety (Crotty *et al*., 2000).

We sought to study the effects of various chemically modified showdomycin derivatives, formycin A, pyrazofurin A, 8-oxoadenine, 8-oxoguanine and ribavirin, on RNA elongation by structurally distinct RNAPs. In these studies, we compared substrate utilization in *Eco* RNAP (bacterial DNA-dependent two-β-barrel RNAP), *Sce* RNAPII (eukaryotic DNA-dependent two-β-barrel RNAP), *Hsa* MT RNAP (mitochondrial DNA-dependent right-hand RNAP) and coxsackievirus B3 (CVB3) RNAP (viral RNA-dependent right-hand RNAP). Our goal was to study how well these compounds are utilized as substrates by different RNAPs and what kind of effects they would have on transcript elongation after incorporation into RNA.

5.3.1 Showdomycin derivatives are uridine analogues that induce pausing at poly-thymidine sequences

When showdomycin is phosphorylated to 5'-triphosphate it isomerizes into isoshowdomycin, which can't be utilized as a substrate by RNAPs. To avoid this, showdomycin monophosphate was modified to produce a series of compounds in which different chemical groups were attached to the thiol group of the maleimide ring (Rosenqvist *et al*., 2022). These compounds could be triphosphorylated without isomerization into isoshowdomycin. We tested six showdomycin derivatives: 4-methylthioshowdomycin (MeS-Sdm), 4-ethylthioshowdomycin (EthS-Sdm), 4-methylseleniumshowdomycin (MeSe-Sdm), 4-bromideshowdomycin (Br-Sdm), and two enantiomers of 4-thiolactoshowdomycin (TLA(R)-Sdm and TLA(S)-Sdm). We first determined in which positions the showdomycin derivatives are added to RNA by various RNAPs. *Eco* RNAP, *Sce* RNAPII and *Hsa* MT RNAP added showdomycin derivatives in place of UMP to the growing RNA chain (Figure 15). *Eco* RNAP incorporated all tested showdomycin derivatives within the time frame of the reaction, although incorporation of the thiolacto-showdomycins (TLA(R)-SMP and TLA(S)-SMP) was incomplete. *Sce* RNAPII and *Hsa* MT RNAP did not incorporate the larger 4-thiolacto-showdomycins, but did incorporate all other tested showdomycin derivatives efficiently. *Eco* RNAP and *Sce* RNAPII did not incorporate the second nucleotide efficiently after incorporating showdomycin derivatives. We also tested if showdomycin derivatives would be incorporated by a viral RNAP (CVB3 RNAP), but found that none of the compounds were utilized in place of uridine or other nucleotides.

We proceeded to study how the showdomycin derivatives would affect transcript elongation over longer templates. Since *Eco* RNAP and *Sce* RNAPII extended RNA poorly after incorporating showdomycin derivatives, we wanted to test the compounds with poly-thymidine tract previously used to test oxazinomycin (Prajapati *et al*. 2019). The scaffold had an 11-nucleotide-long leader sequence that did not contain any thymidines downstream from the RNA primer, followed by seven consecutive T bases from position 12 to 18 (a section in the transcribed sequence called the T-tract). The T-tract was followed by 31 nucleotides with 7 single thymidines at various positions (23, 27, 32, 35, 37, 41, and 47). We compared the overall transcription and pausing of the *Eco* RNAP wild type, *Sce* RNAPII and *Hsa* MT RNAP with all six showdomycin derivatives (Figure 16). We added 2 mM of 5'-triphosphorylated showdomycin derivatives to the reactions, together with 100 μM of NTPs, and incubated the reaction at 25 °C for 5 minutes.

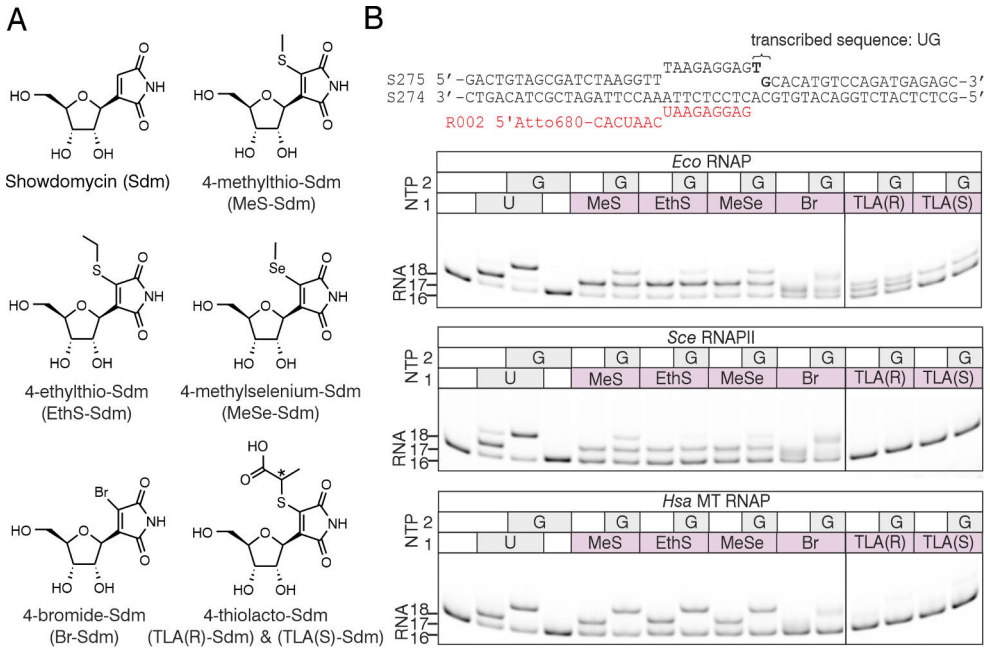


Figure 15. The incorporation of different 5'-triphosphorylated showdomycin derivatives in place of uridine by *Eco* RNAP, *Sce* RNAPII and *Hsa* MT RNAP. (A) Structures of the tested showdomycin derivatives. (B) Incorporation of the showdomycin derivatives in place of uridine. TECs were supplemented with UTP and GTP or with different derivatives of SDM triphosphate and GTP. The nucleic acid scaffold employed in the experiments is depicted above gel panels. NTPs (20 μM) and triphosphorylated showdomycins (100 μM) were added and the reactions were incubated for 1 min at 25 °C. Two lanes between Br-Sdm and TLA(R)-Sdm was spliced out from each gel panel as indicated by black vertical lines. Figure was adapted from the supplementary figures of Rosenqvist *et al.* 2022 (study III).

Eco RNAP was strongly affected by MeS-Sdm, EthS-Sdm, MeSe-Sdm and Br-Sdm, as the TECs paused more at the T-tract and at other positions where either uridine or showdomycin would be incorporated into the RNA. In the presence of 5'-triphosphorylated showdomycin derivatives, RNA products accumulate at positions where UMP (or showdomycin) has been incorporated into the RNA, meaning that TECs pause for longer after showdomycin addition. TLA(R)-Sdm and TLA(S)-Sdm did not have any effect on pausing at the T-tract for any tested RNAP, indicating that thiolacto-showdomycins are not incorporated efficiently in the presence of uridine. *Sce* RNAPII paused strongly at the T-tract at position 17, even in the absence of showdomycin derivatives. However, MeS-Sdm, EthS-Sdm, MeSe-Sdm and Br-Sdm increased *Sce* RNAPII pausing before position 17. *Hsa* MT RNAP was largely unaffected by the showdomycin derivatives in the presence of uridine.

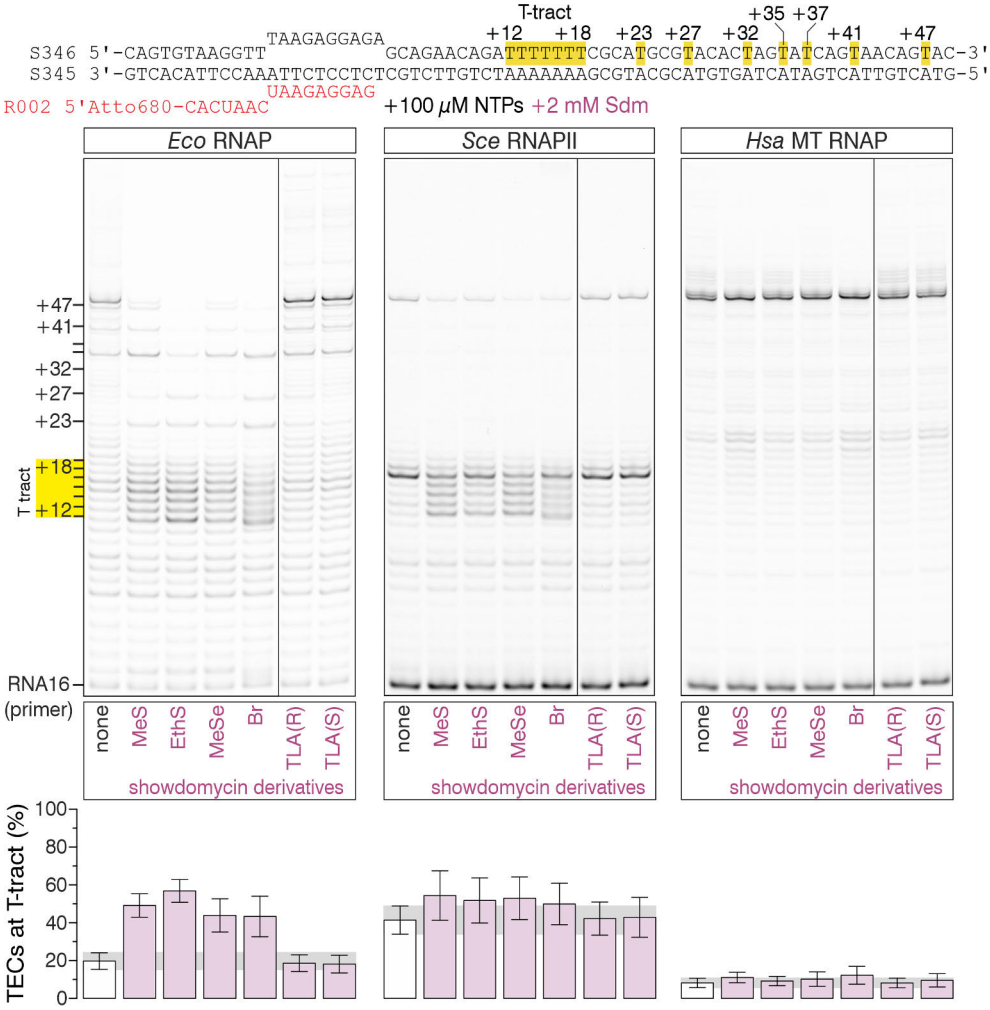


Figure 16. The effects of 5′-triphosphorylated showdomycin derivatives on processive transcription elongation by *Eco* RNAP, *Sce* RNAPII and *Hsa* MT RNAP. The nucleic acid scaffold employed in the experiments is depicted above gel panels. ATP, CTP, GTP, UTP (100 μM each) and showdomycin derivatives (2 mM) were added, and the reactions were incubated for 5 min at 25 °C. Bar graphs below the gel panels display fractions of RNAPs delayed at the polythymidine tract (positions +12 to +18, highlighted in yellow in the scaffold map), white bars represent the control reaction and purple bars represent the showdomycin derivatives. Gray horizontall boxes depict the standard deviation in the control reactions. A single lane between Br-Sdm and TLA(R)-Sdm was spliced out from each gel panel as indicated by black vertical lines. Figure was adapted from Rosenqvist *et al.* 2022 (study III).

5.3.2 Formycin A, pyrazofurin A, 8-oxo-ATP, 8-oxo-GTP and ribavirin as substrates for RNAPs from structurally distinct groups

We then tested formycin A, pyrazofurin A, 8-oxoadenosine, 8-oxoguanine and ribavirin against every template position (Figures 17-20). For these tests we used *Eco* RNAP (DNA-dependent two-β-barrel RNAP), *Hsa* MT RNAP (DNA-dependent right-hand RNAP) and CVB3 RNAP (RNA-dependent right-hand RNAP), as they represent structurally distinct classes of RNAPs. RNA scaffolds for CVB3 RNAP were designed so that template RNAs contained three 2'OMe-NMPs at the 3' end and five 2'OMe-NMPs at the 3' end for stability. In between the 2'OMe-NMP ends was a fourteen-nucleotide long region with a primer binding site (9 nucleotides) and five nucleotides upstream from the primer where the CVB3 RNAP could transcribe the template RNA. 10 μM of NTPs and 100 μM of nucleoside analogues were mixed with TECs and the reaction allowed to proceed for 1 minute at 25 °C to see if the analogues can be incorporated by the RNAPs within a reasonable time frame (Figure 17, Figure 18, Figure 19 and Figure 20).

Formycin A was incorporated in place of adenine and cytosine by *Eco* RNAP and CVB3 RNAP, whereas *Hsa* MT RNAP incorporated it strictly in place of adenine. Pyrazofurin A was incorporated in place of adenine and uridine by *Eco* RNAP, but *Hsa* MT RNAP and CVB3 RNAP incorporated it strictly in place of uridine. Ribavirin, which is structurally similar to pyrazofurin A, was incorporated only by *Eco* RNAP (though inefficiently) in place of adenine and guanine. 8-oxoadenosine was incorporated in place of adenosine by *Eco* RNAP and *Hsa* MT RNAP, and in place of uridine by all the tested RNAPs. 8-oxoguanine was incorporated in place of uridine by *Eco* RNAP and *Hsa* MT RNAP, and in place of guanine by all the tested RNAPs. Based on these results, formycin A, pyrazofurin A, ribavirin, 8-oxoadenosine and 8-oxoguanine are dual coders, as they base pair with multiple tDNA acceptor bases and thus pose as two nucleotides instead of one. *Eco* RNAP, *Hsa* MT and CVB3 RNAP had different preferences for each substrate, underlining the differences in how these RNAPs recognize their base moieties. Interestingly, *Eco* RNAP appeared to be the least strict in incorporation of the nucleoside analogues, as it incorporated all tested analogues against two different acceptor bases.

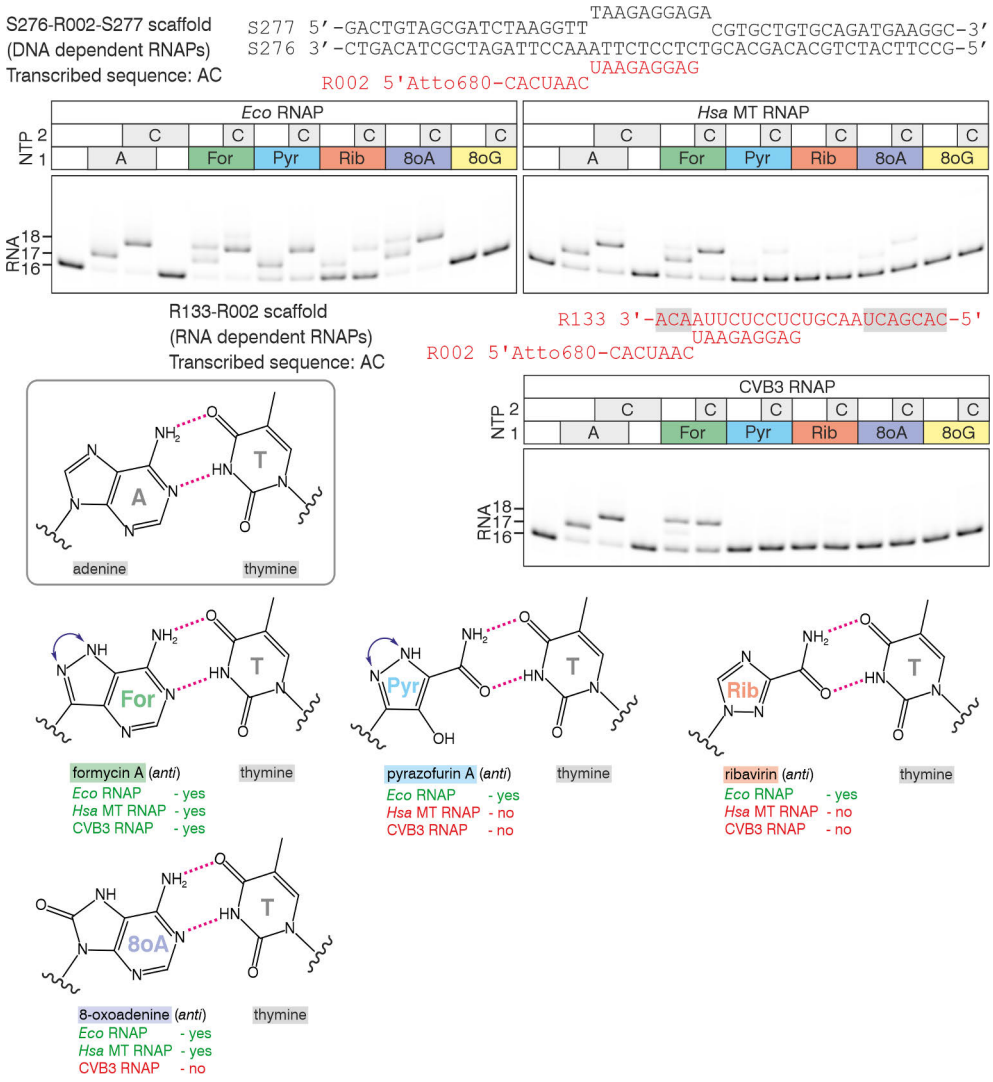


Figure 17. Incorporation of nucleoside analogues in place of adenine by *Eco* RNAP, *Hsa* MT RNAP and CVB3 RNAP. TECs were supplemented with ATP and CTP or with different nucleoside analogues and CTP. The nucleic acid scaffold employed in the experiments is depicted above gel panels. For CVB3 RNAP the RNA-RNA scaffold contained three 2'OMeNMPs at 3' end and seven 2'OMeNMPs at 5' end (highlighted in gray). RNA oligos are colored red, DNA oligos black. NTPs (10 μM) and nucleoside analogues (100 μM) were added and the reactions were incubated for 1 min at 25 °C. Possible hydrogen bonding of adenine analogues with thymine (tDNA acceptor base) are shown below the gel panels. Only the base moieties are shown for clarity. The conformation of the base moiety is written next to the names of the nucleoside analogues. Hydrogen bonds are shown as dashed magenta lines.

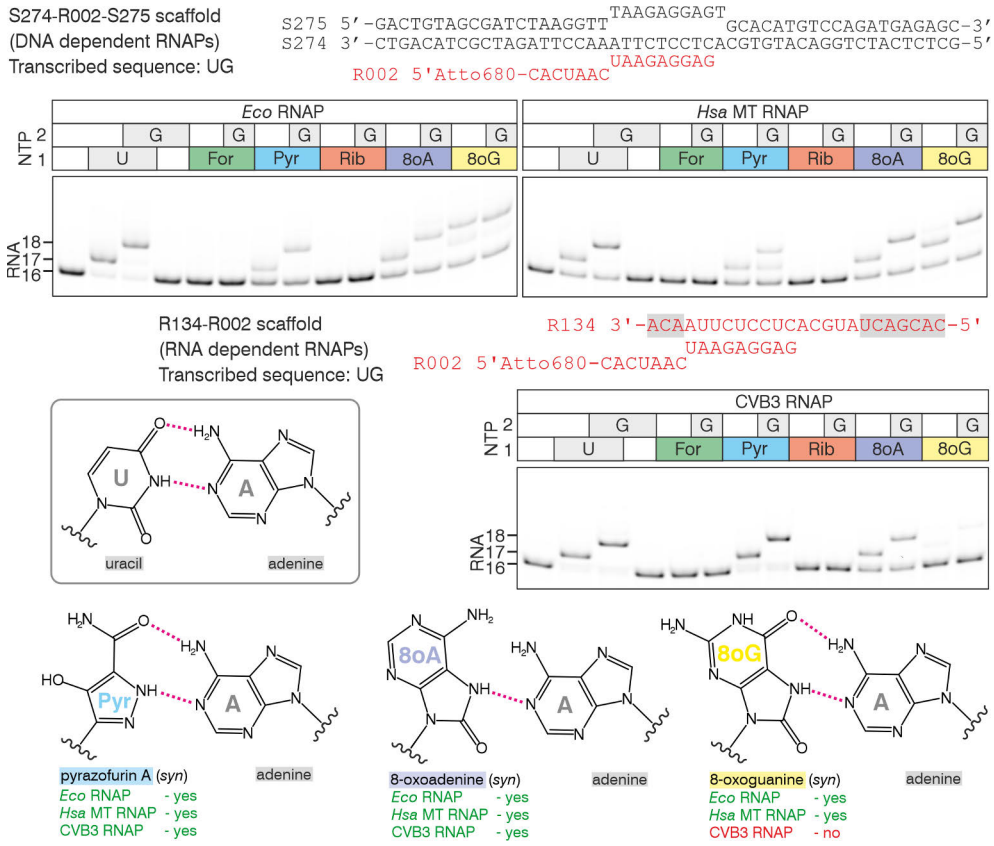


Figure 18. Incorporation of nucleoside analogues in place of uridine by *Eco* RNAP, *Hsa* MT RNAP and CVB3 RNAP. TECs were supplemented with UTP and GTP or with different nucleoside analogues and GTP. The nucleic acid scaffold employed in the experiments is depicted above gel panels. For CVB3 RNAP the RNA-RNA scaffold contained three 2'OMeNMPs at 3' end and seven 2'OMeNMPs at 5' end (highlighted in gray). RNA oligos are colored red, DNA oligos black. NTPs (10 μM) and nucleoside analogues (100 μM) were added and the reactions were incubated for 1 min at 25 °C. Possible hydrogen bonding of uridine analogues with adenine (tDNA acceptor base) are shown below the gel panels. Only the base moieties are shown for clarity. The conformation of the base moiety is written next to the names of the nucleoside analogues. Hydrogen bonds are shown as dashed magenta lines.

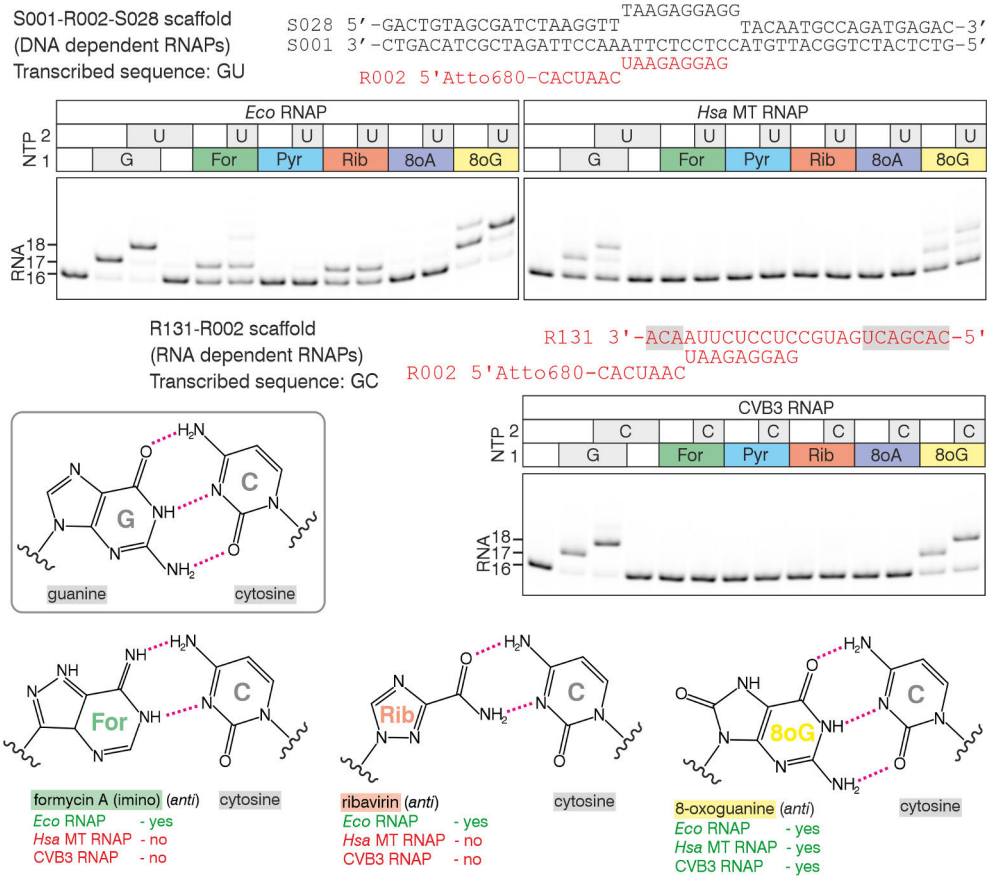


Figure 19. Incorporation of nucleoside analogues in place of guanine by *Eco* RNAP, *Hsa* MTRNAP and CVB3 RNAP. TECs were supplemented with GTP and UTP/CTP or with different nucleoside analogues and UTP/CTP. The nucleic acid scaffold employed in the experiments is depicted above gel panels. For CVB3 RNAP the RNA-RNA scaffold contained three 2'OMeNMPs at 3' end and seven 2'OMeNMPs at 5' end (highlighted in gray). RNA oligos are colored red, DNA oligos black. NTPs (10 μM) and nucleoside analogues (100 μM) were added and the reactions were incubated for 1 min at 25 °C. Possible hydrogen bonding of guanosine analogues with cytidine (tDNA acceptor base) are shown below the gel panels. Only the base moieties are shown for clarity. The conformation of the base moiety is written next to the names of the nucleoside analogues. Hydrogen bonds are shown as dashed magenta lines.

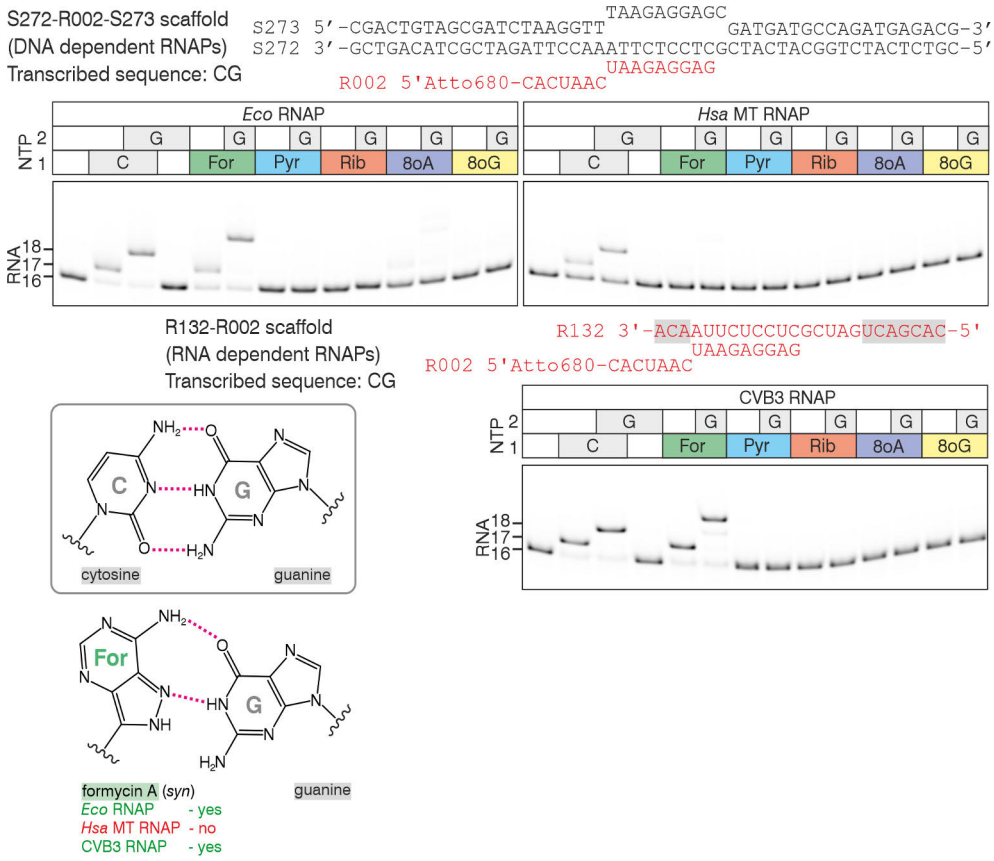


Figure 20. Incorporation of nucleoside analogues in place of cytidine by *Eco* RNAP, *Hsa* MT RNAP and CVB3 RNAP. TECs were supplemented with CTP and GTP or with different nucleoside analogues and GTP. The nucleic acid scaffold employed in the experiments is depicted above gel panels. For CVB3 RNAP the RNA-RNA scaffold contained three 2'OMeNMPs at 3' end and seven 2'OMeNMPs at 5' end (highlighted in gray). RNA oligos are colored red, DNA oligos black. NTPs (10 μM) and nucleoside analogues (100 μM) were added and the reactions were incubated for 1 min at 25 °C. Possible hydrogen bonding of cytidine analogues with guanine (tDNA acceptor base) are shown below the gel panels. Only the base moieties are shown for clarity. The conformation of the base moiety is written next to the names of the nucleoside analogues. Hydrogen bonds are shown as dashed magenta lines.

5.3.3 Utilization of nucleoside analogues in place of cognate ribonucleotides by *Eco* RNAP and *Hsa* MT RNAP

Next, we wanted to test how well formycin A, pyrazofurin A, 8-oxoadenosine, 8-oxoguanine and ribavirin are utilized in place of cognate substrates in longer templates. In these tests the RNAPs were allowed to extend to the end of the DNA template (49 nucleotides) with different sets of nucleotides and analogues. To test *Eco* RNAP and *Hsa* MT RNAP we utilized four different scaffolds, one for each

ribonucleotide (Figure 21 and Figure 22). There were three distinct setups for these experiments: (i) reactions with all four NTPs (100 μM each), (ii) reactions with three NTPs (100 μM), and (iii) reactions with three NTPs (100 μM each) and a nucleoside analogue (100 μM) incorporated in place of the absent NTP. For instance, adenine analogues were tested with a scaffold that had multiple A adding positions in reactions that had the analogues, GTP, CTP and UTP (no ATP). Reactions were allowed to proceed for 5 minutes at 25 °C to ensure efficient extension by the analogues. These tests were qualitative estimates of how well nucleoside analogues can be utilized in place of the nucleotides they pose as.

In the case of *Eco* RNAP, we found that formycin A was a good substitute for adenine as most TECs were able to reach the end of the template with formycin A in place of adenine. The other adenine analogues for *Eco* RNAP (pyrazofurin A, ribavirin and 8-oxoadenosine) did not improve the extension compared to the control with no adenine by much. All uridine analogues for *Eco* RNAP (pyrazofurin A, 8-oxoadenine and 8-oxoguanine) allowed extension past the first U adding positions, although the TECs did not reach the end of the template, indicating that either the incorporated analogues induced pausing, or the extension efficiency with the analogues was so low that the TECs did not extend to the end of the template. 8-oxoguanine was a good guanine analogue for *Eco* RNAP as the TECs did not pause strongly before or after G addition, whereas ribavirin did not improve extension when added in place of GTP. Finally, formycin A improved the extension in the absence of CTP only slightly for *Eco* RNAP, indicating that formycin A is much better utilized as an adenine analogue than as a cytidine analogue by *Eco* RNAP.

Hsa MT RNAP utilized formycin A in place of adenine efficiently, although this RNAP performed relatively poorly with the employed scaffold even in the presence of four NTPs. *Hsa* MT RNAP utilized pyrazofurin A, 8-oxoadenine and 8-oxoguanine efficiently in place of uridine, reaching the end of the template in the presence of 8-oxoguanine. For the *Hsa* MT RNAP, 8-oxoguanine did not work as well in place of guanine as it did in place of uridine, although the RNAP did manage to extend past multiple G adding positions with 8-oxoguanine.

In short, *Eco* RNAP (a two-β-barrel RNAP) and *Hsa* MT RNAP (a right-hand RNAP) both utilized formycin A efficiently in place of adenine, whereas pyrazofurin A was utilized in place of uridine more efficiently by *Hsa* MT RNAP than by *Eco* RNAP. *Eco* RNAP favored 8-oxoguanine in place of guanine more than uridine, whereas *Hsa* MT RNAP favored it more in place of uridine than guanine. These results show how distinctly different classes of DNA-dependent RNAPs have differing preferences for the nucleobases.

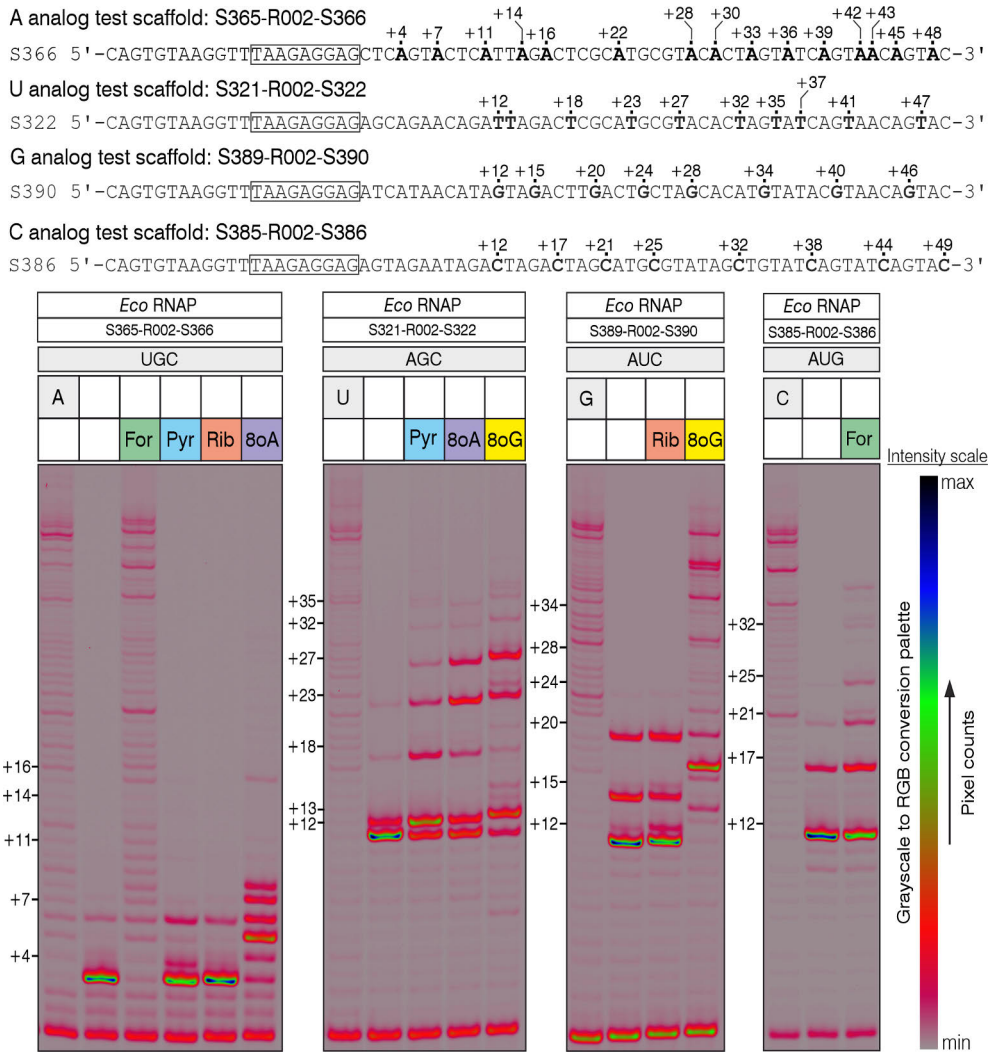


Figure 21. Utilization of nucleoside analogues in processive elongation by *Eco* RNAP. The transcribed sequences (ntDNA sequences of scaffolds) are shown above the gels. The RNA primer binding site is highlighted with a box and the positions downstream from the primer where nucleotides or their analogues are added are bolded and their position number is shown. The corresponding position on the gels are shown at the left edge of the gel images. 100 µM of NTPs and nucleoside analogues were mixed with TEC and incubated for 5 minutes at 25 °C. 16-bit grayscale scans were normalized using max pixel counts within each gel panel and pseudo-colored using inverted Rainbow RGB color palette of ImageJ.

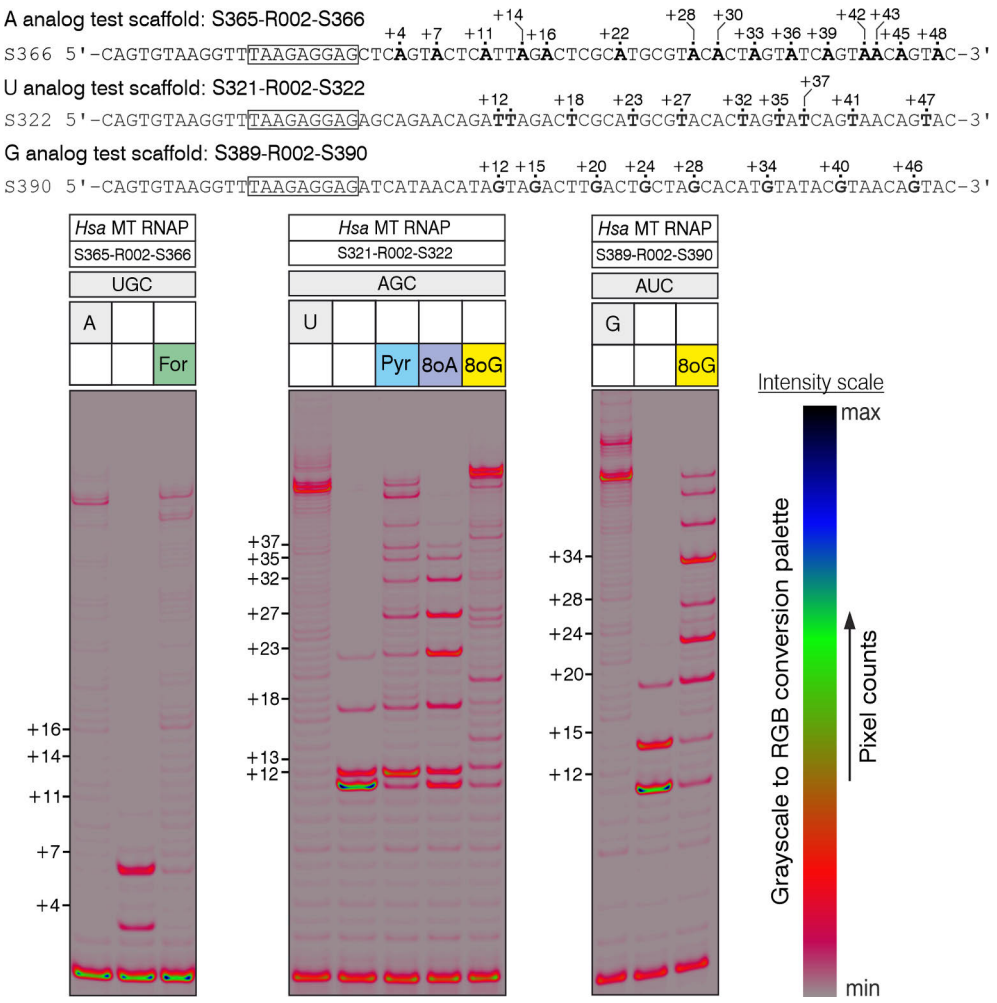


Figure 22. Utilization of nucleoside analogues in processive elongation by *Hsa* MT RNAP. The transcribed sequences (ntDNA sequences of scaffolds) are shown above the gels. The RNA primer binding site is highlighted with a box, the positions downstream from the primer where nucleotides or their analogues are added are highlighted with different colors (A green, T yellow and G red). 100 μM of NTPs and nucleoside analogues were mixed with TEC and incubated for 5 minutes at 25 °C. 16-bit grayscale scans were normalized using max pixel counts within each gel panel and pseudo-colored using inverted Rainbow RGB color palette of ImageJ.

5.3.4 Utilization of nucleoside analogues in place of cognate ribonucleotides by CVB3 RNAP

Because CVB3 RNAP synthesizes RNA using an RNA template, we needed a different approach for the processive elongation experiments, compared to *Eco* RNAP and *Hsa* MT RNAP. Long RNA templates were synthesized with the T7 RNAP variant P266L from chemically synthesized dsDNA templates, purified via anion exchange and concentrated, after which the RNA templates were annealed with an RNA primer forming a 21 base pair duplex. TECs assembled with RNA-RNA scaffolds and CVB3 RNAP were mixed with different sets of NTPs and their analogues (all NTPs, one NTP removed, or one NTP replaced with its analogue, all added at 100 μM) and reactions were incubated for 5 minutes at 25 °C (Figure 23), similarly to how *Eco* RNAP and *Hsa* MT RNAP were tested with the same compounds. The same synthesized RNA template was used to test formycin A (in place of adenine and cytosine), 8-oxoguanine (in place of guanine), pyrazofurin A (in place of uridine) and 8-oxoadenine (in place of uridine). CVB3 RNAP utilized formycin A efficiently in place of adenine and cytosine, extending well past the first positions where the analogues were incorporated. RNA strands were extended relatively well also with pyrazofurin A in place of uridine. 8-oxoguanine and 8-oxoadenine improved RNA extension in the absence of guanine and uridine, respectfully, but the TECs did not reach the end of the template within a reasonable time frame.

We decided to prepare a template with formycin A in place of adenine, to see how that would affect RNA elongation by CVB3 RNAP (Figure 23C). The idea was that because formycin A can hydrogen bond with both uracil and guanine, it could be possible that formycin A in the RNA template could induce errors in the transcripts when transcribed by CVB3 RNAP. We found that in a reaction with only ATP, GTP and CTP, the CVB3 RNAP paused cleanly at the first U adding position when the RNA template was prepared with four rNTPs. When the template had formycin A in place of adenine, CVB3 RNAP continued elongating RNA past the first and other U adding positions, indicating that one of the nucleotides (most likely guanine, as it can hydrogen bond with formycin A) was incorporated in place of uridine, allowing CVB3 RNAP to extend further. This indicates that formycin A can be misread in the RNA template by CVB3 RNAP, resulting in erroneous transcripts. In short, formycin A is a potential mutation inducing nucleoside analogue for viral RNA-dependent RNAPs, similarly to how ribavirin works for poliovirus RNAP and hepatitis C RNAP (Crotty *et al*., 2002; Ortega-Prieto *et al*., 2013).

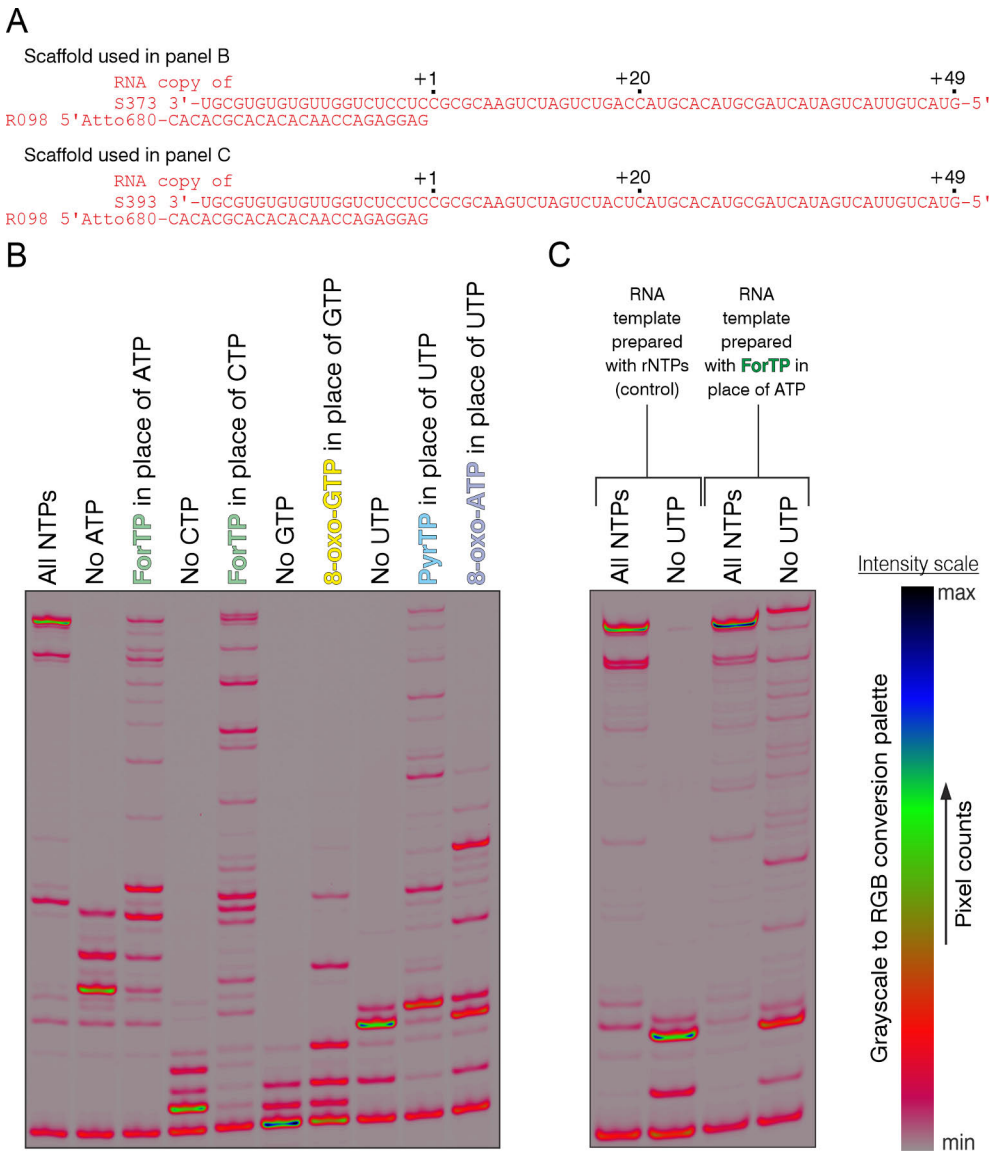


Figure 23. Utilization of nucleoside analogues in processive elongation by CVB3 RNAP. (A) The structures of the RNA-RNA scaffolds used in the experiments. (B) Processive elongation with nucleoside analogues in place of the cognate rNTPs. 100 μM of NTPs and nucleoside analogues were mixed with TEC and incubated for 5 minutes at 25 °C. (C) Comparison of template prepared with rNTPs and template prepared with ForTP in place of ATP. Reactions were chased with all four NTPs or with only ATP, GTP and CTP for 5 min at 25 °C. 16-bit grayscale scans were normalized using max pixel counts within each gel panel and pseudo-colored using inverted Rainbow RGB color palette of ImageJ.

5.3.5 Incorporated 4-ethylthioshowdomycin, formycin A and pyrazofurin A reduce the rate of further RNA elongation by increasing fractional backtracking of TECs

Next, we studied the post-incorporation properties of selected nucleoside analogues with *Eco* RNAP. Showdomycin, formycin A and pyrazofurin A were selected for these tests because they were the focal targets of our research on the transcriptional effects of C-nucleoside antibiotics. 4-ethylthioshowdomycin was chosen out of all the available showdomycin derivatives for further testing due to a strong effect on pausing at polythymidine sequences by *Eco* RNAP and its chemical stability compared to native showdomycin. To see how the nucleoside analogues at the nascent RNA 3' end would affect transcription elongation, we measured the rate of nucleotide addition after nucleoside analogue incorporation with *Eco* RNAP (Figure 24, Table 4). TECs were pre-extended with either cognate NTPs or their respective analogues after which the incorporation rate of the next nucleotide (500 μM) was measured with the HCl quenched method.

Table 4. Rates of nucleotide incorporation after 4-ethylthioshowdomycin, formycin A and pyrazofurin A by *Eco* RNAP. Structures of scaffolds are shown in Figure 24. Shown in the table are the rates and amplitudes of 500 μM NTP addition (GTP, UTP or CTP depending on scaffold). The values were obtained from best-fits of a sum of exponential (corresponds to the fast phase) and stretched exponential (corresponds to the slow phase) functions. The standard deviations are obtained from the ranges of 2 independent experiments.

Scaffold	NMP at 3' end	NTP added	k_{cat} (fast) s^{-1}	k_{cat} (slow) s^{-1}	Fast fraction %
S274-R002-S275	UMP	GTP	22 ± 3.0	0.21 ± 0.076	95 ± 0.15
	EthS-SdmMP	GTP	0.0055 ± 0.0013	0.16 ± 0.012	2.9 ± 1.1
S363-R002-S364	AMP	UTP	27 ± 0.49	7.6 ± 5.7	77 ± 4.6
	ForMP	UTP	25 ± 1.4	0.45 ± 0.039	65 ± 2.5
S276-R002-S277	AMP	CTP	55 ± 0.92	0.17 ± 0.012	83 ± 1.3
	PyrMP	CTP	23 ± 4.5	0.63 ± 0.075	45 ± 3.3

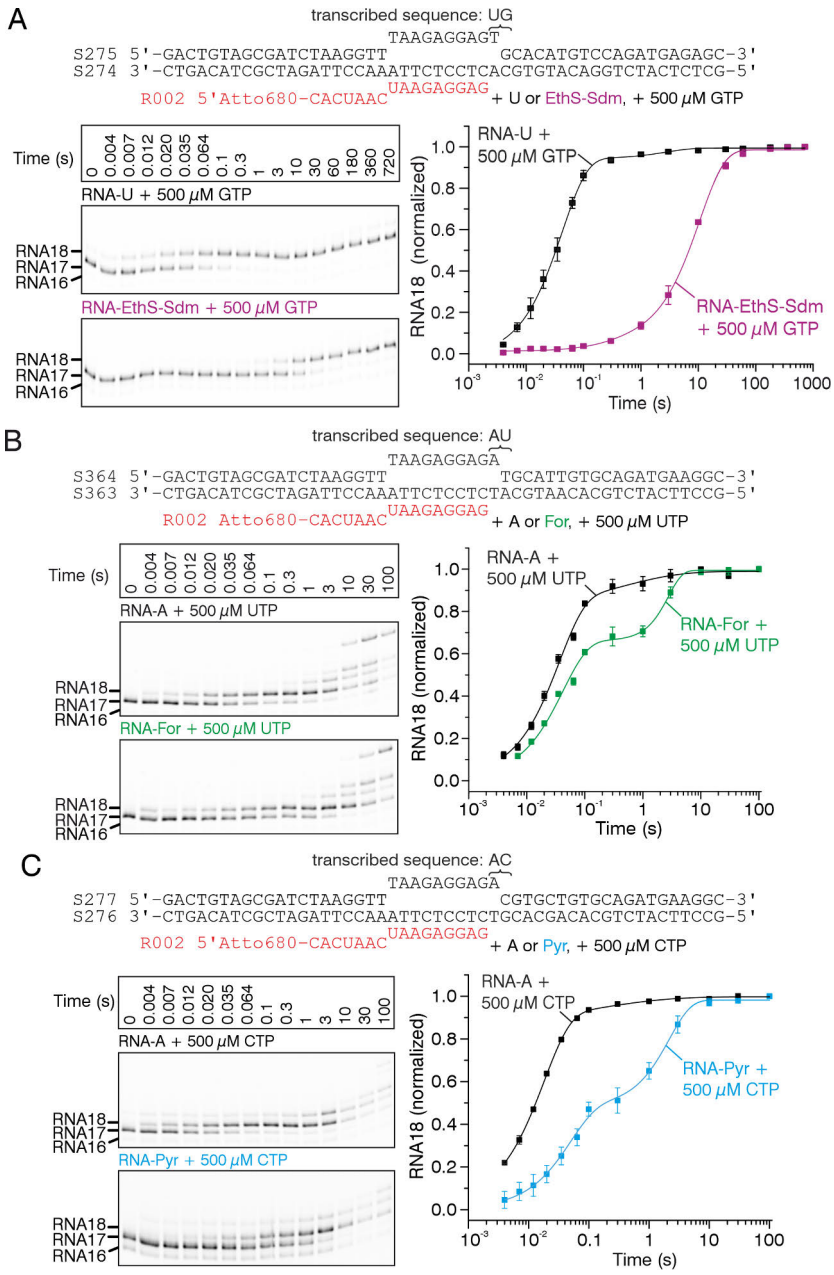
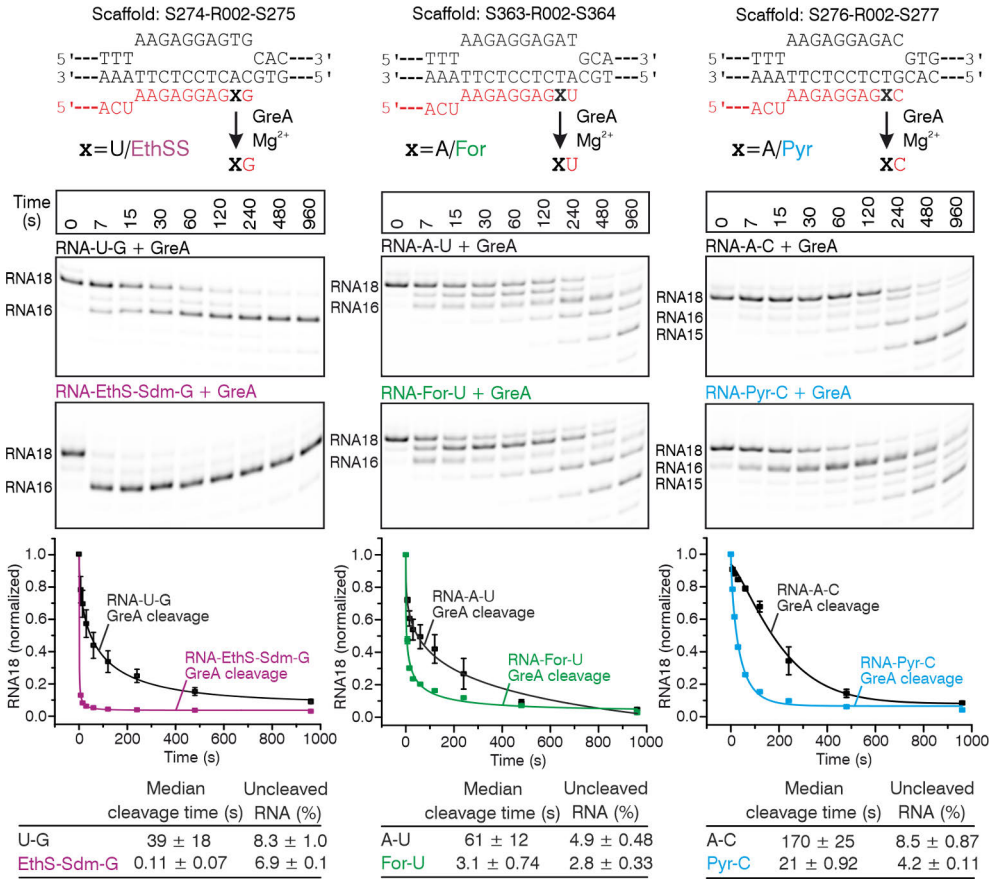


Figure 24. Incorporated 4-ethylthioshowdomycin (A), formycin A (B) and pyrazofurin A (C) slow down further RNA elongation. *Eco* RNAP was used for the experiments. The structures of scaffolds used in the experiments are shown above the gel images and fitted curves. TECs were pre-extended with either 25 μM NTP or 50 μM nucleoside analogue. The rate of second NTP (500 μM) addition was measured with HCl quench method. Error bars are ranges of duplicate measurements and the solid lines are the best-fits to a sum of exponential (corresponds to the fast phase) and stretched exponential (corresponds to the slow phase) functions. The rates and amplitudes obtained from the fits are shown in Table 5.

When EthS-SdmMP was at the RNA 3' end, the incorporation rate of the next nucleotide (GMP) proceeded largely in a slow monophasic reaction, where the rate was ~200-fold slower compared to when UMP was at the RNA 3'end. This explains why TECs paused so strongly at the T-tract when a large excess of EthS-SdmTP was added to the reaction. About 65% of TECs pre-extended with ForMP incorporated the next NMP (UMP) at a similar rate as TECs pre-extended with adenine, while the remaining 35% incorporated UMP more slowly (0.45 s-1). Similarly to ForMP, RNA extension was biphasic when NMP was incorporated after PyrMP. When PyrMP was at the RNA 3'end, there is 42-48% of a fast phase, indicating that the TECs are almost evenly divided between paused (backtracked or pre-translocated) and active (post-translocated) TECs at the start of the reaction. The fast phase is slowed down to less than half the rate (~23 s^{-1}) compared to when adenine is at the RNA 3'end (~55 s^{-1}).

We also measured GreA mediated cleavage of nascent RNA in TECs extended with nucleoside analogues to see if the RNAP can efficiently remove them from RNA after incorporation (Figure 25). Pre-extended TECs were gel filtrated to remove excess nucleotides and $MgCl_2$, after which RNA cleavage was induced by mixing the TECs with 2 μM GreA and 10 mM $MgCl_2$. We found that RNA in TECs pre-extended with EthS-SdmMP and GMP were cleaved ~40-fold faster than TECs pre-extended with UMP and GMP (median cleavage times 0.1 seconds and 39 seconds, respectively). RNA in TECs pre-extended with ForMP and UMP was cleaved 20-fold faster than RNA in TECs pre-extended with AMP and UMP (median cleavage times 3.1 s and 61 s, respectively). GreA cleavage of RNA in TECs pre-extended with PyrMP and CMP was cleaved 8-fold faster than TECs pre-extended with AMP and CMP (median cleavage times 21 seconds and 170 seconds, respectively). In short, all three tested nucleoside analogues (especially 4-ethylthioshowdomycin) were cleaved by GreA much more efficiently than cognate ribonucleotides from the penultimate position. These results indicate that the incorporated nucleoside analogues weaken the RNA-DNA interactions, which results in an increased fractional backtracking of TECs and more efficient proofreading of the incorporated nucleoside analogues.



	Median cleavage time (s)	Uncleaved RNA (%)
U-G	39 ± 18	8.3 ± 1.0
EthS-Sdm-G	0.11 ± 0.07	6.9 ± 0.1

	Median cleavage time (s)	Uncleaved RNA (%)
A-U	61 ± 12	4.9 ± 0.48
For-U	3.1 ± 0.74	2.8 ± 0.33

	Median cleavage time (s)	Uncleaved RNA (%)
A-C	170 ± 25	8.5 ± 0.87
Pyr-C	21 ± 0.92	4.2 ± 0.11

Figure 25. GreA cleavage of incorporated 4-ethylthioshowdomycin, formycin A and pyrazofurin A. TECs were pre-extended with one NTP (25 µM) or 5'-triphosphorylated C-nucleoside (100 µM) and with second NTP (25 µM) to RNA18 (UTP (25 µM) or EtS-STP (100 µM) and GTP (25 µM), ATP (25 µM) or ForTP (100 µM) and UTP (25 µM), or ATP (25 µM) to RNA18. Pre-extended TECs were gel filtrated to remove excess $MgCl_2$ and nucleotides. GreA assisted RNA cleavage reactions were initiated by mixing pre-extended TECs with 2 µM GreA and 10 mM $MgCl_2$. Error bars are ranges of duplicate measurements and the solid lines are the best-fits to a stretched exponential function.

5.3.6 Effects of triphosphorylated formycin A and 4-ethylthioshowdomycin on RNA elongation over an 800 bp long distance

The data presented above shows how the various nucleoside analogues are utilized as a substrate by *Eco* RNAP and how they affect RNA synthesis, in short and carefully designed nucleic acid scaffolds. This data gives some estimate of the inhibitory mechanism of the compounds, but the inhibitory potential on transcription as a whole cannot be observed with these scaffolds alone. To address how triphosphorylated formycin A and 4-ethylthioshowdomycin affect transcription over

longer templates, we utilized an 800 bp long linear template that encodes a fluorescent light-up RNA aptamer (FLAP), which forms a high-fluorescent complex with the fluorophore DFHBI-1T. This setup allows for the measurement of RNA synthesis by bacterial RNAP in real-time using a fluorometer (Huang *et al*., 2022), while the total amount of RNA at the end of the measurement can be estimated by quantifying the RNA from a polyacrylamide gel stained with SYBR® Gold (Figure 26).

RNA synthesis was initiated from a promoter with a holoenzyme (*Eco* RNAP and σ^{70}), after which the RNAPs transcribed through an 819 bp stretch of linear dsDNA. The template contained a Broccoli-FLAP encoding sequence (617-731 bp from the promoter), followed by a hairpin-dependent terminator sequence (759-794 bp from the promoter), and the end of the linear template at 819 bp from the promoter. Transcribing RNAPs either terminate at the terminator hairpin or run off the template, after which they can start transcribing the template again, making the transcription continuous. The fluorescence signal and RNA amount are affected by initiation and elongation, providing a more complete view of how transcription is affected by potential inhibitors *in vitro*. The relatively long transcribed distance also allows for the accumulation of potential pausing effects as the nucleoside analogues are incorporated. Formation of the highly-fluorescent Broccoli-FLAP was observed by measuring the increase in fluorescence over time at a single wavelength (507 nm).

We found that both formycin A and 4-ehtylthioshowdomycin decreased the fluorescence as their concentration was gradually increased. When the ratio of ForTP:NTPs and EthS-SdmTP:NTPs was 1:1, the fluorescence at the end of the measurement was less than the fluorescence observed in the control reaction by ~50 % and ~60 %, respectively. At the highest concentrations of ForTP and EthS-SdmTP (5 mM), fluorescence signals dropped down to ~10 % and ~5 %, respectively. However, when the amount of RNA was quantified from polyacrylamide gels stained with SYBR® Gold, we found that at high concentrations of ForTP the total amount of full length RNA was reduced to ~58 %, whereas EthS-SdmTP affected the amount of full length RNA much more strongly, reducing it to ~4 %, thus inhibiting RNA synthesis almost completely at 5 mM.

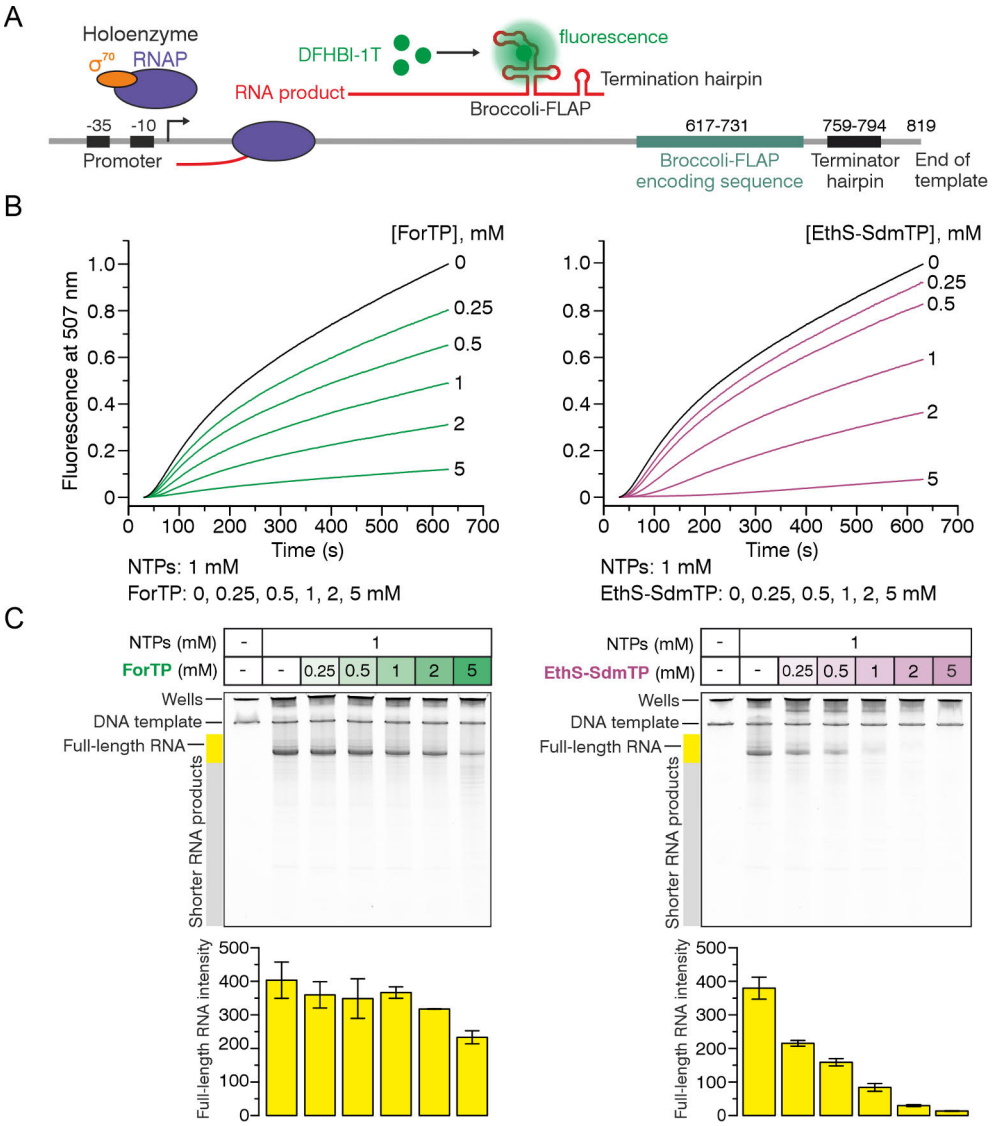


Figure 26. Effects of 5'-triphosphorylated formycin A and 4-ethylthioshowodmycin on RNA synthesis over 800 bp long template. (A) Schematic of the experimental setup. (B) Fluorescence traces of the Broccoli-FLAP formation normalized to the highest observed fluorescence of the control reactions (black curves). Different concentrations of ForTP (green curves) and EthS-SdmTP (purple curves). After 30 seconds from mixing the reactants the fluorescence signal was measured at 507 nm for 600 seconds. Fluorescence traces are averages of two independent experiments. (C) Quantification of RNA products after 700 seconds from mixing the reactants (30 seconds after the end of measurement). RNA products were separated on 6% TBE gel and stained with SYBR® Gold. Gels were scanned at 520 nm and the stained RNA was quantified with ImageJ. The relative intensities of the full-length RNA (yellow bar) are presented as columns. The standard deviations are derived from duplicate experiments.

Despite having similar effects on the observed fluorescence signals, the two nucleoside analogues had very different effects on the accumulation of RNA products. Since formation of the high-fluorescent Broccoli-FLAP is dependent on the 3D structure of the RNA aptamer, it is plausible that incorporated formycin A in RNA affects the hair-pin formation of the folding RNA, producing non-fluorescent RNA aptamers, which reduces the observed fluorescence signal while only modestly affecting the amount of synthesized RNA. This is likely because formycin A can hydrogen bond with uridine and guanine, thus altering the 3D structure of the synthesized RNA aptamer. In contrast, 4-ehtylthioshowdomycin reduces the fluorescence signal by strongly inhibiting RNA synthesis. These results agree with experiments where nucleoside analogues at the RNA 3' end affected further RNA elongation differently (Figure 24, Table 4): formycin A slowed down further elongation of RNA only slightly, whereas 4-ethylthioshowdomycin slowed down further RNA elongation by 200-fold. Over an 800 bp long transcribed sequence EthS-SdmTP would have a large effect on the accumulation of full-length RNA even if it is incorporated only a few times, whereas formycin A can be incorporated many times without majorly affecting the accumulation of full-length RNA. Overall, this experiment displays how the two nucleoside analogues can affect RNA synthesis in different ways.

6 Conclusions

In the first study we sought to answer a very fundamental question about transcription: why do multi-subunit RNAPs synthesize RNA instead of DNA, even though 2'deoxyribonucleotides are smaller in size than ribonucleotides? The solution turned out to be very simple and elegant, as by substituting the invariant arginine residue at the active site with other amino acids we could largely reduce the RNAPs capability to distinguish the two nucleotides from each other. Based on the kinetic analysis and X-ray crystal structures of nucleotides bound at the RNAP active site, we concluded that the invariant arginine traps 2'dNTPs by binding the 3'OH group and stabilizes the 2'endo conformation of the nucleosugar. These interactions limit 2'dNTP incorporation into RNA, explaining how the utilization of a smaller substrate can be disfavored by the multi-subunit RNAP.

In the second study we wanted to look at the transcriptional fidelity from another perspective: how does a multi-subunit RNAP correct the errors it makes during transcription? In this case the solution wasn't as simple, as many domains and clefts within the active site of multi-subunit RNAPs participate in the reaction. However, after screening a library of RNAP variants we made some novel findings that helped to elucidate the mechanism. First, the folding of a mobile domain TL improves proofreading in a 1-nt backtracked state, but it interferes with RNA cleavage at a 2-nt backtracked state, underlining the need for cleavage factors to facilitate an efficient pause release at longer backtracked states. Second, we found multiple variants that had large effects on RNA cleavage, which haven't been previously reported to affect the reaction. Specifically, mutations targeted at the first helical turn of the N-terminal helix of TL, the A-site and E-site affected RNA cleavage greatly at both the 1-nt and 2-nt backtracked states, indicating that they could be involved in RNA cleavage reaction regardless of the backtracking propensity. We argue that TL works as a positional catalyst, directing backtracked RNA into a conformation favorable for the catalysis of endonucleolytic RNA cleavage. Repositioning of the nucleotide at the substrate site is likely needed to allow RNA to adopt a cleavage proficient conformation, as mutations to the A-site substantially modulate the reaction.

In the third and fourth studies we wanted to expand beyond native substrates and multi-subunit RNAPs by studying nucleoside analogues of ribonucleotides with RNAPs from both the two-β-barrel and right-hand RNAP superfamilies. We studied how showdomycin derivatives, formycin A, pyrazofurin A, 8-oxoadenine, 8-oxoguanine and ribavirin are utilized in place of cognate ribonucleotides and found that different RNAPs recognized these nucleoside analogues as different substrates, and in many cases the nucleoside analogues worked as dual coders. Right-hand RNAPs utilized showdomycin derivatives poorly, whereas *Eco* RNAP and *Sce* RNAPII incorporated most of them, and after a more detailed analysis we found that 4-ethylthioshowdomycin greatly slowed down further RNA elongation when incorporated by *Eco* RNAP. *Eco* RNAP, *Hsa* MT RNAP and CVB3 RNAP showed varying preferences for pyrazofurin A, ribavirin, 8-oxoadenine and 8-oxoguanine. Out of the tested nucleoside analogues, the most fascinating was formycin A, which was incorporated efficiently as adenine by *Eco* RNAP, *Hsa* MT RNAP and CVB3 RNAP, and as cytidine by *Eco* RNAP and CVB3 RNAP. CVB3 RNAP misread formycin A incorporated into the RNA template, and formycin A also inhibited the formation of a fluorescent RNA aptamer during RNA synthesis by *Eco* RNAP, indicating that it has potential as a mutagenesis inducing nucleoside analogue for various RNAPs. However, chemical optimization is required to improve its selectivity before it can be used as a viable drug compound, as formycin A is known to be cytotoxic to mammalian cells.

Maintaining a moderate level of accuracy during RNA synthesis is needed for faithful utilization of genetic information in all organisms and reduced transcriptional fidelity has been linked to proteotoxic stress and genomic instability. As RNA synthesis in bacteria and viruses is a sought-after target for drug development, understanding the mechanisms of substrate selection in different RNAPs is essential for developing strategies for the optimization of drugs that efficiently but selectively target RNAPs of pathogens instead of host RNAPs. This work expands upon what is known about the mechanism of substrate selection and transcriptional proofreading, and it provides new insights into nucleoside analogues as potential inhibitors of bacterial and viral RNAPs.

Acknowledgements

This doctoral thesis work was done in the RNA polymerase group at the department of Life Technologies, Faculty of Technology, University of Turku. The work was conducted at Biocity, using the equipment and facilities provided by the Department of Life Technologies and Turku Protein Core unit. The work was financially supported by Academy of Finland, Sigrid Jusélius Foundation, Instrumentarium Science Foundation and the Doctoral Programme of Technology (formerly known as Doctoral Programme of Molecular Life Sciences).

The head of department, Prof. Jyrki Heino, is acknowledged for his leadership. I am deeply grateful to Prof. Lari Lehtiö and Asst. Prof. Tatiana Mishanina for reviewing my thesis and for providing their insight. I offer my humble thanks to Prof. Craig Martin for agreeing to be the opponent in my doctoral dissertation. I thank the members of my thesis advisory committee, Dr. Taina Tyystjärvi and Dr. Anton Zavialov, for the great advice and discussions during my thesis work.

I want to thank my supervisor, Assoc. Prof. Georgiy Belogurov, for the opportunity to participate in many exciting research projects over the years. Your genuine fascination with transcription has inspired me, and I thank you for guiding me at the start of my research career. I thank Dr. Matti Turtola for teaching me the many techniques I employed throughout this work, and I greatly appreciate the advice and example he has given me.

I'm grateful to our scientific collaborators for supporting this work. I thank Dr. Yeonoh Shin and Prof. Katsuhiko Murakami from the Department of Biochemistry and Molecular Biology of Pennsylvania State University, USA, who provided the crystal structures of 2'dCTP and 3'dCTP bound to the active site of *T. thermophilus* RNAP, which were used together with our enzyme kinetic data to verify the mechanism of 2'dNTP discrimination by multi-subunit RNAPs. I am deeply grateful to Dr. Petja Rosenqvist and Prof. Pasi Virta from the Department of Chemistry at the University of Turku, who provided the fascinating 5'-triphosphorylated C-nucleosides for the studies. I wish to thank Kaisa Palmu, Dr. Johanna Jokinen and Prof. Mikko Metsä-Ketelä from the Department of Life Technologies at the University of Turku, who studied the showdomycin derivatives *in vivo*.

I am grateful to Vilma Trapp, Emil Aalto-Setälä and Dr. Sari Paavilainen for the many tea/whisky sessions and fun conversations that made the days in the lab much more enjoyable. I also wish to thank many past co-workers from the RNA polymerase group for their assistance: Eeva Vieras, Oskari Puro, Juho Kotikoski, Ville Levola, Ana Miladinovic, Jelena Pavlovic, Aleksi Nuutila, Klara Kuret, Karina Šapovalovaite, Dr. Anssi Malinen, Dr. Thadée Grocholski, Henri Malmi, Verneri Nissilä, Dr. Ranjit Prajapati and Mila Vainonen.

I wish to thank the many people in the department of Life Technologies who strive to make the department a great workplace. I thank Dr. Jarmo Käpylä, Dr. Anu Salminen, Dr. Pekka Rappu, Dr. Jarmo Niemi and Dr. Jari Nuutila for their tireless efforts in providing me and so many others with the knowledge and motivation to pursue scientific careers. I thank our past and present technicians Jani Sointusalo, Antti Nuutinen, Jari Vehmas and Tapio Ronkainen for the invaluable technical support and equipment maintenance. I am grateful to our laboratory maintenance staff Anu Hirvensalo, Heli Kuusela, Hannele Heinonen and Merja Nylander for their help over the years. I thank the past graduate school coordinators Dr. Sari Järvi and Dr. Nina Lehtimäki for their work. I appreciate the work of our human resources staff Satu Jasu, Katrin Jokinen, Paula Luoma, Marianna Boundouvis and Sanna Laitinen.

I have been fortunate to have family and friends who supported me over the years. I greatly value the peer support and fun times I have shared with the Biostanists: my wife Satu, Dr. Milla Valta, Topi Laaksonen, Marjaana Parikainen, and Salli Talvi. I wish to thank my parents, Timo and Pirjo, my sister Susanna, and my late grandfather Seppo Kirves for their compassion and support over the years. A very special thanks to the late Salme Kirves who always believed in me even when most others wouldn't, and who strongly encouraged me to pursue a career in science. Warmhearted thanks to my mother-in-law Arja, who has been very supportive and a positive influence in my life. My daughter Taru is a boundless source of joy and inspiration for me, giving me the motivation to finalize this thesis. And finally, I wish to thank my beloved wife Satu Silojärvi, who has been with me throughout my career, always cheering me on during the good and bad times. Your love, encouragement and patience has given me the strength to carry on with my research.

30.11.2024

Janne Julius Mäkinen

References

Abbondanzieri EA, Greenleaf WJ, Shaevitz JW, Landick R, Block MS (2005). Direct observation of base-pair stepping by RNA polymerase. *Nature*. 438(7067):460–465.

Abdelkareem M, Saint-André C, Takacs M, Papai G, Crucifix C, Guo X, Ortiz J, Weixlbaumer A (2019). Structural Basis of Transcription: RNA Polymerase Backtracking and Its Reactivation. *Mol Cell*. 75(2):298–309.e4.

Abràmoff MD, Magalhaes PJ, Ram SJ (2004). Image processing with ImageJ. *Biophotonics Int*. 11:36–42.

Aizawa S, Hidaka T, Otake N, Yonehara H, Isono K, Igarashi N, Suzuki S (1965). Studies on a new antibiotic, Laurusin. *Agr Biol Chem*. 29:375–376.

André E, Bastide L, Villain-Guillot P, Latouche J, Rouby J, Leonetti J-P (2004). A multiwell assay to isolate compounds inhibiting the assembly of the prokaryotic RNA polymerase. *Assay Drug Dev Technol*. 2(6):629–635.

Anikin M, Molodtsov D., Temiakov D, McAllister WT (2010). Transcript slippage and recoding, in Recoding: Expansion of Decoding Rules Enriches Gene Expression (Atkins, J. F., and Gesteland, R. F., eds) pp. 409–434, Springer, New York.

Arnold JJ, Sharma SD, Feng JY, Ray AS, Smidansky ED, Kireeva ML, Cho A, Perry J, Vela JE, Park Y, Xu Y, Tian Y, Babusis D, Barauskus O, Peterson BR, Gnatt A, Kashlev M, Zhong W, Cameron CE (2012). Sensitivity of mitochondrial transcription and resistance of RNA polymerase II dependent nuclear transcription to antiviral ribonucleosides. *PLoS Pathog*. 8(11):e1003030.

Artsimovitch I, Svetlov V, Nemetski SM, Epshtein V, Cardozo T, Nudler E (2011). Tagetitoxin inhibits RNA polymerase through trapping of the trigger loop. *J Biol Chem*. 286(46):40395–40400.

Auer GK, Weibel DB (2017). Bacterial Cell Mechanics. *Biochemistry*. 56(29):3710–3724.

Audia JP, Winkler HH (2006). Study of the five Rickettsia prowazekii proteins annotated as ATP/ADP translocases (Tlc): Only Tlc1 transports ATP/ADP, while Tlc4 and Tlc5 transport other ribonucleotides. *J Bacteriol*. 188(17):6261–6268.

Bai L, Shundrovsky A, Wang MD (2004). Sequence-dependent kinetic model for transcription elongation by RNA polymerase. *J Mol Biol*. 344(2):335–349.

Balzarini J, Naesens L, De Clercq E (1998). New antivirals - mechanism of action and resistance development. *Curr Opin Microbiol*. 1(5):535–546.

Bar-Nahum G, Epshtein V, Ruckenstein AE, Rafikov R, Mustaev A, Nudler E (2005). A ratchet mechanism of transcription elongation and its control. *Cell*. 120(2):183–193.

Basu RS, Warner BA, Molodtsov V, Pupov D, Esyunina D, Fernández-Tornero C, Kulbachinskiy A, Murakami KS (2014) Structural basis of transcription initiation by bacterial RNA polymerase holoenzyme. *J Biol Chem*. 289(35):24549–24559.

Batra VK, Shock DD, Beard WA, McKenna CE, Wilson SH (2012). Binary complex crystal structure of DNA polymerase β reveals multiple conformations of the templating 8-oxoguanine lesion. *Proc Natl Acad Sci USA*, 109, 113–118.

Belogurov GA, Vassylyeva MN, Svetlov V, Klyuyev S, Grishin NV, Vassylyev DG, Artsimovitch I (2007). Structural basis for converting a general transcription factor into an operon-specific virulence regulator. *Mol Cell*. 26(1):117–129.

Belogurov GA, Vassylyeva MN, Sevostyanova A, Appleman JR, Xiang AX, Lira R, Webber SE, Klyuyev S, Nudler E, Artsimovitch I, Vassylyev DG (2009). Transcription inactivation through local refolding of the RNA polymerase structure. *Nature*. 457(7227):332–335.

Bochkareva A, Yuzenkova Y, Tadigotla VR, Zenkin N (2012). Factor-independent transcription pausing caused by recognition of the RNA–DNA hybrid sequence. *The EMBO Journal*. 31(3):630–639.

Borukhov S, Sagitov V, Goldfarb A (1993). Transcript cleavage factors from E. coli. *Cell*. 72(3):459–466.

Bradley CC, Gordon AJE, Halliday JA, Herman C (2019). Transcription fidelity: New paradigms in epigenetic inheritance, genome instability and disease. *DNA Repair (Amst)*. 81:102652.

Brieba LG, Eichman BF, Kokoska RJ, Doublié S, Kunkel TA, Ellenberger T (2004). Structural basis for the dual coding potential of 8-oxoguanosine by a high-fidelity DNA polymerase. *EMBO J.*, 23(17), 3452–3461.

Brown JA, Suo Z (2011). Unlocking the sugar "steric gate" of DNA polymerases. *Biochemistry*. 50(7):1135–1142.

Bubunenko MG, Court CB, Rattray AJ, Gotte DR, Kireeva ML, Irizarry-Caro JA, Li X, Jin DJ, Court DL, Strathern JN, Kashlev M (2017). A Cre Transcription Fidelity Reporter Identifies GreA as a Major RNA Proofreading Factor in Escherichia coli. *Genetics*. 206(1):179–187.

Börner T, Aleynikova AY, Zubo YO, Kusnetsov VV (2015). Chloroplast RNA polymerases: Role in chloroplast biogenesis. *Biochim Biophys Acta*. 1847(9):761–769.

Campagnola G, McDonald S, Beaucourt S, Vignuzzi M, Peersen OB (2015). Structure-function relationships underlying the replication fidelity of viral RNA-dependent RNA polymerases. *J Virol*. 89(1):275–286.

Campbell EA, Korzheva N, Mustaev A, Murakami K, Nair S, Goldfarb A, Darst SA (2001). Structural mechanism for rifampicin inhibition of bacterial RNA polymerase. *Cell*. 104(6):901–912.

Campbell EA, Pavlova O, Zenkin N, Leon F, Irschik H, Jansen R, Severinov K, Darst SA (2005). Structural, functional, and genetic analysis of sorangicin inhibition of bacterial RNA polymerase. *EMBO J*. 24(4):674–682.

Carey LB (2015). RNA polymerase errors cause splicing defects and can be regulated by differential expression of RNA polymerase subunits. *Elife*.4:e09945.

Carrano L, Bucci C, De Pascalis R, Lavitola A, Manna F, Corti E, Bruni CB, Alifano P (1998). Effects of bicyclomycin on RNA- and ATP-binding activities of transcription termination factor Rho. *Antimicrob Agents Chemother*. 42(3):571–578.

Castro C, Smidansky E, Maksimchuk KR, Arnold JJ, Korneeva VS, Götte M, Konigsberg W, Cameron CE (2007). Two proton transfers in the transition state for nucleotidyl transfer catalyzed by RNA- and DNA-dependent RNA and DNA polymerases. *Proc Natl Acad Sci U S A*. 104(11):4267–4272.

Castro C, Smidansky ED, Arnold JJ, Maksimchuk KR, Moustafa I, Uchida A, Götte M, Konigsberg W, Cameron CE (2009). Nucleic acid polymerases use a general acid for nucleotidyl transfer. *Nat Struct Mol Biol*. 16(2):212–218.

Cermakian N, Ikeda TM, Cedergren R, Gray MW (1996). Sequences homologous to yeast mitochondrial and bacteriophage T3 and T7 RNA polymerases are widespread throughout the eukaryotic lineage. *Nucleic Acids Res*. 24(4):648–654.

Cheetham GM, Jeruzalmi D, Steitz TA (1999). Structural basis for initiation of transcription from an RNA polymerase-promoter complex. *Nature*. 399(6731):80–83.

Chen FX, Smith ER, Shilatifard A (2018). Born to run: control of transcription elongation by RNA polymerase II. *Nat Rev Mol Cell Biol*. 19(7):464–478.

Cheung ACM, Cramer P (2011). Structural basis of RNA polymerase II backtracking, arrest and reactivation. *Nature*, 471:249–253.

Cheung PP, Jiang B, Booth GT, Chong TH, Unarta IC, Wang Y, Suarez GD, Wang J, Lis JT, Huang X (2020). Identifying Transcription Error-Enriched Genomic Loci Using Nuclear Run-on Circular-Sequencing Coupled with Background Error Modeling. *J Mol Biol*. 432(13):3933–3949.

Chung C, Verheijen BM, Zhang X, Huang B, Coakley A, McGann E, Wade E, Dinep-Schneider O, LaGosh J, Anagnostou M-E, Simpson S, Thomas K, Ernst M, Rattray A, Lynch M, Kashlev M, Benayoun BA, Li Z, Strathern J, Gout J-F, Vermulst M (2023). The fidelity of transcription in human cells. *Proc Natl Acad Sci U S A*. 120(5):e2210038120.

Cramer P, Bushnell DA, Kornberg RD (2001). Structural basis of transcription: RNA polymerase II at 2.8 angstrom resolution. *Science*. 292(5523):1863–1876.

Crotty S, Maag D, Arnold JJ, Zhong W, Lau JY, Hong Z, Andino R, Cameron CE (2000). The broad-spectrum antiviral ribonucleoside ribavirin is an RNA virus mutagen. *Nat. Med.*, 6(12), 1375–1379.

Crotty S, Cameron C, Andino R (2002). Ribavirin's antiviral mechanism of action: lethal mutagenesis? *J Mol Med (Berl)*. 80(2):86–95.

De Clercq E, Balzarini J, Madej D, Hansske F, Robins MJ (1987). Nucleic acid related compounds. 51. Synthesis and biological properties of sugar-modified analogues of the nucleoside antibiotics tubercidin, toyocamycin, sangivamycin, and formycin. *J Med Chem*. 30(3):481–486.

De Clercq E (2016). C-Nucleosides To Be Revisited. *J Med Chem*. 59(6):2301–2311.

De Clercq E, Li G (2016). Approved Antiviral Drugs over the Past 50 Years. *Clin Microbiol Rev*. 29(3):695–747.

Degen D, Feng Y, Zhang Y, Ebright KY, Ebright YW, Gigliotti M, Vahedian-Movahed H, Mandal S, Talaue M, Connell N, Arnold E, Fenical W, Ebright RH (2014). Transcription inhibition by the depsipeptide antibiotic salinamide A. *Elife*. 3:e02451.

Delarue M, Poch O, Tordo N, Moras D, Argos P (1990). An attempt to unify the structure of polymerases. *Protein Eng*. 3(6):461–467.

Doublié S, Tabor S, Long AM, Richardson CC, Ellenberger T (1998). Crystal structure of a bacteriophage T7 DNA replication complex at 2.2 A resolution. *Nature*. 391(6664):251–258.

Dundr M, Hoffmann-Rohrer U, Hu Q, Grummt I, Rothblum LI, Phair RD, Misteli T (2002). A kinetic framework for a mammalian RNA polymerase in vivo. *Science*. 298(5598):1623–1626.

Dutta D, Shatalin K, Epshtein V, Gottesman ME, Nudler E (2011). Linking RNA polymerase backtracking to genome instability in E. coli. *Cell*. 146(4): 533–543.

Eastman RT, Roth JS, Brimacombe KR, Simeonov A, Shen M, Patnaik S, Hall MD (2020). Remdesivir: A Review of Its Discovery and Development Leading to Emergency Use Authorization for Treatment of COVID-19. *ACS Cent Sci*. 6(5):672–683.

Epshtein V, Nudler E, (2003). Cooperation between RNA polymerase molecules in transcription elongation. *Science*. 300(5620):801–805.

Esyunina D, Turtola M, Pupov D, Bass I, Klimašauskas S, Belogurov GA, Kulbachinskiy A (2016). Lineage-specific variations in the trigger loop modulate RNA proofreading by bacterial RNA polymerases. *Nucleic Acids Res*. 44(3):1298–1308.

Feng JY, Xu Y, Barauskas O, Perry JK, Ahmadyar S, Stepan G, Yu H, Babusis D, Park Y, McCutcheon K, Perron M, Schultz BE, Sakowicz R, Ray AS (2016). Role of Mitochondrial RNA Polymerase in the Toxicity of Nucleotide Inhibitors of Hepatitis C Virus. *Antimicrob Agents Chemother*. 60(2):806–817.

Fernandez-Leiro R, Conrad J, Yang J-C, Freund SMV, Scheres SHW, Lamers MH (2017). Self-correcting mismatches during high-fidelity DNA replication. *Nat Struct Mol Biol*. 24(2):140–143.

Ferrero D, Ferrer-Orta C, Verdaguer N (2018). Viral RNA-Dependent RNA Polymerases: A Structural Overview. *Subcell Biochem*. 88:39–71.

Ferrer-Orta C, Arias A, Escarmís C, Verdaguer N (2006). A comparison of viral RNA-dependent RNA polymerases. *Curr Opin Struct Biol*. 16(1):27–34.

Fouqueau T, Zeller ME, Cheung AC, Cramer P, Thomm M (2013). The RNA polymerase trigger loop functions in all three phases of the transcription cycle. *Nucleic Acids Res.* 41 (14), 7048–7059.

Furman R, Tsodikov OV, Wolf YI, Artsimovitch I (2013). An insertion in the catalytic trigger loop gates the secondary channel of RNA polymerase. *J Mol Biol*. 425(1):82–93.

Gabizon R, Lee A, Vahedian-Movahed H, Ebright RH, Bustamante CJ (2018). Pause sequences facilitate entry into long-lived paused states by reducing RNA polymerase transcription rates. *Nat Commun*. 9(1):2930.

Gaillard H, García-Muse T, Aguilera A (2015). Replication stress and cancer. *Nat Rev Cancer*. 15(5):276–289.

Gajos M, Jasnovidova O, van Bömmel A, Freier S, Vingron M, Mayer A (2021). Conserved DNA sequence features underlie pervasive RNA polymerase pausing. *Nucleic Acids Res*. 49(8):4402–4420.

Galburt EA, Grill SW, Wiedmann A, Lubkowska L, Choy J, Nogales E, Kashlev M, Bustamante C (2007). Backtracking determines the force sensitivity of RNAPII in a factor-dependent manner. *Nature*. 446(7137):820–823.

Gnatt, AL, Cramer P, Fu J, Bushnell DA, Kornberg RD (2001). Structural basis of transcription: an RNA polymerase II elongation complex at 3.3 A resolution. *Science*. 292(5523):1876-1882.

Gohara DW, Arnold JJ, Cameron CE (2004). Poliovirus RNA-dependent RNA polymerase (3Dpol): kinetic, thermodynamic, and structural analysis of ribonucleotide selection. *Biochemistry*. 43(18):5149–5158.

Gollnest T, de Oliveira TD, Schols D, Balzarini J, Meier C (2015). Lipophilic prodrugs of nucleoside triphosphates as biochemical probes and potential antivirals. *Nat Commun*. 6:8716.

Gout J-F, Thomas WK, Smith Z, Okamoto K, Lynch M (2013). Large-scale detection of in vivo transcription errors. *Proc Natl Acad Sci U S A*, 110(46):18584–18589.

Gout J-F, Li W, Fritsch C, Li A, Haroon S, Singh L, Hua D, Fazelinia H, Smith Z, Seeholzer S, Thomas K, Lynch M, Vermulst M (2017). The landscape of transcription errors in eukaryotic cells. *Sci Adv*. 3(10):e1701484.

Guillerez J, Lopez JP, Proux F, Launay H, Dreyfus M (2005). A mutation in T7 RNA polymerase that facilitates promoter clearance. *Proc Natl Acad Sci U S A*, 102(17):5958-5963.

Gupta PK, Daunert S, Nassiri MR, Wotring LL, Drach JC, Townsend LB (1989). Synthesis, cytotoxicity, and antiviral activity of some acyclic analogues of the pyrrolo[2,3-d]pyrimidine nucleoside antibiotics tubercidin, toyocamycin, and sangivamycin. *J Med Chem*. 32(2):402–408.

Gutowski GE, Sweeney MJ, DeLong DC, Hamill RL, Gerzon K, Dyke RW (1975). Biochemistry and biological effects of the pyrazofurins (pyrazomycins): initial clinical trial. *Ann N Y Acad Sci* 255:544 –551.

Haberle V, Stark A (2018). Eukaryotic core promoters and the functional basis of transcription initiation. *Nat Rev Mol Cell Biol.* 19(10):621–637.

Haferkamp I, Schmitz-Esser S, Wagner M, Neigel N, Horn M, Neuhaus HE (2006). Tapping the nucleotide pool of the host: novel nucleotide carrier proteins of Protochlamydia amoebophila. *Mol Microbiol*. 60(6):1534–1545.

Heikinheimo P, Pohjanjoki P, Helminen A, Tasanen M, Cooperman BS, Goldman A, Baykov A, Lahti R (1996). A site-directed mutagenesis study of Saccharomyces cerevisiae pyrophosphatase. Functional conservation of the active site of soluble inorganic pyrophosphatases. *Eur J Biochem*. 239(1):138–143.

Hein PP, Palangat M, Landick R (2011). RNA transcript 3'-proximal sequence affects translocation bias of RNA polymerase. *Biochemistry*. 50(32):7002–7014.

Hirao I, Kimoto M, Mitsui T, Fujiwara T, Kawai R, Sato A, Harada Y, Yokoyama S (2006). An unnatural hydrophobic base pair system: site-specific incorporation of nucleotide analogues into DNA and RNA. *Nat Methods*. 3(9):729–735.

Hillen HS, Morozov YI, Sarfallah A, Temiakov D, Cramer P (2017). Structural Basis of Mitochondrial Transcription Initiation. *Cell*. 171(5):1072–1081.

Holmes SF, Santangelo TJ, Cunningham CK, Roberts JW, Erie DA (2006). Kinetic investigation of Escherichia coli RNA polymerase mutants that influence nucleotide discrimination and transcription fidelity. *J. Biol. Chem.* 281(27), 18677–18683.

Hori M, Takita T, Koyama G, Tadeuchi T, Umezawa H (1964). A new antibiotic, formycin. *J Antibiot (Tokyo)*. 17:96–99.

Huang Y, Eckstein F, Padilla R, Sousa R (1997). Mechanism of ribose 2'-group discrimination by an RNA polymerase. *Biochemistry*. 36(27):8231–8242.

Huang J, Brieba LG, 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(38):11571–11580.

Huang Y, Trapp V, Puro O, Mäkinen JJ, Metsä-Ketelä M, Wahl MC, Belogurov GA (2022). Fluorogenic RNA aptamers to probe transcription initiation and co-transcriptional RNA folding by multi-subunit RNA polymerases. *Methods Enzymol*. 675:207–233.

Imashimizu M, Oshima T, Lubkowska L, Kashlev M (2013). Direct assessment of transcription fidelity by high-resolution RNA sequencing. *Nucleic Acids Res*. 41(19):9090–9104.

Imashimizu M, Takahashi H, Oshima T, McIntosh C, Bubunenko M, Court DL, Kashlev M (2015). Visualizing translocation dynamics and nascent transcript errors in paused RNA polymerases in vivo. *Genome Biol*. 16(1):98.

Ishizuka M, Sawa T, Hori S, Takayama H, Takeuchi T (1968). Biological studies on formycin and formycin B. *J Antibiot (Tokyo)*. 21:5–12.

Isken O, Maquat LE (2007). Quality control of eukaryotic mRNA: safeguarding cells from abnormal mRNA function. *Genes & Development*, 21:1833–1856.

Islam S, Kjällquist U, Moliner A, Zajac P, Fan J-B, Lönnerberg P, Linnarsson S (2011). Characterization of the single-cell transcriptional landscape by highly multiplex RNA-seq. *Genome Res*. 21(7):1160–1167.

James K, Gamba P, Cockell SJ, Zenkin N (2017). Misincorporation by RNA polymerase is a major source of transcription pausing in vivo. *Nucleic Acids Res*. 45(3):1105–1113.

Jin J, Bai L, Johnson DS, Fulbright RM, Kireeva ML, Kashlev M, Wang MD (2010). Synergistic action of RNA polymerases in overcoming the nucleosomal barrier. *Nat Struct Mol Biol*. 17(6):745–752.

Johnson KA (2009). Fitting enzyme kinetic data with KinTek Global Kinetic Explorer. *Methods Enzymol.*, 467, 601–626.

Hein PP, Palangat M, Landick R (2011). RNA transcript 3'-proximal sequence affects translocation bias of RNA polymerase. *Biochemistry*. 50(32):7002–7014.

Ka Man Tse C, Xu J, Xu L, Sheong FK, Wang S, Chow HY, Gao X, Li X, Cheung PP, Wang D, Zhang Y, Huang X (2019). Intrinsic Cleavage of RNA Polymerase II Adopts a Nucleobase-independent Mechanism Assisted by Transcript Phosphate. *Nat Energy.* 2(3):228–235.

Kamiya H, Miura H, Murata-Kamiya N, Ishikawa H, Sakaguchi T, Inoue H, Sasaki T, Masutani C, Hanaoka F, Nishimura S (1995). 8-Hydroxyadenine (7,8-dihydro-8-oxoadenine) induces misincorporation in in vitro DNA synthesis and mutations in NIH 3T3 cells. *Nucleic Acids Res.*, 23(15), 2893–2899.

Kang JY, Olinares PDB, Chen J, Campbell EA, Mustaev A, Chait BT, Gottesman ME, Darst SA (2017). Structural basis of transcription arrest by coliphage HK022 Nun in an Escherichia coli RNA polymerase elongation complex. *Elife*. 6:e25478.

Kang JY, Mishanina TV, Bellecourt MJ, Mooney RA, Darst SA, Landick R (2018). RNA Polymerase Accommodates a Pause RNA Hairpin by Global Conformational Rearrangements that Prolong Pausing. *Mol Cell*. 69(5):802–815.

Kang JY, Mishanina TV, Bao Y, Chen J, Llewellyn E, Liu J, Darst SA, Landick R (2023). An ensemble of interconverting conformations of the elemental paused transcription complex creates regulatory options. *Proc Natl Acad Sci U S A*. 120(8):e2215945120.

Kaplan CD, Larsson K-M, Kornberg RD (2008). The RNA polymerase II trigger loop functions in substrate selection and is directly targeted by alpha-amanitin. *Mol Cell*. 30(5):547–556.

Kashkina E, Anikin M, Brueckner F, Pomerantz RT, McAllister WT, Cramer P, Temiakov D (2006). Template misalignment in multisubunit RNA polymerases and transcription fidelity. *Mol Cell*. 24(2):257–266.

Kiefer JR, Mao C, Braman JC, Beese LS (1998). Visualizing DNA replication in a catalytically active Bacillus DNA polymerase crystal. *Nature*. 391(6664):304–307.

Kireeva ML, Nedialkov YA, Cremona GH, Purtov YA, Lubkowska L, Malagon F, Burton ZF, Strathern JN, Kashlev M (2008). Transient reversal of RNA polymerase II active site closing controls fidelity of transcription elongation. *Mol Cell*. 30(5):557–566.

Koag M-C, Jung H, Lee S (2019). Mutagenic Replication of the Major Oxidative Adenine Lesion 7,8-Dihydro-8-oxoadenine by Human DNA Polymerases. *J. Am. Chem. Soc.*, 141(11), 4584–4596.

Komatsu Y, Tanaka K (1968). Mechanism of Action of Showdomycin: Part I. Effect of Showdomycin on the Syntheses of Nucleic Acids and Proteins in Escherichia coli K-12. *Agricultural and Biological Chemistry*. 32(8):1021–1027.

Komissarova N, Kashlev M (1997a). RNA polymerase switches between inactivated and activated states by translocating back and forth along the DNA and the RNA. *J Biol Chem*. 272(24):15329–15338.

Komissarova N, Kashlev M (1997b). Transcriptional arrest: Escherichia coli RNA polymerase translocates backward, leaving the 3' end of the RNA intact and extruded. *Proc Natl Acad Sci U S A*. 94(5):1755–1760.

Komissarova N, Kireeva ML, Becker J, Sidorenkov I, Kashlev M (2003). Engineering of elongation complexes of bacterial and yeast RNA polymerases. *Methods Enzymol*. 371:233–251.

Kotsantis P, Silva LM, Irmscher S, Jones RM, Folkes L, Gromak N, Petermann E (2016). Increased global transcription activity as a mechanism of replication stress in cancer. *Nat Commun.* 7:13087.

Kukko E, Heinonen J (1982). The intracellular concentration of pyrophosphate in the batch culture of Escherichia coli. *Eur J Biochem*. 127(2):347–349.

Kukko-Kalske E, Lintunen M, Inen MK, Lahti R, Heinonen J (1989). Intracellular PPi concentration is not directly dependent on amount of inorganic pyrophosphatase in Escherichia coli K-12 cells. *J Bacteriol*. 171(8):4498–4500.

Landick R (2001). RNA polymerase clamps down. *Cell*. 105(5):567–570.

Lang GI, Murray AW (2008). Estimating the per-base-pair mutation rate in the yeast Saccharomyces cerevisiae. *Genetics*. 178(1):67–82.

Lange U, Hausner W (2004). Transcriptional fidelity and proofreading in Archaea and implications for the mechanism of TFS-induced RNA cleavage. *Mol Microbiol*. 52(4):1133–1143.

Larson MH, Zhou J, Kaplan CD, Palangat M, Kornberg RD, Landick R, Block SM (2012). Trigger loop dynamics mediate the balance between the transcriptional fidelity and speed of RNA polymerase II. *Proc Natl Acad Sci U S A*. 109(17):6555–6560.

Larson MH, Mooney RA, Peters JM, Windgassen T, Nayak D, Gross CA, Block SM, Greenleaf WJ, Landick R, Weissman JS (2014). A pause sequence enriched at translation start sites drives transcription dynamics in vivo. *Science*. 344(6187):1042–1047.

Lassila JK, Zalatan JG, Herschlag D (2011). Biological phosphoryl-transfer reactions: understanding mechanism and catalysis. *Annu Rev Biochem*. 80:669–702.

Le TT, Yang Y, Tan C, Suhanovsky MM, Fulbright RM Jr., Inman JT, Li M, Lee J, Perelman S, Roberts JW, Deaconescu AM, Wang MD (2018). Mfd Dynamically Regulates Transcription via a Release and Catch-Up Mechanism. *Cell*. 172(1-2):344–357.e15.

van Leeuwen FW, de Kleijn DP, van den Hurk HH, Neubauer A, Sonnemans MA, Sluijs JA, Köycü S, Ramdjielal RD, Salehi A, Martens GJ, Grosveld FG, Peter J, Burbach H, Hol EM (1998). Frameshift mutants of beta amyloid precursor protein and ubiquitin-B in Alzheimer's and Down patients. *Science*. 279(5348):242–247.

Li W, Lynch M (2020). Universally high transcript error rates in bacteria. *eLife*. 2020; 9: e54898.

Lin W, Mandal S, Degen D, Liu Y, Ebright YW, Li S, Feng Y, Zhang Y, Mandal S, Jiang Y, Liu S, Gigliotti M, Talaue M, Connell N, Das K, Arnold E, Ebright RH (2017). Structural Basis of Mycobacterium tuberculosis Transcription and Transcription Inhibition. *Mol Cell*. 66(2):169–179.e8.

Liu B, Zuo Y, Steitz TA (2016). Structures of E. coli σS-transcription initiation complexes provide new insights into polymerase mechanism. *Proc Natl Acad Sci U S A*. 113(15):4051–4056.

Lynch M (2010). Evolution of the mutation rate. *Trends in Genetics*, 26:345–352.

Ma C, Yang X, Kandemir H, Mielczarek M, Johnston EB, Griffith R, Kumar N, Lewis PJ (2013). Inhibitors of bacterial transcription initiation complex formation. *ACS Chem Biol*. 8(9):1972–1980.

Ma C, Yang X, Lewis PJ (2016a). Bacterial Transcription Inhibitor of RNA Polymerase Holoenzyme Formation by Structure-Based Drug Design: From in Silico Screening to Validation. *ACS Infect Dis*. 2(1):39–46.

Ma C, Yang X, Lewis PJ (2016b). Bacterial Transcription as a Target for Antibacterial Drug Development. *Microbiol Mol Biol Rev*. 80(1):139–160.

Maffioli SI, Zhang Y, Degen D, Carzaniga T, Gatto GD, Serina S, Monciardini P, Mazzetti C, Guglierame P, Candiani G, Chiriac AI, Facchetti G, Kaltofen P, Sahl H-G, Dehò G, Donadio S, Ebright RH (2017). Antibacterial Nucleoside-Analogue Inhibitor of Bacterial RNA Polymerase. *Cell*. 169(7):1240–1248.

Makeyev EV, Bamford DH (2002). Cellular RNA-dependent RNA polymerase involved in posttranscriptional gene silencing has two distinct activity modes. *Mol Cell*. 2002. 10(6):1417–1427.

Malagon F, Kireeva ML, Shafer BK, Lubkowska L, Kashlev M, Strathern JN (2006). Mutations in the Saccharomyces cerevisiae RPB1 gene conferring hypersensitivity to 6-azauracil. *Genetics*. 172(4):2201–2209.

Malinen AM, Turtola M, Parthiban M, Vainonen L, Johnson MS, Belogurov GA (2012). Active site opening and closure control translocation of multisubunit RNA polymerase. *Nucleic Acids Res*. 40: 7442–7451.

Malinen AM, Nandymazumdar M, Turtola M, Malmi H, Grocholski T, Artsimovitch I, Belogurov GA (2014). CBR antimicrobials alter coupling between the bridge helix and the β subunit in RNA polymerase. *Nat Commun*. 5:3408.

Marr MT, Roberts JW (1997). Promoter recognition as measured by binding of polymerase to nontemplate strand oligonucleotide. *Science*. 276(5316):1258–1260.

Marietta & Brooks (2007). Transcriptional bypass of bulky DNA lesions causes new mutant RNA transcripts in human cells. *EMBO Rep*. 8:388–393.

Matsuura S, Shiratori O, Katagiri K (1964). Antitumor activity of showdomycin. *J Antibiot (Tokyo)*. 17:234–237.

Mayer A, Landry HM, Churchman LS (2017). Pause & go: from the discovery of RNA polymerase pausing to its functional implications. *Curr Opin Cell Biol*. 46:72–80.

Mehellou Y, Rattan HS, Balzarini J (2018). The ProTide Prodrug Technology: From the Concept to the Clinic. *J Med Chem*. 61(6):2211–2226.

Miropolskaya N, Esyunina D, Kulbachinskiy A (2017). Conserved functions of the trigger loop and Gre factors in RNA cleavage by bacterial RNA polymerases. *Biol Chem*. 292(16):6744–6752.

Mishanina TV, Palo MZ, Nayak D, Mooney RA, Landick R (2017). Trigger loop of RNA polymerase is a positional, not acid-base, catalyst for both transcription and proofreading. *Proc Natl Acad Sci U S A*. 114(26):E5103–E5112.

Molodtsov V, Fleming PR, Eyermann CJ, Ferguson AD, Foulk MA, McKinney DC, Masse CE, Buurman ET, Murakami KS (2015). X-ray crystal structures of Escherichia coli RNA polymerase with switch region binding inhibitors enable rational design of squaramides with an improved fraction unbound to human plasma protein. *J Med Chem*. 58(7):3156–3171.

Mosaei H, Molodtsov V, Kepplinger B, Harbottle J, Moon CW, Jeeves RE, Ceccaroni L, Shin Y, Morton-Laing S, Marrs ECL, Wills C, Clegg W, Yuzenkova Y, Perry JD, Bacon J, Errington J, Allenby NEE, Hall MJ, Murakami KS, Zenkin N (2018). Mode of Action of Kanglemycin A, an Ansamycin Natural Product that Is Active against Rifampicin-Resistant Mycobacterium tuberculosis. *Mol Cell*. 72(2):263–274.e5

Mosaei H, Zenkin N (2021). Two distinct pathways of RNA polymerase backtracking determine the requirement for the Trigger Loop during RNA hydrolysis. Nucleic Acids Res. 49(15):8777–8784.

Mukhopadhyay J, Sineva E, Knight J, Levy RM, Ebright RH (2004). Antibacterial peptide microcin J25 inhibits transcription by binding within and obstructing the RNA polymerase secondary channel. *Mol Cell*. 14(6):739–751.

Mukhopadhyay J, Das K, Ismail S, Koppstein D, Jang M, Hudson B, Sarafianos S, Tuske S, Patel J, Jansen R, Irschik H, Arnold E, Ebright RH (2008). The RNA polymerase "switch region" is a target for inhibitors. *Cell*. 135(2):295-307.

Mustaev A, Kozlov M, Markovtsov V, Zaychikov E, Denissova L, Goldfarb A (1997). Modular organization of the catalytic center of RNA polymerase. *Proc Natl Acad Sci U S A*. 94(13):6641–6645.

Mäkinen JJ, Shin Y, Vieras E, Virta P, Metsä-Ketelä M, Murakami KS, Belogurov GA (2021). The mechanism of the nucleo-sugar selection by multi-subunit RNA polymerases. *Nat Commun* 12(1):796.

Nakamura T, Zhao Y, Yamagata Y, Hua Y, Yang W (2012). Watching DNA polymerase η make a phosphodiester bond. *Nature*. 487(7406):196–201.

Nedialkov YA, Nudler E, Burton ZF (2012). RNA polymerase stalls in a post-translocated register and can hyper-translocate. *Transcription*. 3(5):260–269.

Nedialkov YA, Opron K, Assaf F, Artsimovitch I, Kireeva ML, Kashlev M, Cukier RI, Nudler E, Burton ZF (2013). The RNA polymerase bridge helix YFI motif in catalysis, fidelity and translocation. *Biochim Biophys Acta*. 1829(2):187–198.

Nishimura H, Mayama M, Komatsu Y, Kato H, Shimaoka N, Tanaka Y (1964). Showdomycin, a new antibiotic from a streptomyces SP. *J Antibiot (Tokyo)*. 17:148–155.

Nishimura H, Komatsu Y (1968). Reversal of inhibiting action of showdomycin on the proliferation of Escherichia coli by nucleosides and thiol compounds. *J Antibiot (Tokyo)*. 21(4):250–254.

Nudler E, Mustaev A, Lukhtanov E, Goldfarb A (1997). The RNA-DNA hybrid maintains the register of transcription by preventing backtracking of RNA polymerase. *Cell*. 89(1):33–41.

Nudler E (2012). RNA polymerase backtracking in gene regulation and genome instability. *Cell*. 149(7):1438–1445.

Orlova M, Newlands J, Das A, Goldfarb A, Borukhov S (1995). Intrinsic transcript cleavage activity of RNA polymerase. *Proc Natl Acad Sci U S A*. 92(10):4596–4600.

Ortega-Prieto AM, Sheldon J, Grande-Pérez A, Tejero H, Gregori J, Quer J, Esteban JI, Domingo E, Perales C (2013). Extinction of hepatitis C virus by ribavirin in hepatoma cells involves lethal mutagenesis. *PLoS One*. 8(8):e71039.

Outten CE, O'Halloran TV (2001). Femtomolar sensitivity of metalloregulatory proteins controlling zinc homeostasis. *Science*, 292, 2488–2492.

Palo MZ, Zhu J, Mishanina TV, Landick R (2020). Conserved Trigger Loop Histidine of RNA Polymerase II Functions as a Positional Catalyst Primarily through Steric Effects. *Biochemistry*. 60(44):3323–3336.

Pelechano V, Chávez S, Pérez-Ortín JE (2010). A complete set of nascent transcription rates for yeast genes. *PLoS One*. 5(11):e15442.

Perreault J, Weinberg Z, Roth A, Popescu O, Chartrand P, Ferbeyre G, Breaker RR (2011). Identification of hammerhead ribozymes in all domains of life reveals novel structural variations. *PLoS Comput. Biol*. 7, e1002031.

Petryk N, Kahli M, d'Aubenton-Carafa Y, Jaszczyszyn Y, Shen Y, Silvain M, Thermes C, Chen C-L, Hyrien O (2016). Replication landscape of the human genome. *Nat Commun*. 7:10208.

Pomerantz RT, O'Donnell M (2008). The replisome uses mRNA as a primer after colliding with RNA polymerase. *Nature*. 456(7223):762–766.

Pomerantz RT, O'Donnell M (2010). Direct restart of a replication fork stalled by a head-on RNA polymerase. *Science*. 327(5965):590–592.

Prajapati RK, Rosenqvist P, Palmu K, Mäkinen JJ, Malinen AM, Virta P, Metsä-Ketelä M, Belogurov GA (2019). Oxazinomycin arrests RNA polymerase at the polythymidine sequences. *Nucleic Acids Res*. 47(19):10296–10312.

Proshkin S, Rahmouni AR, Mironov A, Nudler E (2010). Cooperation between translating ribosomes and RNA polymerase in transcription elongation. *Science*. 328(5977):504–508.

Reid-Bayliss KS, Loeb LA (2017). Accurate RNA consensus sequencing for high-fidelity detection of transcriptional mutagenesis-induced epimutations. *Proc Natl Acad Sci U S A*. 114(35):9415–9420.

Reines D, Ghanouni P, Li QQ, Mote J Jr. (1992). The RNA polymerase II elongation complex. Factor-dependent transcription elongation involves nascent RNA cleavage. *J Biol Chem*. 267(22):15516–15522.

Riaz-Bradley A, James K, Yuzenkova Y (2020). High intrinsic hydrolytic activity of cyanobacterial RNA polymerase compensates for the absence of transcription proofreading factors. *Nucleic Acids Res*. 48(3):1341–1352.

Robson F, Khan KS, Le TK, Paris C, Demirbag S, Barfuss P, Rocchi P, Ng W-L (2020). Coronavirus RNA Proofreading: Molecular Basis and Therapeutic Targeting. *Mol Cell.* 79(5):710–727.

Roßbach S, Ochsenfeld C (2017). Quantum-Chemical Study of the Discrimination against dNTP in the Nucleotide Addition Reaction in the Active Site of RNA Polymerase II. *J Chem Theory Comput*. 13(4):1699–1705.

Rosenqvist P, Mäkinen JJ, Palmu K, Jokinen J, Prajapati RK, Korhonen HJ, Virta P, Belogurov GA, Metsä-Ketelä M (2022). The role of the maleimide ring system on the structure-activity relationship of showdomycin. *Eur. J. Med. Chem*. 237, 114342.

Roston D, Demapan D, Cui Q (2019). Extensive free-energy simulations identify water as the base in nucleotide addition by DNA polymerase. *Proc Natl Acad Sci U S A.* 116(50):25048-25056.

Roth A, Weinberg Z, Chen AGY, Kim PB, Ames TD, Breaker RR (2014). A widespread self-cleaving ribozyme class is revealed by bioinformatics. *Nat. Chem. Biol*. 10, 56–60.

Rudd MD, Izban MG, Luse DS (1994). The active site of RNA polymerase II participates in transcript cleavage within arrested ternary complexes. *Proc Natl Acad Sci U S A*. 91(17):8057–8061.

Saecker RM, Record MT Jr., Dehaseth PL (2011). Mechanism of bacterial transcription initiation: RNA polymerase - promoter binding, isomerization to initiation-competent open complexes, and initiation of RNA synthesis. *J Mol Biol*. 412(5):754–771.

Scheuermann R, Tam S, Burgers PM, Lu C, Echols H (1983). Identification of the epsilon-subunit of Escherichia coli DNA polymerase III holoenzyme as the dnaQ gene product: a fidelity subunit for DNA replication. *Proc Natl Acad Sci U S A*. 80(23):7085–7089.

Schwanhäusser B, Busse D, Li N, Dittmar G, Schuchhardt J, Wolf J, Chen W, Selbach M (2011). Global quantification of mammalian gene expression control. *Nature.* 473:337–342.

Sekine S, Murayama Y, Svetlov V, Nudler E, Yokoyama S (2015). The ratcheted and ratchetable structural states of RNA polymerase underlie multiple transcriptional functions. *Mol. Cell*. 57(3):408–421.

Seo YJ, Matsuda S, Romesberg FE (2009). Transcription of an expanded genetic alphabet. *J Am Chem Soc*. 131(14):5046–5047.

Shamoo Y, Steitz TA (1999). Building a replisome from interacting pieces: sliding clamp complexed to a peptide from DNA polymerase and a polymerase editing complex. *Cell*. 99(2):155–166.

Singh J, Padgett RA (2009). Rates of in situ transcription and splicing in large human genes. *Nat Struct Mol Biol*. (11):1128–1133.

Smee DF, Hurst BL, Egawa H, Takahashi K, Kadota T, Furuta Y (2009). Intracellular metabolism of favipiravir (T-705) in uninfected and influenza A (H5N1) virus-infected cells. *J Antimicrob Chemother*. 64(4):741–746.

Sosunov V, Sosunova E, Mustaev A, Bass I, Nikiforov V, Goldfarb A (2003). Unified two-metal mechanism of RNA synthesis and degradation by RNA polymerase. *EMBO J*. 22(9):2234–2244.

Sosunov V, Zorov S, Sosunova E, Nikolaev A, Zakeyeva I, Bass I, Goldfarb A, Nikiforov V, Severinov K, Mustaev A (2005). The involvement of the aspartate triad of the active center in all catalytic activities of multisubunit RNA polymerase. *Nucleic Acids Res*. 33(13):4202–4211.

Sosunova E, Sosunov V, Kozlov M, Nikiforov V, Goldfarb A, Mustaev A (2003). Donation of catalytic residues to RNA polymerase active center by transcription factor Gre. *Proc Natl Acad Sci U S A*. 100(26):15469–15474.

Sosunova E, Sosunov V, Epshtein V, Nikiforov V, Mustaev A (2013). Control of transcriptional fidelity by active center tuning as derived from RNA polymerase endonuclease reaction. *J Biol Chem*. 288(9):6688–6703.

Sousa R, Chung YJ, Rose JP, Wang BC (1993). Crystal structure of bacteriophage T7 RNA polymerase at 3.3 Å resolution. *Nature*. 364(6438):593–599.

Sousa R, Padilla R (1995). A mutant T7 RNA polymerase as a DNA polymerase. *EMBO J*. 14(18):4609–4021.

Srivatsan A, Tehranchi A, MacAlpine DM, Wang JD (2010). Co-orientation of replication and transcription preserves genome integrity. *PLoS Genet*. 6(1):e1000810.

Steitz TA (1998). A mechanism for all polymerases. *Nature*. 391(6664):231–232.

Strathern J, Malagon F, Irvin J, Gotte D, Shafer B, Kireeva M, Lubkowska L, Jin DJ, Kashlev M (2013). The fidelity of transcription: RPB1 (RPO21) mutations that increase transcriptional slippage in S. cerevisiae. *J Biol Chem*. 288(4):2689–2699.

Svetlov V, Vassylyev DG, Artsimovitch I (2004). Discrimination against deoxyribonucleotide substrates by bacterial RNA polymerase. *J Biol Chem*. 279(37):38087–38090.

Svetlov V, Belogurov GA, Shabrova E, Vassylyev DG, Artsimovitch I (2007). Allosteric control of the RNA polymerase by the elongation factor RfaH. *Nucleic Acids Res*. 35(17):5694–5705.

Sydow JF, Brueckner F, Cheung ACM, Damsma GE, Dengl S, Lehmann E, Vassylyev D, Cramer P (2009). Structural basis of transcription: mismatch-specific fidelity mechanisms and paused RNA polymerase II with frayed RNA. *Mol. Cell*. 34(6):710–721.

Temiakov D, Patlan V, Anikin M, McAllister WT, Yokoyama S, Vassylyev DG (2004). Structural basis for substrate selection by T7 RNA polymerase. *Cell*. 116(3):381–391.

Temiakov D, Zenkin N, Vassylyeva MN, Perederina A, Tahirov TH, Kashkina E, Savkina M, Zorov S, Nikiforov V, Igarashi N, Matsugaki N, Wakatsuki S, Severinov K, Vassylyev DG (2005). Structural basis of transcription inhibition by antibiotic streptolydigin. *Mol Cell*. 19(5):655–666.

Tjaden J, Winkler HH, Schwöppe C, Van Der Laan M, Möhlmann T, Neuhaus HE (1999). Two nucleotide transport proteins in Chlamydia trachomatis, one for net nucleoside triphosphate uptake and the other for transport of energy. *J Bacteriol*. 181(4):1196–1202.

Traut TW (1994). Physiological concentrations of purines and pyrimidines. *Mol Cell Biochem*. 140(1):1–22.

Traverse CC, Ochman H (2016). Conserved rates and patterns of transcription errors across bacterial growth states and lifestyles. *PNAS*. 113:3311–3316.

Tupin A, Gualtieri M, Leonetti J-P, Brodolin K (2010). The transcription inhibitor lipiarmycin blocks DNA fitting into the RNA polymerase catalytic site. *EMBO J*. 29(15):2527–2537.

Turtola M, Belogurov GA (2016). NusG inhibits RNA polymerase backtracking by stabilizing the minimal transcription bubble. *Elife*. 5:e18096.

Turtola M, Mäkinen JJ, Belogurov GA (2018). Active site closure stabilizes the backtracked state of RNA polymerase. *Nucleic Acids Res*. 46(20):10870–10887.

Ulrich S, Kool ET (2011) Importance of steric effects on the efficiency and fidelity of transcription by T7 RNA polymerase. *Biochemistry*. 50(47):10343–10349.

Van Rompay AR, Johansson M, Karlsson A (2003). Substrate specificity and phosphorylation of antiviral and anticancer nucleoside analogues by human deoxyribonucleoside kinases and ribonucleoside kinases. *Pharmacol Ther*. 100(2):119–139.¨

Vassylyev DG, Svetlov V, Vassylyeva MN, Perederina A, Igarashi N, Matsugaki N, Wakatsuki S, Artsimovitch I (2005). Structural basis for transcription inhibition by tagetitoxin. *Nat Struct Mol Biol*. 12(12):1086–1093.

Vassylyev DG, Vassylyeva MN, Perederina A, Tahirov TH, Artsimovitch I (2007a). Structural basis for transcription elongation by bacterial RNA polymerase. *Nature*. 448(7150):157–162.

Vassylyev DG, Vassylyeva MN, Zhang J, Palangat M, Artsimovitch I, Landick R (2007b), Structural basis for substrate loading in bacterial RNA polymerase. *Nature*. 448(7150):163–168.

Vermulst M, Denney AS, Lang MJ, Hung C, Moore S, Moseley MA, Thompson JW, Madden V, Gauer J, Wolfe KJ, Summers DW, Schleit J, Sutphin GL, Haroon S, Holczbauer A, Caine J, Jorgenson J, Cyr D, Kaeberlein M, Strathern JN, Duncan MC, Erie DA (2015). Transcription errors induce proteotoxic stress and shorten cellular lifespan. *Nat Commun*. 6:8065.

Viswanathan A, Doetsch PW (1998). Effects of nonbulky DNA base damages on Escherichia coli RNA polymerase-mediated elongation and promoter clearance. *J Biol Chem*. 273(33):21276–21281.

Vogel U, Jensen KF (1994). The RNA chain elongation rate in Escherichia coli depends on the growth rate. *J Bacteriol*. 176(10):2807–2813.

Walmacq C, Kireeva ML, Irvin J, Nedialkov Y, Lubkowska L, Malagon F, Strathern JN, Kashlev M (2009). Rpb9 subunit controls transcription fidelity by delaying NTP sequestration in RNA polymerase II. *J Biol Chem*. 284(29):19601–19612.

Wang D, Bushnell DA, Westover KD, Kaplan CD, Kornberg RD (2006). Structural basis of transcription: role of the trigger loop in substrate specificity and catalysis. *Cell*. 127(5):941–954.

Wang D, Bushnell DA, Huang X, Westover KD, Levitt M, Kornberg RD (2009). Structural basis of transcription: backtracked RNA polymerase II at 3.4 angstrom resolution. *Science*. 324(5931):1203–1206.

Wang HY, Elston T, Mogilner A, Oster G (1998). Force generation in RNA polymerase. *Biophys J*. 74(3):1186–1202.

Wang J, Sattar AK, Wang CC, Karam JD, Konigsberg WH, Steitz TA (1997). Crystal structure of a pol alpha family replication DNA polymerase from bacteriophage RB69. *Cell*. 89(7):1087–1099.

Wang W, Walmacq C, Chong J, Kashlev M, Wang D (2018). Structural basis of transcriptional stalling and bypass of abasic DNA lesion by RNA polymerase II. *Proc Natl Acad Sci U S A*. 115(11):E2538–E2545.

Webb, C.-H.T. & Lupták, A (2011). HDV-like self-cleaving ribozymes. *RNA Biol*. 8, 719–727.

Wee LM, Tong AB, Ariza AJF, Cañari-Chumpitaz C, Grob P, Nogales E, Bustamante CJ (2023). A trailing ribosome speeds up RNA polymerase at the expense of transcript fidelity via force and allostery. *Cell*. 186(6):1244–1262.e34.

Westover KD, Bushnell DA, Kornberg RD (2004). Structural basis of transcription: nucleotide selection by rotation in the RNA polymerase II active center. *Cell*. 119(4):481–489.

Wilusz CJ, Wormington M, Peltz SW (2001). The cap-to-tail guide to mRNA turnover. *Nat Rev Mol Cell Biol*. 2(4):237–246.

Windgassen TA, Mooney RA, Nayak D, Palangat M, Zhang J, Landick R (2014). Trigger-helix folding pathway and SI3 mediate catalysis and hairpin-stabilized pausing by Escherichia coli RNA polymerase. *Nucleic Acids Res*. 42(20):12707–12721.

Winkler HH, Neuhaus HE (1999). Non-mitochondrial ATP transport. *Trends Biochem Sci*. 24(2):64–68.

Yamagami R, Bingaman JL, Frankel EA, Bevilacqua PC (2018). Cellular conditions of weakly chelated magnesium ions strongly promote RNA stability and catalysis. *Nat. Commun.* 9(1), 2149.

Yin YW, Steitz TA (2004). The structural mechanism of translocation and helicase activity in T7 RNA polymerase. *Cell*. 116(3):393–404.

Yuzenkova Y, Bochkareva A, Tadigotla VR, Roghanian M, Zorov S, Severinov K, Zenkin N (2010). Stepwise mechanism for transcription fidelity. *BMC Biol*. 8:54.

Yuzenkova Y, Zenkin N (2010). Central role of the RNA polymerase trigger loop in intrinsic RNA hydrolysis. *Proc Natl Acad Sci U S A*. 107(24):10878–10883.

Zenkin N, Yuzenkova Y, Severinov K (2006). Transcript-assisted transcriptional proofreading. *Science*. 313(5786):518–520.

Zhang J, Palangat M, Robert Landick R (2010). Role of the RNA polymerase trigger loop in catalysis and pausing. *Nat Struct Mol Biol*. 17(1):99–104.

Zhang Y, Degen D, Ho MX, Sineva E, Ebright KY, Ebright YW, Mekler V, Vahedian-Movahed H, Feng Y, Yin R, Tuske S, Irschik H, Jansen R, Maffioli S, Donadio S, Arnold E, Ebright RH (2014). GE23077 binds to the RNA polymerase 'i' and 'i+1' sites and prevents the binding of initiating nucleotides. *Elife*. 3:e02450.

TURUN
YLIOPISTO
UNIVERSITY
OF TURKU

Painosalama, Turku, Finland 2024

ISBN 978-951-29-9927-9 (PRINT)
ISBN 978-951-29-9928-6 (PDF)
ISSN 0082-7002 (Print)
ISSN 2343-3175 (Online)