Turun yliopisto
University of Turku

TRANSLOCATION MECHANISM OF THE MULTISUBUNIT RNA POLYMERASE

Matti Turtola

TURUN YLIOPISTON JULKAISUJA – ANNALES UNIVERSITATIS TURKUENSIS
Sarja - ser. AI osa - tom. 584 | Astronomica - Chemica - Physica - Mathematica | Turku 2018

Turun yliopisto
University of Turku

TRANSLOCATION MECHANISM OF THE MULTISUBUNIT RNA POLYMERASE

Matti Turtola

TURUN YLIOPISTON JULKAISUJA – ANNALES UNIVERSITATIS TURKUENSIS
Sarja - ser. AI osa - tom. 584 | Astronomica - Chemica - Physica - Mathematica | Turku 2018

University of Turku
Faculty of Science and Engineering
Department of Biochemistry
The National Doctoral Programme in Informational and Structural Biology (ISB) and
Doctoral Programme in Molecular Life Sciences (DPMLS)

Supervised by

Docent Georgiy A. Belogurov
Department of Biochemistry
University of Turku
Turku, Finland

Reviewed by

Doctor Roman Tuma
Astbury Centre for Structural Molecular Biology
School of Molecular and Cellular Biology
University of Leeds
Leeds, United Kingdom

Professor Andre Sanches Ribeiro
BioMediTech Institute
Tampere University of Technology
Tampere, Finland

Opponent

Professor Finn Werner
Division of Biosciences
University College London
London, United Kingdom

Sarja AI 584

ISBN 978-951-29-7274-6 (PRINT)
ISBN 978-951-29-7275-3 (PDF)
ISSN 0082-7002 (Print)
ISSN 2343-3175 (Online)

Juvenes Print - Helsinki 2018

It is always advisable to perceive clearly our ignorance.
– Charles Darwin

Aatokselle ja Sofialle

TIIVISTELMÄ

Elämä perustuu geneettisen perimän ilmentymiseen ja kopioitumiseen. Kaikissa eliöissä geenien ilmenemisen ensimmäinen vaihe on DNA:n kopioituminen RNA:ksi eli transkriptio. Transkriptiota katalysoivat RNA-polymeraasientsyymit (RNAP). Transkription aikana RNAP erottaa kaksinauhaisen DNA:n juosteet toisistaan ylläpitämällä 10 nukleotidiä sisältävää ”transkriptiokuplaa”, jolloin toista DNA-juostetta voidaan käyttää RNA:n synteesin ohjeena. RNAP:n entsyymikoneisto kytkee yhteen RNA-synteesin nukleosiditrifosfaatti-substraateista, DNA:n kaksoiskierteen avaamisen ja liikkeen (translokaation) DNA:ta pitkin. Jokaisen nukleotidin liittämisen jälkeen RNAP on ”pre-translokaatiotilassa”, josta se lämpöliikkeen vaikutuksesta siirtyy joko eteen- tai taaksepäin. Eteenpäin suuntautuvan liikkeen jälkeen RNAP päätyy ”post-translokaatiotilaan”, jossa se voi sitoa seuraavan substraatin ja katalysoida RNA:n synteesiä. Taaksepäin suuntautuva liike puolestaan synnyttää ”peruutustilan”, jossa RNA:n synteesi estyy. RNAP:iin sitoutuvat toiset proteiinit ja DNA:n signaalit voivat ohjata näitä liikkeitä ja siten säädellä geenien ilmenemistä. Vaikka RNAP on perusteellisesti kuvattu sekä rakenteellisesti että biokemiallisesti translokaation eri vaiheissa, näiden tilojen välinen dynamiikka sekä niitä välittävät konformaatiomuutokset ovat edelleen huonosti tunnettuja. Joukko synteettisiä ja luonnosta eristettyjä antibiootteja ehkäisee mikrobien kasvua estämällä RNAP:n liikkumista DNA:ta pitkin. Uusien antibioottien kehittämiseksi on tärkeää ymmärtää RNA-synteesin ja translokaation mekanismeja sekä selvittää, miten tunnetut antibiootit estävät RNAP:a katalysoimasta yllä mainittuja reaktiovaiheita. Transkription säätelyyn liittyvien molekyylitason periaatteiden selvittämistä on hidastanut olennaisesti puute menetelmistä, joilla translokaatiota voidaan suoraan havainnoida.

Väitöskirjassani kehitin biofysikaalisia menetelmiä, joilla voitiin mitata *Escherichia coli* -bakteerista eristetyn RNAP:n translokaatiota millisekuntien aikaskaalassa yhden emäsparin (~3.4 Å) tarkkuudella. Menetelmät perustuvat fluoresoivien emäsanalogien sijoittamiseen DNA:han. Sopivissa kohdissa transkriptiokuplaa fluoroforien ympäristö muuttuu translokaation seurauksena, ja tämä voidaan havaita fluoresenssisignaalin kasvuna. Tutkimme RNAP:n katalyysimekanismia mittaamalla translokaation lisäksi nukleotidin sitoutumista aktiiviseen keskukseen, nukleotidin liittymistä RNA:han ja pyrofosfaatin (PP_i) irtoamista. Näin pystyimme ensimmäisinä erottamaan toisistaan RNAP:n katalyysisyklin erilliset reaktiovaiheet ja mittaamaan translokaation kestoksi 30-80% katalyysisyklin kokonaisajasta.

Menetelmien avulla tutkimme translokaation termodynaamista perustaa. Keskeinen havaintomme oli, että translokaatiota säätelee RNAP:n aktiivisen keskuksen avautuminen ja sulkeutuminen. Aktiivisen keskuksen sulkeutuminen stabiloi pre-translokaatiotilaa, kun taas sen avautuminen mahdollisti eteenpäin suuntautuvan liikkeen muutoin stabiilimpaan post-translokaatiotilaan. PP_i:n sitoutuminen ja irtoaminen kytkeytyivät translokaatioon, edellinen suosimalla aktiivisen keskuksen sulkeutumista ja jälkimmäinen sen avautumista. Havaintojemme vahvistukseksi tunnistimme RNAP:sta rakenteellisia mutaatioita, jotka vaikuttivat translokaatioon ja PP_i:n irtoamiseen. Nämä tutkimukset liittyivät myös selvitykseen mekanismeista, joilla pienmolekyylit estävät RNAP:n katalyysiä. Havaitsimme, että tagetitoksiini esti translokaatiota stabiloimalla suljettua aktiivista keskusta, kun taas CBR703 vaikutti päinvastaisesti ja edisti translokaatiota. Jälkimmäisessä tapauksessa RNAP kuitenkin muuttui

samalla yliherkäksi DNA:n transkriptiota säätteleville signaaleille, mikä hidasti RNA:n synteesinopeutta.

Selvitimme translokaation aikana tapahtuvia DNA:n ja RNA:n liikkeitä kineettisten mittausten lisäksi biokemiallisilla menetelmillä, joilla voitiin analysoida transkriptiokuplan rakennetta. Tulosten mukaan translokaatioon sisältyy useita vaiheita. Nukleotidin liittymistä seurasi aktiivisen keskuksen avautuminen ja RNA:n translokaatio, mutta DNA ei kuitenkaan translokoitunut tässä vaiheessa vielä kokonaan viitaten siihen, että tyypillisten pre- ja post-translokaatiotilojen lisäksi on olemassa niiden välivaihe. Tämän jälkeen seurasi katalyysisyklin päättävä DNA:n täydellinen translokaatio. Mielenkiintoista oli se, että DNA:n emäsjärjestys vaikutti DNA:n translokaationopeuteen jopa toisinaan siinä määrin, että siitä tuli koko katalyysisyklin nopeutta rajoittava vaihe. Kokeelliset tuloksemme tukevat kaksivaiheista translokaatiomallia ja osoittavat, miten malli voi selittää DNA:n emäsjärjestyksen erilaiset vaikutukset RNA-synteesiin.

Tutkimme myös taaksepäin suuntautuvaa translokaatiota ja sen säätelyä mittamalla RNAP:n peruutustilan muodostumista. Tässä hyödynsimme Gre-proteiineja, jotka edistävät RNA:n leikkautumista RNAP:n peruutustilassa. DNA:n kaksoiskierteen avautuminen transkriptiokuplan perässä edisti taaksepäin suuntautuvaa translokaatiota. Evolutiivisesti hyvin konservoitunut säätelyproteiini NusG vähensi peruutustilaan joutumista stabiloimalla DNA:n emäspareja transkriptiokuplan perässä. Tämä tulos selittää molekyylitasolla, mihin perustuu NusG:n kyky nopeuttaa RNA:n synteesiä. Toisessa osatyössä tutkimme RNAP:n peruutustilan rakennetta sijoittamalla fluoresoivan emäsanalogin RNA:n päähän. Fluoresenssi kasvoi, kun RNA:n pää erkani DNA:sta peruutustilan muodostuessa. Yllättäen aktiivisen keskuksen sulkeutuminen stabiloi peruutustilaa ja näin pitkitti RNA-synteesin pysähdystä. Havaitsimme, että peruutustilaan joutunut RNA sitoutui aktiivisen keskuksen lähellä sijaitsevaan taskuun, joka sivuaa PP_i:n ja tagetitoksiinin sitoutumiskohtia. Yhdessä nämä vuorovaikutukset hidastivat RNAP:n eteenpäin suuntautuvaa translokaationopeutta noin sadasosaan alkuperäisestä. Vaikka RNA:n pitkittynyt sitoutuminen peruutustilaan edisti RNA:n oikolukua, sulkeutunut aktiivinen keskus esti Gre-välitteistä RNA:n leikkautumista. Kuvaamamme peruutustilan vuorovaikutukset voivat täten ensisijaisesti vaikuttaa RNA-synteesin nopeuteen, ja edelleen geenien ilmenemisen säätelyyn.

Väitöskirjani tulokset edistävät kvantitatiivista ymmärrystä RNAP:n katalyysimekanismeista. Havaintomme valaisevat RNAP:n molekyylikoneiston toimintaa ja periaatteita, joilla nukleiinihappo-rakenteet, transkriptiota säätelevät proteiinit ja pienmolekyylit vaikuttavat RNAP:n aktiivisuuteen. Esitetyt tutkimusmenetelmät ovat sovellettavissa myös arkkibakteerien ja eukaryoottien transkriptiokoneistojen tutkimiseen.

ABSTRACT

The expression and replication of genetic information forms the basis of life. The first step of gene expression, the transcription of DNA into RNA, is carried out in all organisms by multisubunit RNA polymerase enzymes (RNAP). During transcription, RNAP maintains 10 nucleotides of unwound DNA, the "transcription bubble", allowing the use of one DNA strand as a template for RNA synthesis. The enzymatic machinery of RNAP couples the synthesis of RNA from nucleoside triphosphate substrates to the unwinding of the double helical DNA and the movement (translocation) along the DNA. After each transcribed nucleotide RNAP first occupies the "pre-translocated state", after which RNAP and the transcription bubble thermally fluctuate either in the forward or backward direction. Upon forward movement, RNAP occupies the "post-translocated state" where it can bind the next nucleoside triphosphate substrate and catalyze the synthesis of RNA. Backward movement, in turn, forms the "backtracked state", which blocks the active site in the enzyme and pauses the RNA synthesis. The molecular transitions during the translocation motion also serve as targets for regulating gene expression by protein factors and signals in the transcribed DNA. The RNAP translocation states have been characterized both structurally and biochemically, yet, the conformational dynamics of the RNAP and nucleic acids that control the translocation of RNAP still remain poorly understood. This gap in understanding is partially due to a lack of methods to directly monitor translocation with sufficient spatial and temporal resolution. Consequently, the detailed principles of the regulation of transcript synthesis remain to be discovered. Furthermore, a number of natural and synthetic antibiotics inhibit transcription by interfering with the movement of RNAP along the DNA. Thus, understanding the mechanisms of RNA synthesis and RNAP translocation as well as how they are inhibited by existing antibiotics provide routes and strategies for developing new antimicrobial drugs.

In this thesis, I present the development of biophysical methods to measure *Escherichia coli* RNAP translocation in a time scale of milliseconds and with a resolution of one base pair (~3.4 Å). The methods rely on labelling the transcribed DNA with fluorescent nucleotide base analogues. When incorporated into suitable positions in the transcription bubble, the environment of the fluorophores changes with RNAP translocation, which can be detected as a change in the fluorescence intensity. We dissected the catalytic mechanism of RNAP by performing the fluorescent translocation measurements in parallel with assays that monitored the binding of the nucleoside triphosphate substrate into the active site of RNAP, the incorporation of the nucleotide into the RNA and the liberation of the by-product pyrophosphate (PP_i). With this approach we were able to kinetically separate, for the first time, the individual steps of the RNAP catalytic cycle and established that RNAP spends a 30-80% share of the total time of the catalytic cycle in the translocation step.

We employed the methods described above to study the thermodynamic and conformational basis of RNAP translocation. A key finding was that translocation is controlled by the opening and closing of the RNAP active site. The closing of the active site stabilized the pre-translocated state, whereas the opening of the active site facilitated forward translocation to the intrinsically more stable post-translocated state. PP_i binding and release modulated translocation by favoring active site closing and opening, respectively. These findings were

corroborated by the identification of RNAP variants with altered translocation and/or PP_i release properties. Moreover, these investigations were inseparably connected to the identification of the inhibitory mechanisms of small molecules. We found that the RNAP inhibitor tagetitoxin prevented forward translocation by stabilizing the closed active site, whereas CBR703 destabilized the closed active site facilitating forward translocation, which, however, made RNAP hypersensitive to DNA encoded regulatory signals.

We further dissected the movements of the DNA and the RNA during translocation with a combination of kinetic assays and the structural analysis of the transcription bubble using nucleic acid photo-crosslinking techniques. Our observations indicated that translocation is a multi-step process. Nucleotide addition by RNAP was followed by the opening of the active site and the translocation of the RNA, however, the DNA translocated only partially at this stage, suggesting the presence of an intermediate translocation state between the canonical pre- and post-translocation states. The RNAP catalytic cycle was then completed by the translocation of the DNA. Interestingly, the transcribed sequence could modulate the rate of DNA translocation to an extent that it became the rate-limiting step of the RNAP catalytic cycle. We provide experimental evidence for the two-step translocation mechanism and demonstrate how it can explain the control of RNA synthesis by the DNA sequence.

Finally, we studied the mechanism and regulation of RNAP backtracking by measuring the kinetics of entry into the backtracked state by Gre-factor mediated RNA cleavage. RNAP backtracking was found to be facilitated by the melting of the DNA at the rear end of the transcription bubble. The universally conserved transcription factor NusG inhibited RNAP backtracking by stabilizing the DNA base pairs at the rear end of the transcription bubble, explaining at the molecular level, how NusG speeds up the rate of RNA synthesis. In another study, we investigated the structural basis of backtracking by incorporating a base analogue fluorophore into the RNA. Upon entry into the backtracked state the end of the RNA separated from the DNA giving rise to a fluorescence signal. Unexpectedly, the closing of the active site stabilized the backtracked RNA and prolonged the life-time of the backtracked pause. We found that the backtracked RNA binds in a pocket near the RNAP active site (the so-called E-site) that overlaps with the PP_i and tagetitoxin binding sites. Together these interactions inhibited the forward translocation of RNAP by a factor of ~100. Although the prolonged occupancy of the RNA in the backtracked state facilitated transcriptional proofreading, the closed active site interfered with the RNA cleavage by the Gre-factors. Thus, our observations indicate that the binding of the backtracked RNA in the E-site may contribute to gene regulation by modulating the rate of RNA synthesis.

The results present advances in the quantitative description of the catalytic mechanism of RNAP. Our observations reveal important features of RNAP operation as a molecular machine controlled by nucleic acids, transcription factors and small molecules, and open new mechanistic questions to be pursued. The methods presented here are applicable to similar studies in Archaeal and Eukaryotic transcription machineries.

TABLE OF CONTENTS

ABBREVIATIONS

2´dNTP	2´ deoxynucleoside triphosphate
2-AP	2-aminopurine
6-MI	6-methylisoxanthopterin
6-TG	6-thioguanine
8-MP	8-methoxy psoralen
A	adenine
AMP	adenosine monophosphate
ATP	adenosine triphosphate
BH	bridge helix
bp	base pair
C	cytosine
CMP	cytidine monophosphate
CMPcPP	cytidine-5'-[(α,β)-methyleno]triphosphate
CTP	cytidine triphosphate
CBR703	*N*-hydroxy-*N*´- phenyl-3-trifluoromethyl-benzamidine
dA	2´ deoxyadenosine
dC	2´ deoxycytidine
dG	2´ deoxyguanosine
dsDNA	double stranded DNA
dT	2´ deoxythymidine
DNA	deoxyribonucleic acid
EDTA	ethylenediaminetetraacetic acid
FL	fork loop 2
G	guanosine
GL	gate loop
GMP	guanosine monophosphate
GMPcPP	guanosine-5'-[(α,β)-methyleno]triphosphate
GTP	guanosine triphosphate
HIV	human immunodeficiency virus
ITC	initially transcribing complex
ITP	inosine triphosphate
kb	kilo base pair
LL	lid loop
MD	molecular dynamics
MDCC	*N*-[2-(l-maleimidyl)ethyl]-7-diethylamino)coumarin-3-carboxyamide
mRNA	messenger RNA
NMPcPP	nucleoside-5'-[(α,β)-methyleno]triphosphate
NMP	nucleoside monophosphate
NTP	nucleoside triphosphate
ops	operon polarity suppressor
PAGE	polyacrylamide gel electrophoresis

PBP	phosphate binding protein
PCR	polymerase chain reaction
PDB	protein data bank
P_i	phosphate
PPase	inorganic pyrophosphatase
PP_i	pyrophosphate
(p)ppGpp	guanosine (penta)tetraphosphate
RL	rudder loop
RNA	ribonucleic acid
RNAP	RNA polymerase
rRNA	ribosomal RNA
SBHM	sandwich barrel hybrid motif
ssDNA	single stranded DNA
STL	streptolydigin
SW2	switch 2
T	thymine
TEC	transcription elongation complex
TGT	tagetitoxin
THB	three helical bundle
TL	trigger loop
tRNA	transfer RNA
U	uracil
UV	ultra violet
UTP	uridine triphosphate
UTR	untranslated region
Y	pyrimidine

Abbreviations of amino acid residues

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

LIST OF ORIGINAL STUDIES

This thesis is based on five original publications and two manuscripts, which are referred to by the following Roman numerals in the text:

I Malinen AM, **Turtola M**, Belogurov GA (2015) Monitoring Translocation of Multisubunit RNA Polymerase Along the DNA with Fluorescent Base Analogues. In: Artsimovitch I, Santangelo T (eds) Bacterial Transcriptional Control. *Methods in Molecular Biology*. Vol 1276. Humana Press, New York, NY

II 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 Research.* 40:7442-7451

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

IV 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 Research.* 44:1298-1308

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

VI **Turtola M**, Mäkinen JJ, Belogurov GA. Trigger loop stabilizes the backtracked state of RNA polymerase. *Manuscript*

VII **Turtola M**, Mäkinen JJ, Malinen AM, Gregorova P, Belogurov GA. A two-step translocation mechanism of the multisubunit RNA polymerase. *Manuscript*

1. INTRODUCTION

The transfer of genetic information from *DNA to RNA to protein* is the basis of life. The genetic instructions for building organisms are contained in chromosomes written in long chains of DNA polymers of alternating four types of nucleotides. The sequence of different nucleotides encodes the sequence of the amino acids in the polypeptide chains that make up a protein. To translate the DNA code into a chain of amino acids, the nucleotide sequence is first transcribed into an RNA molecule, which then serves as the instruction for the protein synthesis machinery (**Figure 1**). DNA transcription is carried out by the enzyme RNA polymerase (RNAP), which is composed of multiple *protein* chains folded into an intricate molecular machinery. RNAP accesses the genetic code by unwinding the *DNA* double helix and exposes the other strand to serve as a template for making *RNA*. Thus, the act of DNA transcription fascinatingly manifests all three players of the genetic information: DNA, RNA and protein.

Transcription, being the first and often the decisive step of gene expression, is under strict control by genetic regulatory programs. Transcription activity is always restricted to a subset of genes within a genome and it changes in response to environmental challenges or as the organism goes through different stages of development. A gene constitutes a start site, the coding region and an end. Respectively, RNAP must be able to recognize the start, read the code, and then stop the synthesis of RNA. Thus, the act of transcription is conceptually and mechanistically divided into three phases: initiation, elongation and termination (**Figure 1**). In the initiation phase RNAP recognizes the start of the gene, the promoter, with the help of transcription factors. Binding of RNAP to the promoter leads to the separation of the two DNA strands; one strand enters the RNAP to serve as a template for RNA synthesis. A successful initial transcription allows RNAP to break the contacts with the promoter and transition to the elongation phase. RNAP synthesizes the RNA from ribo-nucleoside triphosphate (NTP) substrates that it chemically joins together to form an extended polynucleotide chain. The genetic code in DNA is defined by the sequence of four bases: adenine (A), thymine (T), guanine (G) and cytosine (C). RNAP extends the RNA by selecting different substrates (ATP, GTP CTP and UTP (uridine triphosphate)) based on their ability to align with the DNA according to the Watson-Crick base pairing rules: ATP aligns with thymine, GTP with cytosine, CTP with guanine and UTP with adenine. The accuracy of biological information transfer is intimately linked to the transcription mechanism. Finally, when the full length RNA has been synthesized, RNAP must end the transcription correctly. Synthesis is terminated by releasing RNAP and the nascent RNA from the DNA. All the three phases in the transcription cycle are subject to extensive regulation.

A typical bacterial protein coding gene covers a 300 nanometer stretch of DNA (~900 nucleotides) – a distance over 20 times longer than the diameter of a single RNAP enzyme. This poses the physical problem for how the enzymatic center of RNAP can access the full information of the DNA code. A solution that must have been conceived by evolution at the dawn of life, is that the DNA runs through RNAP and the enzyme moves, or *translocates*, stepwise along the DNA reading the code one nucleotide at a time. With an average synthesis rate of 30 nucleotides per second, the elongation phase takes about half a minute for an average

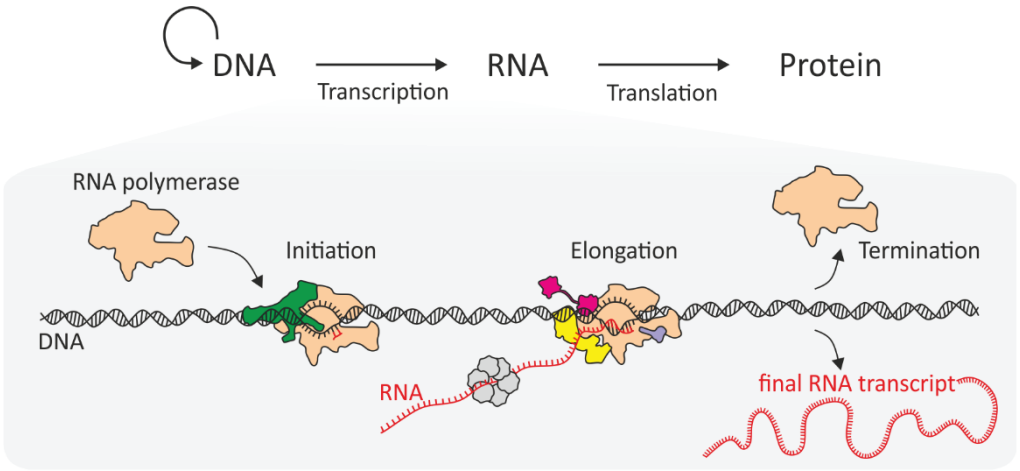


Figure 1. Transcription of DNA into RNA is the first step of gene expression. The genetic instructions contained and copied in DNA are expressed by making a RNA copy of the DNA (a transcript). The transcript can work as a structural RNA, such as the ribosomal RNA or transfer RNA, or as a messenger RNA carrying the genetic information to the translation machinery to synthesize proteins. Transcription is carried out by the enzyme RNA polymerase (brown). The synthesis of each individual RNA molecule is divided mechanistically into three phases: initiation, elongation and termination. The activity of RNA polymerase is regulated in each phase by transcription factors, the transcribed DNA sequence and secondary structures in the nascent RNA. Here depicted are some of the protein regulators of the bacterial RNA polymerase: the initiation factor Sigma (green), the elongation factors NusG (magenta), NusA (yellow) and GreA (violet), and the termination factor Rho (grey).

sized bacterial gene. However, the genes in eukaryotic organisms are typically longer and individual genes can be up to millions of nucleotides long making the feat several magnitudes longer in time and distance. The ability to create unidirectional movement is a remarkable property shared among motor proteins, of which RNAP is a great example. While it is generally assumed that the locomotion of RNAP is fueled by the thermal energy and biased forward by the condensation of the nucleotide triphosphates into the nascent RNA, there remains several unresolved questions about the contribution of the movement step to the overall transcription rate and the roles of nucleotide triphosphates binding and condensation steps in enforcing the overall directionality of RNAP movement (Shimamoto, 2013). Investigations on this topic not only help to understand the mechanistic basis of transcription, but shed light on the general principles of how these nanometer-sized molecular machines create movement.

Encoded in their structure, the DNA and the transcribed RNA contain regulatory signals that modulate the enzymatic activity of the elongating RNAP through elaborate non-covalent interactions. The signals that cause transcription to halt temporarily (transcriptional pausing) or permanently (transcription termination) are part of feedback systems that control the expression of the genetic material. Furthermore, a number of protein factors control the activity of the RNA polymerase by tuning the parameters of the elementary steps in the nucleotide addition cycle (reviewed in Belogurov & Artsimovitch, 2015) (**Figure 1**). The outcome of transcription, i.e. whether a given gene is expressed or not, often depends on the response of the elongating RNAP to these signals and elongation factors. Although these phenomena have been described long ago, their molecular basis and quantitative significance remain poorly

understood. Thereby, while the RNA product of transcription can be predicted from the DNA sequence alone, the outcome of transcription events are challenging to predict.

This thesis focuses on the biochemical and biophysical aspects of transcription elongation and aims to increase our understanding of how the bacterial RNAP works as a molecular machine when synthesizing RNA and translocating along the DNA. A mechanistic understanding of the transcription machinery forms the foundation to link RNAP function to transcriptional regulation by protein factors and nucleic acid sequences. Importantly, as the enzymatic activity of RNAP is a target for a number of natural and synthetic antibiotics (Ma *et al*, 2016), defining the RNAP mechanism and the basis for its inhibition provide routes and strategies for developing new antimicrobials. This description must not only include the main structural states of the RNAP machinery, but also the dynamics that connect these states. Indeed, the wealth of structural information gained in recent years depicting RNAP in various conformations has outstripped our understanding of the dynamics between different conformations. The relationships between states are described by thermodynamics, but this information is insufficient for describing the complex multi-step reactions, like those catalyzed by RNAPs. To decipher the dynamics of the RNAP machinery one must measure the rates of transitions between the states. While not all rates in the multistep mechanism manifest themselves during transcription in the living cell, the determination of the minimal reaction mechanism (which is biologically relevant) relies on the principles of chemical kinetics and accurate measurements of the forward and backward rates (Shimamoto, 2013). This "kinetic approach" is adopted throughout the experimental part of the thesis. Ultimately, a description of the transcription machinery requires the integration of kinetic and structural information. To this end, the RNAP mechanisms remain incompletely characterized despite nearly 60 years of investigation since the discovery of RNAP in the early 1960´s (Hurwitz, 2005).

In the literature review, I have aimed to cover the current state of understanding of the catalytic properties of RNAP. At the heart of the catalytic activity of RNAP is the nucleotide addition cycle, which is dissected step by step and where emphasis is placed on translocation, the main target of investigation in the experimental part of the thesis. Bifurcation of elongation between active and inactive states is also discussed with emphasis on backward translocation (backtracking). Examples of how the activity of RNAP is modulated by protein factors and inhibitors are given. Translocation and backtracking have remained among the least understood properties of RNAP, due to both their complexity at the molecular level and the lack of methods to directly monitor translocational motion with high spatial and temporal resolution. The experimental part of the thesis describes new methods that enable direct monitoring of RNAP translocation. These methods are combined with enzyme kinetic approaches, protein mutagenesis and other biochemical assays to delineate the translocation and backtracking mechanisms. The thermodynamic and kinetic understanding of these processes formed the basis for uncovering the inhibitory mechanisms of two known RNAP inhibitors, tagetitoxin and CBR703, and the elucidation of the regulatory mechanism for a universal transcription factor NusG. As a culmination of these studies, a two-step model for RNAP translocation is presented. Because

some of these studies were published some time ago and the field of mechanistic transcription has developed since, it felt imperative to occasionally refer to our own published works that are part of this thesis in the introductory literature review. These studies are also later described in detail in the results section.

2. LITERATURE REVIEW

2.1 Overall structure of the transcription elongation complex

Bacterial RNAP is a $\alpha_2\beta\beta'\omega$ assembly of four different subunits (**Figure 2**). This structure is evolutionarily highly conserved and forms the core in both archaeal and eukaryotic RNAPs, which then also contain additional subunits (Murakami, 2015; Werner & Grohmann, 2011; Lane & Darst, 2010a, 2010b). The outline of the two largest subunits β and β′ resembles a crab claw; the subunits form a large nucleic acid binding cleft (the main channel) that is wide in the absence of nucleic acids but narrows upon DNA binding (Chakraborty *et al*, 2012; Landick, 2001; Gnatt *et al*, 2001; Cramer *et al*, 2001) and undergoes dynamic opening and closing during transcription elongation in response to nucleic acid encoded signals and transcription elongation factors (Tagami *et al*, 2010; Sevostyanova *et al*, 2011; Weixlbaumer *et al*, 2013; Kolb *et al*, 2014; Hein *et al*, 2014; Sekine *et al*, 2015; Schulz *et al*, 2016; Ray-Soni *et al*, 2016). The main channel harbors the enzymatic active site including the absolutely conserved aspartate triad responsible for the binding of the catalytic magnesium ion and the RNA 3′ end, and two mobile β′ elements, the bridge helix (BH) and the trigger loop (TL) **(Figure 3)**. The active site is buried deep inside the RNAP structure and is connected to the outside solution through the secondary channel, a ~40 Å long negatively charged pore that narrows near the active site.

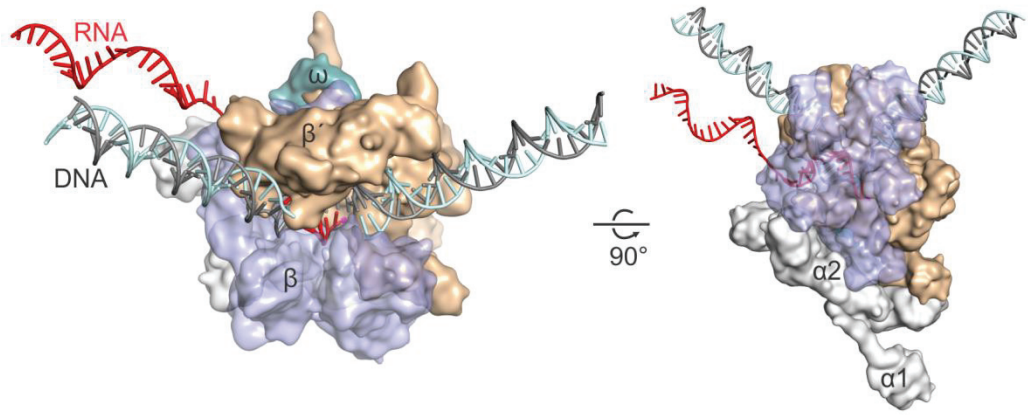


Figure 2. The overall structure of a bacterial transcription elongation complex. RNAP transcribes from left to right. The substrate NTPs bind to the active site located next to the Mg^{2+}-ion (magenta sphere). DNA template strand (grey), DNA non-template strand (cyan), RNA (red), RNAP subunits β (purple, transparent), β′ (brown), α1 and α2 (light grey), ω (teal). The single stranded region of the non-template DNA within the transcription bubble is absent from the model. The TEC model is from Turtola & Belogurov 2016.

In the transcription elongation complex (TEC) RNAP covers ~35 nucleotides of DNA **(Figure 2)**. The DNA helix unwinds just before the active site and rewinds ≥10 base pairs upstream forming the "transcription bubble" **(Figures 2 and 3)**. The template DNA strand enters deep inside into the main channel and pairs with the nascent RNA forming a 9-10 base pair (bp) long RNA:DNA hybrid, whereas the non-template strand threads single stranded near the enzyme surface. The nascent RNA is peeled away from the template and exits through a narrow RNA exit channel. The separation of the product strand from the template is a distinguishing property of the transcribing RNAP compared to DNA polymerases (Westover *et al*, 2004a; Kettenberger *et al*, 2004) and replicative RNA polymerases.

The high structural conservation of bacterial and eurarchaeal transcription machineries suggests that their core catalytic mechanisms are conserved. Therefore, it is common in the field of mechanistic transcription to integrate, yet with caution, structural and functional data gathered from RNAPs from evolutionarily distant organisms like *Saccharomyces cerevisiae* and *Escherichia coli* (e.g. Landick, 2001; Martinez-Rucobo *et al*, 2011; Kireeva *et al*, 2010).

2.2 A unit of transcription elongation: The nucleotide addition cycle

During the transcription elongation phase RNA polymerase repetitiously and stepwise adds nucleotides to the growing RNA chain. The nucleotide addition cycle consists of the following elementary steps (**Figure 3**): *1*) The substrate NTP complexed with a Mg^{2+}-ion enters the RNAP active site and binds to the A-site (a.k.a. the *i+1*-site). This includes an additional accommodation step (see later); *2*) RNAP catalyzes the reaction between the NTP and the nascent RNA forming a covalent phosphodiester bond; *3*) The reaction by-product pyrophosphate (PP_i) dissociates out of the active site; *4*) RNAP translocates one nucleotide along the DNA to the next coding position so that the nascent RNA 3´ end occupies the P-site (a.k.a. the *i*-site) (reviewed in (Belogurov & Artsimovitch, 2015)).

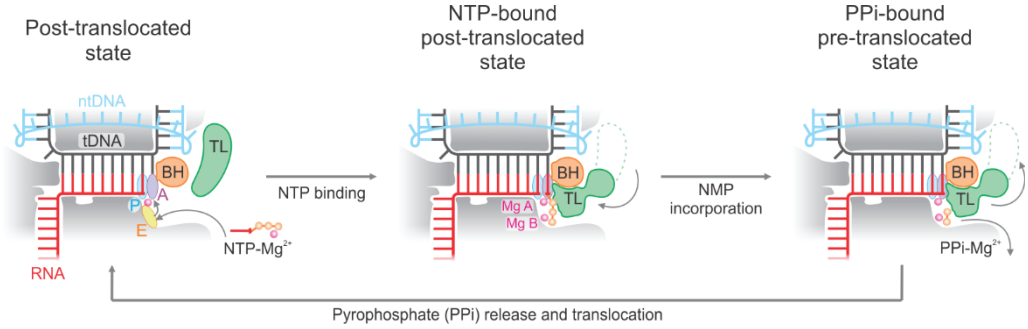


Figure 3. Overview of the nucleotide addition cycle. See text for details. TL (green), trigger loop; BH (orange), bridge helix, A-site (purple oval), P-site (cyan oval); E-site (yellow, oval); DNA template strand (black); DNA non-template strand (cyan); RNA (red); magnesium ions (magenta spheres); phosphate moiety of the NTP (yellow spheres). The small arrows display the diffusion of the substrate NTP-Mg^{2+} to the active site through the secondary channel (left panel), the folding of the TL into the helical state (middle panel), unfolding of the TL into the unfolded state (right panel), and the diffusion of the pyrophosphate (PP_i-Mg^{2+}) out of the active site (right panel). Adapted from (Belogurov & Artsimovitch, 2015).

Each elementary step is a succession of coordinated conformational transformations induced by the non-covalent interactions between nucleic acid and protein residues (**Figure 4**). Transcription accuracy depends on these interactions and conformational transitions. The high ratio of selection for correct nucleotides over incorrect ones is maintained by thermodynamic and kinetic discrimination.

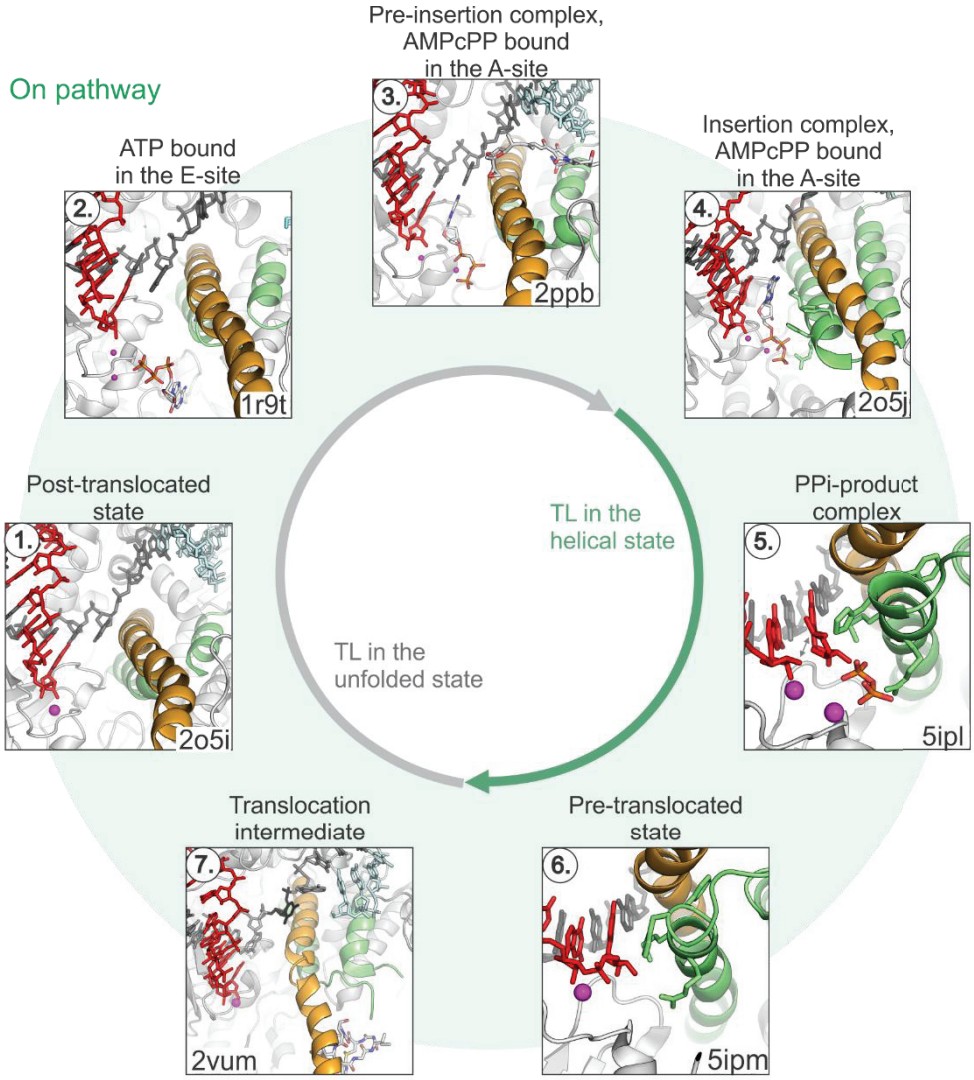
2.2.1 Substrate binding to the entry site and the active site

Substrate NTPs enter the RNAP active site through the secondary channel (Batada *et al*, 2004; Landick, 2005) which may contain multiple non-specific binding sites for NTPs (Wang *et al*, 2017). NTPs then bind non-selectively to an entry site (E-site) through ionic interactions of the triphosphate moiety (Sosunov *et al*, 2003; Westover *et al*, 2004b) (**Figure 3** and **Figure 4, state 2**). Substrate rotation around the triphosphate moiety to the A-site (Westover *et al*, 2004b) allows NTP to make base selective interactions in an open active site (pre-insertion complex, **Figure 4, state 3**), but the phosphates remain too far from the magnesium A (Mg^{2+} (A)) for catalytic activity (Vassylyev *et al*, 2007; Kettenberger *et al*, 2004, Temiakov *et al*, 2005). The base, sugar and the triphosphate moiety of the NTP interact extensively with the active site residues and nucleic acids. The NTP base stacks against the RNA 3´ residue and interacts with the templating DNA base through hydrogen bonding (Kettenberger *et al*, 2004; Vassylyev *et al*, 2007b).

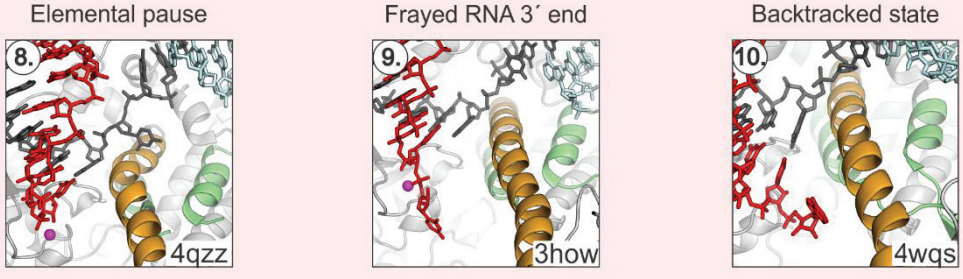
2.2.2 Insertion complex: The closure of the active site

The exchange of the TL between two major conformations enables the active site to load substrates and carry out catalysis: An unfolded TL keeps the secondary channel wide allowing the diffusion of NTPs and their initial binding in the E- and A-sites (**Figure 4, states 1-3**). On the other hand, the folding of the TL into a helical state allows it to form induced fit -type interactions with the NTP and BH, which stabilizes the catalytically active state in the insertion complex (Wang *et al*, 2006; Vassylyev *et al*, 2007b; Kaplan *et al*, 2008) (**Figure 4, state 4**).

Figure 4. Conformational states of the RNAP active site. The on-pathway states forming the nucleotide addition cycle are along the green circle; off-pathway (inactive) states are on the red rectangle. For clarity, only the β´ subunit (*grey*) is visible. Other structural parts are highlighted as follows: TL, *green*; BH, *orange*; RNA, *red*; template DNA, *dark grey*; non-template DNA, *cyan*; Mg^{2+}, *magenta.* Ligands including NTP (2), NMPcPP (3 & 4), streptolydigin (3), PP_i (5) and α-amanitin (7) are shown as sticks. In panels 4-6, the side chains of the TL amino acids β´Met^{932}, Arg ^{933}, Phe^{935} and His^{936} (*Eco* numbering) are shown as sticks. The arrow (5) indicates a discontinuous electron density between the NMP in the A-site and the rest of the RNA which is due to equilibrium of nucleotide addition and pyrophosphorolysis or intrinsic cleavage of RNA in the crystals. PDB codes are indicated for each structure. The source organism and references for each structure are: post-translocated state (*Thermus thermophilus, Tth*) (Vassylyev *et al*, 2007a); ATP bound to the E-site (*Saccharomyces cerevisiae, Sce*) (Westover *et al*, 2004b); pre-insertion complex with a non-hydrolyzable AMPcPP, bound to the A-site, stabilized by STL (*Tth*) (Vassylyev *et al*, 2007b); insertion complex with a non-hydrolyzable AMPcPP bound to the A-site (*Tth*) (Vassylyev *et al*, 2007b); PP_i-bound to the active site of the initially transcribing complex (ITC) (*E. coli, Eco*) (Liu *et al*, 2016); pre-translocated state of the ITC (Eco) (Liu *et al*, 2016); α-amanitin stabilized translocation intermediate state (*Sce*) (Brueckner & Cramer, 2008); elemental paused state (*Tth*) (Weixlbaumer *et al*, 2013); frayed RNA 3´ end (*Sce*) (Sydow *et al*, 2009); backtracked state (*Tth*) (Sekine *et al*, 2015).


On pathway
Pre-insertion complex, AMPcPP bound in the A-site
3.
2ppb
ATP bound in the E-site
2.
1r9t
Insertion complex, AMPcPP bound in the A-site
4.
2o5j
Post-translocated state
1.
2o5i
TL in the helical state
TL in the unfolded state
PPi-product complex
5.
5ipl
Translocation intermediate
7.
2vum
Pre-translocated state
6.
5ipm

Off pathway


Elemental pause
8.
4qzz
Frayed RNA 3´ end
9.
3how
Backtracked state
10.
4wqs

The ends of the foldable TL tip are connected to two anti-parallel alpha helices that are referred to as the TL "base". The folding of the unstructured loop into the alpha-helical state extends both of the helices. In the folded helical state the TL forms a three-helical bundle (THB) with the BH (**Figure 4, states 4-6**). The formation of the THB adds to interactions with the NTP and prevents it from dissociating by narrowing the secondary channel pore. The TL residues β´Met^{932} and β´Phe^{935} form a hydrophobic cavity that sandwiches the correctly base paired nucleoside between the TL and 3´ end of the RNA (**Figure 5**, right). Furthermore, the TL residues β´Arg^{933} and β´His^{936} are two out of a total of six residues that stabilize the substrate triphosphates by 11 hydrogen bonds. These interactions rotate the phosphates 2.5 Å from the pre-insertion conformation to reach the catalytic configuration (Vassylyev *et al*, 2007b) (**Figure 5**, left). The formation of the THB also facilitates the dehydration of the active site, which is necessary for the correct alignment of the base paired NTP (Seibold *et al*, 2010).

The THB is a meta-stable structure which is stabilized by the correctly paired NTP in the active site (Wang *et al*, 2013). NTP binding and active site closing are enthalpically driven: in the helical state the TL makes ~30 more contacts with the rest of the RNAP and the NTP. Calculations based on molecular dynamics (MD) simulations estimate that, in the presence of an NTP, the helical conformation of the TL is favored over the unfolded conformation by 8 kJ/mol. (Wang *et al*, 2013). Thereby, a fraction of the free energy contained in the NTP drives the closing of the active site and is harnessed before the incorporation of the NTP into the RNA chain. In turn, the relaxation of the THB to the ground state is driven by the increase in entropy (Wang *et al*, 2013).

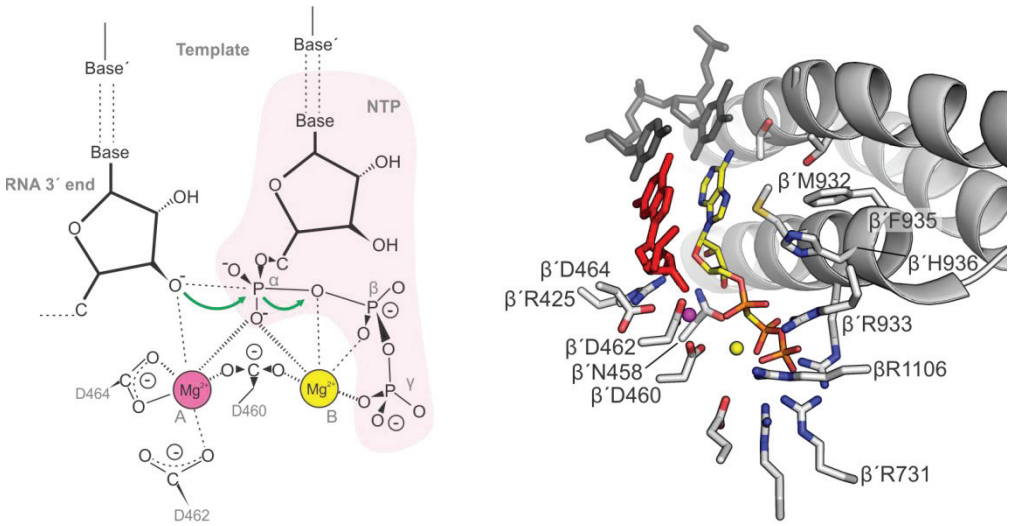


Figure 5. The catalysis of the nucleotidyl transfer reaction. *Left*: The two-metal catalysis of the nucleotidyl transfer reaction (Figure adapted to *Eco* RNAP from (Steitz, 1998)). The electron transfers occurring during the S_N2-type nucleophilic substitution are indicated by green arrows. The substrate NTP is highlighted with a pink shade. A water molecule coordinating the Mg^{2+} (B) and another one coordinating both (A) and (B) (Vassylyev *et al*, 2007b) are omitted for clarity. *Right*: The *Tth* insertion complex (2O5J (Vassylyev *et al*, 2007b)). The amino acids stabilizing the NTP (C-atoms in yellow), Mg^{2+}-ions (colored as in the figure on the left) and the template base are indicated with sticks; the amino acids are numbered according to *Eco* RNAP. The RNA 3´-terminal nucleotide (red) occupying the P-site and the template-DNA (black) occupying the *i* and *i+1-sites* are shown.

The closed RNAP active site provides the interactions that stabilize the transition state of the nucleotidyl transfer reaction effectively coupling the formation of the THB with the chemical reaction (Temiakov *et al*, 2005; Wang *et al*, 2006; Vassylyev *et al*, 2007b). The principal effect of the THB on catalysis is the stabilization of the correct substrate orientation that creates the proper geometry for the nucleophilic attack (Wang *et al*, 2006; Vassylyev *et al*, 2007b) (discussed further in section 2.2.3). Accordingly, changes in THB stability affect the catalytic rate: mutations that facilitate the formation of THB or increase its dynamics, accelerate the rate of catalysis (Tan *et al*, 2008; Jovanovic *et al*, 2011; Nedialkov *et al*, 2013; Malinen *et al*, 2014), whereas prevention of the formation of the THB by either deleting the TL, destabilizing the THB through mutations or by introducing the inhibitor streptolydigin slows down the catalysis rate three to four orders of magnitude (Yuzenkova *et al*, 2010; Zhang *et al*, 2010; Vassylyev *et al*, 2007b; Temiakov *et al*, 2005; Tuske *et al*, 2005; Windgassen *et al*, 2014). Also, salinamide A (Degen *et al*, 2014) and CBR-series of compounds (Malinen *et al*, 2014; Bae *et al*, 2015; Feng *et al*, 2015) inhibit RNAP by interfering with the formation of the THB. The BH may also facilitate catalysis independent of the TL because the THB targeting antibiotics salinamide A and CBR703 inhibit the catalytic rate of RNAP variants lacking the TL. Possibly the BH bending dynamics (discussed in section 2.3.4) help the NTP to acquire conformations close to the transition state (Degen *et al*, 2014; Feng *et al*, 2015; Bae *et al*, 2015).

RNAP elements surrounding the THB contribute to its stability and mediate allosteric signals that control the catalytic properties of RNAP. The “cap” surrounding the N-terminal end of the BH consists of an F-loop, connector, fork loop 2 and a β-DloopII (see Figure 7 p. 27 and Figure 25 p. 99). For example, variations in the F-loop structure govern the temperature adaptivity of RNAPs in thermophilic and mesophilic bacteria (Miropolskaya *et al*, 2010, 2014 and 2009), and the β´P750L mutation in the F-loop slows down the unfolding of the TL (Malinen *et al*, 2014). Mutations in the fork loop 2 alter the catalytic rate of RNAP by disrupting a hydrogen bonding network that involves residues from the βDloopII and the THB and extends into the catalytic site (Seibold *et al*, 2010; Kireeva *et al*, 2011). Finally, the movements of the β´ clamp domain (Chakraborty *et al*, 2012) may affect the folding energetics of the THB through the “anchor”, which consists of switch 1 and 2 regions that surround the C-terminal end of the BH (Nayak *et al*, 2013) (see Figure 7 p. 27). Furthermore, in many bacterial lineages the TL harbors an insertion (SI3 or β´in6, insertion site is shown in Figure 25, p. 99) that is composed of two or more SBHM domains (sandwich barrel hybrid motif). Interestingly, SI3 is the only bacterial sequence insertion that is located in an otherwise structurally conserved region (Lane & Darst, 2010a). SI3 interacts with the jaw domain near the downstream DNA (Chlenov *et al*, 2005; Liu *et al*, 2016) and disfavors the folding of the TL into the helical state (Windgassen *et al*, 2014; Miropolskaya *et al*, 2010). These connections could hypothetically allow the downstream DNA sequence to mediate signals to the active site (Palangat *et al*, 2004). The numerous allosteric connections underscore the complexity of the factors that modulate the catalytic activity of RNAP.

2.2.3 Catalysis: the chemistry of the nucleotidyl transfer reaction

RNAP incorporates nucleotides to the nascent RNA 3´ end by a two-metal-ion mechanism, which is common to nucleotide polymerases (Steitz, 1998). The reaction proceeds with an S_N2-type nucleophilic substitution, where the 3´ OH group of the nascent RNA attacks the alpha-phosphate of the NTP (**Figure 5**, left). The role of the first metal ion, Mg^{2+} (A), is to lower the affinity of the RNA 3´ oxygen toward the proton, which facilitates the O^- attack to the α-phosphate. The second metal, Mg^{2+} (B), binds to the leaving PP_i group, stabilizing and partially neutralizing it. Both metals stabilize and neutralize the negatively charged penta-covalent transition state. The Mg^{2+} (A) is bound in the RNAP active site by the absolutely conserved aspartate triad DFDGD (Sosunov *et al*, 2005; Zaychikov *et al*, 1996; Cramer *et al*, 2001; Zhang *et al*, 1999) with high affinity (K_D ~100 μM) (Mustaev *et al*, 1997) (**Figure 5**, right). The Mg^{2+} (B) is brought to the active site bound to the triphosphate moiety of the NTP (Sosunov *et al*, 2003) and additional coordination is provided by one aspartate of the triad and two water molecules, one of them shared with Mg^{2+} (A) (Westover *et al*, 2004b; Vassylyev *et al*, 2007b). Recently, the catalysis of the nucleotidyl transfer reactions has been proposed to involve a third magnesium ion (Mg^{2+} (C)) (Freudenthal *et al*, 2013, 2015; Vyas *et al*, 2015, Nakamura *et al*, 2012; Gao & Yang, 2016).

Nucleotide incorporation involves two proton transfer reactions: deprotonation of the 3´ OH group of the nascent RNA and protonation of the leaving group PP_i (Castro *et al*, 2007). Proton transfers limit the rate of the nucleotidyl transfer reaction in single subunit RNA and DNA polymerases (Castro *et al*, 2009, 2007), but their contribution to the catalytic rate of multi-subunit RNA polymerases remains unknown. Also, the acceptor and donor of these protons are unknown. The invariant histidine (β´His^{936} in *E. coli*) at the tip of the TL (**Figure 5**, right) has been suggested to participate in proton transfers (Castro *et al*, 2009). However, the role of the histidine in acid/base catalysis seems unlikely, because its substitutions did not affect the catalysis rate (Mishanina *et al*, 2017) or reduced it only modestly between one (Zhang *et al*, 2010; Kaplan *et al*, 2008) and two (Yuzenkova *et al*, 2010) orders of magnitude. Rather, the histidine appears to function primarily as a positional catalyst (Mishanina *et al*, 2017) by stabilizing the NTP through hydrogen bonding (Vassylyev *et al*, 2007b; Huang *et al*, 2010). Finally, the histidine can interact with the PP_i in the product complex (Liu *et al*, 2016), and based on the results on single subunit nucleotide polymerases, it was suggested that the histidine would donate a proton to PP_i to facilitate its release (Castro *et al*, 2009, 2007).

2.2.4 Pyrophosphate release

RNA extension by the NMP generates PP_i as a side product (**Figure 4, state 5**). PP_i participates in the coordination of the Mg^{2+} (B) and comes into contact with at least four RNAP residues: β´Arg^{933} and β´His^{936} at the tip of the helical TL, and two arginines at the end of the secondary channel (Liu *et al*, 2016). The release of PP_i coincides with the rotation of β´His^{936} (**Figure 4, state 6**) and the unresolved SI3 insertion domain in the TL becomes visible in the crystal structure. The F-loop and the rim helices at the edge of the secondary channel move towards the TH-SI3 linkers and tighten the connection between SI3 and the jaw domains (Liu *et al*, 2016). The

molecular events between PPi formation and release have been investigated by MD simulations. The simulations suggested that upon formation of PP_i from the NTP, the TL tip residues would undergo larger fluctuations that in turn would destabilize the contact between the TL and the nascent RNA end. The dissociating PP_i might initially maintain some contacts and "drag" the TL tip towards the secondary channel. This would expose other TL residues to the solvent and initiate the unfolding process. Ultimately, PP_i diffuses out of RNAP through the secondary channel (Da *et al*, 2012, 2013).

To measure the rate of PP_i release in a multi-subunit RNAP, (Malinen *et al*, 2012) employed the MDCC-labeled phosphate binding protein that has been validated for use in rapid kinetic measurements (Brune *et al*, 1994, 1998) and coupled it to yeast inorganic pyrophosphatase (PPase). The release of PP_i occurred 5-7 milliseconds after the incorporation of the NMP (rate ~120 s^{-1}) indicating that PP_i had a substantial retention time in the active site (see the experimental part of the thesis). Importantly, the release of PP_i preceded the translocation of RNAP measured in parallel experiments. As translocation was shown to require opening of the active site, the release of PPi was suggested to coincide with the opening of the active site (Malinen *et al*, 2012).

In principle, PP_i release could thermodynamically drive conformational changes. However, the ~0.5 mM intracellular PP_i concentration in *E. coli* (Heinonen, 2001) could possibly saturate the active site in RNAP and prohibit this step from performing any work for the system. An observation of transcription taking place at 0.5 mM PP_i (albeit at a reduced rate) indirectly suggests that TL is capable of unfolding even in the presence of a high concentration of PPi (this work and (Hein *et al*, 2011)). The questions if – and how – the release of PP_i, the opening of the active site and translocation are coupled, are addressed in the experimental part of this thesis.

2.2.5 Transcription fidelity

Incorporation of a incorrectly aligned nucleotide into the RNA chain results in an erroneous transcript of the DNA that can produce several defective protein copies, which in turn can cause damage to the cell and the organism (Imashimizu *et al*, 2013b; Strathern *et al*, 2012) or produce phenotypic diversity (Gordon *et al*, 2009). Misincorporation events also cause transcriptional pausing and RNAP backtracking *in vivo*, and consequently, the stalling of RNAP may lead to collisions with the replisome that can cause genotoxic DNA double strand breaks (Imashimizu *et al*, 2013b, 2015; James *et al*, 2016; Dutta *et al*, 2011). Multiple mechanisms have evolved to allow RNAP to select the correct nucleotides, detect the misincorporation events and remove the errors (Sydow & Cramer, 2009). As a result, the frequency of transcription errors found in mRNAs is only one per 10^5-10^6 nucleotides (Gout *et al*, 2013; Imashimizu *et al*, 2013b; Irvin *et al*, 2014). For unknown reasons, CpG dinucleotides are hotspots for transcription errors introducing frequent G-to-A misincorporation events (Imashimizu *et al*, 2015; James *et al*, 2016). Another type of transcription error is "slippage" in which RNAP confuses the DNA register by reading one template nucleotide multiple times or skipping it altogether. This leads to frame-shift mutations in the final mRNA. During slippage the RNA strand slides upstream or

downstream, while the DNA register remains unchanged. Slippage can happen in homo-polymeric DNA template regions containing weak RNA:DNA hybrids (for example dT:A), where the base pairing remains unchanged upon moving the RNA register (Zhou *et al*, 2013).

The selection of the correct NTP depends on the non-covalent interactions between the substrate and the transcription elongation complex. NTPs initially bind to the open active site with high affinity and specificity (Temiakov *et al*, 2005; Vassylyev *et al*, 2007b), but the energy from NTP base pairing and stacking in the pre-insertion configuration alone are insufficient to achieve the low frequency of incorrect nucleotide incorporation (Wang *et al*, 2006; Blank *et al*, 1986). Thus, the initial binding of the NTP in the pre-insertion configuration is an effective fidelity checkpoint only for few nucleotide mismatches (Yuzenkova *et al*, 2010). On the other hand, the closing of the active site through the formation of the THB is a major checkpoint for substrate selection and incorporation (Wang *et al*, 2006; Yuzenkova *et al*, 2010; Kellinger *et al*, 2012; Kireeva *et al*, 2008; Kaplan *et al*, 2008). The correct Watson-Crick pairing between the NTP and the template DNA orients the nucleotide in a position that optimally interacts with the TH in the catalytic configuration (Kellinger *et al*, 2012). Shape complementarity between substrate and the closed active site (**Figure 5**) is critical for the stability of the THB and the sensitivity of interactions allow only the correct NTP to stabilize the THB. This way nucleotide recognition is coupled to catalysis (Wang *et al*, 2006). An incorrect nucleotide incorporation reaction may actually have a different – and catalytically less optimal – transition state structure (Hwang *et al*, 2016). Furthermore, an incorrectly paired NTP is misaligned and may clash with the helical TL, which promotes its ejection (Yuzenkova *et al*, 2010).

RNAP discriminates between ribonucleotides (NTPs) and 2´deoxyribonucleotides (2´dNTPs) 100-1200-fold (Svetlov *et al*, 2004). β´Asn^{458} at the base of the A-site (**Figure 5**, right) hydrogen bonds with the 3´ OH and 2´ OH groups of the NTP both in the pre-insertion and insertion complexes (Vassylyev *et al*, 2007b; Kettenberger *et al*, 2004), but its role in discriminating between different sugars remains debated (Yuzenkova *et al*, 2010; Svetlov *et al*, 2004). On the other hand, β´Arg^{425} (**Figure 5**, right) bridges the NTP´s 2´ OH group and the RNA 3´ end, and a substitution mutation in this residue was lethal in yeast, suggesting it may play a role in discrimination (Vassylyev *et al*, 2007b; Wang *et al*, 2006). The active site retains significant 2´dNTP discrimination in the absence of the TL (Zhang *et al*, 2010), but the formation of the THB brings additional discriminatory power (Yuzenkova *et al*, 2010). The formation of the THB may allow the β´Gln^{929} (shown in Figure 25, p. 99) to interact with the 2´ OH group. Accordingly, a substitution to the corresponding glutamine in *Tth* RNAP decreased the 2´dNTP discrimination by 25-fold (Yuzenkova *et al*, 2010). Overall, β´Arg425 and β´Gln929 may create a molecular surface that can discriminate between the 3' endo conformation of ribose in NTPs and 2' endo conformation of deoxyribose in 2´dNTPs as well as the overall shapes of the DNA:DNA and RNA:DNA helices (Wang *et al*, 2006; Yuzenkova *et al*, 2010).

The high ratio of selection of correct nucleotides over incorrect ones is also maintained by kinetic discrimination in which the active site closing serves a central role. The active site may

probe interactions with the NTP over multiple cycles of TL folding and unfolding before catalysis takes place because the formation of THB is reversible and presumably faster than the formation of the chemical bond (Kireeva *et al*, 2008): Estimates based on the kinetic modelling of nucleotide addition in *S. cerevisiae* RNA polymerase II suggest that with the correctly paired NTP (here CTP), the formation of THB (k_{THB} 1200 s^{-1}) is faster than the dissociation of the NTP (k_{off} 100 s^{-1}). Once the THB was formed, the NTP could not dissociate before the relaxation of the THB (k_{rel} 300 s^{-1}). As a result, the active site effectively sequestrated the bound NTP, eventually initiating the catalysis (k_{pol} 100 s^{-1}). In this analysis the nucleotide incorporation step (k_{pol}) was 3-fold slower than the relaxation of the THB (k_{rel}) implying that the THB could form and unfold several times before the catalytic reaction commenced. On the other hand, an incorrectly paired NTP (here UTP) had a higher dissociation rate (k_{off} 10 000 s^{-1}) than the THB formation rate (k_{THB} 13 s^{-1}) and also a lower catalysis rate (k_{pol} 0.2 s^{-1}) because of suboptimal interactions in the active site and its inability to stabilize the THB. These combined effects prevent the stable retention of the mismatched NTP in the active site and make its ejection more likely than its incorporation (Kireeva *et al*, 2008). Accordingly, RNAP mutations that slow down the opening of the active site make the enzyme more error-prone (Kireeva *et al*, 2008; Nedialkov *et al*, 2013; Larson *et al*, 2012). However, while the above rate estimates are consistent with many observations, it should be emphasized that the rate of THB formation and relaxation were not directly measured experimentally and remain undetermined to this date.

In case a mismatched nucleotide evades the selection and gets incorporated into the RNA it can be detected and corrected though the post-catalytic proof-reading pathway involving RNAP backtracking and the endonucleolytic cleavage of RNA. These mechanisms are discussed in section 2.4.

2.3 Translocation

2.3.1 Structural transitions during RNAP translocation

To access the information in linear DNA molecules, RNA polymerases translocate along the template DNA strand in the 3´ to 5´ direction unveiling the DNA sequence one base at a time (Abbondanzieri *et al*, 2005). The TEC cycles between the pre-translocation and the post-translocation states (**Figure 6**) that are well defined both structurally and biochemically: The pre-translocation state is formed after the incorporation of the ribonucleotide (or following translocation backwards from the post-translocated state) when the nascent RNA 3´ end occupies the nucleotide binding site (the A-site) (Gnatt *et al*, 2001; Liu *et al*, 2016) (**Figure 4, state 6**). The RNA 3´ end makes contacts with the helical TL similarly as the substrate NTP in the insertion complex (Malinen *et al*, 2012; Basu *et al*, 2014; Liu *et al*, 2016). In the pre-translocation register, the nascent RNA can react with PP_i and excise a nucleotide from the RNA to generate a NTP (**Figure 4, state 5**). The tendency of TECs to undergo pyrophosphorolysis has previously been the standard biochemical method for measuring occupancy of the pre-translocated state (Kashkina *et al*, 2006; Hein *et al*, 2011; Artsimovitch & Landick, 2000). A one-register translocation of the RNA:DNA hybrid forward from the pre-translocated state into the post-translocated state frees space for the entry of the new template DNA nucleotide into the active

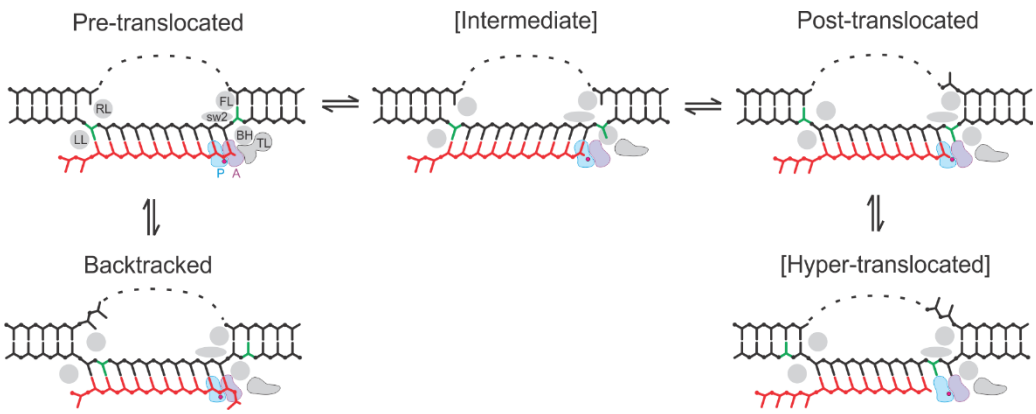


Figure 6. RNAP translocation states. Transcription goes from left to right. The DNA is black, RNA is red. The catalytic Mg^{2+} (A) (red sphere) is shown in the active site with the A- and P-sites colored purple and cyan, respectively. Template DNA bases that melt during the forward translocation at the upstream RNA:DNA hybrid and the downstream DNA are colored green. The single stranded non-template DNA is depicted with a dashed line. RNAP structures described in the text are colored in grey and abbreviated as follows: BH, bridge helix; TL, trigger loop; FL, fork loop 2; sw2, β´ switch 2; RL, rudder loop; LL, lid loop. TL is depicted in two alternative forms corresponding to the helical (pre-translocated state) and unfolded conformations (other translocation states). The backtracked state is depicted to form from the pre-translocated state although this has not been demonstrated directly. The intermediate translocation state is hypothetical (brackets) and combines features from (Brueckner & Cramer, 2008; Liu *et al*, 2011; Cheung *et al*, 2011). Likewise, the hyper-translocation state is hypothetical.

site (**Figure 4, state 1**). The RNA 3´ end now occupies the nucleophilic site (P-site) where it does not contact the BH or the TL. Instead, the RNA 3´ OH group coordinates the Mg^{2+} (A) at the P-site (Vassylyev *et al*, 2007a). This interaction stabilizes the RNA in the post-translocated register (Malinen et al. 2012). The templating DNA base directs the binding of the complementary NTP in the active site, which culminates in the structural isomerization of the RNAP into the catalytic conformation.

Translocation beyond the pre- and post-registers leads to the "off-pathway" backtracked (**Figure 6** and **Figure 4, state 10**) and possibly hyper-translocated states (discussed in sections 2.4 and 2.5). The register and contacts of the RNA 3´ end vary noticeably in different translocation states, whereas the overall trajectory of the template DNA is more uniform, with the exception of the putative translocation intermediate states (see below).

Translocation entails rearrangements of non-covalent interactions in DNA, RNA and RNAP. RNA polymerase unwinds one base pair of double stranded DNA at the downstream site (Wang *et al*, 2006; Kettenberger *et al*, 2004; Kashkina *et al*, 2007), while simultaneously one base pair is re-established when the DNA strands anneal 10 bp upstream (Turtola & Belogurov, 2016) (**Figure 6**). During the translocation motion, extensive ionic interactions between the RNAP and the backbones of nucleic acids are broken and re-established (Gnatt *et al*, 2001). DNA melts at *i+2*, which is just before the template DNA base enters the active site (Kashkina *et al*, 2007; Vassylyev *et al*, 2007a; Kireeva *et al*, 2011) (**Figure 7**). A 90 degree kink in the template DNA separates the

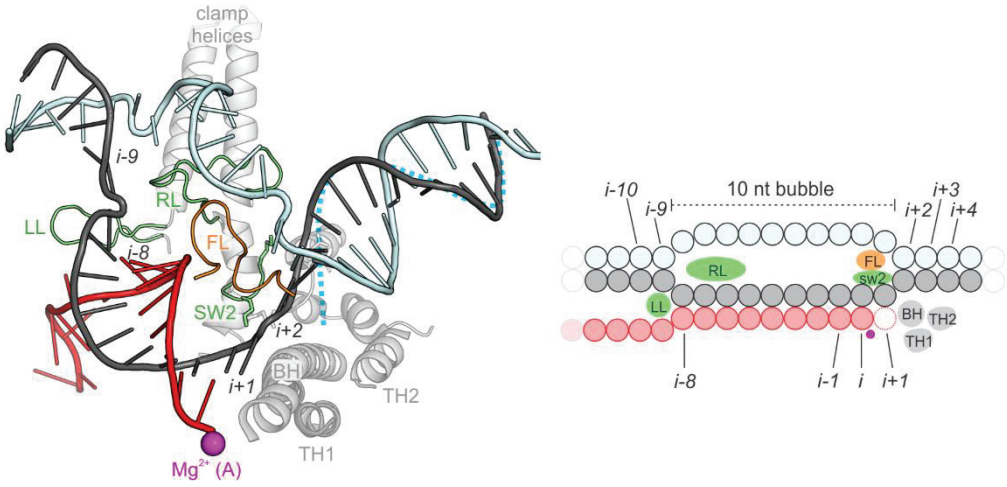


Figure 7. The overall view of the transcription bubble and the unwinding of the downstream DNA. The interactions of RNAP with the template strand induce distortion of the DNA duplex (between *i+2* and *i+4*) and facilitate its unwinding. The approximate path of the canonical B-form DNA is shown as a dashed blue line aligned with the downstream template DNA strand. The RNA:DNA hybrid separates at *i-8* and the DNA strands re-anneal at *i*-9. The template DNA strand is black, non-template DNA is cyan, RNA is red. The catalytic Mg^{2+} (A) is shown in the active site as a purple sphere. Structural β´ regions discussed in the text are colored green (sw2, switch 2; RL, rudder loop; LL, lid loop), the β fork loop 2 (FL2) is colored orange. Critical positively charged residues in switch 2 are shown as sticks; β´Lys^{334} discussed in the Results section 5.4.4 contacts the DNA backbone between the *i+2* and *i+1* sites. BH, bridge helix; TH1, trigger helix 1; TH2, trigger helix 2. The transcription bubble schematics with nomenclature for nucleic acid sites relative to RNAP are shown on the right. These sites are independent of the translocation state of the TEC. PDB code: 6alf (Kang *et al*, 2017).

double stranded downstream DNA and the RNA:DNA hybrid (Gnatt *et al*, 2001; Vassylyev *et al*, 2007a). DNA unwinding is facilitated by the distortion of the B-form DNA conformation between the residues located at the *i+2* and *i+4* sites (see the dashed line in **Figure 7**). The deviation in the DNA path is caused by the positively charged residues in the switch 2 domain that attract the template strand. Furthermore, the fork loop 2 obstructs the path of the non-template strand in double helical DNA preventing the re-association of the DNA (Kettenberger *et al*, 2004; Wang *et al*, 2006).

Biochemical and structural data suggest that the template DNA anneals with the non-template strand immediately after emerging from the main channel between the lid and the rudder loops (at *i-9*) (Turtola & Belogurov, 2016; Kang *et al*, 2017) (**Figure 7**). Within the moving DNA bubble, RNA polymerase maintains a constant 9-10 bp RNA:DNA hybrid that screws through the enzyme (Nudler *et al*, 1997). The RNA:DNA hybrid is 10 bp in length after nucleotide incorporation but shortens to 9 bp following translocation (Turtola & Belogurov, 2016; Kang *et al*, 2017; Vassylyev *et al*, 2007a). The separation of the RNA:DNA hybrid is facilitated by several interactions: The RNA:DNA hybrid clashes with the lid loop (Kent *et al*, 2009; Toulokhonov & Landick, 2006; Naryshkina *et al*, 2006) and its hydrophobic residues may provide a hydrophobic surface for separating the RNA and DNA bases (Westover *et al*, 2004a; Kettenberger *et al*, 2004; Kang *et al*, 2017; Turtola & Belogurov, 2016). The switch 3 forms a pocket that can trap the first single

stranded RNA nucleotide (Vassylyev *et al*, 2007a), and furthermore, the non-template DNA strand promotes the separation of the RNA:DNA hybrid by re-annealing with the template DNA strand (Toulokhonov & Landick, 2006; Turtola & Belogurov, 2016; Naryshkina *et al*, 2006). The rudder loop is critical for the overall stability of the TEC but, it is not required for the separation of the RNA:DNA hybrid (Kuznedelov *et al*, 2002; Turtola & Belogurov, 2016).

Translocation intermediates (**Figure 6**) between the canonical pre- and post-translocated states have been observed in structural studies (Brueckner & Cramer, 2008; Cheung *et al*, 2011; Weixlbaumer *et al*, 2013) and MD simulations (Silva *et al*, 2014; Feig & Burton, 2009). The proposed models collectively suggest that the RNA:DNA hybrid and the downstream DNA translocate asynchronously: The RNA:DNA hybrid first translocates to the post-translocated position accompanied by melting of the downstream DNA and partial translocation to the "pre-templating" position (see **Figure 6** and **Figure 4, state 7**). Then at a later step the downstream DNA slides forward to complete the translocation process. In the intermediate conformation where the RNA occupies the post-translocated register but the DNA resides in an incompletely translocated position the RNA:DNA hybrid bases are tilted relative to the long axis of the helix (Cheung & Cramer, 2011; Cheung *et al*, 2011; Weixlbaumer *et al*, 2013).

As a closer inspection of such states, an intermediate translocation state of the RNA polymerase II TEC was trapped with the transcription inhibitor α-amanitin (Brueckner & Cramer, 2008) (**Figure 4, state 7**). In the absence of the inhibitor the nucleic acids were observed to oscillate within one register, whereas the addition of α-amanitin biased the RNA:DNA hybrid towards the post-translocated state in the crystals. The downstream DNA occupied an intermediate translocation state wherein the template DNA acceptor base (the base that would pair with NTP in the post-state) was unpaired and stacked against the BH between the pre-translocated and post-translocated registers in a "pre-templating position". The central BH was stabilized in an upstream shifted position by a wedged TL, which blocked the entry of the template DNA into the actives site. Similar blocking of the active site by the bent BH was observed in the paused TEC structure (Weixlbaumer *et al*, 2013). It was hypothesized that the completion of TL unfolding from the wedged conformation would destabilize the shifted BH and move it downstream to allow the templating base to enter the active site. Then, the sliding of the downstream DNA would complete the translocation step (Brueckner & Cramer, 2008). Later, MD simulations suggested that crossing this type of an intermediate state would be the rate limiting step along the translocation pathway. A conserved tyrosine in the BH was suggested to that stack with the base of the translocating nucleotide and thereby accelerate the rate of translocation (Silva *et al*, 2014).

It should be emphasized that the presence of above mentioned intermediate conformations in the non-paused transcription elongation complexes remains unknown, and therefore the "on-pathway" translocation intermediates are hypothetical. However, the observation and kinetic characterization of the intermediate translocation state(s) is presented in the experimental part of this thesis.

2.3.2 Thermodynamic basis of translocation

RNAP is a powerful motor protein that tracks long distances along the DNA and can transcribe against a ~30 pN opposing force (Neuman *et al*, 2003; Dalal *et al*, 2006; Wang *et al*, 1998). The RNAP movement during the nucleotide addition cycle entails coordinated conformational transitions including changes in multiple non-covalent interactions between RNAP and nucleic acids, as well as the NTP and PP_i ligands. Conformational transitions are energized by the incorporation of NTPs into the growing RNA chain and the liberation of PP_i. In motor proteins, the conformational transitions typically have a Gibbs free energy change (ΔG) of 20-30 kJ/mol, which is low enough for the ΔG=-40 kJ/mol of NTP hydrolysis to readily drive the reaction under physiological conditions (Purich, 2010). The NTP hydrolysis that releases pyrophosphate (PP_i) has an even larger ΔG =-50-60 kJ/mol (Heinonen, 2001). Irrespective of at which elementary reaction step this energy is harnessed and released, the directional translocation over multiple nucleotides is ultimately powered by the incorporation of NMPs into the growing RNA chain.

In the aqueous microenvironment the vibrating solvent molecules create thermal noise with the energy k_BT, which is small compared to the energy consumption of RNAP (10-20 k_BT from NTP hydrolysis). Due to the thermal noise, the nanoscale movements of motor proteins are described as "random walk". In addition, the viscous drag caused by the surrounding solution is so large that the system is brought to mechanical equilibrium immediately. In effect, motor proteins create coordinated conformational changes by utilizing the energy from NTPs, while borrowing the momentum from the thermal energy (Purich, 2010; Astumian, 1997).

The central questions regarding the translocation mechanism are if, and how, is the chemical energy in the NTP coupled to the disruption or creation of new non-covalent interactions during RNAP translocation? The "power stroke" model states that translocation is thermodynamically and mechanically coupled to the release of PP_i (Shimamoto, 2013) (**Figure 8A**). For example, the structures of the single subunit T7 RNA polymerase in different stages of the nucleotide addition cycle suggested that PP_i release was linked to a conformational change – movement of the fingers domain – that was coupled to translocation (Yin & Steitz, 2004). In this model the transition state of the translocation motion coincides with PP_i release. However, experiments and theoretical considerations support the view that the translocation motion of RNAP *per se* does not require additional energy input, except for the thermal energy borrowed from the solvent (Komissarova & Kashlev, 1997a; Bar-Nahum *et al*, 2005; Abbondanzieri *et al*, 2005; Bai *et al*, 2007; Herbert *et al*, 2008; Feig & Burton, 2010; Shimamoto, 2013; Imashimizu *et al*, 2014; Silva *et al*, 2014; Zhang *et al*, 2016; Guajardo & Sousa, 1997). RNAP translocation is described to operate by a "thermal ratchet" mechanism, which implies that interconversions between the translocation states are driven by thermal motion (discussed further in section 2.3.4).

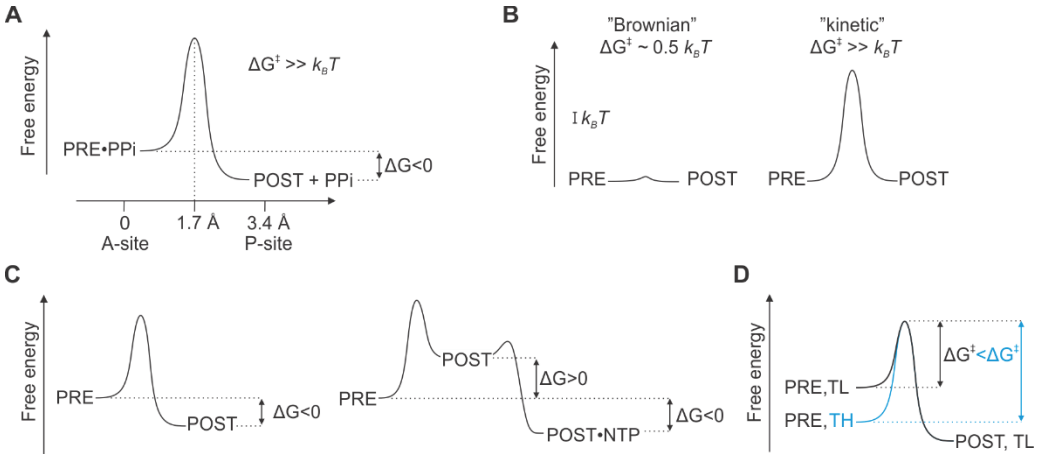


Figure 8. Thermodynamic basis of translocation. A) The power stroke translocation mechanism. Here, by definition, the $\Delta G^{\ddagger} >> k_B T$ and $\Delta G<0$. The x-axis shows the trajectory of the RNA 3´-terminus migrating from the A-site to the P-site. PP_i release coincides with the translocation transition state. Here, the translocation transition state is shown half-way between the A- and P-sites as is typically assumed; however, the most unstable state along the translocation trajectory does not necessarily need to be at the half distance. **B)** In the thermal ratchet translocation mechanism the oscillation between the pre-translocated and post-translocated states is driven by the energy borrowed from the thermal noise. The mechanism is viewed as "Brownian" when the $\Delta G^{\ddagger}$ is within the energy of Brownian motion ($\Delta G^{\ddagger} > \sim 0.5\ k_B T$) and "kinetic" when $\Delta G^{\ddagger} >> k_B T$. The distinction is important because the Brownian ratchet mechanism cannot be described by the transition state theory. **C)** The translocation bias. In the left hand side graph translocation is biased towards the post-translocated state; the graph on the right shows a situation where NTP binding is required to bias translocation towards the post-translocated state. **D)** Stabilization of the pre-translocated state by the helical trigger loop (TH). The unfolded trigger loop (TL) cannot stabilize the pre-translocated state which lowers the energy barrier ($\Delta G^{\ddagger}$) for forward translocation. This does not affect the stability of the post-translocated state

It is important to differentiate between two versions of the ratchet type of translocation mechanism, which are distinguished by the heights of their energy barriers (**Figure 8B**). A thermal ratchet mechanism (typically referred as the "Brownian ratchet") implies that the translocation oscillation happens within the regime of the Brownian motion (~0.5 $k_B T$), that is, the translocation states are not separated by a definable energy barrier (Gibbs free energy of activation, $\Delta G^{\ddagger}$). In this version, the outcome of the translocation motion is dependent on the individual molecular trajectories, and classical kinetics relying on the transition state theory do not apply. When the $\Delta G^{\ddagger}$ is >2-3 $k_B T$ the rate constants can be assigned to transitions. This is the kinetic version of the ratchet mechanism. These principles apply to all models that assume thermally driven motion and are independent of their mechanistic details (Shimamoto, 2013).

Any translocation model must not only describe the basis of overcoming the energy barrier but also explain the molecular basis for the direction of translocation. The key concept, translocation bias (**Figure 8C**), refers to the relative thermodynamic stabilities of the pre-translocation and the post-translocation states. The negative value for the change in Gibbs free energy ($\Delta G<0$) implies that translocation is biased towards the post-translocated state, whereas

ΔG>0 implies a bias for the pre-translocated state. The binding of ligands (for example NTP, PP_i or inhibitors) can affect the ΔG. As will be explained in the following sections, the active site of RNAP interacts with the substrate (NTP) with higher affinity than the product (extended RNA), therefore thermodynamically favoring forward translocation (ΔG<0). Thereby, the chemical energy in NTP is converted into a bias between forward and backward motion. In order to understand the molecular and thermodynamic basis of how RNA extension is coupled to RNAP translocation, it is necessary to describe the main structural states along the translocation trajectory, their relative thermodynamic stabilities and the kinetic barriers that separate the states.

2.3.3 Overcoming the translocation energy barrier

To translocate forward, RNAP must cross an energy barrier that prevents it from freely diffusing along the DNA track. The height of the barrier, and hence the contribution of the translocation step to the overall elongation rate has been addressed in several studies that mainly utilized single-molecule techniques and force-velocity relationships. When the thermally driven translocation represents a rate-contributing step in the nucleotide addition pathway, the overall (pause-free) elongation rate should be sensitive to force. In this scenario, an opposing force inhibits NTP binding in a manner similar to a competitive inhibitor (Dangkulwanich *et al*, 2013). Indeed, the transcription elongation rate was observed to depend on force even at saturating NTP concentrations (Abbondanzieri *et al*, 2005; Bai *et al*, 2007; Larson *et al*, 2012; Dangkulwanich *et al*, 2013). In contrast, other single molecule studies found negligible or no effect of force on the elongation rate, arguing that translocation is not a rate-contributing step (Wang *et al*, 1998; Forde *et al*, 2002; Neuman *et al*, 2003; Galburt *et al*, 2007; Mejia *et al*, 2015). However, these inferences are weakened by the technical limitations of the current single molecule methodologies: translocation is not measured directly, and further, while the measurements of the pause-free elongation rate are obtained through filtering out the long pauses, the short pause are left undetected, which unavoidably affects the inferred parameters.

On the other hand, direct measurements of translocation kinetics with a fluorescence-based assay indicated that translocation takes place at a millisecond time scale and its rate is comparable to the nucleotide addition rate (Malinen *et al*, 2012, 2014), favoring the kinetic version of the ratchet mechanism. Similar rate estimates were inferred from the comparison of elongation rates of single RNA pol II molecules that transcribed either bare or nucleosomal DNA (Dangkulwanich *et al*, 2013). Furthermore, MD simulations also detected distinct kinetically definable translocation states, but the interconversions between the pre-translocated and post-translocated states happened in a time scale of microseconds (Silva *et al*, 2014), which is three orders of magnitude faster than observed in experiments (Malinen *et al*, 2012, 2014). The current downside of these simulations is that they were limited to existing TEC structures that lacked the full non-template DNA. They also failed to capture major conformational changes associated with the translocation pathway including the TL folding dynamics and changes in DNA:DNA and RNA:DNA base pairing, which may limit the rate of translocation. These missing features may accelerate the rates directly or allow flexibility that is not possible in complete

TECs. Nevertheless, these studies collectively demonstrate that kinetically definable energy barrier separates the pre-translocated and post-translocated states (Malinen *et al*, 2012; Dangkulwanich *et al*, 2013), thereby providing strong support for the kinetic version of the ratchet mechanism (see the previous section 2.3.2).

It remains unestablished what molecular events limit the translocation rate. PP_i release preceded or coincided with translocation, indicating that opening of the active site may limit the rate of translocation (Malinen *et al*, 2012) (discussed further in section 2.3.6 below). This issue is centrally addressed in the experimental part of this thesis. Furthermore, because the translocation energy barrier that separates the pre-translocation and post-translocation states appears to be well above the regime of the Brownian motion (>>2-3 k_bT), it is conceivable that translocation intermediates analogous to those discussed in section 2.3.1 could be detected also kinetically. The observation of such intermediates in the experimental part of this thesis suggests, that RNAP translocation encompasses at minimum two separate energy barriers (section 5.4).

2.3.4 A two-pawl ratchet model of translocation

Molecular motor proteins have structural features that allow them to bias Brownian fluctuations into a net movement in one direction. The molecular biasing device at the heart of such movement is a type of a ratchet connected to a pawl. The pawl interacts asymmetrically with the ratchet so that the ratchet moves more favorably in one direction. However, as the pawl itself is subject to thermal fluctuations, such device cannot create unidirectional movement without an external energy source. Molecular motors couple the ratchet´s movement to a non-equilibrium chemical reaction, such as NTP hydrolysis and condensation, and thereby create the molecular asymmetry required for biasing the movement. In effect, the energy from the chemical reaction is used to harness the thermal noise to create directional motion (Astumian, 1997; Shimamoto, 2013).

In a TEC, the ratchet is the RNA:DNA hybrid, which thermally fluctuates within the RNAP between pre-translocated and post-translocated states. The BH has been suggested to be the structural analogue of the pawl (Bar-Nahum *et al*, 2005; Epshtein *et al*, 2002). The BH forms the central element of the active site that spans the main channel at the interface between the RNA:DNA hybrid and the downstream DNA. The BH was observed in straight (Gnatt *et al*, 2001) and bent forms (Zhang *et al*, 1999; Vassylyev *et al*, 2002) and the 3-4 Å oscillations between these forms were proposed to promote translocation (Tuske *et al*, 2005; Cramer *et al*, 2001; Gnatt *et al*, 2001; Epshtein *et al*, 2002; Bar-Nahum *et al*, 2005; Vassylyev *et al*, 2002). The bent BH occludes the active site and in effect blocks the RNA:DNA hybrid from entering the pre-translocated register. According to the early translocation model, the straightening of the BH without a concomitant hybrid backward motion would create the space required for the translocation of a new template DNA base into the active site (Gnatt *et al*, 2001). Oscillations due to bending of the BH were observed in nucleic acid-protein crosslinking experiments (Epshtein *et al*, 2002) and in MD simulations (Kireeva *et al*, 2012; Silva *et al*, 2014). However, it

should be emphasized that in a fluctuating system the operation of the ratchet and pawl are stochastic without one-to-one correspondence between the motions of the pawl and ratchet (Silva *et al*, 2014; Shimamoto, 2013).

The binding of NTP in the active site biases translocation in the forward direction (Bar-Nahum *et al*, 2005; Malinen *et al*, 2014; Dangkulwanich *et al*, 2013) and can trap the post-translocated state (Nedialkov *et al*, 2003; Kireeva *et al*, 2008; Nedialkov *et al*, 2013). A "two-pawl ratchet" model of translocation was envisaged, which included the NTP binding step as the major thermodynamic forward driving component (Bar-Nahum *et al*, 2005). In this model there are two thermally oscillating systems, the RNA:DNA hybrid and the BH. Both of these take two major conformations: the RNA:DNA hybrid oscillates between the pre-translocated and post-translocated registers (see Figures 3 and 5), and the BH oscillates between the bent and straight conformations. This makes four different states. Because the bent BH and the RNA:DNA hybrid in the pre-translocated register prevent the NTP from binding, the NTP can bind to only one state which has a straight BH and the RNA:DNA hybrid in the post-translocated register. The NTP bound post-translocated complex makes the fifth state, which can commence catalysis (Shimamoto, 2013).

2.3.5 Translocation bias

The two-pawl ratchet mechanism implies that NTP binding is necessary to bias translocation forward (**Figure 8C,** right). A single molecule analysis of *S. cerevisiae* RNA polymerase II transcription concluded that translocation was backward-biased (Dangkulwanich *et al*, 2013), supporting the model. However, another single-molecule study claimed that translocation was forward-biased (Larson *et al*, 2012), and all experimental TECs translocated forward in the absence of the next incoming nucleotide (Malinen *et al*, 2012). Furthermore, a number of experiments concluded that the resting and catalytically active TECs were predominantly post-translocated (Malinen *et al*, 2012; Nedialkov *et al*, 2012; Kashkina *et al*, 2006; Malinen *et al*, 2014; Esyunina *et al*, 2016b; Turtola & Belogurov, 2016). These experiments directly probed the translocation state with methods relying on fluorescence, exonuclease foot-printing or catalytic activities of the TEC. In addition, the majority of the transcription elongation complexes characterized by X-ray crystallography have been post-translocated, both in the absence and presence of nucleotides, as surveyed in (Liu *et al*, 2016). Therefore, the majority of recent studies suggests that TECs possess an inherent thermodynamic bias towards the post-translocated state. Therefore, while NTP binding necessarily biases RNAP forward, such an action may influence the overall elongation rate only at a subset of transcribed positions.

The translocation bias is influenced by how the downstream edge of the RNA:DNA hybrid interacts with RNAP in the A and P-sites (**Figure 9**). The open active site forms an asymmetric environment that thermodynamically (by ~3 k_bT) biases the TEC to occupy the post-translocated register in the absence of NTPs, suggesting that the RNA:DNA hybrid preferably interacts with the P-site (Malinen *et al*, 2012, 2014). Importantly, the translocation bias is lost when the interaction between the RNA 3′ OH group and the Mg^{2+} (A) is disrupted (Malinen *et al*, 2012,

2014) (see the experimental part of the thesis). Apparently, the preferable interaction between the post-translocated RNA and Mg^{2+} (A) at the P-site slows the rate of backward translocation below that of the rate of forward translocation causing the forward translocation bias. As discussed in the next section, closing of the active site inverts the translocation bias towards the pre-translocation state.

Interestingly, the translocation bias is influenced by the sequence of the RNA 3´ end (Hein *et al*, 2011; Malinen *et al*, 2012, 2014). The interaction preference of the RNA 3´ end with the P-site declines in the order GMP > AMP > CMP > UMP. On the other hand, the forward translocation rate was largely independent of the identity of the RNA 3´ end, suggesting that the interaction at the A-site is less influenced by the RNA sequence (Malinen *et al*, 2012).

2.3.6 Control of the translocation ratchet by closing and opening of the active site

The overall transcription elongation rate is altered by amino acid substitutions that change the folding/unfolding equilibrium or rate of the TL, or both (Mejia *et al*, 2015; Larson *et al*, 2012; Bar-Nahum *et al*, 2005; Kireeva *et al*, 2012; Tan *et al*, 2008; Nedialkov *et al*, 2013; Kireeva *et al*, 2008; Dangkulwanich *et al*, 2013). These observations are partly explained by effects on TL controlled NTP sequestration and on catalytic steps (see sections 2.2.2 and 2.2.3). Additionally, TL folding dynamics affect the rates of the post-catalytic steps (PP_i release and translocation) that constitute 30-50% of the catalytic cycle (Malinen *et al*, 2012, 2014; Dangkulwanich *et al*, 2013). Accordingly, the pre-translocated state becomes more populated and translocation is slowed down when the TL is stabilized in the helical state by amino acid substitutions or cysteine-pair crosslinking (Malinen *et al*, 2014; Windgassen *et al*, 2014). Counterintuitively though, while the stabilization of the helical TL decreases the translocation rate, the overall RNAP elongation rate increases due to alterations in RNAP pausing (Malinen *et al*, 2014).

The interaction strength between RNAP and the RNA:DNA hybrid in the pre-translocated register is dictated by the folding state of the TL (**Figure 8D**). In structural terms, the helical TL contacts the pre-translocated RNA 3´ end and restricts its forward translocation (Feig & Burton, 2009; Kireeva *et al*, 2012; Malinen *et al*, 2012; Feig & Burton, 2010) (**Figure 9,** left). The longer retention of the RNA 3´ end in the pre-translocated register in comparison to the post-translocated register biases the equilibrium towards the pre-translocation state. Release of PP_i is suggested to destabilize the helical TL and initiate the unfolding process (Malinen *et al*, 2012; Da *et al*, 2012, 2013; Wang *et al*, 2013). In the unfolded state the contacts between the TL and the RNA 3´ end are lost. This destabilizes the RNA in the pre-translocated register and allows forward translocation to take place (Feig & Burton, 2009; Kireeva *et al*, 2012; Malinen *et al*, 2012; Feig & Burton, 2010) (**Figure 9,** right).

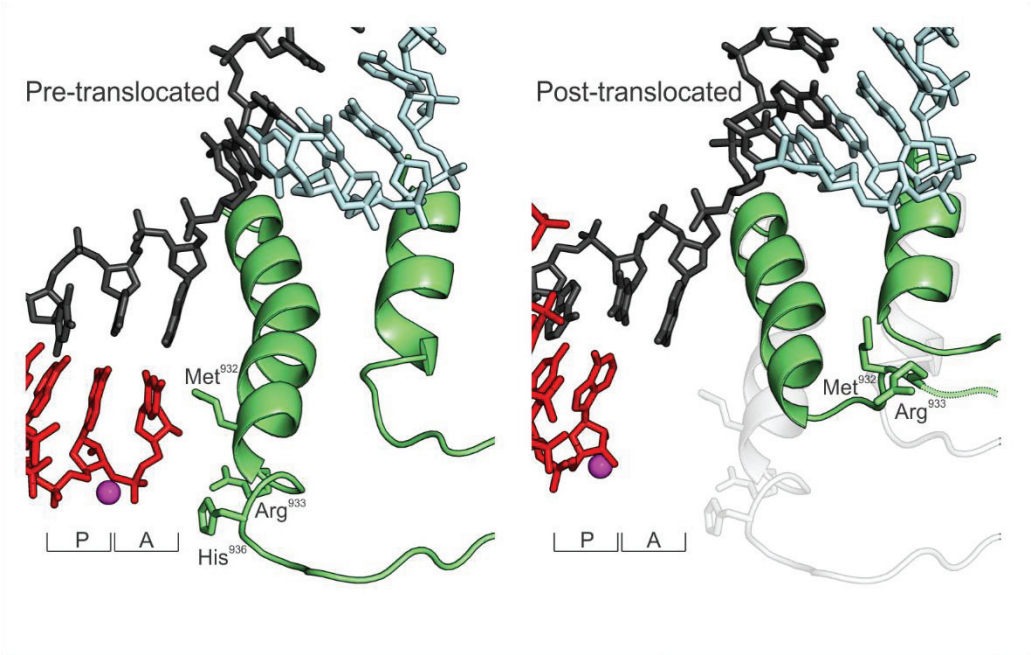


Figure 9. The alternative TL conformations control RNAP translocation. *Left*: the helical TL (*green*) stabilizes the pre-translocated state by projecting β´Met932 and His936 towards the A-site where they interact with the RNA (*red*) 3´-terminal NMP (PDB code: 5ipm (Liu *et al*, 2016)). *Right*: The unfolding of the TL (*green*; the helical state in *grey* is depicted for comparison) removes the contacts with the RNA 3´-terminus allowing the RNA:DNA hybrid to relax into the post-translocated state. The post-translocated state is stabilized by the interaction between the Mg^{2+} (A) -ion (*purple* sphere) and the RNA 3´-terminus in the P-site. PDB code: 6alf (Kang *et al*, 2017).

Small molecules can inhibit RNAP translocation by targeting the TL folding dynamics (Ma *et al*, 2016). STL stabilizes the post-translocated state by obstructing the formation of the helical TL (Tuske *et al*, 2005; Temiakov *et al*, 2005; Vassylyev *et al*, 2007b), whereas compounds of the CBR-series promote forward translocation by destabilizing the helical TL (Malinen *et al*, 2014; Bae *et al*, 2015; Feng *et al*, 2015). On the other hand, TGT stabilizes the closed active site and thereby stabilizes the pre-translocated state (Artsimovitch *et al*, 2011; Malinen *et al*, 2012; Yuzenkova *et al*, 2013). It is yet unknown whether cellular or viral elongation factors can also change the translocation properties of RNAP by targeting the TL folding dynamics.

2.3.7 The effect of the DNA sequence on translocation

It has long been appreciated that the RNAP machinery can "sense" the transcribed nucleotide sequence: the signals encoded within the RNA:DNA hybrid, the downstream DNA and the exiting RNA modulate the enzymatic activity of RNAP, assumedly by affecting translocation (Landick, 2001; Herbert *et al*, 2006; Hein *et al*, 2011; Bochkareva *et al*, 2011).

Genome-wide sequencing of the nascent RNA 3´ end positions within the transcribing RNAP revealed a common consensus motif that pauses transcription by interfering with the translocation of RNAP (Larson *et al*, 2014; Vvedenskaya *et al*, 2014; Imashimizu *et al*, 2015). The

three components of this motif are the G:dC and C:dG/U:dA pairs at the upstream and downstream ends of the RNA:DNA hybrid, respectively, and a dC in the template strand directing binding of the next incoming nucleotide (corresponding to the RNA sequence $G_{-10}Y_{-1}G_{+1}$; Y stands for pyrimidine). The pause sequence presumably induces an "elemental paused" state involving structural changes in the TEC that inhibit nucleotide addition: a kinked BH, an unfolded TL, an open clamp and possibly a frayed RNA 3´ end in the pre-translocated register (Weixlbaumer *et al*, 2013; Artsimovitch & Landick, 2000; Toulokhonov *et al*, 2007; Malinen *et al*, 2014; Ray-Soni *et al*, 2016; Nayak *et al*, 2013) (**Figure 4, state 8**).

The translocation state of the TEC in this elemental paused state remains unclear. Biochemical experiments indicate that RNAP is stabilized in the pre-translocated register (Larson *et al*, 2014; Vvedenskaya *et al*, 2014). However, a spacing of 9, 10 or 11 nucleotides between the upstream G and the downstream $Y_{-1}G_{+1}$ flanking the active site can stabilize the pause in the post- or pre-translocated or backtracked states, respectively (Imashimizu *et al*, 2015). On the other hand, structural characterization of a paused TEC revealed the RNA of the RNA:DNA hybrid to be in the post-translocated register whereas the template DNA was only partially translocated, indicating that the paused state could involve an intermediate translocation state (Weixlbaumer *et al*, 2013) (**Figure 4, state 8**). Overall, it remains unknown how the pause inducing sequences prevent RNAP progression or cause the above mentioned conformational changes, and whether the block in translocation is a cause or a consequence of pausing.

Interestingly, the pause components locate to positions where the bases un-pair in the course of translocation. During a forward translocation step, two base pairs are broken (in the downstream DNA and in the RNA:DNA hybrid) and one new is formed (in the upstream DNA). It is likely that these changes in base-pairing contribute to the translocation energy barriers. In support of this, the artificial disruption of the downstream DNA base pairing facilitated forward translocation in *S. cerevisiae* RNA polymerase II (Kireeva *et al*, 2011). To what extent the melting of base pairs affects the translocation kinetics is not known, but this must depend on hydrogen bonding and stacking energies (which depend on the base sequence) as well as the possible sequence specific interactions between nucleic acids and RNAP. Indeed, the pause components locate to positions where the bases can possibly directly interact with the protein chains of RNAP (Wang *et al*, 2015; Vvedenskaya *et al*, 2014; Petushkov *et al*, 2015). The effect of the transcribed sequence on RNAP translocation is addressed in the experimental part of the thesis.

2.4 Translocation in reverse: RNAP backtracking

2.4.1 Structural basis of backtracking

The backwards translocation of RNAP beyond the pre-translocated state disengages the based paired RNA 3´ end from the template DNA and extrudes the nascent RNA into the secondary channel (Nudler *et al*, 1997; Komissarova & Kashlev, 1997b, 1997a) (**Figure 6**). In the resulting backtracked state RNA extension is prevented by the nascent RNA that blocks the active site. Backtracking pauses transcription until the RNA 3´ end is re-aligned into the active site of RNAP.

Crystal structures of backtracked TECs show that the backtracked state has a discrete and stable structure (Wang *et al*, 2009; Cheung & Cramer, 2011; Sekine *et al*, 2015). The RNA backbone is sharply bent between the last two 3´ terminal nucleotides. The last base paired nucleotide is positioned in the A-site, but the plane of the base pair is tilted 15-30° relative to the hybrid axis. The first unpaired nucleotide occupies the "proofreading site" located at the base of the secondary channel (**Figure 4, state 10**). This site is walled by the BH, a loop (βDloopII) and a helix from the β-subunit, and the RNA itself. In a one-nucleotide backtracked TEC the TL adopted a conformation between the open and closed states (Wang *et al*, 2009; Sekine *et al*, 2015) allowing the RNA to occupy its position in the proofreading site (the fully helical TL would clash with the RNA nucleotide in this site). More extensively backtracked RNA traps the TL in an unfolded state. RNA is stabilized inside the secondary channel by several ionic and a few hydrophobic interactions (Cheung & Cramer, 2011). Although RNAP pausing may be associated with the opening of the clamp domain (Kolb *et al*, 2014; Hein *et al*, 2014; Weixlbaumer *et al*, 2013), the clamp was closed in the crystallized backtracked TECs (Sekine *et al*, 2015; Tagami *et al*, 2010; Cheung & Cramer, 2011; Wang *et al*, 2009).

2.4.2 Formation of the backtracked state

The backtracked state is thought to originate from a common pause intermediate, the elemental pause (Artsimovitch & Landick, 2000) (**Figure 4, state 8**). At these elemental pause sites RNAP undergoes a structural isomerization possibly involving the RNA 3´ end fraying in the pre-translocated register (**Figure 4, state 9**), BH bending and clamp opening (Artsimovitch & Landick, 2000; Toulokhonov *et al*, 2007; Zhang *et al*, 2010; Weixlbaumer *et al*, 2013). This isomerization may happen at sites that inhibit downstream DNA translocation (Belogurov & Artsimovitch, 2015; Ray-Soni *et al*, 2016). However, it cannot be excluded that backtracking does not require any specific and prior off-pathway isomerization *per se* and only necessitates the slowdown of the transcription to take place. In general, the slowdown of transcription elongation rate increases the occurrence of pauses because pausing and elongation are in kinetic competition (Forde *et al*, 2002; Dangkulwanich *et al*, 2013; Mejia *et al*, 2015).

Certain template sites cause more frequent backtracking (Nudler *et al*, 1997; Komissarova & Kashlev, 1997a, 1997b; Artsimovitch & Landick, 2000; Kireeva & Kashlev, 2009) indicating that the underlying DNA sequence affects the backtracking tendency. Backtracking has been viewed to take place in template positions which form thermodynamically unstable RNA:DNA hybrids. These would allow TECs to diffusively equilibrate towards another template position that forms a more stable RNA:DNA hybrid placing the backtracked TEC in a deep energy well (Komissarova & Kashlev, 1997b). In general, the thermodynamic stability of the TEC at each nucleotide position is determined by three free energy components: the destabilizing component from the opening of the transcription bubble, the stabilizing component from the hybridization of the RNA and the DNA, and the overall stabilizing component from interactions between RNAP and nucleic acids (Yager & von Hippel, 1991; Greive & von Hippel, 2005). As a result, different translocation states of the same TEC may have different thermodynamic stabilities.

While it has been posited that TECs capable of incorporating nucleotides are functionally homogeneous (Pasman & von Hippel, 2002; von Hippel & Pasman, 2002), it is unclear whether TECs can thermally equilibrate between the paused and catalytically active states at each nucleotide position. The models attempting to derive transcription elongation rates (and backtracking tendency) based on the thermodynamic stabilities of TECs at different translocation registers have assumed that the translocation oscillations are many-fold faster than the nucleotide addition rate and hence the TECs can equilibrate between the activated and inactivated (for example backtracked) states at each nucleotide position (Bai *et al*, 2004; Tadigotla *et al*, 2006). However, the observation of infrequent translocation oscillations (Malinen *et al*, 2012, 2014; Dangkulwanich *et al*, 2013) implies that during processive transcription some translocation states may be kinetically restricted albeit being thermodynamically favored. This is overall consistent with the low probability of backtracking (Shaevitz *et al*, 2003; Abbondanzieri *et al*, 2005; Galburt *et al*, 2007; Herbert *et al*, 2010; Turtola & Belogurov, 2016) and with the mechanistic view that efficient backtracking is facilitated by the elemental pause isomerization that itself delays the addition of nucleotides (Artsimovitch & Landick, 2000). Therefore, the relative rates of forward and backward translocation, backtracking and nucleotide addition determine how TECs partition between active and paused states at each nucleotide position (Greive & von Hippel, 2005; von Hippel & Pasman, 2002).

Single molecule experiments can distinguish backtracking from other pauses when RNAP reverse translocates ~3-5 nucleotides (Shaevitz *et al*, 2003; Abbondanzieri *et al*, 2005; Galburt *et al*, 2007). The half-time of these pauses is >20 s and they occur infrequently about 0.4-1 kb^{-1}. However, these "long pauses" constitute only ~5% of all detected pauses. The remaining "short pauses" have frequencies of about 10 kb^{-1} and half-times of 1.5-7 s (Shaevitz *et al*, 2003; Herbert *et al*, 2010). The conformation of RNAP at these "ubiquitous pauses" may resemble the elemental paused state (Neuman *et al*, 2003; Larson *et al*, 2011; Herbert *et al*, 2006; Kireeva & Kashlev, 2009) but experimental and theoretical considerations also suggest that they may constitute RNAPs backtracked by one or two nucleotides (Galburt *et al*, 2007; Depken *et al*, 2009). Clearly, new methods that could both spatially and temporally resolve the step(s) leading to the backtracked state formation would be required for uncovering the kinetic barriers that determine the partitioning of RNAP between active and paused states.

2.4.3 Factors affecting the formation of the backtracked state

The energetic barrier separating active and backtracked TECs can be lowered by RNA:DNA base pair mismatch in the RNA 3′ end (Zenkin *et al*, 2006; Yuzenkova & Zenkin, 2010; Sydow & Cramer, 2009; Larson *et al*, 2012; Turtola & Belogurov, 2016; Da *et al*, 2016; Toulokhonov *et al*, 2007; Komissarova & Kashlev, 1997a). Indeed, lowering the barrier for backtracking is thought to be a universal mechanism for recognizing transcription errors and correcting them by RNA editing in the backtracked state (Imashimizu *et al*, 2015; Sydow & Cramer, 2009; Wang *et al*, 2009; Irvin *et al*, 2014). Mismatches facilitate backtracking by promoting RNA fraying in the pre-translocated register (Sydow *et al*, 2009) (**Figure 4, state 9**). This (*i*) directs the unpaired RNA 3′ end towards the secondary channel, and (*ii*) prevents the folding of the TL into the helical form.

It is generally believed that the helical TL prevents backtracking as the alterations in the folding dynamics of the TL by mutations change the backtracking propensity (Larson *et al*, 2012; Dangkulwanich *et al*, 2013; Mejia *et al*, 2015; Nedialkov *et al*, 2013; Bar-Nahum *et al*, 2005). Further, BH bending was proposed to facilitate the backward translocation of a mismatched RNA 3´ end (Da *et al*, 2016).

The likelihood of RNA 3´ end misalignment and backtracking is increased by factors that prolong the RNAP´s dwell time in the pre-translocated state, either by inhibiting the forward translocation or NTP binding (or both) (Imashimizu *et al*, 2013a; Dangkulwanich *et al*, 2013). The increased dwell time may be caused by RNA:DNA mismatches which delay the incorporation of the next NTP (Sydow *et al*, 2009; Sydow & Cramer, 2009; Irvin *et al*, 2014), low levels of substrates (Larson *et al*, 2012; Forde *et al*, 2002; Dangkulwanich *et al*, 2013), inhibition of the forward progression of RNAP by chromatin binding proteins, including H-NS in bacteria (Kotlajich *et al*, 2015) and nucleosomes in eukaryotes (Kireeva *et al*, 2005; Bintu *et al*, 2012; Dangkulwanich *et al*, 2013), or a DNA lesion (Charlet-Berguerand *et al*, 2006; Xu *et al*, 2017). Also, DNA supercoiling generates torque forces in front and behind of RNAP that can stall transcription (Ma & Wang, 2014) by inducing backtracking (Kotlajich *et al*, 2015).

Several components interacting with the transcribing RNAP are known to affect its backtracking properties. An essential transcription elongation factor NusG, that binds the clamp helices of the β´ subunit (Martinez-Rucobo *et al*, 2011) (see Figure 7 p.27 and Figure 23 on p. 94), reduces the formation of the backtracked state by stabilizing the upstream DNA and preventing its melting in the pre-translocated state (Herbert *et al*, 2010; Turtola & Belogurov, 2016). The translating ribosome or trailing RNAP molecules transcribing head to tail the same DNA region can physically prevent RNAP backtracking (Proshkin *et al*, 2010; Epshtein & Nudler, 2003). Furthermore, the secondary structures forming in the nascent RNA behind RNAP during transcription can also physically prevent backtracking (Komissarova & Kashlev, 1997b; Reeder & Hawley, 1996; Artsimovitch & Landick, 2000; Zamft *et al*, 2012).

On the other hand, the transcription initiation factor Sigma70 (σ^{70}) can induce backtracking of the early elongating RNAP in the promoter-proximal regions. Interactions between σ^{70} and the -10-like or -35-like DNA sequence elements and the transcribing TEC cause the formation of a stressed intermediate, where the DNA is “scrunched” meaning that single strands of DNA are looped out in the absence of space for forming the double helix. This intermediate relaxes either by dissociating the σ^{70} or by backtracking (Perdue & Roberts, 2010, 2011; Zhilina *et al*, 2012). During the early phase of the transcription cycle the absence of RNA secondary structures and translating ribosomes may further favor RNAP backtracking. σ^{70} binds to the clamp helices of the β´ subunit and is exchanged with the transcription elongation factor NusG during the transition from initiation to elongation (Mooney *et al*, 2009). As σ^{70} and NusG have opposite effects on backtracking, it is plausible to assume that the exchange of these factors during the transition from initiation to elongation phase is accompanied by decreased tendency of RNAP to enter the backtracked state.

2.4.4 Recovery from the backtracked state

Although the overall backtracking frequency is low, backtracking has a great impact on the overall elongation rate due to the long duration of these pauses (Dangkulwanich *et al*, 2013; Larson *et al*, 2012; Herbert *et al*, 2010; Yuzenkova *et al*, 2014; de Maddalena *et al*, 2016). For transcription to continue, the backtracked RNAP must re-align with the RNA 3´ end into the active site by either cleaving the RNA (see the next section) or translocating forward. In the absence of cleavage factors some backtracked pauses can last hundreds of seconds (Shaevitz *et al*, 2003) implying that the barrier for forward translocation is high. The recovery from the backtracked state is very sensitive to force and this effect defines the overall opposing force that the transcribing RNAP can tolerate (Galburt *et al*, 2007). Interestingly, the observation that *E. coli* RNAP is less sensitive to the opposing force than *S. cerevisiae* RNA polymerase II (Neuman *et al*, 2003; Wang *et al*, 1998) may hint to different structural properties of the backtracked state in these RNAPs. However, little is known about factors that govern the stability of the backtracked state, and this issue is addressed in the experimental part of the thesis.

The recovery from backtracked pauses typically does not exhibit single step kinetics (Artsimovitch & Landick, 2000; Galburt *et al*, 2007; Turtola & Belogurov, 2016). The distribution of pause lifetimes follows a power law suggesting that the paused RNAP diffuses between several backtracked states (Galburt *et al*, 2007). Structural studies have proposed that the interaction of the RNA with the P-site stabilizes the 1 nucleotide backtracked state and forms a barrier for further backtracking. Thereby the 1 nucleotide backtracked state would be readily reversible, whereas longer backtracks would cause long duration pauses due to the extensive contacts of RNA with the protein in the RNAP secondary channel (Wang *et al*, 2009; Cheung & Cramer, 2011). Consistently, the backtracking distance follows exponential distribution and in ~60% of instances RNAP backtracks less than 3 nucleotides before returning to elongation (Galburt *et al*, 2007). Finally, it was concluded that TL folding dynamics do not play a role in pause recovery (Mejia *et al*, 2015), but in this study we have reached the opposite conclusion (see Results).

The rescue of the backtracked RNAP can be facilitated by the translocase Mfd that is able to release the RNAP from the DNA template (Park *et al*, 2002). Mfd is part of the transcription coupled DNA repair machinery and it is responsible for removing stalled transcription complexes from sites of DNA damage (Adebali *et al*, 2017). DNA lesions prevent transcription elongation and induce backtracking effectively blocking the damage site. Mfd binds to the DNA residing immediately upstream of the TEC and employs its ATPase and translocase activities to move RNAP forward. The translocase and dissociation activities are related to the DNA rewinding at the upstream edge of the transcription bubble which may be coupled to the unwinding of the RNA:DNA hybrid and the forward translocation of RNAP (Park & Roberts, 2006). A second DNA repair protein, the DNA helicase UvrD was proposed to act in an opposite manner by inducing RNAP backtracking at sites of DNA damage in order to expose the damaged site to the DNA repair machinery (Epshtein *et al*, 2014). However, the role of UvrD in transcription coupled DNA

repair has been claimed to involve activities not related to RNAP displacement (Adebali *et al*, 2017).

2.4.5 Backtracked state catalyzes the endonucleolytic hydrolysis of the nascent RNA

RNAP has an intrinsic RNA cleavage activity, which enables the release of the backtracked portion of the RNA and re-alignment of the RNA 3´ OH in the active site (Orlova *et al*, 1995). This RNase activity hydrolyzes the phosphodiester bond located between the A and P-sites and therefore the cleavage happens in the same active site as the nucleotide addition reaction. The S_N2 type nucleophilic attack of a water or hydroxyl ion on the phosphodiester bond is catalyzed by the two-metal ion mechanism. The reaction is at least partially limited by the retention of the Mg^{2+} (B) ion (Sosunov *et al*, 2003, 2005). The RNA itself has been suggested to catalyze its own cleavage by stabilizing the Mg^{2+} (B) ion through interactions provided by the backtracked nucleotide (Zenkin *et al*, 2006).

The role of TL in the cleavage reaction is controversial. While in certain reaction conditions the catalysis of the cleavage reaction does not require TL (Zhang *et al*, 2010), several other studies found a central role of TL in accelerating the reaction (Yuzenkova & Zenkin, 2010; Esyunina *et al*, 2016b; Temiakov *et al*, 2005). It has been suggested that the conserved histidine at the tip of the TL acts as a general base by deprotonating the attacking water molecule (Yuzenkova & Zenkin, 2010). However, the relatively mild effects of histidine substitutions argue that also other functional groups participate in acid/base catalysis (Mishanina *et al*, 2017). Furthermore, in light of the current structural view of the backtracked state (Wang *et al*, 2009; Cheung & Cramer, 2011; Sekine *et al*, 2015) it is unclear how the tip of the helical TL would be able to fold and reach the scissile bond, while the RNA occupies the proofreading site. Apparently the RNA needs to change its conformation for the intrinsic cleavage to take place or the TL tip has to adopt a conformation distinct from the one supporting nucleotide addition.

The RNase activity is greatly stimulated by the RNA cleavage factors GreA and GreB (Borukhov *et al*, 1992; 1993). The cleavage factors insert an acidic domain through the secondary channel and stabilize the Mg^{2+} (B) ion in the RNAP active site (Laptenko *et al*, 2003; Sosunova *et al*, 2003; Opalka *et al*, 2003). The preferred cleavage product is a dinucleotide indicating that the accommodation of the RNA and cleavage factors is easier when the TEC has backtracked by just one nucleotide. GreB can also release longer RNA oligonucleotides from TECs backtracked by several nucleotides (Borukhov *et al,* 1993). The cleavage factors can only fit in the secondary channel when the active site is open, and accordingly, they are thought to substitute the helical TL in the backtracked TEC (Roghanian *et al*, 2011) and their binding may require clamp opening (Sekine *et al*, 2015; Tagami *et al*, 2010). Despite the cleavage factors are often seen as acting on the backtracked TECs, they may associate with the periphery of the RNAP continuously throughout the transcription cycle (Kusuya *et al*, 2011; Tetone *et al*, 2017).

The importance of the cleavage reaction is exemplified by several ways GreA/B factors affect RNAP transcription: They reduce transcription errors (Bubunenko *et al*, 2017; Imashimizu *et al*,

2015; Erie *et al*, 1993), assist in the recovery of arrested TECs (Marr & Roberts, 2000; Adelman *et al*, 2005), increase the overall transcription output *in vivo* and *in vitro* (Toulmé *et al*, 2000; Yuzenkova *et al*, 2014; de Maddalena *et al*, 2016; Sigurdsson *et al*, 2010), affect the distribution of RNAP along the transcription units (Kusuya *et al*, 2011), and allow RNAP to transcribe against larger opposing forces (Galburt *et al*, 2007) and to pass through chromatin bound barriers (Toulmé *et al*, 2000; Kireeva *et al*, 2005; Kotlajich *et al*, 2015).

2.5 Translocating too much: Hyper-translocation

Hyper-translocation ensues when post-translocated RNAP moves forward along the DNA template without the concomitant addition of nucleotides (**Figure 6**). The strong interaction between the RNA 3′ OH and the Mg^{2+} (A) at the nucleophilic site needs to be broken, which probably typically limits hyper-translocation. The structure of the hyper-translocated state has not been determined. Furthermore, its occurrence and dynamics have not been properly characterized, although one study interpreted the RNAP protected exonuclease DNA digestion patterns being consistent with the hyper-translocated state (Nedialkov *et al*, 2012). The hyper-translocated state has no known biochemical activity, but it has been hypothesized to be a transient intermediate in the intrinsic or factor-mediated transcription termination process (Larson *et al*, 2008; Ray-Soni *et al*, 2016; Roberts & Park, 2004).

2.6 Principles of RNAP regulation during transcription elongation

During transcript elongation, RNAP recognizes and responses to a variety of regulatory signals (Mooney *et al*, 1998; Nudler, 2012; Belogurov & Artsimovitch, 2015; Ray-Soni *et al*, 2016; Mustaev *et al*, 2017). The DNA encoded pause signals and RNAP binding elongation factors work together to modulate the efficiency of pause recognition and duration. Pauses are off-pathway events that can be mechanistically divided into three classes: the elemental, backtracked and hairpin-stabilized pauses. It has long been assumed that the backtracked and hairpin-stabilized pauses isomerize from an elemental paused state (Artsimovitch & Landick, 2000). The elemental pause-like state is also believed to be the first step in transcription termination. Therefore, the mechanisms that reduce or enhance RNAP pausing typically have similar effects on termination (Ray-Soni *et al*, 2016; Mustaev *et al*, 2017).

The elemental paused state (discussed in the section 2.3.7) involves conformational changes in the TEC that inhibit nucleotide addition: the DNA resides in an intermediate translocation state, the BH is in the bent form, TL is unfolded, clamp domain is open and the RNA 3´ end may be frayed (Weixlbaumer et al, 2013; Artsimovitch & Landick, 2000; Toulokhonov et al, 2007; Malinen et al, 2014; Ray-Soni et al, 2016; Nayak et al, 2013). Stabilization of the paused state by backtracking or by a RNA hairpin involves further conformational changes in the TEC, although the preceding conformational changes may also differ in different pause types (Windgassen et al, 2014). In the backtracked pauses the RNA 3´ end enters to the secondary channel and blocks the active site (Nudler et al, 1997; Komissarova & Kashlev, 1997b, 1997a) (section 2.4). In the hairpin-stabilized pauses the upstream nascent RNA folds into a hairpin, which inserts into the RNA exit channel. This directly inhibits forward translocation and promotes the open clamp domain conformation, which in turn loosens the interaction of the clamp with the RNA:DNA hybrid and favors a constrained TL conformation that cannot promote nucleotide addition (Hein *et al*, 2014; Kolb *et al*, 2014; Nayak *et al*, 2013). Thereby, RNAP activity is controlled by allosteric connections between the nucleic acids, the RNA exit channel, main channel, secondary channel, the clamp domain, and the THB of the active site (Hein *et al*, 2014; Kolb *et al*, 2014; Nayak *et al*, 2013; Weixlbaumer *et al*, 2013). Cellular elongation factors and bacteriophage anti-termination factors regulate RNAP pausing and termination propensities by targeting these allosteric connections (**Table 1**).

Regulation during the transcription elongation phase is often linked to coupled biochemical events downstream of RNA transcription, like RNA folding and RNA translation (reviewed in (Belogurov & Artsimovitch, 2015; Pan & Sosnick, 2006)). The reciprocal crosstalk between transcription and RNA processing is enabled by modulating the propensities of RNAP pausing and termination. A classical example is the regulation of amino acid biosynthetic operons by the attenuation mechanism where RNAP pausing at the 5´UTR couples transcription with translation allowing RNAP to sense the tRNA availability through the rate of translation of the leader peptide (Yanofsky, 1981). Whether the ribosome stalls or not at the leader peptide determines whether the nascent RNA folds into an anti-terminator hairpin, that allows the operon to be expressed, or to an alternative RNA structure, terminator hairpin, that terminates RNA

Table 1. Proteins that regulate RNAP elongation activity by allosteric mechanisms

Cellular proteins		
NusG	Affects RNAP pausing through contacts to the DNA; binds to the β´ clamp helices	(Yakhnin *et al*, 2016; Turtola & Belogurov, 2016)
RfaH	Affects RNAP pausing through contacts to the DNA and stabilizes the closed clamp; binds to the β´ clamp helices and β gate loop	(Svetlov *et al*, 2007; Sevostyanova *et al*, 2011; Kolb *et al*, 2014)
σ^{70}	Induces DNA scrunching and backtracking by binding to the β´ clamp helices and -10 or -35-like DNA elements downstream of promoters	(Perdue & Roberts, 2011; Zhilina *et al*, 2012)
NusA	Enhances pausing by stimulating RNA hairpin formation and inhibiting translocation synergistically with RNA hairpins; binds to the β-flap near the RNA exit channel	(Hein *et al*, 2014; Zhou *et al*, 2011)
DksA	Inhibits elongation and enhances intrinsic termination in a TL dependent manner synergistically with (p)ppGpp; binds to the secondary channel	(Furman *et al*, 2012)
GreA/B	Stimulate RNA endonucleolytic cleavage of RNA by stabilizing the catalytic Mg^{2+}-ions; bind to the secondary channel and replace the helical TL; binding may require RNAP clamp opening or ratcheting	(Laptenko *et al*, 2003; Sosunova *et al*, 2003; Opalka *et al*, 2003; Roghanian *et al*, 2011; Sekine *et al*, 2015)
Gfh	Prolongs pausing and enhances termination by stabilizing the open clamp and a kinked BH (in *Dra* and *Tth*) in a "ratcheted state"; binds to the secondary channel of a paused TEC	(Tagami *et al*, 2010; Esyunina *et al*, 2016a)
H-NS	Nucleoid-associated protein; forms topologically restricted chromatin filaments that create torsional stress on the DNA sensitizing RNAP to backtracking and Rho-dependent termination	(Kotlajich *et al*, 2015; Peters *et al*, 2012)
Phage proteins		
λQ	Anti-termination factor of the *E. coli* bacteriophage λ; reduces pausing and termination; binds to β-flap	(Deighan *et al*, 2008; Deighan & Hochschild, 2007)
λN	Anti-termination factor of the *E. coli* bacteriophage λ; renders RNAP resistant to termination; binds to λnut in the nascent RNA and near the RNA exit channel; recruits NusABEG factors and redirects the RNA	(Said *et al,* 2017)
gp39	Anti-termination factor of the *T. thermophilus* bacteriophage P23-45; reduces pausing and termination; binds to the β-flap	(Berdygulova *et al*, 2012)
p7	Anti-termination factor p7 of *Xanthomonas oryzae* (rice pathogen) bacteriophage Xp10; biases forward translocation making RNAP resistant to intrinsic terminators; binds to the N-terminus of the β´-subunit, the β-flap and the upstream DNA	(Zenkin *et al*, 2015)
Nun	Arrest factor of the *E. coli* bacteriophage HK022; blocks translocation and arrests RNAP by binding to the main channel, presumably during transient clamp opening; recruited at the bacteriophage λ nut RNA; competes with and antagonizes the function of λN	(Vitiello *et al*, 2014; Kang *et al*, 2017)

synthesis, respectively. In the case of the *trp* operon, the ~600-fold dynamic range in the expression level originates from the repression of transcription initiation by TrpR (70-fold) and the repression of elongation through attenuation (8-10-fold), whereas in several other amino acid biosynthetic operons attenuation is the sole control mechanism (Yanofsky, 1981).

While the transient transcriptional pauses ensure that the pioneering ribosome translates in close proximity with RNAP, conversely, the trailing ribosome physically prevents extensive RNAP backtracking (Proshkin *et al*, 2010; McGary & Nudler, 2013). This activity, together with other anti-backtracking factors including GreA/B, DksA (Zhang *et al*, 2014) and possibly NusG (Turtola & Belogurov, 2016) prevent RNAP stalling thereby reducing the chances that RNAP would collide with the replisome, which can cause in-del mutations and DNA double stranded breaks (Sankar *et al*, 2016; Tehranchi *et al*, 2010; Dutta *et al*, 2011). Hence, the extent of backtracking needs to be carefully balanced.

When transcription is uncoupled from translation at intergenic regions, poorly translated horizontally transferred genes, or at antisense transcription units, RNAP is typically destined to termination. RNA synthesis without downstream RNA processing exposes the nascent RNA to the helicase Rho that translocates along the RNA and, aided by NusG at a subset of terminators, terminates transcription (Peters *et al*, 2009, 2012; Cardinale *et al*, 2008). Because Rho patrols unprocessed transcripts, the transcription of non-coding RNA genes, like those for ribosomal RNA (rRNA), requires protection from the termination factor Rho by the anti-termination complex containing NusABGE and S4 proteins. However, the conservation of the anti-termination complex, but not Rho, among all bacteria suggest that functionalities other than the protection from Rho may be more important, for example the chaperoning of rRNA folding by the NusA S1 and KH-domains (Pan *et al*, 1999; Bubunenko *et al*, 2013; Belogurov & Artsimovitch, 2015).

Interestingly, transcription elongation rates in different parts of the rRNA genes vary greatly (Dennis *et al*, 2009). This kinetic aspect may be important for the folding of the ribosomes, because the variations in RNAP elongation rates are known to affect the co-transcriptional RNA folding (Pan *et al*, 1999; Pan & Sosnick, 2006). As described above using the attenuation mechanism as an example, the timing when the RNA emerges out of the RNA exit channel of RNAP affects the folding landscape of the nascent transcript. In this respect it is interesting that the NusABGE anti-termination complex accelerates the RNAP elongation rate two-fold (Zellars & Squires, 1999), possibly facilitating the correct folding of the rRNA.

Another way RNAP pausing regulates transcription is it facilitates the recruitment of elongation factors. In effect, pausing provides a time window for regulator binding at a fixed genomic position. For example, pausing at the *ops* (operon polarity suppressor) sequence gives time to recruit RfaH to the TEC (Artsimovitch & Landick, 2002, 2000). RfaH is a NusG paralogue that through its N-terminal domain recognizes the *ops* sequence in the exposed single stranded non-template DNA (Artsimovitch & Landick, 2002) and binds to the β´ clamp helices (Belogurov *et al*,

2007). RfaH reduces polarity effects in long poorly transcribed operons by suppressing pausing and increasing RNAP processivity, by recruiting the translating ribosome (Burmann *et al*, 2012) and by preventing the association of Rho through the occlusion of NusG binding (Sevostyanova *et al*, 2011; Belogurov *et al*, 2009). Another example is the recruitment of the bacteriophage lambda anti-termination factor Q. The binding of Q to the TEC requires a σ^{70} induced backtracked pause and a Q-recruiting DNA sequence at a promoter proximal site (Perdue & Roberts, 2010). Modification of the TEC by Q allows the read-through of several viral terminators allowing the expression of late genes, which is critical for the completion of the lytic cycle (Deighan & Hochschild, 2007).

Lastly, RNAP pausing and stalling at sites of DNA damage is an efficient way to recognize DNA damage and recruit the components of the DNA damage response (Howan *et al*, 2014; Belogurov & Artsimovitch, 2015). Indeed, a damaged site in DNA is repaired faster in the transcribed DNA strand than in the non-transcribed complementary strand (Mellon & Hanawalt, 1989). This process – the transcription-coupled DNA repair – requires the removal of the stalled RNAP from the site of damage. The stalled RNAP is removed by the action of the ATPases Mfd or UvrD that push or pull, respectively, RNAP from the upstream edge of the transcription bubble (Belogurov & Artsimovitch, 2015; Epshtein *et al*, 2014; Roberts & Park, 2004).

The regulators also enable the RNAP activity to respond to cellular metabolic states. In addition to its prominent role in regulating transcription initiation, the stringent response regulator (p)ppGpp reduces the overall elongation rate, possibly by affecting the conformation of RNAP and sensitizing it to pauses (Kingston *et al*, 1981). Additionally, (p)ppGpp strongly enhances the otherwise modest elongation inhibition effect by DksA (Furman *et al*, 2012). Interestingly, *in vivo* (p)ppGpp reduced the elongation rate of coding genes in response to amino acid starvation, but it did not have this effect on an untranslated rRNA gene (Vogel & Jensen, 1994a). This possibly reflects the need to maintain transcription-translation coupling when translation is slowed down by a shortage of amino acids. Finally, several examples indicate that the regulatory potential of the secondary channel binding factors DksA and Gfh1 can adjust in response to cellular stress, including low pH (Furman *et al*, 2015; Laptenko *et al*, 2006), the presence of manganese ions (Esyunina *et al*, 2016a) and reactive oxygen and nitrogen species (Crawford *et al*, 2016).

3 AIMS OF THE STUDY

The ability to translocate is a fundamental property of all RNA polymerases. The intricate molecular transitions during the translocation motion not only allow the transcription of genes encoded in long DNA strings but also serve as a target to regulate gene expression. A number of studies that measured the overall or local rates of RNAP elongation have reached a consensus that RNA polymerases translocate via a thermal ratchet mechanism. However, as these measurements do not temporally distinguish the intrinsic translocation step from other steps of the nucleotide addition cycle, some central questions remain obscure. Specifically, it is unclear whether the energetic barriers separating the RNAP translocation states are low, i.e. fluctuate with the frequency of Brownian motions, or high, such that the translocation steps can be distinguished from other steps in the RNAP nucleotide addition cycle. In other words, the intrinsic rate of RNAP translocation is unknown. This seemingly minor detail has larger implications on the mechanism of translocation, and on the other hand, on the principles how RNAP translocation can control transcription elongation and how it can be subject to regulation. Furthermore, it remains unclear what structural features of RNAP control translocation, what is the specific role of the trigger loop in translocation and whether translocation and PP_i release are coupled. In general, the lack of methods to monitor translocation at a relevant spatial and temporal resolution has precluded the functional testing of predictions from structural studies. The first two aims of the thesis are to:

i) **Develop methods to monitor RNAP on-pathway translocation in time-resolved and equilibrium assays.**

ii) **Characterize the translocation mechanism thermodynamically and kinetically. Address the question of whether PP_i release and translocation are coupled processes.**

The overall trajectories of nucleic acids movements during translocation are known approximately, but their dynamics remain poorly characterized. Predictably, linking the dynamics of nucleic acids and conformational changes in RNAP will be necessary to understand how the signals encoded in the DNA and RNA control the enzymatic activity of RNAP. While it has been anticipated that such signals can target the translocation step, it remains to be shown how the sequence specific effects manifest themselves. The third aim of the thesis is to:

iii) **Elucidate the major conformational transitions in RNAP and nucleic acids during translocation. Investigate how the transcribed sequence affects translocation.**

Bacterial RNAPs are molecular targets for a few clinically used anti-microbial drugs and remain as targets for the development of new antibiotics. Additionally, provided their mechanism of action is well understood, inhibitors are invaluable tools for mechanistic studies. Two such molecules, tagetitoxin and CBR703, inhibit *E. coli* RNAP elongation but their inhibitory mechanisms have not been investigated in detail. Thus, the fourth aim of the thesis is to:

***iv)* Investigate how the enzymatic activity of RNAP is modulated by the small molecule inhibitors tagetitoxin and CBR703.**

Finally, RNAP elongation is frequently interrupted by off-pathway pauses. Backward translocation, or backtracking, is one of the main types of pauses and enables transcriptional regulation as well as RNA proofreading, but under certain circumstances may cause toxic effects in the cell. To understand how RNAP enters and exits the backtracked state, new kinetic methods that monitor these steps are required. On the other hand, the elongation activity of all multi-subunit RNA polymerases is modulated by transcription elongation factors but the general principles for how translocation is regulated are largely unknown. The universally conserved elongation factor NusG has been previously suggested to enhance RNAP processivity by biasing translocation forward and thereby inhibiting backtracking. However, the molecular basis for these activities has remained largely speculative. Therefore, the final aims of the thesis are to:

***v)* Develop methods to monitor RNAP backtracking in time-resolved and equilibrium assays.**

***vi)* Study the molecular basis of RNAP backtracking and its regulation by the transcription elongation factor NusG.**

Here, RNAP from the bacterium *Escherichia coli* is used for investigations of the translocation mechanism due to the ease of expression of large quantities of wild-type and variant RNAP and its history as a central model enzyme for mechanistic transcription studies. Among the multisubunit RNA polymerases, the bacterial RNAP (ααββ´ω) is structurally the simplest enzyme, while it shares a homologous catalytic core with more complex RNA polymerases from Eukarya and Archaea. Thereby, it is anticipated that the results obtained with the bacterial RNAP will contribute to the understanding of multisubunit RNA polymerases in general. Technically, these aims will be addressed with a combination of fluorescence, nucleic acid crosslinking and classic *in vitro* transcription techniques. As the most reductionist approach to investigate RNAP enzymatic activity, the majority of analyses focus on a single nucleotide addition cycle. To this end, the transcription elongation complex is reconstituted *in vitro* from synthetic oligonucleotides and purified proteins, which allow the extensive manipulation of structure and experimental variables as well as rigorous interpretation of the results. A large number of mutations that change the functional properties of *E. coli* RNAP are known. These and new rationally designed mutations will be analyzed in search of structural changes that will disturb translocation or backtracking.

4 MATERIALS AND METHODS

4.1 RNAP mutagenesis

Mutations to the $His_{6/10}$-tagged RNAP β' or β-subunits were introduced directly into expression plasmids bearing all RNAP subunits under the control of the T7 RNA polymerase promoter and containing cloning sites every 0.5-1 kb within the *rpoC* (pMT2/pMT32: T7p-α-β-β'-TEV-His_{10}-T7p-ω, derivatives of pIA981) or *rpoB* genes (pMT23: T7p-α-His_6-β-β'-ω), respectively. The DNA fragments containing the substitutions were designed and ordered from Thermo Fisher Scientific GENEART GmbH (Regensburg, Germany) as linear DNA strings or as DNA fragments cloned into plasmids, and inserted into the expression vectors using restriction enzymes and DNA ligase. A subset of single amino acid substitutions to the β' subunit were introduced by inverse PCR to accessory plasmids containing a fragment of the *rpoC* gene (pMT1, derivative of pIA458). PCR was carried out with Phusion High Fidelity DNA polymerase (Thermo Fischer Scientific) using back-to-back design of primers that contained the mutations. The linear PCR products were purified, phosphorylated and circularized by ligation, and the parent plasmid was digested with DpnI. The plasmids were amplified and the fragments containing the substitution were sub-cloned into the expression plasmid pIA981 (T7p-α-β-β'-TEV-His_{10}-T7p-ω). The plasmid constructs were verified by DNA sequencing the newly inserted region. The plasmids are specified in each publication. The plasmids used in section 5.2.6 (unpublished work) are listed in **Table 2**.

Table 2. *E. coli* protein expression vectors used in section 5.2.6

Name	Description	Source or reference
wild-type *E. coli* RNAP	pVS10 (T7p-α-β-β'_His6-T7p-ω)	(Belogurov *et al*, 2007)
β´F773V	pVS48 (T7p-α-β-β'[F773V]_His6-T7p-ω)	(Svetlov *et al*, 2007)
β´P750L	pTG12 (T7p-α-β-β'[P750L]_TEV_His10-T7p-ω)	(Malinen *et al*, 2014)
β´G1136S	pJM1 (T7p-α-β-β'[G1136S]_TEV_His10-T7p-ω)	Study VI
β´F935Y	pGB128 β'F935Y (T7p-α-β-β'[F935Y]_TEV_His10-T7p-ω)	This work
β´H936Q	pJM3 (T7p-α-β-β'[H936Q]_TEV_His10-T7p-ω)	Study VI
β´H936A	pGB130 (T7p-α-β-β'[H936A]_TEV_His10-T7p-ω)	(Malinen *et al*, 2012)
β´R933A	pIA846 (T7p-α-β-β'[R933A]_His6-T7p-ω)	(Artsimovitch *et al*, 2011)

4.2 Protein expression

Expression plasmids were transformed into an Xjb (BL21 (DE3) derivative that expresses λ endolysin under an arabinose-inducible promoter) (Zymo Research, Irvine, CA, USA) or T7 Express (BL21 with T7 RNA polymerase integrated in the *lac* operon and additionally expressing $lacI^q$ (lac repressor) and lysY (T7 lysozyme)) (New England Biolabs, Ipswich, MA, USA) *E. coli* expression strains. Cell cultures in LB medium containing an antibiotic (typically carbenicillin 100 µg/ml or kanamycin 30 µg/ml) for plasmid selection were grown under shaking at 37 °C to OD_{600}≈0.6 at which point the protein expression was induced by the addition of 1 mM IPTG (and 0.1 % w/vol arabinose when culturing the Xjb cells). Cells were collected after 3-5 h by centrifugation and frozen at -80 °C. Cells were suspended in Lysis buffer 1 (50 mM Tris-HCl pH 6.9, 0.5 M NaCl, 5 % glycerol, 1 mM β-mercaptoethanol, 0.1 % Tween-20) containing 0.1 mM EDTA and a protease inhibitor cocktail (Roche Diagnostics, Mannheim, Germany) and lysed by sonication. The lysate was clarified by centrifugation at 42 000 g for 45 mins.

Table 3. Expression and purification of other proteins

Protein (organism)	Expression plasmid	Purification protocol and other specifications
NusG (*E. coli*)	pGB43: T7p-His_6-TEV-NusG	-Ni-sepharose in Lysis buffer 1 -TEV-protease (3 µg/ml) over-night -Dialysis over-nigh in Lysis buffer 1 -Removal of un-cleaved NusG, the cleaved tag and the TEV-protease by capture to Ni-sepharose matrix
GreA and GreB (*E. coli*)	pIA578: T7p-GreA-His_6 pTG1: T7p-His_6-TEV-GreA[D41N] pIA577: T7p-GreB-His_6 pIA689: T7p-GreB[D41N]-His_6	-Ni-sepharose in Lysis buffer 1 containing 0.1 mM EDTA and 1 M NaCl -Gel filtration in Chromatography buffer containing 1 M NaCl [for GreA D41N: TEV-protease (3 µg/ml) over-night, removal of un-cleaved GreA D41N, the cleaved tag and the TEV-protease by capture to Ni-sepharose matrix] -Dialysis against the Storage buffer containing 1 M NaCl
$Sigma^{70}$ (*E. coli*)	pET28-His_6-σ^{70}	-Like RNAP purification with the difference that the buffers during Ni-sepharose contained 1 M NaCl
Inorganic pyrophosphatase (*S. cerevisiae*)	pKW9: T7p-PPase	- Expression strain Xjb RIL (Zymo Research, Irvine, CA, USA) -Lysis in French press in Lysis buffer 2 (25 mM Tris-HCl pH 8.5, 15 mM $MgCl_2$) -DEAE-sepharose anion-exchange chromatography -Gel filtration (HiPrep 16/60 Sephacryl S-200 HR (GE Healthcare)) with Elution buffer (50 mM Tris-HCl pH 6.9, 5 % glycerol, 1 mM β-mercaptoethanol, 0.1 mM EDTA, 0.75 M NaCl, 15 mM $MgCl_2$) -Analysis for protein by SDS-PAGE and for PPase activity by a colorimetric molybdate-acid assay
Phosphate binding protein (*E. coli*)	pAM8: T7p-PBP[A197C]-His_8	-Lysis buffer 3 (50 mM Tris-HCl pH 6.9, 1 M NaCl, 5 % glycerol, 1 mM 1 mM TCEP, 0.1 % Tween-20, 0.1 mM EDTA) with protease inhibitors cocktail (Roche) -Ni-sepharose, wash and elution with Buffer A (20 mM Tris-HCl pH 7.9, 250 mM NaCl, 1 mM TCEP) supplied with 20 and 250 mM imidazole, respectively -Pi removal by dialysis (20 h) against a custom-made FeO paper in Buffer A.

4.3 Purification of RNAP

RNAP with a His_6- or His_{10} tag was captured from the cleared lysate onto a Ni-sepharose matrix (GE Healthcare) in the presence of 20 mM imidazole. Bound proteins were washed and eluted by Lysis buffer 1 supplemented with 20 mM and 200 mM imidazole, respectively. The eluted sample was diluted with the Chromatography buffer (50 mM Tris-HCl pH 6.9, 5 % glycerol, 1 mM β-mercaptoethanol, 0.1 mM EDTA) and applied to the HiTrap heparin column (GE Healthcare). Proteins were eluted with a NaCl gradient and the peak fractions were analyzed by SDS-PAGE. Fractions containing all RNAP subunits were pooled and treated over-night at 4 °C with TEV-protease (3 µg/ml) (when applicable); the un-cleaved RNAP, the cleaved tag and the TEV-protease were captured onto a Ni-sepharose matrix. The sample was further purified with Resource Q (GE Healthcare) anion exchange chromatography. The collected fractions were concentrated and dialyzed against the Storage buffer ((10 mM Tris-HCl pH 7.9, 50 % glycerol, 0.1 mM EDTA, 100 mM NaCl, 0.1 mM DTT). The protein concentration was determined from the sample absorbance at 280 nm and the calculated RNAP extinction coefficient. Purified RNAPs were stored at -20 °C. The purification of other proteins is presented in **Table 3**.

4.4 DNA and RNA oligonucleotides

DNA and RNA oligonucleotides were synthesized by IBA GmbH (Gottingen, Germany), Fidelity systems (Gaithersburg, MD, USA) and Eurofins MWG Operon (Ebensburg, Germany). The

oligonucleotides were purchased in HPLC (most of non-template DNAs except for those with base analogue fluorophores) or PAGE (most of template DNAs, modified oligos and RNA) purified form. The nucleotide sequences are specified in each publication. The oligonucleotides used in section 5.2.6 (unpublished work) are listed in **Table 4**.

Table 4. Oligonucleotides used in section 5.2.6

Template strand oligonucleotide		**Manufacturer**
S041M	GCTACTCTACTGACATGATGCCTCCTCTXGAACCTTAGATCGCTACAAGT	Fidelity systems
Non-template strand oligonucleotides		
S042	ACTTGTAGCGATCTAAGGTTCCAGAGGAGGCATCATGTCAGTAGAGTAGC	MWG
S166b	Biotin-5´ACTTGTAGCGATCTAAGGTTCCAGAGGAGGCATCATGTCAGTAGAGTAGC	MWG
RNA oligonucleotides		
R024	Atto680-CUCACAACCAGAGGAG	IBA
R030	CUCACAACCAGAGGAG	Fidelity systems

X=6-methylisoxanthopterin

4.5 RNAP inhibitors and other reagents

Tagetitoxin was purchased from Epicentre (Madison, WI, USA) but is no longer commercially available. CBR703 was from Maybridge (Tintagel, UK) and STL from Sourcon-Padena (Tübingen, Germany). NTPs, 3´deoxy NTPs, cytidine-5´-[(α,β)-methyleno]triphosphate and guanine-5´-[(α,β)-methyleno]triphosphate were from Jena Bioscience (Jena, Germany). 2-aminopurine-5´-triphosphate was from TriLink Biotechnologies (San Diego, USA) or the TriLink European distributor tebu-bio (Roskilde, Denmark). 2´deoxy-NTPs were from Bioline (London, UK).

4.6 TEC assembly

The *in vitro* assembly of TEC was based on the principles developed by (Sidorenkov *et al*, 1998; Komissarova *et al*, 2003). Unless otherwise specified, the pre-annealed RNA:template DNA hybrid (1.4 µM:1 µM) was mixed with purified RNAP (1.5 µM) in Transcription buffer (TB; 40 mM HEPES-KOH pH 7.5, 80 mM KCl, 5% glycerol, 0.1 mM EDTA, and 0.1 mM DTT, 0-10 mM $MgCl_2$; final $MgCl_2$ concentrations in mM are specified by a number after TB), incubated for 10 min at 25 °C and supplemented with an excess of the non-template DNA (2 µM) following a further 20 min incubation at 25 °C and dilution with TB to the final concentration (typically 0.05-0.4 µM). In a subset of experiments (RNA analysis in PP_i titration experiments, section 5.2.6), TECs were immobilized on streptavidin coated magnetic beads and purified from unbound nucleic acids. 60 pmoles of TEC containing a biotin linker at the 5´ end of the non-template DNA strand was incubated for 15 min with 20 µl of streptavidin beads (Dynabeads MyOne Streptavidin C1, Invitrogen) in a 300 µl total volume, and washed three times with TB10 buffer. In the experiments in section 5.6 the TEC assembly was often followed by pre-extension with 2-AP or buffer change from TB1 to TB0. In such cases the TEC samples were further gel filtrated in a Zeba 40K spin desalting column (Thermo Fischer Scientific) pre-equilibrated with the desired TB buffer.

4.7 Equilibrium fluorescence measurements

Equilibrium levels of the 6-MI and 2-AP fluorescence were determined by continuously recording the light emission at the 420 and 380 nm wavelengths (excitation at 340 and 320 nm), respectively, with an LS-55 spectrofluorometer (PerkinElmer, Waltham, MA, USA) in a 16.160-F/Q/10 quartz cuvette (Starna Scientific) at 25 °C. The final TEC concentration was 100-200 nM. Where indicated, NTPs (5-50 μM), 2´ deoxy-NTP (100 μM), 3´ deoxy-NTP (500 μM), NMPcPP (500 μM) and ligands (TGT, STL, PP_i, Mg^{2+}) were directly pipetted into the cuvette and the resulting dilution of the fluorescent species was taken into account during the data analysis. The fluorescence was monitored for at least 2 mins to ensure that the fluorescence reached the equilibrium level. RNA extension was checked by withdrawing 4 μl aliquots from the cuvette into 6 μl of Gel loading buffer (94% formamide, 20 mM Li_4-EDTA and 0.2% Orange G).

4.8 Analysis of Atto680 labeled RNAs and DNAs

Atto680-labeled RNA/DNA samples in the formamide containing quench buffers were denatured at 95 °C for 2 min prior to loading onto a 16% denaturing urea-polyacrylamide gel and analyzed by gel electrophoresis (PAGE) in 1xTBE buffer (100 mM Tris, 90 mM boric acid, 1 mM EDTA, pH 8.4). The gel was imaged with an Odyssey Infrared Imager (Li-Cor Biosciences, Lincoln, NE, USA) using the 700 nm channel. The band intensities were quantitated with the Image J software (Abràmoff *et al*, 2004).

4.9. Photo-crosslinking experiments

Photo-crosslinking experiments were carried out in a custom made thermally controlled chamber with a UV LED (P8D1: 365 nm, optical power 40 mW, dome-type radiation pattern; Seoul Viosys, Ansan, Korea) placed in the top center (height=17 mm). The 8-MP mono-adduct and 6-TG have absorption maxima at 342 nm and 340 nm, respectively. The samples (5 μl) were placed on an 18-well circular tray (Ø=26 mm, all wells equidistant from the center) and exposed to UV-light for 30 mins at 25 ˚C. Sample aliquots (4 μl) were quenched with 6 μl of the Gel loading buffer (see section 4.7). In 6-TG crosslinking experiments the reaction mixture contained 1 μM TEC in TB10 buffer. In 8-MP crosslinking the reaction mixture contained 1 μM TEC, 0.92 mM 8-MP, 6.3% DMSO in TB10 buffer. NTPs (5 μM), NMPcPP (500 μM) or TGT (5 μM) were added where indicated. The Atto680-labeled crosslink products were separated on 14% denaturing urea-polyacrylamide gels and imaged as described in section 4.8.

4.10 Time-resolved nucleotide addition and RNA cleavage measurements

Study I is devoted to the detailed description of the nucleotide addition measurements. Time resolved measurements of nucleotide addition were performed in an RQF 3 quench-flow instrument (KinTek Corporation, Austin, TX). The reaction was initiated by the rapid mixing of 14 μl of solution 1 with 14 μl of solution 2. The reaction solutions are specified in **Table 5**. The reaction was allowed to proceed for 0.004–1000 s at 25˚C, quenched with 86 μl of 0.5 M HCl and immediately neutralized by adding 171 μl of Neutralizing-loading buffer (94% formamide, 290 mM Tris base, 13 mM Li4-EDTA, 0.2% Orange G). EDTA quench experiments (Kireeva *et al*, 2009) were carried out in the same way except that the quenching was done with a 0.45 M Na-

EDTA solution prepared in TB buffer and the quenched samples were supplemented with 171 µl of Formamide buffer (97% formamide, 0.2% Orange G). RNA cleavage (Figure 21) was monitored by the manual mixing of 50 µl of 200 nM TEC in TB0 buffer with 50 µl of 16 µM GreA in TB2 buffer at 25˚C. The aliquots (8 µl) were withdrawn at the indicated time points and quenched with 12 µl of the Gel loading buffer (see section 4.7). The TEC solutions were supplemented with 4 µM NusG where indicated. RNAs were separated on 16% denaturing polyacrylamide gels and imaged as described in section 4.8.

4.11 Time-resolved translocation, backtracked state recovery and dinucleotide release measurements

Study I is devoted to the detailed description of the protocols for translocation measurements. Measurements were performed in an Applied Photophysics (Leatherhead, UK) SX.18 MV stopped flow instrument at 25˚C. The reactions were initiated by the rapid mixing of 60 µl of solution 1 with 60 µl of solution 2 and the sample fluorescence was continuously monitored for 10-1000 s. The reaction solutions are specified in **Table 5**. The 6-MI and 2-AP fluorophores were excited at the 340/320 nm wavelength and the emitted light was collected through the 400/375 nm long pass filters, respectively. Three to seven individual time traces were averaged for each reported curve.

4.12 PP_i release measurements and PBP labelling

The assays for the time-resolved detection of Pi and PP_i were described in detail by Brune *et al.* (Brune *et al*, 1994) and Pais *et al.* (Pais *et al*, 2005) and were here adapted to transcription reactions. Phosphate was detected with *E. coli* phosphate binding protein (PBP) labeled site-specifically at the substituted A197C residue by the MDCC-fluorophore (*N*-[2-(l-maleimidyl)ethyl]-7-diethylamino)coumarin-3-carboxyamide). For PBP labeling, 16 ml of Buffer B (20 mM Tris-HCl pH 7.9, 250 mM NaCl) was pre-incubated for 10 min with "P_i-mop" (200 µM 7-methylguanosine and 0.2 U ml^{-1} purine nucleoside phosphorylase (Sigma-Aldrich)) in order to remove Pi. 50 µM of purified PBP[A197] (see **Table 3**) was added and incubated for an additional 10 min. The labelling was initiated by the addition of 150 µM MDCC (Sigma-Aldrich) dissolved in DMSO. The reaction tube was protected from light and incubated under mild shaking at room temperature for 30 min. The MDCC-labeled PBP (MDCC-PBP) was purified from the unbound label by Ni-sepharose in Buffer C (50 mM Tris-HCl pH 6.9, 250 mM NaCl, 1 mM TCEP) and eluted in Buffer C supplemented with 250 mM imidazole. The imidazole was removed by dilution-concentration cycles using Amicon Ultra-15 centrifugal filters (10 kDa molecular weight cutoff, Merck Millipore) in Buffer C. The MDCC-PBP was stored at -80 °C.

PPi release was detected by the rapid hydrolysis of PP_i into P_i with inorganic pyrophosphatase from *S. cerevisiae* (PPase) and binding of P_i to MDCC-PBP. Yeast PPase was used because of its lower NTP hydrolysis activity compared to the PPase from *E. coli*. The measurements were performed in an Applied Photophysics (Leatherhead, UK) SX.18MV stopped-flow instrument at 25 °C. To reduce ubiquitous P_i contamination, the sample syringes and lines of the stopped-flow system as well as the NTP stock solutions were treated with P_i-mop. The reaction was initiated

Table 5. Experimental setups of the time-resolved measurements

Experiment	Conc. (μM) during TEC assembly				Solution 1	Solution 2	Instrument*
	tDNA	RNA	ntDNA	RNAP			
TEC16+NTP, ***nucleotide addition***, Figures 11, 14, 15, 19	1.4	1	2	1.5	TEC16 200 nM, TB10	NTP(1) 400 μM, TB10 (+PP_i 0-1 mM)	QF (HCl 0.5 M)
TEC16+NTP, ***translocation*** Figures 11, 14, 15, 19	1	1.4	2	1.5	TEC16 100-200 nM, TB10	NTP(1) 400 μM, TB10 (+PP_i 0-1 mM)	SF (340/400)
TEC16+NTP, ***PP_i release*** Figures 11, 14	1	1.4	2	1.5	TEC16 400 nM PBP-MDCC 5 μM PPase 20 μM, TB10	NTP(2) 400 μM PBP-MDCC 5 μM, PPase 100 pM, TB10	SF (430/455)
TEC17+PP_i, ***PP_i induced backward translocation*** Figure 15	1	1.4	2	1.5	TEC16 200 nM, NTP(3) 50 μM, TB10	PPi 1 mM, NTP(3) 50 μM, TB10	SF (340/400)
TEC17+NTP, ***nucleotide sequestration*** Figure 19	1.4	1	2	1.5	TEC16 200 nM, NTP(4) 5 μM, TB10	NTP(5) 1 mM, TB10	QF (Na-EDTA 0.45 M)
TEC18+GreA, ***backtracking*** Figure 22	1	1.4	2	1.5	TEC18 200 nM ±NusG 4 μM, TB0	GreA 16 μM TB2	SF (340/400)
TEC18+GreA, ***RNA cleavage*** Figure 22	1.4	1	2	1.5	TEC18 200 nM ±NusG 4 μM, TB0	GreA 16 μM TB2	manual quench (HCl 0.5 M)
TEC18+GreA, ***dinucleotide release*** Figure 22	1.4	1	2	1.5	TEC18 200 nM ±NusG 4 μM, TB0	GreA 16 μM TB2	SF (320/375)
TEC-2AP+CTP, ***backtracked state recovery*** Figure 24	1.4	1	2	1.5	TEC-2AP 100 nM, TB1	CTP 400 μM, TB1	SF (320/375)

Solution 1 and Solution 2 were mixed in 1:1 ratio

*QF = Quench Flow (quench solution), SF = Stopped Flow (fluorescence excitation/emission wavelengths, nm)

(1) transcribed nucleotide of the TEC16: G/A/U/C (Figure 10); G (Figures 14 & 15); G in the GC-system, C in the CG system (Figure 19)

(2) transcribed nucleotide of the TEC16: G/A/U/C (Figure 11), G (Figure 14 & 15)

(3) transcribed nucleotide of the TEC16: G

(4) transcribed nucleotide of the TEC16: G in the GC-system, C in the CG system (Figure 19)

(5) transcribed nucleotide of the TEC17: C in the GC-system, G in the CG system (Figure 19)

by mixing 60 μl of a 400 μM NTP solution containing 5 μM MDCC-PBP and 100 pM PPase with 60 μl of a 0.4 μM TEC solution containing 5 μM MDCC-PBP and 20 μM PPase. The reaction solutions are also specified in **Table 5**. The 100 pM PPase in the NTP solution removed the contaminating PP_i without causing notable NTP hydrolysis and the resulting P_i was sequestered by MDCC-PBP before the reactions started. On the other hand, the high concentrations of MDCC-PBP and Y-PPase in the final reaction mixture compensated for the low concentrations of PP_i (200 nM) and P_i (400 nM) released by 200 nM TEC during a single nucleotide addition cycle effectively reducing the system delay time to 6-8 ms (**Study II, Figure S3**). All solutions were prepared in TB10 buffer. The MDCC fluorophore was excited at 430 nm and the emitted light was collected through a 455 nm long-pass filter. After mixing the solutions, the sample fluorescence was continuously monitored for 10 s. At least six individual traces were averaged for each reported curve.

4.13 Transcription over a long DNA template

The DNA template was amplified by PCR from the pIA146 plasmid. The resulting 1.4 kb long transcription template contained the T7A1 promoter followed by a fragment of the *rpoB* gene (Svetlov *et al*, 2007). The RNAP holoenzyme stock was prepared by incubating 10 µM RNAP and 50 µM sigma^{70} in storage buffer at 30 °C for 20 min and stored at -20 °C until use. The A29 halted complex was formed by incubating the RNAP-sigma^{70} holoenzyme (RNAP 335 nM) in the presence of 250 nM of DNA template, Atto680- ApU initiating dinucleotide (50 µM) and ATP, GTP and CTP (each 5 µM) in TB10 buffer for 15 min at 37 °C. The sample was gel filtrated in a Zeba 40K spin desalting 0.5 ml column (Thermo Fischer Scientific) to remove excess Atto680-ApU. Elongation was initiated by the addition of all four NTPs (final concentration 200 µM), PP_i (0 or 500 µM) and rifapentin (200 µg/ml; for prevention of transcription re-initiation). The elongation was carried out at 25 °C and aliquots of samples were quenched by the addition of Gel loading buffer (see section 4.7).

4.14 Kinetic data analyses

The analysis routines are detailed in the supplementary material accompanying each publication and are elaborated in **Study I**. Translocation and PP_i release curves were individually fit to sequential two-step models where the nucleotide addition curve was used for fitting the first step. Additionally, the model for fitting the PP_i release curves included a third step that accounted for the 6-8 ms signal delay originating from PP_i hydrolysis and P_i binding to MDCC-PBP (**Study II, Figure S3**). The models also postulated that the TECs isomerized between active and inactive states before the nucleotide addition step. The inactive state typically accounted for 5-15% of the total TEC. The kinetic data was fitted using the numerical integration capabilities of the KinTek Explorer software (Johnson, 2009) (KinTek Corporation, Austin, TX, USA). The "pre-catalytic" measurements in section 5.4 were fitted into bi-exponential functions with the Origin2015 software (OriginLab Corporation, Northampton, USA), or explicit kinetic models shown in Figure 18 with KinTek Explorer. These models contained reversible translocation and NTP binding steps and an irreversible NMP incorporation step. Additionally, the un-translocated TEC could isomerize into an off-pathway state. EDTA quench experiments were explicitly modeled using the pulse chase/substrate trap routine of KinTek Explorer. TGT titration data were fit to the dissociation equilibrium equations that accounted for changes in the TEC and TGT concentrations upon formation of the TEC–TGT complex using the Scientist 2.01 software (Micromath, Saint Louis, MO, USA). GreA cleavage experiments in Study V and the CMP addition experiments in Study VI were fitted with Origin2015 using the stretched exponential function (aka Weibull function, specified in Study V). The analysis of the PP_i titration experiments (section 5.2.6) are detailed below. The normalized 6-MI fluorescence (*F*) was fitted using the

$$F = \frac{F_{PPi} + \frac{K_F^{PPi} \times F_{TEC17}}{[PPi]}}{1 + \frac{K_F^{PPi}}{[PPi]}} \quad \text{equation (1)},$$

where $[PP_i]$ was the concentration of PP_i and the fitted parameters were: K_F^{PPi}, apparent equilibrium constant for PP_i interaction with TEC (at 50 µM GTP); F_{PPi}, fluorescence at the

saturating PP_i (at 50 µM GTP); F_{TEC17}, fluorescence in the absence of added PP_i (at 50 µM GTP). Fluorescence (*F*) was normalized using the

equation (2)
$$F = \frac{F_{meas} - F_{TEC16}}{F_{TEC17\text{-}2'dNMP} - F_{TEC16}},$$

where F_{meas}, measured fluorescence of TEC17 in the presence of 50 µM GTP and 0-1000 µM PP_i; F_{TEC16}, fluorescence of TEC16; $F_{TEC17\text{-}2'dNMP}$, fluorescence of 2´dGMP extended TEC17 (fully post-translocated TEC17). The fraction of RNA17 was fitted using the

equation (3)
$$RNA17 = \frac{RNA17_{PPi} + \frac{K_{RNA}^{PPi}}{[PPi]}}{1 + \frac{K_{RNA}^{PPi}}{[PPi]}},$$

where the fitted parameters were: K_{RNA}^{PPi}, apparent equilibrium constant for pyrophosphorolysis (at 50 µM GTP); $RNA17_{PPi}$, fraction of RNA17 at saturating PP_i (at 50 µM GTP). *RNA17* was normalized using the

equation (4)
$$RNA17 = \frac{RNA17_{meas}}{RNA17_{meas} + RNA16_{meas} - RNA16_0},$$

where $RNA17_{meas}$, the measured RNA17 band intensity; $RNA16_{meas}$, the measured RNA16 band intensity; $RNA16_0$, the amount of RNA16 at 50 µM GTP and no added PP_i. All RNA band intensities were normalized for the total amount of RNA. The fitting was done with the Origin2015 software.

4.15 Structural modeling

The modeling of the NusG bound TEC (Figure 22) and the backtracked TEC (Figure 24) were performed manually using the PyMOL Molecular Graphics System software, RRID: SCR_000305 (Schrödinger, New York, NY, USA) as specified in **Studies V** and **VI**, respectively. The model geometry was evaluated using MolProbity (Chen *et al*, 2010).

5 RESULTS AND DISCUSSION

5.1 Positional probing of the transcription elongation complex (Studies **I, II, V**)

The transcribing RNAP moves stepwise along the DNA by separating the two helical strands and maintaining a short unwound DNA region, the transcription bubble. Within the bubble, the templating DNA strand and the synthesized nascent RNA form a one helical turn RNA:DNA hybrid. At each transcribed nucleotide, RNAP melts one DNA base pair at the front edge of the transcription bubble and one RNA:DNA base pair at the upstream edge of the RNA:DNA hybrid, while one DNA:DNA base pair is formed at the rear edge of the transcription bubble. These rearrangements enable RNAP to translocate in the forward direction and to bind the next NTP, whereas opposite rearrangements take place when RNAP translocates backwards. In different translocation states the nascent RNA 3´ end occupies different positions relative to the RNAP active site. While the conformational changes associated with RNAP translocation are large and involve extensive rearrangements of non-covalent interactions, measuring translocation is challenging because it does not leave a permanent mark on the transcribed DNA.

Determining the translocation state of a TEC with current methods relies on correlating the RNA 3´ end with the position of the RNAP on the DNA template. The position of RNAP has been conventionally determined by various foot-printing techniques using either DNA endonuclease (Krummel & Chamberlin, 1992), DNA exonuclease (Nedialkov *et al*, 2013; Bar-Nahum *et al*, 2005; Kireeva & Kashlev, 2009; Toulokhonov *et al*, 2007; Komissarova & Kashlev, 1997a, 1997b; Nudler *et al*, 1997) or RNase (Komissarova & Kashlev, 1997a, 1997b) for determining the region protected by RNAP, or chemicals for mapping the unpaired/unstacked DNA bases of the transcription bubble (Komissarova & Kashlev, 1997a, 1997b; Zaychikov *et al*, 1995). When combined with the analysis of the nascent RNA by gel electrophoresis or RNA sequencing (Imashimizu *et al*, 2015), these methods can sometimes detect the RNAP translocation state with a 1 bp resolution. However, these measurements are semi-quantitative and are limited to equilibrium conditions or paused TECs. On the other hand, single molecule technologies that are based on optical or magnetic tweezers can track the movement of RNAP with a 1 bp accuracy under non-equilibrium conditions and exert forces on the TEC (reviewed in (Herbert *et al*, 2008)), but they cannot directly monitor the position of the RNA 3´ end and therefore are unable to temporally separate nucleotide addition and translocation (see however (Herbert *et al*, 2006; Dangkulwanich *et al*, 2013)). Additionally, activity measurements have been used for inferring translocation states: The sensitivity towards pyrophosphorolysis has been correlated with the occupancy of the pre-translocation state (Hein *et al*, 2011; Kashkina *et al*, 2006; Artsimovitch & Landick, 2000), whereas the sensitivity towards intrinsic, Gre-mediated or Fe^{2+}-induced RNA cleavage has been correlated with the tendency for backtracking (Nudler *et al*, 1997; Bochkareva *et al*, 2011; Kent *et al*, 2009; Larson *et al*, 2014). While these assays can under certain conditions be used for extracting translocation rates (see section 5.5), they are subject to confounding factors unrelated to translocation and do not directly probe the conformational changes taking place during translocation. Finally, the “running start, two bond, two quench” pre-steady state kinetic measurements allow the indirect detection of at least one translocation event (Kireeva *et al*, 2009). Here, the translocation is inferred from the interval between

nucleotide incorporation and the subsequent stable loading of the next NTP-Mg^{2+} into the active site (see section 5.4). However, translocation is measured indirectly and the method is not applicable for measurements under equilibrium conditions.

Here, we present a direct method for detecting translocation. It is based on probing the conformational changes that take place in the transcription bubble and the RNA:DNA hybrid during the translocational motions by monitoring the fluorescence of the environmentally sensitive probes. As such fluorescent probes may report conformational changes not necessarily related to translocation, we also provide cross-validation with complementary techniques, namely by site-specific cross-linking of nucleic acids, and quenching the reaction with EDTA (introduced in section 5.4). Fluorescence monitoring enables translocation measurements in equilibrium and time-resolved assays. Parallel kinetic measurements of nucleotide addition and PP_i release further allow the separation of translocation from these steps, which is absolutely essential for determining the minimal mechanism of the RNAP nucleotide addition cycle. Additionally, the site-specific nucleic acid cross-linking methods enable the mapping of the boundaries in the RNA:DNA hybrid and the transcription bubble. Together, the fluorescence and nucleic acid crosslinking methods allow the probing of the structures of the upstream (section 5.3) and downstream fork junctions (section 5.4) providing information on how the transcribed DNA sequence and transcription elongation factor NusG (sections 5.4 & 5.5) control the structure of the transcription bubble.

5.1.1 Monitoring RNAP translocation with 6-methyl-isoxanthopterin fluorescent beacon

Fluorescent base analogues embedded in the transcribed DNA have been used to map the transcription bubble of the T7 RNA polymerase (Liu & Martin, 2001) and the DNA melting by multisubunit RNAPs (Kashkina *et al*, 2007). Such probes are sensitive to the structure of DNA and could analogously be used for measuring conformational changes that take place during translocation. 6-methyl-isoxantopterin (6-MI) (**Figure 10A**) is a fluorescent analogue of guanine, which has been utilized as a DNA conformation specific fluorescent beacon. The melting temperatures of 6-MI-containing oligonucleotides paired to complementary strands are almost identical to those of control oligonucleotides where the 6-MI label is replaced by a guanine (Hawkins, 2008). By placing 6-MI near or at positions that undergo strand separation or base unstacking the probe can display either low fluorescence when in dsDNA or high fluorescence when in ssDNA (Hawkins *et al*, 1997). Considering these properties, we first placed the 6-MI probe in the upstream fork junction of the TEC where the RNA:DNA hybrid separates and the DNA strands reanneal. Upon walking the TEC along the DNA template by additions of subsets of NTPs (TEC16-19), 6-MI moved relative to RNAP and yielded distinguishable alterations in the amplitude of the fluorescence intensity upon single nucleotide additions (**Figure 10A** and **Study II, Figure S1**). In TEC16, the 6-MI was positioned 8 nucleotides upstream from the RNA 3´ end. In this position the complete fluorescent beacon, consisting of a 3´-GXT-5´ (X=6-MI) sequence, was paired with the RNA. The upstream guanine was presumably responsible for the low 6-MI fluorescence, because the neighboring stacked purines effectively quench the 6-MI fluorescence through non-radiative energy transfer (Hawkins, 2008). Upon formation of the

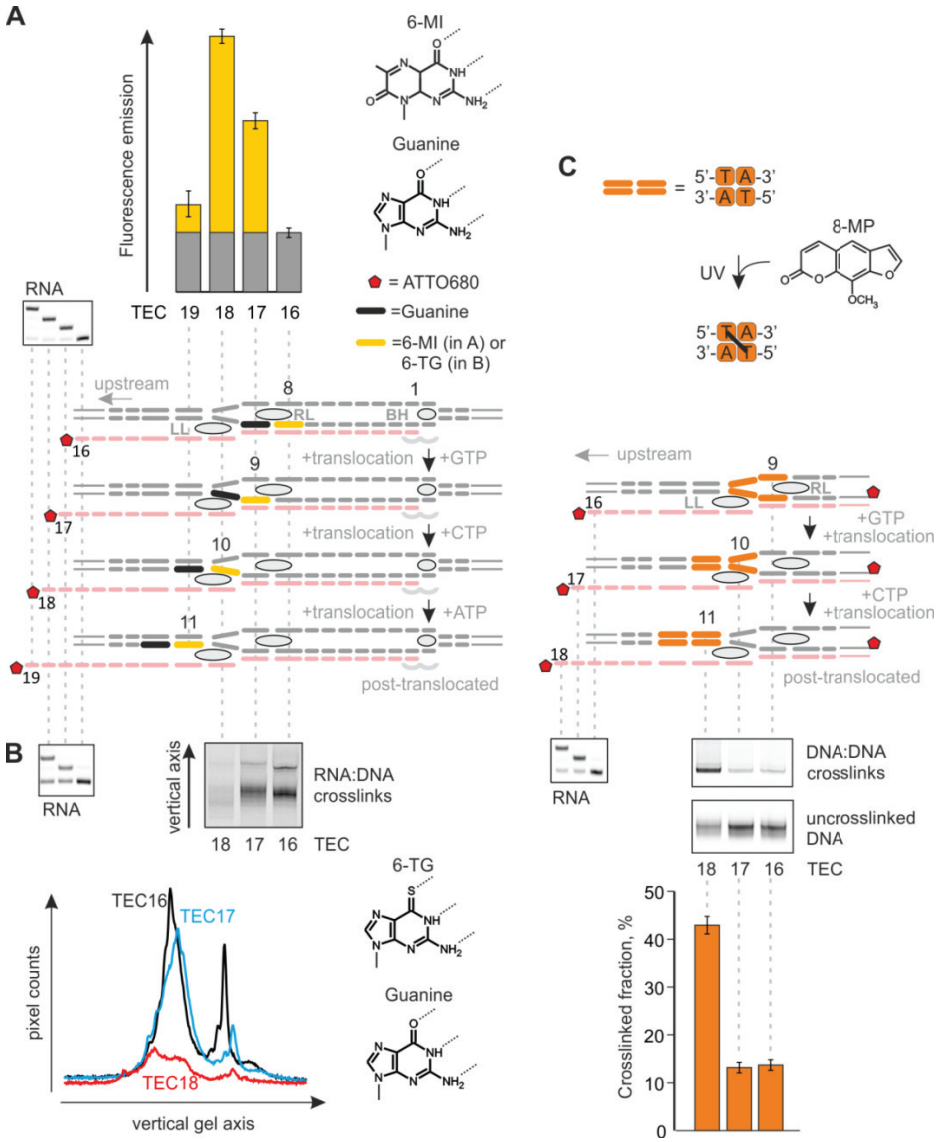


Figure 10. Complementary methodologies for monitoring the position of RNAP along the DNA template. A) Probing translocation with the 6-MI fluorescent beacon. The 6-MI structure and comparison to guanine is shown on the right. TECs were walked by up to three nucleotides which positioned the 6-MI probe (yellow bar in the template DNA) 8-11 nucleotides from the RNA 3´ end. The samples were illuminated at a 340 nm wavelength and the fluorescence emission was recorded at 420 nm (bars). Schematic representation of the transcription bubble is shown for each position in the post-translocated register; see Figure 12 for distinguishing between the translocation states. The RNA extension was monitored by analyzing the Atto-680 labeled RNAs by PAGE (gel panel). **B)** Mapping the RNA:DNA interaction by 6-TG UV-crosslinking. The 6-TG structure and comparison to guanine are shown on the right. The gel panels that were cropped from the same gel show the Atto680-scans of free RNA (left) and the slower migrating RNA:DNA crosslinks (right). The quantitation of the crosslink products is shown below. **C)** DNA:DNA interaction mapped by 8-MP mediated TA-site photo-crosslinking. The principle of the method and the structure of 8-MP are shown above. The fraction of the crosslinked template DNA relative to the total amount of template DNA (labeled with Atto-680) was quantitated (bar graph). The Atto-680 scan of RNAs (left) was cropped from the same gel. The figure was adapted from Turtola & Belogurov 2016.

TEC17 by nucleotide addition, the 6-MI fluorescence increased 2.5-fold. The increase in the 6-MI fluorescence originated from the translocation of the TEC17 complex, because the TEC16 was stabilized in the post-translocated state by the 9 bp RNA:DNA hybrid (Kashkina *et al*, 2006) and the 5´-AG-3´ dinucleotide sequence of the RNA 3´ end (Hein *et al*, 2011). Presumably the 6-MI fluorescence was unquenched as a result of the separation of the upstream guanine from the RNA:DNA hybrid (**Study II, Figure S1**). In the next position (TEC18), the fluorescence increased nearly 4-fold from the initial level. Here, the 6-MI was separated both from the RNA and the upstream guanine. In turn, in TEC19 the fluorescence returned to close to the initial level, presumably because the 6-MI entered the dsDNA.

As is described in the following sections, the fluorescence change associated with translocation could be temporally separated from the preceding nucleotide addition step (TEC16 to TEC17) in time-resolved assays, which allowed the determination of the translocation rates. Furthermore, the 6-MI fluorescent beacon system was used in various equilibrium translocation assays and for the structural probing of the upstream fork junction, thus making it here the most versatile tool for measuring RNAP translocation. We also positioned the 6-MI probes in the downstream DNA, which allowed independent monitoring of the downstream DNA translocation (section 5.4). One should be aware, that while the 6-MI probes in general have minimal effects on the DNA structure and the melting temperature (Hawkins *et al*, 1997), in some assays the 6-MI affected the kinetic properties of the TECs (see section 5.4). This indicates that the 6-MI is not a completely neutral probe despite its close structural similarity with guanine.

5.1.2 RNA:DNA photo-crosslinking with 6-thioguanine

During translocation the RNA:DNA hybrid moves one register within the RNAP´s nucleic acid binding channel by rotation along its helical axis. Translocation into the forward direction breaks one base pair at the rear end of the RNA:DNA hybrid. The clash of the upstream edge of the hybrid with the lid loop directs the DNA outside of the nucleic acid binding channel and the RNA towards the RNA exit channel (see section 2.3.1). Upon backward translocation this RNA:DNA base pair is presumably re-formed. To directly monitor the RNA:DNA interaction and the hybrid separation, we placed a photo-activated crosslinker 6-thioguanine (6-TG) into the template DNA (**Figure 10B**). Upon UV-illumination, 6-TG forms short-range covalent crosslinks with nucleic acids and amino acid side chains (Favre *et al*, 1998; Temiakov *et al*, 2003). Because of close structural similarity with guanine, 6-TG is expected to minimally influence the nucleic acid structure. The RNA:DNA interaction was probed by walking a single 6-TG probe through the upstream fork junction. The oligonucleotides had otherwise the same sequence as in Figure 9A except that the 6-MI in the template strand was replaced with 6-TG. TECs were walked by up to two nucleotides and illuminated with UV-light. The Atto680-labeled samples were analyzed by PAGE where the cross-linked RNA:DNA products migrated slower than the non-cross-linked RNA. As can be seen from the gel in **Figure 10B**, the template DNA and RNA formed crosslinks when the 6-TG was 8 (TEC16) and 9 (TEC17) nucleotides upstream from the RNA 3´ end. However, crosslinking did not occur when 6-TG was 10 nucleotides upstream from the RNA 3´

end (TEC18). Thus, our measurements with 6-TG crosslinking and the fluorescent 6-MI beacon together demonstrate that the RNA:DNA hybrid was 9 bp long after translocation.

In contrast to many other photo-activated groups that are bulky, 6-TG can be placed in the RNA:DNA hybrid and inside the tightly spaced RNAP nucleic acid binding channel. Furthermore, the short crosslinking distance restricts the possible crosslinks to a small area, which further enhances specificity. However, the interpretation of the 6-TG crosslinking results is complicated by the observation that the crosslinking does not simply follow the RNA:DNA base pairing, but rather reflects the physical accessibility of the probe with the RNA (**Study V, Figure 5 supplement 1**). Also, the protein residues can compete with RNA for crosslinking. Therefore, the structural changes of RNAP in the vicinity of the 6-TG probe can alter the crosslinking efficiency of 6-TG towards the RNA in several ways: by changing the RNA:DNA pairing, by affecting the physical accessibility of RNA to DNA or by changing the 6-TG crosslinking efficiency towards the protein groups.

5.1.3 DNA:DNA photo-crosslinking with 8-methoxypsoralen

Following the separation of the RNA:DNA hybrid, the template DNA emerges out of the main nucleic acid binding channel and reanneals with the non-template DNA behind RNAP. We probed the annealing of DNA by crosslinking utilizing the double stranded DNA (dsDNA) specific crosslinking agent 8-methoxypsoralen (8-MP) (**Figure 10C**). 8-MP intercalates into the dsDNA and upon UV-illumination covalently attaches to two thymines on opposite strands. Importantly, the crosslinking is limited to 5´-TA-3´ or 5´-AT-3´ dinucleotide motifs (Cimino *et al*, 1985), (**Study V, Figure 4, supplement 1**), allowing site-specific probing of the DNA structure. We placed a single TA-motif in the transcribed DNA at the upstream fork junction of the TEC. The TECs were walked by up to two nucleotides and illuminated with UV-light. The Atto680-labeled samples were analyzed by PAGE, where the cross-linked DNA:DNA products migrated slower than the non-cross-linked DNA and the RNA. In the initial TEC16, the downstream border of the TA-motif was positioned 9 nucleotides upstream of the RNA 3´ end. Extension of the RNA by one (TEC17) or two nucleotides (TEC18) moved the TA-motif 10 and 11 nucleotides upstream of the RNA 3´ end. TEC16 and TEC17 yielded little crosslinked products, whereas TEC18 crosslinked strongly and with comparable efficiency to that of the double stranded DNA without RNAP (**Figure 10C** and **Study V, Figure 4, supplement 1**). This indicated that the conformation of the TA-motif became suitable for 8-MP photo-binding 11 nucleotides upstream of the RNA 3´ end, marking the stable upstream border of the transcription bubble.

Importantly, the lack of TA-site crosslinking does not alone prove that the DNA is single stranded because the distortion of the DNA structure by e.g. protein contacts may disfavor 8-MP photo-binding in the double stranded TA-site. For example, the crosslinking efficiency in the TA-site was greatly reduced downstream of the transcription bubble where the double stranded DNA makes extensive contacts with RNAP (see section 5.4.4). However, in some instances the sensitivity of the crosslinking in the TA-site towards subtle conformational changes in the DNA offered a unique tool for probing the TEC structure (see sections 5.3, 5.4 and 5.5)

5.2 The thermodynamic basis of RNAP translocation and its modulation by the inhibitors TGT and CBR703 (Studies **II, III** and unpublished observations)

5.2.1 Time-resolved measurements of individual steps along the RNAP nucleotide addition cycle

Knowledge of the sequence of the reaction steps and their duration is essential for defining a minimal mechanism for an enzymatic reaction (Shimamoto, 2013). In this respect the RNAP nucleotide addition cycle is poorly characterized, and for example the kinetics and relative order of translocation and PP_i release are unknown (Erie *et al*, 1992; Kireeva *et al*, 2010). We studied the kinetics of the nucleotide addition cycle by performing parallel time-resolved measurements of the individual reaction steps (**Figure 11A**). To measure the nucleotide addition reaction, TEC16 and the next incoming NTP were mixed in a quenched flow apparatus. The reaction was stopped with HCl after variable reaction times and the samples were analyzed by PAGE to resolve the amount of extended RNA as a function of time. To allow detection of the reaction products, TECs were assembled using an RNA that was 5´ labeled with the Atto680 fluorophore. Translocation was measured with the 6-MI based fluorescent beacon assay (described in the previous section) by mixing TEC16 and NTP in a stopped flow device and continuously monitoring the increase in fluorescence. Time-resolved assays for the detection of PP_i and P_i were reported by (Brune *et al*, 1994; Pais *et al*, 2005) and were adapted here to transcription reactions (see methods). TEC16 and NTP supplemented with inorganic pyrophosphatase (PPase) and a fluorescently labeled phosphate binding protein (MDCC-PBP) were mixed in the stopped flow device and the increase in fluorescence was monitored continuously. PPase hydrolyzed the released PP_i to orthophosphates (P_i), which bound to MDCC-PBP increasing the fluorescence of MDCC.

Mixing TEC with the next incoming NTP initiated the reaction in a synchronous fashion lengthening the RNA by one nucleotide (**Figure 11A**). PP_i release and translocation were clearly delayed relative to nucleotide incorporation, allowing the inference of their rates from the delay between the RNA extension curve and the respective fluorescence traces (see Materials and Methods). The resolution of the individual reaction steps revealed important aspects of the catalytic properties of RNAP. First, whereas the rate of nucleotide incorporation varied for different NTPs, the rates of the post-catalytic steps were largely independent of the transcribed nucleotide (**Figure 11B**). This small sample suggests that the sequence specific effects on the catalytic rate may not manifest at PP_i release or RNA:DNA hybrid translocation steps. Second, the post-catalytic steps were clearly delayed relative to the nucleotide addition and comprised 30-50 % of the total duration of the cycle. Therefore, the nucleotide addition cycle does not have a single rate-determining step, but minimally two rate-contributing steps. In effect, the post-catalytic steps significantly contribute to the overall elongation rate of RNAP, and may become subject to modulation by transcriptional regulators. Third, the release of PP_i always preceded or coincided with translocation. This suggests that the PP_i bound TEC is stabilized in the pre-translocation state and the release of PP_i triggers the translocation of the RNA:DNA hybrid. Further studies on the connection between translocation and PP_i release will be presented in section 5.2.6.

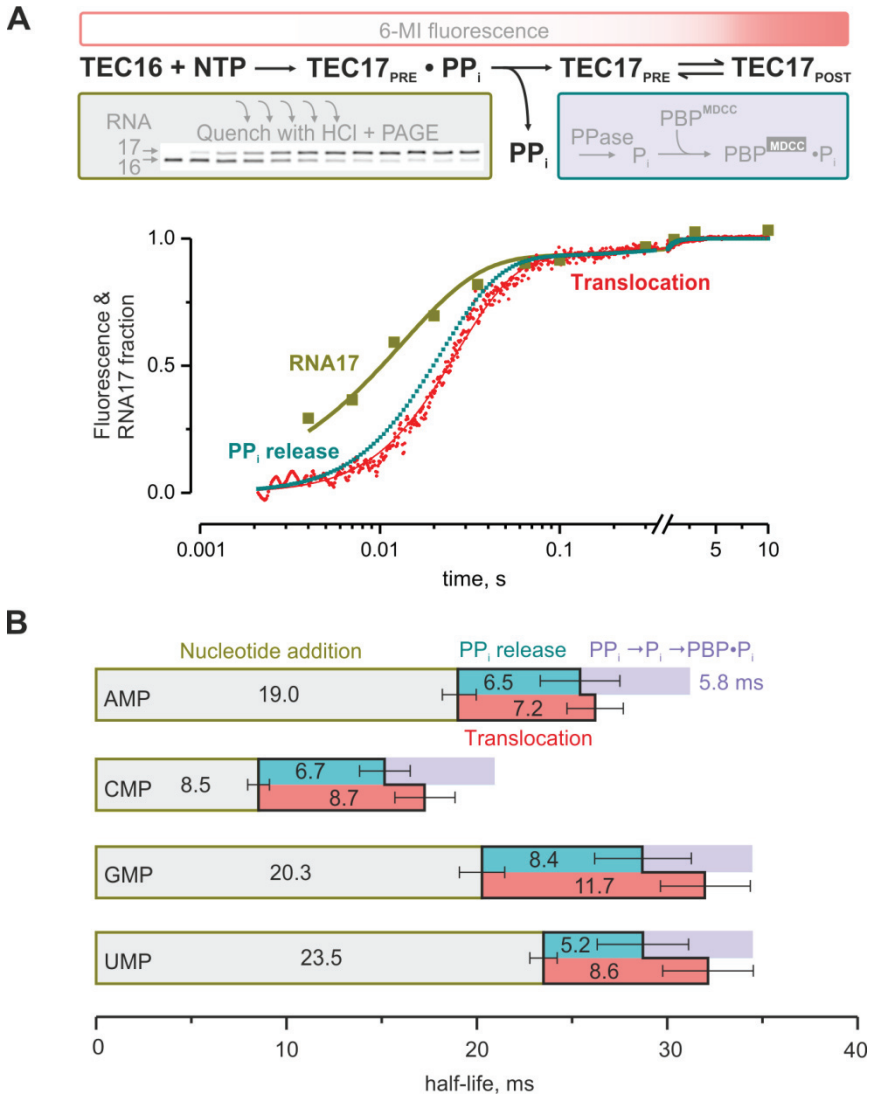


Figure 11. Time resolved dissection of the RNAP nucleotide addition cycle. A) Measurements of the individual reaction steps. The scheme of the nucleotide addition cycle and the principle of the methods are shown above. TEC16 and NTP were mixed together in a quenched flow (nucleotide addition measurement) or stopped flow (PP_i release and translocation measurements) instruments. The graph shows data from a parallel nucleotide addition, PP_i release and translocation measurements for the CMP incorporation reaction. The PP_i release trace was kinetically modelled from the experimental trace to remove the signal delay originating from PPi hydrolysis by PPase and P_i binding to PBP-MDCC. **B)** Comparison of the nucleotide addition and translocation kinetics between different incorporated NMPs. Figures were adapted from (Malinen et al. 2012).

The major fraction of the assembled TECs completed the nucleotide addition cycle within 100 ms (half-life 20-30 ms) after the addition of NTPs, which was overall consistent with the *E. coli* transcription elongation rates determined in vivo (Vogel & Jensen, 1994b). However, in most experiments a small 5-15% fraction of TECs incorporated NTPs ~50-fold slower than the majority of the TECs (**Figure 11A**). The same slow fraction was observed in NMP incorporation,

translocation and PP_i release assays, indicating that it originated from TEC16. This fraction may represent misassembled TECs that react via a different pathway or alternatively paused TECs that isomerize back to the on-pathway before the addition of a nucleotide can take place. We have not been able to identify the nature of these complexes but we prefer the latter explanation because the kinetic characteristics of these TECs are indistinguishable from paused TECs.

5.2.2 Mechanism of RNAP inhibition: Tagetitoxin traps TEC in the pre-translocated state by stabilizing the helical TL

Tagetitoxin (TGT) is produced by the plant pathogen *Pseudomonas syringae* pathovar *tagetis* and it inhibits the RNA synthesis activity of chloroplast and *E. coli* RNAPs (Mathews & Durbin, 1989). The TGT binding site in the RNAP:TGT binary complex partially overlaps with the active site of RNAP, but TGT is not thought to prevent substrate binding completely (Vassylyev *et al*, 2005). To elucidate the molecular mechanism for how TGT inhibits RNAP catalysis, we performed time-resolved translocation experiments in the presence of TGT. Notably, TGT did not affect the translocation rate up to 100 milliseconds at which point forward translocation and PP_i release were completed (**Study II, Figure 3**). However, at a time scale of seconds, TGT caused a fluorescence decrease close to the initial level. This suggested that TGT did not affect the rates of nucleotide addition, forward translocation or PP_i release but induced net backward translocation.

We corroborated the effect of TGT on translocation by equilibrium measurements (**Figure 12**): First, the addition of TGT to TEC17 lowered the 6-MI fluorescence near to the level of TEC16 (**Figure 12A**, left), whereas the addition of TGT to TEC18 lowered the fluorescence near to the level of TEC17 (**Figure 12A**, right), confirming that RNAP translocation was biased backwards by one nucleotide register. Second, TGT had an identical effect on the 2-aminopurine (2-AP, fluorescent analogue of adenine) positioned in the template DNA at the RNAP active site (**Study II, Figures S1 and S2**), demonstrating that the RNA:DNA hybrid and the downstream DNA moved uniformly. Third, the addition of TGT to TEC18 restored the 6-TG crosslinking to the level of TEC17 (**Figure 12B**), reaffirming the fluorescence observations. Fourth, the addition of TGT to TEC18 reduced the TA-crosslinking near to the level of TEC17 (**Figure 12C**), indicating that TGT induced the backward movement of the entire transcription bubble. These results suggest that TGT stabilizes the TEC into the pre-translocated state.

As proposed in the previous section, the binding of the natural ligand PP_i to the RNAP active site stabilizes the pre-translocated state. It would be structurally plausible that PP_i stabilizes the helical TL that in turn restricts forward translocation. Therefore, we explored the hypothesis that TGT stabilizes the pre-translocation state in an analogous manner by stabilizing the helical TL. Consistently, TGT failed to stabilize the pre-translocation state in the TEC assembled with an RNAP bearing a deleted TL (ΔTL) (**Figure 13A**). As the RNAP:TGT binary complex formed without molecular contacts between TL and TGT (Vassylyev *et al*, 2005), TGT presumably retained the ability to bind to the ΔTL RNAP. Next, substitutions were made to amino acid residues that were

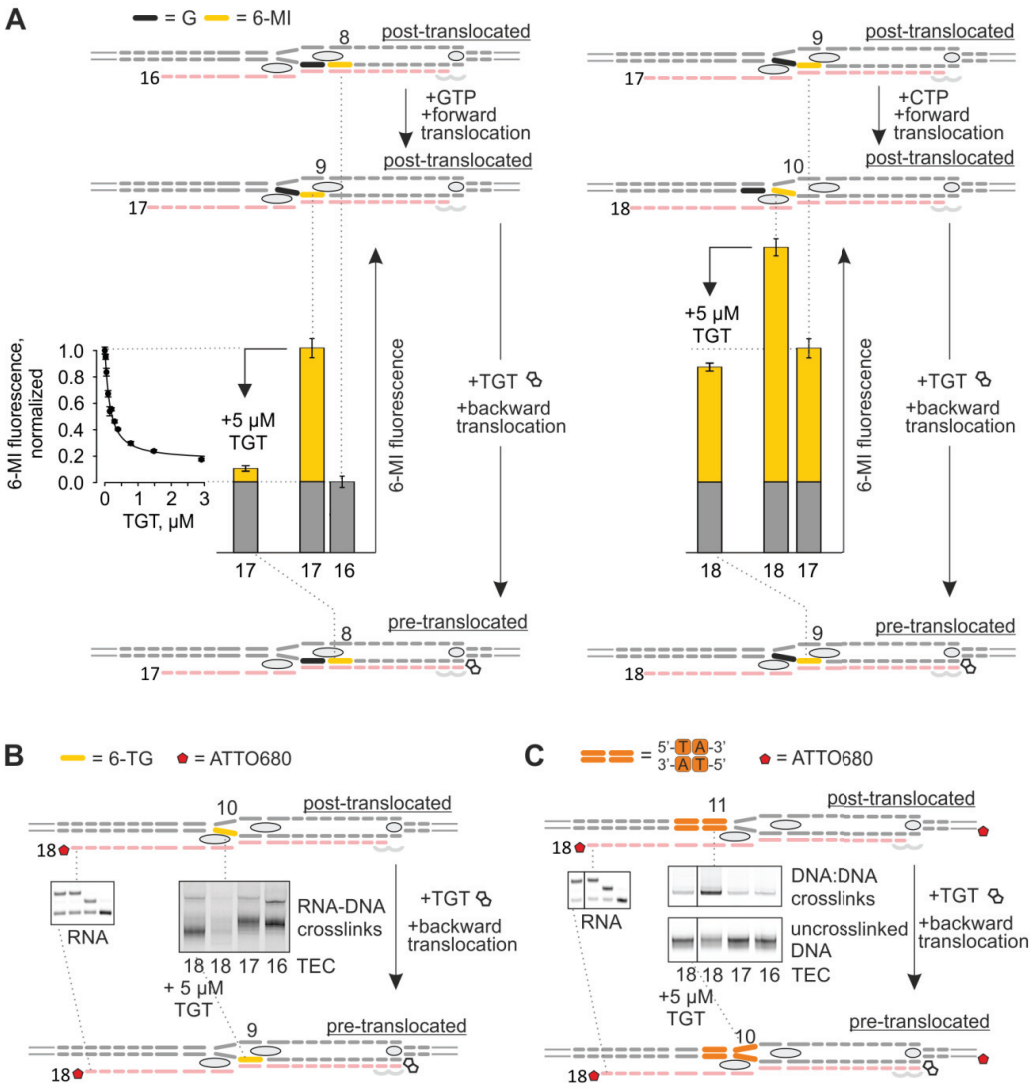


Figure 12. The effect of TGT on translocation. A) Addition of TGT to TEC17 (*left*) and TEC18 (*right*) decreases 6-MI fluorescence corresponding to the reversal of the TECs position by one nucleotide. B) TGT restores the RNA:DNA crosslinking in TEC18. C) TGT abolishes the DNA:DNA crosslinking in TEC18. The interpretations of the results are shown in schemes in which the predominantly post-translocated TECs become stabilized in the pre-translocated state in the presence of TGT. The figure was adapted from Turtola & Belogurov 2016.

expected to contact the pre-translocated RNA 3′ end in the *i+1* site. The TL residues β′Met^{932}, β′Arg^{933} and β′His^{936} interact with the RNA 3′ end only when the TL is helical (closed active site), whereas β′Asn^{458} interacts with the RNA 3′ regardless of whether the active site is open or closed. Substitutions to these residues reduced the effectiveness of TGT to bias TECs to the pre-translocated state (**Figure 13A**) arguing that interactions between RNA and the helical TL, as well as the helical TL and TGT were necessary for the TGT action. As a further support for the hypothesis, incubating TECs with steptolydigin (STL) reduced the ability of TGT to bias

translocation (**Study III, Figure 2**). STL is known to stabilize the TL in the unfolded state by binding ~20 Å away from the TGT binding site (Temiakov *et al*, 2005; Tuske *et al*, 2005; Vassylyev *et al*, 2007b). Finally, the non-hydrolysable nucleotide analogue complementary to the next base in the template DNA competed with TGT (**Study II, Figure 6**). By binding to the active site, the nucleotide analogue forward-biased translocation counteracting the TGT effect.

In order to investigate the structural basis of TGT inhibition, we modelled the TEC in the closed pre-translocation state and docked it with TGT from the binary RNAP:TGT complex (Vassylyev *et al*, 2005) (**Figure 13B** and **Study II, Figure S4**). The modelling supported the hypothesis that TGT could bind in the closed active site and interact with the helical TL. This direct stabilization of the TL is responsible for the TGT´s allosteric effect on the RNA and the translocation equilibrium: the helical TL stabilized by TGT contacts the RNA 3´ end in the nucleotide addition site (A-site) effectively biasing the translocation equilibrium towards the pre-translocated state. Interestingly, a phosphate and a carboxyl moiety of TGT overlapped with the putative binding site of PP_i, suggesting that TGT may act as a high affinity PP_i -mimic. This is mechanistically analogous to the action of the broad spectrum antiviral drug phosphonoformic acid (foscarnet). Foscarnet is a PP_i -analogue that inhibits the reverse transcriptase of the human immunodeficiency virus (HIV) and viral family B DNA polymerases by trapping the pre-translocated closed state (product complex) of the enzyme (Marchand *et al*, 2007; Zahn *et al*, 2011).

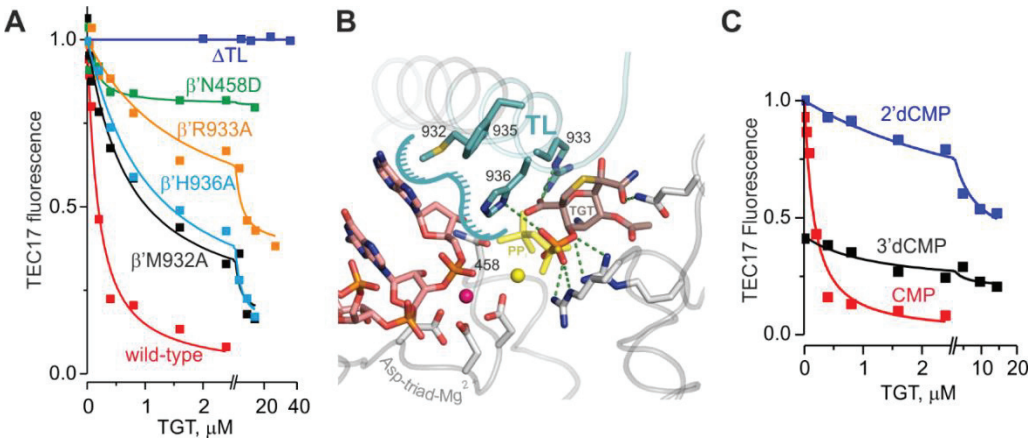


Figure 13. TGT traps TEC in the pre-translocated state by stabilizing the helical TL and can be used as a probe for RNAP translocation. A) TGT titration of AMP extended TECs assembled with RNAPs bearing substitutions in the active site. **B)** Model of pre-translocated TEC with bound TGT. The modeling was based on the *T. thermophilus* insertion complex (see Figure 4, state 4). TL residues β´Met^{932}, β´Phe^{935}, β´His^{936} and β´Arg^{933} (residues numbered according to *E. coli* RNAP) and their interaction surface with the RNA 3´end are colored with teal. The residues of the aspartate triad (β´Asp^{460}, Asp^{462}, Asp^{464}) and the β´Asn^{458} are also shown. Mg^{2+}-ions bound in the aspartate triad (magenta) and complexed with PP_i (yellow) are shown as spheres. **C)** TGT titration of ribo-, 2´deoxy- and 3´deoxy-CMP extended TEC17. The 3´OH is required for stabilizing both the post- and pre-translocation states whereas 2´OH is critical for the stability of the pre-translocation state. In effect, the fluorescence level of a 2´dNMP extended TEC can be used as a reference for the complete translocation. Figures were adapted from Malinen *et al.* 2012.

In summary, we found strong evidence that TGT traps the RNAP in the pre-translocated state by stabilizing the helical TL. Other laboratories have also reached similar conclusions (Artsimovitch *et al*, 2011; Yuzenkova *et al*, 2013). Importantly, the amount of TGT needed to convert TEC17 into a low fluorescent state could be utilized as a quantitative measure to compare the stabilities of the pre-translocation state in different TECs. The identification of the TGT inhibition mechanism increases its utility as a tool to probe the conformational states of RNAP.

5.2.3 RNA:DNA hybrid translocation is intrinsically biased towards the post-translocated state

RNAP translocation has been suggested to operate by a thermal ratchet mechanism, whereby the RNA:DNA hybrid freely oscillates between pre- and post-translocation states driven by thermal motion, and the post-translocated state is captured by binding of the next incoming NTP (section 2.3.4 of the literature review). However, we observed near identical levels of fluorescence in the absence and presence of the next incoming NTP analogue (non-hydrolysable CMPcPP) (**Study II, Figures 6 and S2**), suggesting that TEC was intrinsically biased towards the post-translocation state.

Next, we investigated the molecular basis for the translocation bias by blocking the RNA-protein interactions in the active site of RNAP. The pre-translocated RNA occupies the A-site where also the substrate NTPs bind. RNAP effectively discriminates 2´ deoxy-NTPs in the A-site (Svetlov *et al*, 2004; Yuzenkova *et al*, 2010) (see section 2.2.5) and therefore we anticipated that the 2´ deoxy-NMPs in the RNA 3´ end would similarly discriminate the A-site. As a result, we expected that the 2´ deoxy-NMP extended TECs would disfavor the pre-translocation state. Consistently, the 2´ deoxy-NMP extended TECs were resistant to TGT (**Figure 13C & Study II, Figures 3 & S2**), suggesting that they were more strongly biased towards the post-translocation state than the NMP extended TECs. Thus, the 2´ deoxy extended TECs offered a quantitative reference for a fully post-translocated TEC (RNA 3´ end residing in the P-site). Interestingly, the fluorescence levels of 2´ deoxy-NMP and NMP extended TECs were always the same. Considering the sensitivity of the method to measurement errors, these observations suggest that the RNA:DNA hybrid was biased >85% towards the post-translocation state after the extension by NMPs.

Based on the structural data (Vassylyev *et al*, 2007a), the post-translocated RNA 3´ OH group interacts with the Mg^{2+} (A) bound in the Asp-triad. We investigated how this interaction affects the translocation bias by blocking it in two ways: by incorporation of a 3´ deoxy-AMP at the RNA 3´ end, and by sequestration of Mg^{2+} with EDTA. Notably, in both cases the preference of the RNA:DNA hybrid to occupy the post-translocated state disappeared (**Study II, Figures 3 & 7,** see also Figure 13C), suggesting that this interaction was critical for the observed translocation bias. On the other hand, the addition of EDTA to the 2´ deoxy-AMP extended TEC had a negligible effect on the translocation bias, reinforcing the view that the removal of the 2´OH group from the RNA end biased translocation strongly towards the post-translocation state.

In conclusion, we found that the active site in RNAP forms an anisotropic environment that favors the interaction of the RNA:DNA hybrid with the P-site and biases translocation towards the post-translocated state in the absence of the next NTP. This bias is largely explained by the strong interaction between the 3´ OH group of the post-translocated RNA and the Mg^{2+} (A) bound in the Asp-triad.

5.2.4 Identification of translocation deficient RNAP mutants

The conformational dynamics of the tripartite BH/TL/F-loop have been suggested to allosterically control the active site in RNAP (Bar-Nahum *et al*, 2005; Miropolskaya *et al*, 2009, 2010; Kireeva *et al*, 2008; Tan *et al*, 2008; Jovanovic *et al*, 2011; Zhang *et al*, 2010). RNAPs with single amino acid substitutions in this region, specifically β´F773V (in BH), β´P750L (in F-loop), and β´G1136S (in TL), transcribe long DNA templates at a higher rate than WT RNAP due their reduced propensity for pausing (Bar-Nahum *et al*, 2005; Svetlov *et al*, 2007; Nedialkov *et al*, 2013). We characterized the kinetic properties of these variants by measuring the individual reaction steps of the nucleotide addition cycle. Counterintuitively, the mutations inhibited forward translocation (**Figure 14**). First, the TECs assembled with variant RNAPs were ~30-50% pre-translocated under equilibrium conditions, and unlike the WT RNAP, they were forward-biased with the next incoming nucleotide analogue (**Figure 14A** and **Study III, Figure 2**). Secondly, the mutants had considerably slower forward translocation rates than the WT RNAP (**Figure 14B** and **Study III, Figure 3**).

β´Phe^{773} and β´Pro^{750} are located closely near the region where the tip of the helical TL contacts the BH and F-loop (**Figure 14C**). As we proposed in the previous sections, the unfolding of the helical TL limits forward translocation. Therefore, we suspected that the mutations may inhibit translocation by stabilizing the helical TL. In agreement with this proposal, β´F773V had increased affinity for TGT, and furthermore, STL, which stabilizes the unfolded state of the TL, biased it towards the post-translocated state (**Publication III, Figure 2**). On the other hand, β´G1136 is located at the hinge region of the TL that undergoes a loop-to-helix transition during the folding of the TL into the three helical bundle (THB) with the BH. As noted in (Mejia *et al*, 2015), the Gly to Ser mutation is expected to promote the helical form at the hinge and increase the overall stability of the THB.

In WT RNAP, the high rate of the forward translocation (k_{forw}) of the RNA:DNA hybrid (60-80 s^{-1}) masks the rate of backward translocation (k_{back}) in ensemble measurements (see **Note 1**). In contrast, the slow forward translocation rate observed in mutant RNAPs (10-30 s^{-1}) allowed the more accurate inference of the backward translocation rate from the combined analysis of translocation bias and the apparent translocation rate (**Figure 14A-B**, see **Note 1**). The backward translocation rate was ~5-10 s^{-1} for TECs bearing GMP at the RNA 3´ end, and ~20 s^{-1} for TECs bearing CMP at the RNA 3´ end (**Publication III, Figure 3**). Interestingly, the tendency

Note 1 The apparent rate (observed rate) of a reversible reaction is the sum of the forward and backward rates ($k_{app} = k_{forw} + k_{back}$) (Purich, 2010). The translocation bias is determined by the ratio k_{forw}/k_{back}.

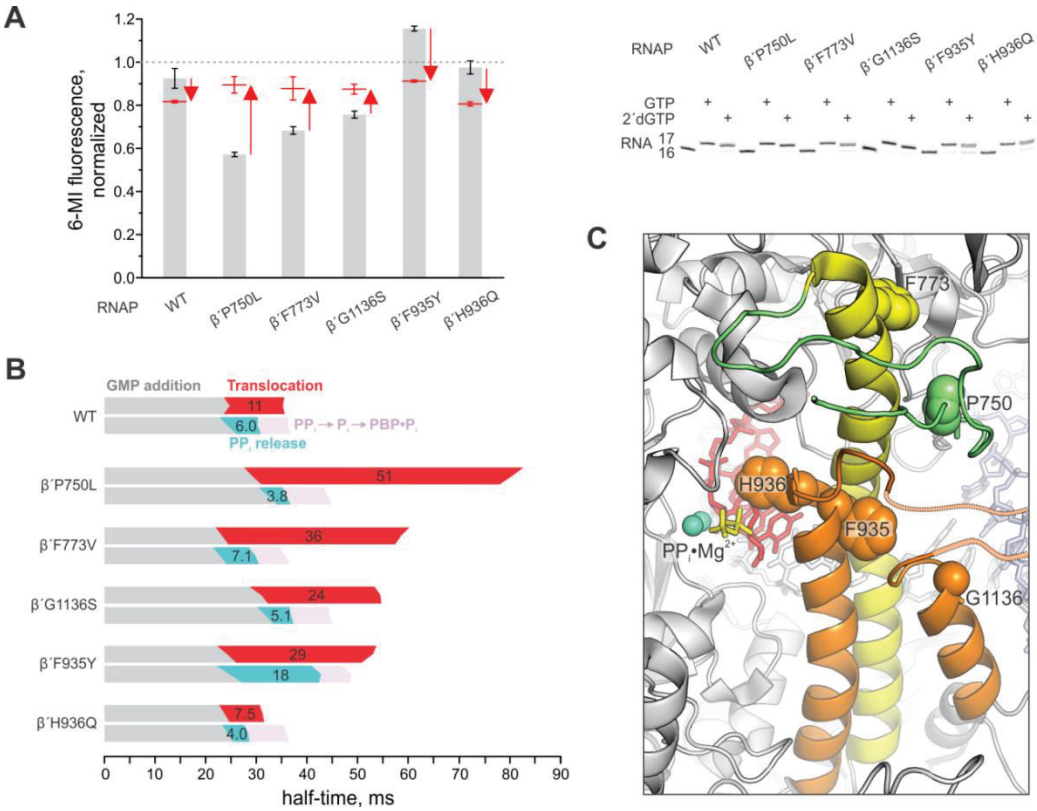


Figure 14. Identification of RNAP variants with alterations in translocation and PP_i release properties. A) Translocation bias of WT and variant RNAPs. The GMP extended TEC17 fluorescence levels (grey bars) were normalized to the fluorescence of 2´dGMP extended TEC17s (fully post-translocated reference, dashed line). The fluorescence levels following the addition of the next incoming non-hydrolysable nucleotide analogue, CMPcPP (500 μM), are shown by red lines and arrows. The gel panels on the right show the RNA extension controls (GTP 5 μM, 2´dGTP 100 μM). **B)** Kinetic analysis of nucleotide addition, forward translocation and PP_i release (GMP addition). Translocation and PP_i release reaction half-times (best-fit values) are indicated. The middle of the bar indicates the best-fit value whereas the upper and lower corners indicate lower and upper bounds, respectively, of parameters calculated at a 10% increase in Chi2 over the best-fit value using FitSpace routine of Kintek Explorer. The TECs were assembled from S041M-R024/R030-S042 oligonucleotides (see Materials & Methods). **C)** Structural view of the closed pre-translocated *E. coli* RNAP (PDB ID: 5IPL (Liu *et al*, 2016)) indicating the substituted amino acid residues. Part of the β´-subunit is omitted for clarity. Note that in the closed active site of *T. thermophilus* RNAP the Phe^{935} sticks towards the active site (Figure 13B). Other highlighted elements include: helical TL, *orange*; BH, *yellow;* F-loop, *green*; PP_i, yellow sticks; RNA, red sticks; DNA, *light purple*; Mg^{2+}, cyan spheres. The unresolved SI3 domain protruding from the TL tip is marked by a dashed orange line.

of the RNA:DNA hybrid to backward translocate appears to be dependent on the transcribed sequence. These rates depend on the stability of the post-translocated TEC and are presumably independent of the folding dynamics of the TL. Therefore, WT RNAP is expected to possess the same backward translocation rates.

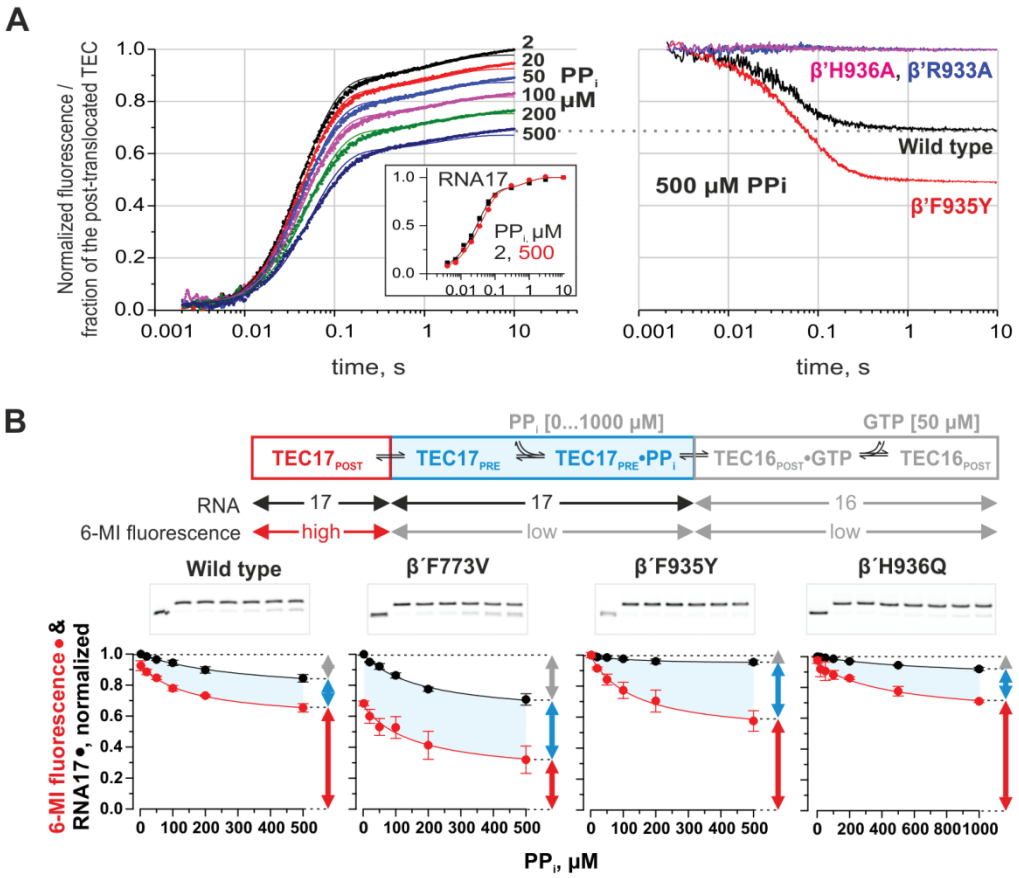
5.2.5 Mechanism of RNAP inhibition: CBR703 allosterically destabilizes the helical TL

CBR703 was discovered by the high throughput screening of a small molecule library against the *E. coli* RNA polymerase holoenzyme (Artsimovitch *et al*, 2003). CBR703 inhibits nucleotide

addition, but not translocation, and its inhibition potency varies along the transcribed template. The β´F773V and β´P750L mutations described above were initially identified by their resistance to the CBR703 inhibitor. Fascinatingly, cells bearing these RNAP mutations could not grow in the absence of the inhibitor (Artsimovitch *et al*, 2003). As we demonstrated in **Figure 14B**, the forward translocation rates of the β´F773V and β´P750L mutants were reduced. The mutants were also deficient in recognizing transcriptional regulatory pauses (**Study III, Figure 4**). Interestingly, in the presence of CBR703, the translocation defects of the β´F773V and β´P750L RNAPs´ were partially restored and they were able to recognize the transcriptional pauses (**Study III, Figures 2-4**). These observations provided a molecular explanation for why *E. coli* cells bearing β´F773V and β´P750L RNAP mutations are dependent on the CBR703 antimicrobials: the mutations deregulate transcription elongation preventing cell growth whereas these defects are compensated by CBR703. In the WT RNAP, CBR703 slowed the rates of nucleotide addition and translocation two-fold. However, CBR703 dramatically reduced the overall elongation rate by hyper-sensitizing RNAP to transcriptional pauses. It is likely that the impeded transcription elongation causes the CBR703´s toxic effect on WT cells.

CBR703 forward biased the translocation equilibrium of the WT and β´F773V RNAPs (**Study III, Figure 2**). An action similar to that of STL and opposite to that of TGT suggested that CBR703 could destabilize the helical TL. The location of the CBR703 resistance mutations (Artsimovitch *et al*, 2003) and our structural modelling suggested that CBR703 binds to the interface between the BH and the β-subunit ~20-30 Å away from the active site (**Study III, Figures 1 and S4**). Later, this binding site was confirmed by a structural analysis (Feng *et al*, 2015; Bae *et al*, 2015). We propose that the inhibitory effect of CBR703 on the catalysis is mediated allosterically by interfering with the interaction between the BH and the β-subunit. In effect, this destabilizes the contacts between the BH cap (involving the F-loop) and the helical TL. Thereby, CBR703 accelerates the translocation of the β´F773V and β´P750L RNAPs by destabilizing the helical TL. At the same time, these allosteric interactions mediate RNAP isomerization to off-pathway states. CBR703 enhances pausing by disfavoring the conformations of BH and TL that mediate processive elongation. Furthermore, as demonstrated by (Feng *et al*, 2015; Bae *et al*, 2015), part of CBR703's inhibitory effect may be related to direct blocking of the BH bending dynamics.

5.2.6 Translocation is thermodynamically coupled to PP_i release (unpublished observations)
The measurements of PP_i release and translocation rates suggested that PP_i release took place before translocation (**Figure 11**). In order to clarify the causal relationship between the two processes, we performed the PP_i release measurements for the translocation deficient RNAP variants. PP_i release took place with a similar or faster rate in β´F773V, β´P750L, and β´G1136S variants compared to the WT RNAP (**Figure 14B**). Due to the slow translocation in these variants, PP_i release unarguably preceded translocation. Furthermore, the lack of a one-to-one correspondence between PP_i release and translocation steps argues against the power-stroke mechanism (section 2.3.2 of the literature review). Satisfyingly, we identified two other RNAP variants with altered PP_i release kinetics. First, the β´F935Y mutant displayed a 3-fold longer retention of PP_i compared to the WT RNAP, and strikingly, it also translocated slower (**Figure**



C

RNAP	K_F^{PPi} (µM)	F^{TEC17}	F^{PPi}	K_{RNA}^{PPi} (µM)	$RNA17^{PPi}$
WT	153 ±22	0.93 ±0.01	0.57 ±0.02	328 ±34	0.74 ±0.01
β´F773V	157 ±77	0.67 ±0.03	0.22 ±0.08	179 ±29	0.60 ±0.03
β´F935Y	132 ±31	1.00 ±0.02	0.48 ±0.04	115 ±31	0.94 ±0.01
β´H936Q	464 ±227	0.95 ±0.01	0.60 ±0.07	798 ±79	0.86 ±0.01

Figure 15. PP_i backward biases RNAP translocation. A) Effect of PP_i on RNAP forward (left graph) and backward (right graph) translocation. In the left graph, forward translocation was monitored following the mixing of TEC16 (WT RNAP) with 200 µM GTP and varying concentrations of PP_i. The inset shows the parallel nucleotide addition experiments in the presence of 200 µM GTP and 0 or 500 µM PP_i. In the right graph, backward translocation of TEC17 (formed by incubation of TEC16 with 50 µM GTP) was monitored following the addition of 500 µM PP_i. GTP was kept at 50 µM after mixing with PP_i in order to prevent complete pyrophosphorolysis. The TECs were assembled from S041M-R030-S042 oligonucleotides (see Materials & Methods). **B)** Effect of pyrophosphate on the translocation equilibrium. The scheme above shows the equilibrium system for monitoring the TEC species and translocational states. Concentrations of PP_i and GTP affect the equilibrium; here the PP_i concentration was varied whereas GTP was kept constant at 50 µM. The graphs depict best-fit lines to the normalized 6-MI fluorescence (red dots) and the fraction of RNA17 (black dots).

14B). Secondly, the β´H936Q mutant had a faster PP_i release rate compared to the WT RNAP, and it translocated faster than the WT RNAP. These observations indicate that PP_i release and translocation are coupled in these RNAP variants. The amino acid substitutions were located in conserved residues at the tip of the TL (see Figures 13B & 14C), which pointed towards functional coupling between PP_i binding, TL folding dynamics and translocation.

If PP_i release and translocation were strictly coupled (one-to-one correspondence), an increasing PP_i concentration should progressively inhibit the rate of forward translocation. To test this prediction, TEC16 assembled with WT RNAP was mixed with GTP in the presence of 0-500 μM PP_i and the forward translocation was monitored. Noticeably, neither the nucleotide addition rate nor the apparent translocation rate (k_{app}) changed with an increasing PPi concentration (**Figure 15A,** left), meaning that the sum of the forward (k_{forw}) and backward translocation rates (k_{back}) remained unchanged (according to $k_{app} = k_{forw} + k_{back}$, see **Note 1, p. 69**). However, the final level of fluorescence decreased with an increasing PP_i concentration, and this effect was more pronounced in the β´F935Y mutant RNAP. Therefore, PP_i shifted the translocation bias without affecting the translocation rates. This paradoxical result can be reconciled if PP_i affected the equilibrium of a step that precedes translocation. This argument is enforced by the observation that a slower PP_i release rate in β´F935Y was accompanied by a *proportionally* slower forward translocation rate, and in β´H936Q the faster PP_i release rate was accompanied by a *proportionally* faster forward translocation rate (**Figure 14B**). If PP_i release and translocation were not connect at all, then translocation should take place after nucleotide addition always with the same delay. However, because this delay was proportional to the PP_i retention time, PP_i must modulate an equilibrium step that precedes translocation. Logically, this step could involve the folding and unfolding of the TL.

The decrease in the final fluorescence levels in the presence of PP_i indicated that PP_i induced backward translocation. We measured the backward translocation directly by mixing the pre-formed TEC17 with 500 μM PP_i. To protect the TEC17 bearing a GMP at the RNA 3´ from processive RNA pyrophosphorolysis, experiments were performed in the presence of 50 μM GTP. As expected, PP_i induced net backward movement in a millisecond timescale (**Figure 15A,** right). The amplitude of the time-resolved fluorescence decrease was dependent on the concentration of PP_i (data not shown) and was larger in the β´F935Y mutant RNAP. Importantly, the backward translocation was dependent on the TL, because amino acid substitutions to the β´His^{936} and β´Arg^{933} (see Figure 13) residues prevented the PP_i induced backward translocation.

Continued figure 15 legend from the previous page. Error bars are SDs of at least three independent measurements. The area between RNA17 and 6-MI fluorescence is colored blue to highlight the fraction of pre-translocated TECs. The arrows on the side of each graph depict the equilibrium occupancy of TECs at 500 μM PP_i. The gel panels above the graphs show examples of RNA samples analyzed by PAGE. Samples were (from left to right): TEC16, TEC17 (50 μM GTP) with 0, 20, 50, 100, 200, 500 μM PP_i (and 1000 μM for β´H936Q). The TECs were assembled from S041M-R030-S042 (fluorescence analysis) or S041M-R024-S166b (RNA analysis) oligonucleotides (see Materials & Methods). **C)** Fit parameters from PP_i equilibrium titrations (see methods). Parameters shown in red and black are from fits to 6-MI and RNA17 data, respectively. *K* is the apparent PP_i binding constant. F^{TEC17} and F^{PP} are fluorescence levels for PP_i-free and PP_i-saturated TECs, respectively. $RNA17^{PPi}$ is the fraction of RNA17 in the PP_i-saturated TEC.

PP_i bound in the pre-translocated state can catalyze pyrophosphorolysis, the reverse reaction of a nucleotide addition. The shift in the equilibrium between nucleotide addition and pyrophosphorolysis towards the latter could also decrease the 6-MI fluorescence. To separate the pre-translocated and pyrophosphorolyzed TECs, we combined the 6-MI translocation equilibrium measurements with the analysis of RNA by PAGE. Titration of TEC17 with PP_i in the presence of 50 µM GTP decreased the 6-MI fluorescence, but this was only partially accounted for by the pyrophosphorolysis of RNA17 to RNA16 (**Figure 15B**). This demonstrated that the fractional amount of pre-translocated TEC17 increased with an increasing PP_i concentration. The ability of PP_i to shift the translocation equilibrium towards the pre-translocated state suggests coupling between translocation and PPi release. However, fitting the equilibrium data to a hyperbolic function (modeling PPi binding to one site) indicated that at the saturating PP_i concentration (which was not experimentally achievable at 10 mM Mg^{2+} due to PP_i precipitation at concentrations above 1 mM) 57% of the TEC is in the post-translocation register (**Figure 15C**). The same experiments (forward and backward translocation, PP_i titration) with similar observations were acquired with the WT RNAP by assaying a TEC that incorporated CMP as the next nucleotide (data not shown). These results suggest that PP_i release is only weakly coupled to translocation and further strengthen the conclusion that PP_i affects the TL folding equilibrium that in turn controls the translocation equilibrium.

As expected, the pre-translocation state biased β´F773V RNAP was more sensitive to pyrophosphorolysis than WT RNAP (**Figures 15B & C**). Interestingly, the decrease in the 6-MI fluorescence upon PP_i titration was completely accounted for by pyrophosphorolysis and the relative amount of the pre-translocation state remained unchanged, consistent with the uncoupling of PP_i release and translocation in this mutant. The β´H936Q variant had a lower affinity towards PP_i compared to WT RNAP, explaining its faster PP_i release rate. On the other hand, the β´F935Y mutant with a slow PP_i release rate also exhibited a higher apparent PP_i binding affinity. Despite this mutant´s apparent sensitivity to PP_i, the amount of pyrophosphorolyzed RNA remained low. This paradoxical observation suggests that PP_i may have at least two binding modes in the RNAP active site: The first site may be occupied after nucleotide incorporation and from here PP_i can react back with the RNA 3´ end NMP in the pyrophosphorolysis reaction. The second site, which is preferred in the β´F935Y mutant, is formed by a small movement of PP_i. In this configuration, PP_i retains its ability to inhibit the unfolding of the helical TL, but cannot react with the RNA 3´ end.

These results collectively suggest that RNAP translocation is thermodynamically coupled to PP_i release through the TL folding dynamics. PP_i binding in the active site provides additional stability to the helical TL. This is exemplified both by the β´F935Y RNAP, which binds PP_i tighter and delays translocation as well as β´H936Q, which dissociates PP_i faster and allows faster translocation. When the helical TL is allosterically stabilized (β´F773V, β´P750L, β´G1136S), PP_i release may take place without immediate collapse of the helical TL and forward translocation. However, these results argue against a concerted model wherein the sequence of PP_i release, TL unfolding and translocation steps are obligatorily determined. Instead, there are three

equilibrium constants that control the translocation of the RNA:DNA hybrid: one for PP_i binding and release, the second for TL folding and unfolding, and the third for forward and backward translocation. Furthermore, β´His^{936} does not appear to function in acid/base catalysis, because the substitution to glutamine – which cannot exchange protons – retains a fast nucleotide addition rate. Thus, our observations are fully consistent with recent results by Mishanina *et al*. 2017 (Mishanina *et al*, 2017). On the other hand, our observation that the PP_i release rate is accelerated in β´H936Q does not provide support for the earlier proposals that PP_i release would be limited by its protonation (Castro *et al*, 2009, 2007) (see section 2.2.3).

Finally, while the backward bias of translocation induced by PP_i is only marginal at each sequence position, the effects translate into an approximately twofold inhibition of the transcription elongation rate over a long *rpoB* template at a physiologically relevant PP_i concentration (**Figure 16**). Accordingly, variations in the cytoplasmic PP_i concentration may play a role in the regulation of transcription elongation.

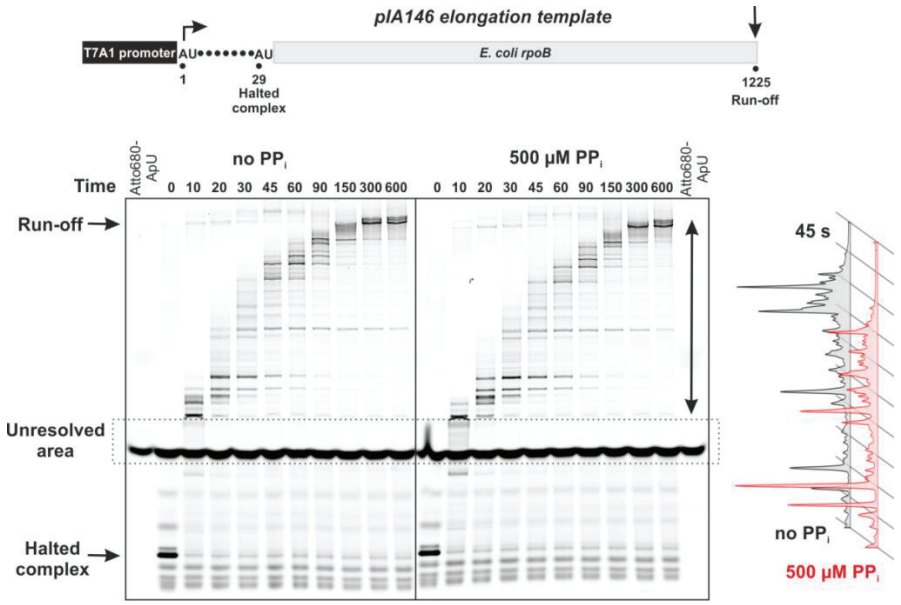


Figure 16. The effect of PP_i on transcription elongation on a long DNA template. The halted complex was formed by incubating the RNAP-$sigma^{70}$ holoenzyme with a PCR amplified pIA146 "elongation template" (Svetlov *et al*, 2007), an Atto680-labeled ApU initiating dinucleotide and a subset of NTPs (ATP, GTP and CTP). The sample was gel filtrated to remove excess Atto680-ApU. Transcription was initiated by the addition of all four NTPs (final concentration 200 µM), PP_i (0 or 500 µM) and rifapentin in order to prevent re-initiation. The intensity profiles along the vertical axis (quantitated area is marked by a line with two arrowheads) are shown for 45 s elongation time points. Note that the residual Atto680-labeled ApU migrates slowly on the gel and disturbs band separation immediately above (unresolved area).

5.3 Structure of the upstream fork junction (Study **V**)

The separation of the RNA:DNA hybrid and the re-annealing of the DNA strands at the rear end of the transcription bubble (the upstream fork junction) play important roles in maintaining the processivity of the TEC. The stability of the TEC depends on the length of the RNA:DNA hybrid (Sidorenkov *et al*, 1998; Komissarova *et al*, 2002; Kireeva *et al*, 2000a) and its separation modulates the activity of RNAP possibly by affecting translocation (Larson *et al*, 2014; Vvedenskaya *et al*, 2014; Imashimizu *et al*, 2015; Nudler *et al*, 1997) (see section 5.4). It has been proposed that DNA annealing is regulated by elongation factors (Klein *et al*, 2011; Zenkin *et al*, 2015) and promotes transcription termination (Park & Roberts, 2006; Ryder & Roberts, 2003), whereas prevention of upstream DNA annealing changes the properties of the TEC (Bochkareva *et al*, 2011; Toulokhonov & Landick, 2006; Kireeva *et al*, 2000a, 2000b). Consequently, an artificial change in the site of RNA:DNA separation or prevention of upstream DNA annealing may provide ways to perturb the translocation of RNAP (this strategy was utilized in sections 5.4, 5.5 and 5.6). Biochemical probing has detected 8-10 bp RNA:DNA hybrids (Nudler *et al*, 1997; Sidorenkov *et al*, 1998; Kyzer *et al*, 2007; Kashkina *et al*, 2006; Kent *et al*, 2009), whereas in the crystallized TECs the RNA:DNA hybrid is 9 bp in the post-translocated state (Vassylyev *et al*, 2007a). However, at the time our experiments were carried out, the upstream DNA was missing from all TEC structures, raising the question of where exactly does the RNA:DNA hybrid separate in complete TECs? Furthermore, because foot-printing techniques tend to detect variable transcription bubble sizes due to the lateral oscillation of RNAP along the DNA (Komissarova & Kashlev, 1997a; Nudler *et al*, 1994; Zaychikov *et al*, 1995), the location where the DNA strands pair and thereby the actual size of the transcription bubble have remained ambiguous (for example, compare the schematics in (Larson *et al*, 2014; Vvedenskaya *et al*, 2014)). For these reasons we investigated the structure of the upstream fork junction with several complementary biochemical techniques.

Figure 17 summarizes the results from 6-MI fluorescence, 8-MP and 6-TG crosslinking (Figures 10 & 12, and Study V, Figures 4-7) and Gre-factor mediated RNA cleavage assays (results described in the section 5.5). Critically, the employment of TGT in equilibrium measurements allowed the assignment of translocation states for each TEC (Figure 12) and thereby the unambiguous structural interpretation of the results. The RNA:DNA hybrid length was 9 bp in the post-translocated state and 10 bp in the pre-translocated state (**Figure 12B**). Interestingly, a careful comparison of the TEC17 and TEC18-TGT crosslinks in Figure 12B suggest that the conformation of the upstream edge of the RNA:DNA hybrid is different in pre-translocated and post-translocated states. However, the functional significance of this observation remains unclear. The B-form DNA started 11 nucleotides upstream from the RNA 3´ end (**Figure 12 C**). That said, the DNA strands appeared to base pair already 10 nucleotides upstream from the RNA 3´ end as evidenced by the effects of the DNA:DNA mismatches on the6-MI fluorescence intensity, 6-TG/8-MP crosslinking efficiency (Study V, Figure 7) and the rate of Gre-factor mediated RNA cleavage. The high fluorescence of 6-MI and the lack of 8-MP crosslinking at this position further argue that the -10 base pair deviates from the canonical B-form DNA or is very dynamic, or both (see section 5.3.3). In effect, the size of the translocation bubble is 10 bp. This

model is fully consistent with the cryo-electron microscopy structure of *E. coli* TEC (Kang *et al*, 2017), that was released after we published the upstream junction model derived from biochemical data.

The removal of the lid (LL) and the rudder loops (RL) compromised the annealing of the upstream DNA (**Study V, Figure 4**), confirming the structurally plausible hypothesis (Zhang *et al*, 1999; Korzheva *et al*, 2000; Gnatt *et al*, 2001) that these clamp-domain elements direct the path of the template DNA out from the main channel. The removal of the lid loop (LL) allowed the RNA:DNA interaction beyond the normal separation point (**Study V, Figure 5**), consistent with the role of the lid in promoting the separation of the RNA:DNA hybrid (Toulokhonov & Landick, 2006; Naryshkina *et al*, 2006; Vassylyev *et al*, 2007a).

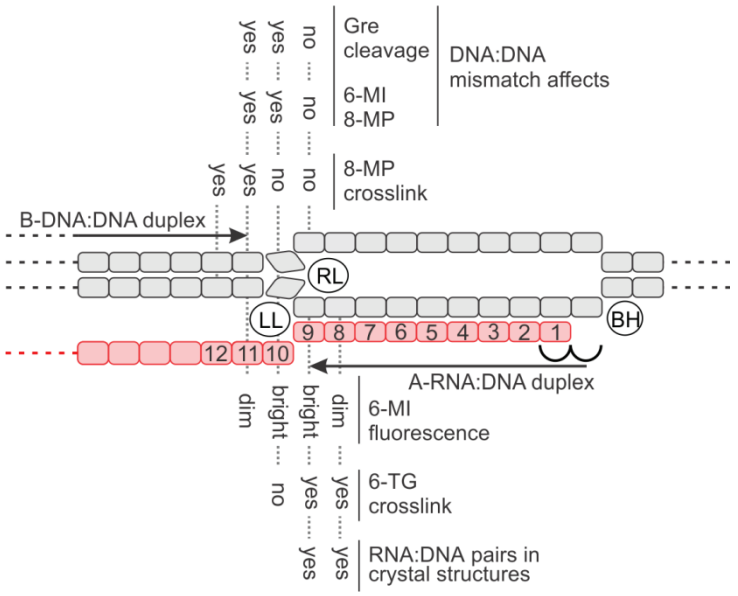


Figure 17. Summary of the results from experiments probing the structure of the upstream fork junction. 6-MI fluorescence, 6-TG and 8MP crosslinking results are shown in Figures 9 and 11. The effect of DNA:DNA mismatches on 6-MI and 8-MP crosslinking is shown in Study V, Figure 7 (including its supplements). Gre cleavage experiments are presented in Figure 21. The figure was adapted from Turtola & Belogurov 2016.

5.4 A two-step translocation mechanism (Study **VII**)

5.4.1 Measuring downstream DNA translocation with the 6-MI fluorescent beacons

The detection of intermediate translocation states in structural and simulation studies has led to the proposal that the RNA:DNA hybrid and the downstream DNA translocate in an asynchronous fashion (Brueckner & Cramer, 2008; Cheung *et al*, 2011; Silva *et al*, 2014) (section 2.3.1 of the literature review). To date the asynchronous translocation of RNAP has not been detected in biochemical experiments, and thus the existence of intermediate translocation

states during normal elongation let alone their implications on RNA synthesis remain unexplored. It also remains unclear whether intermediate translocation states represent intermediates on the normal translocation trajectory (on-pathway) or whether they prevent the progression of RNAP (off-pathway). In order to elucidate the motions of nucleic acids during translocation and to complement the previously described translocation experiments that monitored the movement of the RNA:DNA hybrid, we investigated the translocation of the downstream DNA rigorously with a combination of strategies. First, the translocation of the downstream DNA was monitored with the fluorescent beacon system. To this end, we placed the 6-MI fluorophores in the downstream non-template DNA strand (*i+3* in TEC16). Upon the addition of the next nucleotide (to form TEC17) and forward translocation, 6-MI moves to the upstream edge of the downstream DNA (*i+2*), where it is spatially separated from the upstream bases in the template and non-template DNA strands resulting in increased fluorescence (**Figure 18**). We assessed two types of TECs that transcribe different nucleotides: first G and second C (the GC-system) or first C and second G (the CG-system). The two systems differ in their downstream sequences but the upstream DNAs and RNA:DNA hybrids of the TECs are identical.

First, we assessed the extent of translocation by equilibrium fluorescence measurements (**Figure 19**). In both systems the extension by NMP (GMP in the GC-system and CMP in the CG-system) resulted in a lower fluorescence level than the extension by the corresponding 2´deoxy-NMP indicating that translocation was incomplete. Consistently, the NMP extended TECs were biased towards the post-translocated state by the addition of the non-hydrolysable nucleotide analogue NMPcPP (CMPcPP or GMPcPP, respectively). These results suggest that ~70% of GC-TEC17 and ~20% of CG-TEC17 were post-translocated. Interestingly, the CG-TEC17 fluorescence could not be saturated with GMPcPP under the experimental conditions (**Study VII, Figure S1**) demonstrating a low affinity towards the substrate (apparent K_D ~0.8 mM).

Next, we investigated the kinetics of translocation. Translocation was triggered by mixing TEC16 with 200 µM of the next incoming NTP, and the RNA extension and translocation were monitored in parallel experiments (**Figure 19A & E**). We now refer to this type of experiments as "post-catalytic" translocation measurements, because the monitored translocation step is preceded by the chemical nucleotide addition step (**Figure 18**). The apparent translocation rates (k_{app}) were inferred from the delay between the RNA extension and the 6-MI fluorescence curves. Assuming a simple one-step translocation mechanism, the apparent translocation rate could be split into forward (k_{forw}) and backward translocation rates (k_{back}) based on the assessment of the completeness of translocation as deduced from the fluorescence levels at equilibrium (see **Note 1** on page 69). This analysis revealed that the CG-TEC had a lower k_{forw} and a higher k_{back} than the GC-TEC (compare the "post-catalytic" schemes in **Figure 19D & H**). These results contrast with the previous observations with the TECs that harbored the 6-MI at the upstream end of the RNA:DNA hybrid. However, the results described in the next section confirmed that the downstream 6-MI labels specifically slowed down the forward translocation rate by about two-fold, possibly by stabilizing the base pair upstream of the probe.

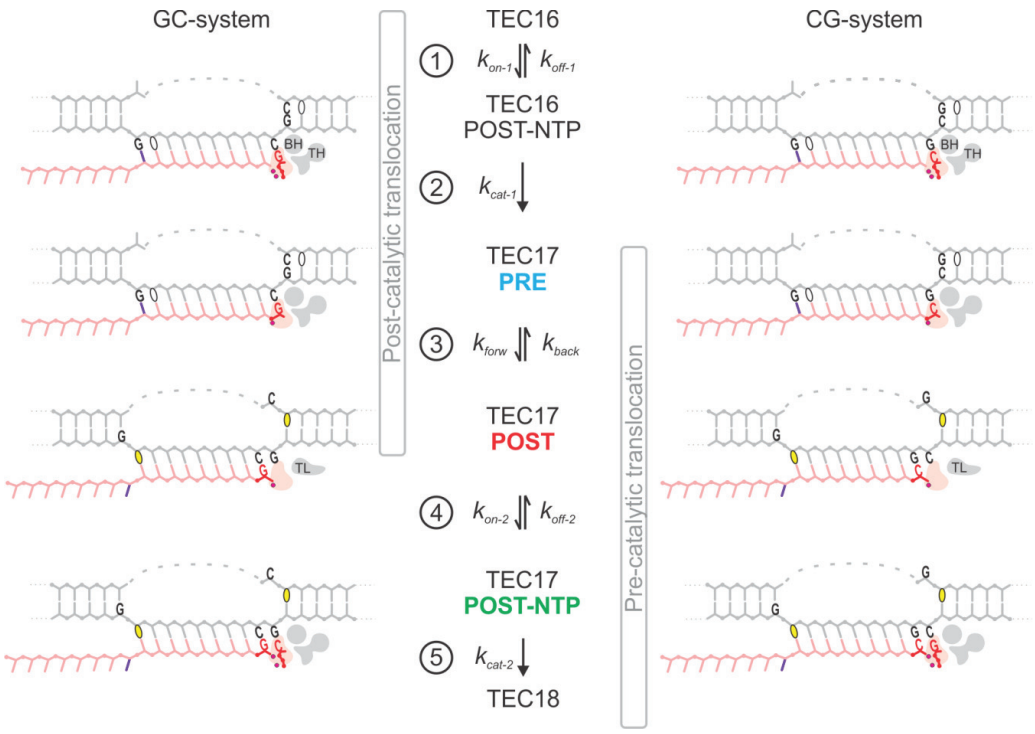


Figure 18. Experimental systems for monitoring translocation and nucleotide sequestration. The essential nucleic acid scaffold elements of the CG (left) and GC (right) transcribing TECs are displayed in different reaction states. The 6-MI probes are drawn as ovals either in the quenched (white) or the bright (yellow) states. For saving space, both the upstream and the downstream positions of the probes are included in the same figure, although experimental TECs contained only one probe (or no probe at all). The forward translocation increased the fluorescence intensity due to the spatial separation of the quenching guanine from the 6-MI fluorophore. The quenching guanine was placed upstream of the 6-MI fluorophores in the same strand (upstream and downstream positions in the GC-system) or in the non-template DNA (the downstream position in the CG-system). In a subset of experiments the complementarity of the RNA:DNA hybrid was shortened by a base substitution in the RNA (purple base). The oligonucleotide sequences and their combinations are specified in Tables 1 and 2. In the post-catalytic translocation measurements translocation was triggered by incorporating a NMP into the TEC16 (steps 1 and 2) and the subsequent translocation was recorded by measuring the 6-MI fluorescence (step 3). The apparent translocation rate ($k_{app} = k_{forw} + k_{back}$) was determined as the delay between the nucleotide addition (combined steps 1 and 2) and the fluorescence signal. In turn, the ratio k_{forw}/k_{back} in the TECs after NMP incorporation was inferred from the comparison of the 6-MI fluorescence levels with the fully translocated references (TEC17 formed by extension with 2´deoxy-NMPs). In the pre-catalytic translocation measurements, the reaction between the TEC17 and the next NTP was quenched with large excess of EDTA that chelates Mg^{2+} from solution and prevents un-sequestrated NTPs from reacting. The fraction of the TEC that proceeds to catalysis (TEC18) is dependent on the translocation (step 3) and NTP binding rates (k_{on-2} step 4), as well as the ratio of NTP dissociation to catalysis rates (k_{off-2} /k_{cat-2}).

As a result, the translocation equilibrium was shifted backwards, which usefully, however, revealed the backward translocation rates. These conclusions notwithstanding, the translocation equilibrium was more backward-biased in the CG-system in comparison to the GC-system.

We also perturbed the translocation step by shortening the RNA:DNA hybrid to 9 bp in TEC17 (the mismatched base is colored purple in Figure 18). This way the upstream edge of the RNA:DNA hybrid was melted already at the beginning of the translocation process. Importantly, this forward-biased the translocation equilibrium in both systems (**Figure 19B & F**, columns on the right). Curiously, in both systems the upstream mismatch increased the apparent rate of translocation only marginally, indicating that the pre-melting of the upstream base pair both increased the forward translocation rate *and* decreased the backward translocation rate. The implications of this finding is discussed below (section 5.4.3).

5.4.2 Pre-catalytic translocation measurements by EDTA quenching reveal the effects of the transcribed sequence and the fluorescent probes

The results described so far reported the translocation of the TEC17 *after* nucleotide addition (post-catalytic translocation, Figure 18). To establish that the translocation of the downstream DNA was completed to the extent that the TEC17s could bind the next NTP we employed the EDTA quench-method. Quenching with EDTA differentiates the post-translocated conformation capable of sequestering a NTP from other conformations that are unable to stably bind the substrate (Kireeva *et al*, 2009; Nedialkov *et al*, 2003; Kireeva *et al*, 2008). While quenching with HCl inactivates the enzyme, quenching with EDTA traps the substrate. The principle of the assay is based on the fact that the second catalytic Mg^{2+} ion binds to the active site with the substrate as a NTP-Mg^{2+} complex. Depletion of Mg^{2+} from the solution by the high affinity Mg^{2+}-chelator EDTA does not inhibit the incorporation of NMPs into TECs that have already sequestered the NTP-Mg^{2+} into their active sites, whereas TECs that have not sequestered NTP-Mg^{2+} by the time of the EDTA addition are depleted of the substrate and do not react. As the sequestration of NTP-Mg^{2+} requires base pairing with the template DNA nucleotide in the A-site (Vassylyev *et al*, 2007b) the EDTA quench can be used to probe the conformation of the active site and to make inferences about the translocation step. Importantly, these measurements monitor the translocation process of the TEC17 complex *before* the next nucleotide addition step. We refer to these types of experiments as “pre-catalytic” translocation measurements (**Figure 18**).

The TEC17 used in the kinetic experiments was formed by the pre-extension of TEC16 by the next NMP (GMP in the GC-system and CMP in the CG-system). TEC17 was then mixed with 500 µM of the next incoming NTP (CTP in the GC-system and GTP in the CG-system) in a quench flow device and after variable reaction times, EDTA was added. The RNA reaction products were resolved by PAGE and detected with the Atto680 label attached to their 5´ end. The nucleotide sequestration curves generally displayed burst kinetics because the binding of an NTP (at 500 µM) to the post-translocated TEC was faster than what could be resolved by the earliest 4 ms sampling time point (**Figure 19C & G**). The nucleotide sequestration experiments revealed markedly different kinetics of the GC- and GC-systems (compare **Figure 19C** with **19G,** see also **Study VII, Figure 2**). Firstly, the unresolved burst phase was larger in the GC-system compared to the CG-system. Secondly, the first resolved phase had a four-to-five-fold higher rate in the GC-system compared to the CG-system. On the other hand, the rate and the amplitude of the second resolved phase did not markedly differ between the two systems. The amplitude of this

phase was largely invariant in our analyses suggesting, and was treated as, a separate population not in equilibrium with the rest of the TEC (see Study VII for details). We performed the nucleotide sequestration measurements at two NTP concentrations (25 µM and 500 µM, only the 500 µM curves are shown in Figure 19), which allowed the estimation of the nucleotide binding rate. Interestingly, the apparent nucleotide binding rates were different in the two systems: 1.6-2.5 $\mu M^{-1} s^{-1}$ for CTP binding in the GC-system and a surprisingly low 0.3-0.9 $\mu M^{-1} s^{-1}$ for GTP binding in the CG-system (**Figure 19**). These values indicated that the NTP binding rates did not limit the first resolved phase, which thus represented a step preceding NTP binding, presumably translocation. Notably, the pre-melting of the upstream base pair of the RNA:DNA hybrid increased the rates of nucleotide sequestration (**Figure 19C & G**), presumably by increasing the rate of forward translocation.

The EDTA-quench offered also a way to directly measure the effect of the 6-MI labels on the kinetic behavior of the TECs (**Study VII, Figure 2**). Crucially, a comparison of the nucleotide sequestration profiles of the TECs with and without 6-MI labels showed that while the labels affect the kinetic properties of TEC17, the labeled and unlabeled TECs with the same transcribed sequence manifest qualitatively similar behavior. Specifically, the upstream 6-MI promoted nucleotide sequestration whereas the downstream 6-MI inhibited it, presumably through the effects to the translocation step (described in more detail in Study VII). This indicates that the otherwise identical TECs bearing 6-MI either in their downstream or upstream DNA cannot be directly compared. Serendipitously, the kinetics of the unlabeled TECs are in between the TECs bearing the upstream and the downstream 6-MI labels suggesting that the translocation measurements with both the upstream and downstream systems may provide good approximate boundaries of the translocation in the native TECs. While this observation complicates some of the conclusions, it underscores the sensitivity of the TECs to small structural changes in the DNA bases at the upstream and downstream fork junctions, which may enable TECs to sense damaged DNA bases and regulatory signals encoded in the DNA sequence.

5.4.3 Comparison of the post-catalytic and pre-catalytic translocation kinetics suggests a multi-step translocation mechanism

We next compared the results from the post-catalytic and pre-catalytic translocation measurements. Note that we performed these experiments also for TECs bearing 6-MI probes in the upstream end of the RNA:DNA hybrid (see results in Study VII), but discuss here only the results with the systems bearing the downstream 6-MI labels. In the GC-TECs, the kinetics of the nucleotide sequestration were consistent with the post-catalytic translocation measurements with the rate estimates for the translocation step: k_{forw} 30-40 s^{-1} and k_{back} 13-16 s^{-1}. (**Figure 19A-C**). Therefore, these data were consistent with a one-step translocation model (**Figure 19D**).

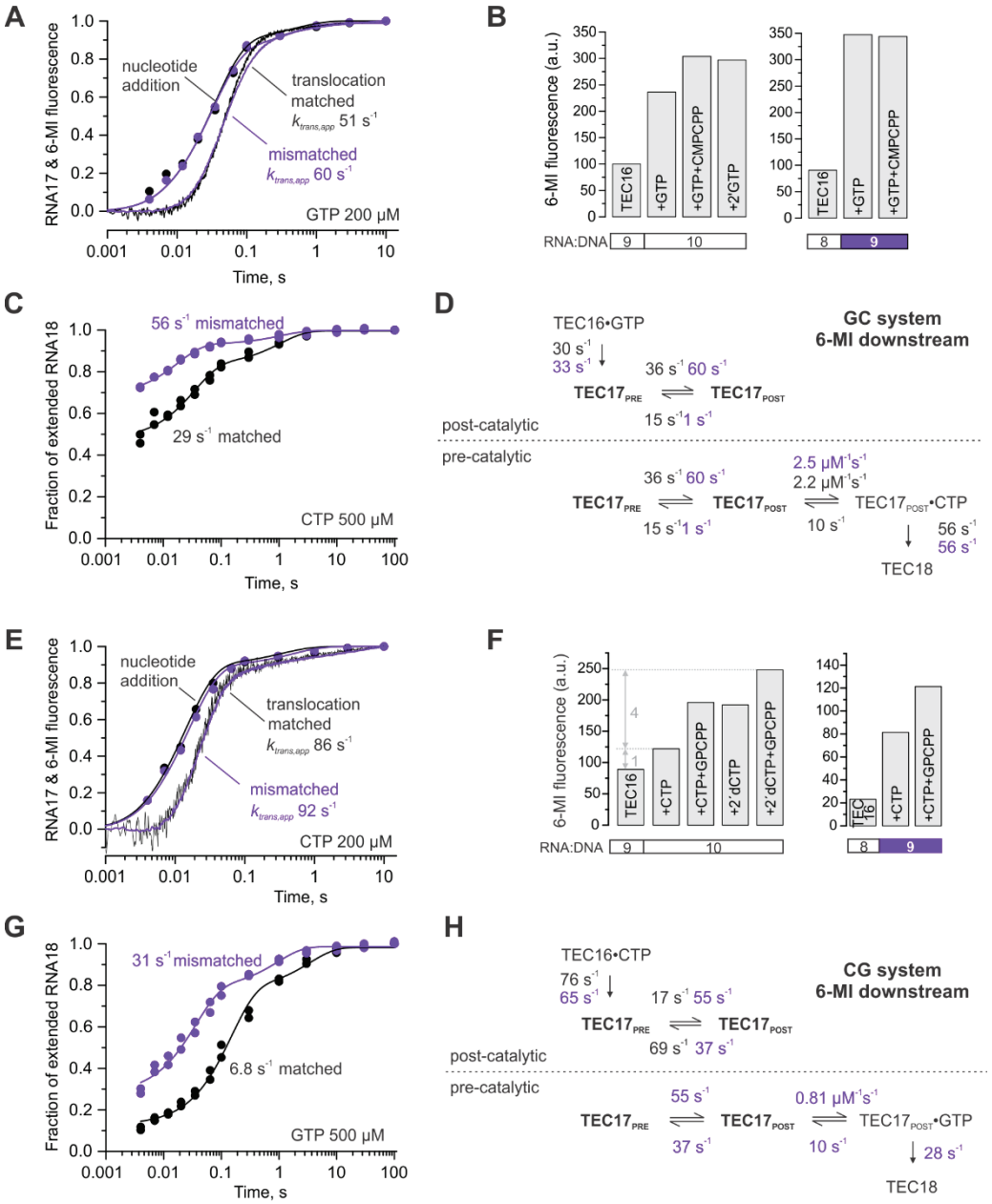


Figure 19. Comparison of post-catalytic and pre-catalytic translocation measurements. The upper four panels (**A-D**) show the data for the GC-system, the lower four panels (**E-H**) show the data for the CG-system. (**A & E)** Measurements of the forward translocation (continuous time traces) triggered by nucleotide addition (discrete time points) in the matched (black) and upstream mismatched TECs (purple). The apparent translocation rates were derived from the delay between nucleotide addition and fluorescence curves. **(B & F)** Assessment of the completeness of translocation by measuring the equilibrium levels of the 6-MI fluorescence. Representative experiments are shown. The left and right graphs present the "matched" and the "upstream mismatched" (RNA base at *i-8* substituted) TEC data, respectively. **(C & G)** NTP sequestration (pre-catalytic translocation measurement) in the matched (black) and upstream mismatched TEC17s (purple). The reactions were initiated by mixing GC-TEC17 with 500 µM CTP (C) or CG-TEC17 with 500 µM GTP (G), quenched after variable time points with EDTA following the quantitation of the extended RNA (≥18 nts).

The pre-melting of the base pair at the upstream edge of the RNA:DNA hybrid (the upstream mismatch) moderately increased the apparent translocation rate to k_{app} = 60 s^{-1} (**Figure 19A**, purple curve). At the same time, the CTP sequestration rate nearly doubled to 56±5 s^{-1} (**Figure 19C)**, indicating clearly that the pre-melting of the upstream base pair increased the forward translocation rate. While this result is in line with the equilibrium fluorescence results (**Figure 19B**), in closer inspection the magnitude of the increase in forward translocation rate is not sufficient to explain the near 100% occupancy of the post-translocation state in this TEC and simultaneously account for the apparent translocation rate of k_{app} = 60 s^{-1}. Therefore, the pre-melting of the upstream base pair must also decrease the backward translocation rate. This in turn implies that the formation of the upstream base pair modulates the rate of backward translocation. This interpretation is not consistent with a one-step translocation model because such model assumes that backward translocation rate is a property of the post-translocated state where the tenth upstream base pair of the RNA:DNA hybrid cannot form. This analysis therefore implicitly suggest the presence of an additional translocation step that manifests itself as an effect of the upstream mismatch on the translocation rates but not in the individual properties of the matched and mismatched GC-TECs.

On the other hand, in the TEC transcribing the CG-sequence, inconsistencies arise between the post-catalytic and pre-catalytic translocation measurements making it impossible to fit all observations into one-step translocation models. In particular, the GTP sequestration experiments give much slower rate estimates for translocation than do the 6-MI probes: the GTP sequestration data required a slow translocation step (6-7 s^{-1}), whereas the fluorescence data suggested an apparent translocation rate of 86 s^{-1} that was split into k_{forw} = 17 s^{-1} and k_{back} = 69 s^{-1} based on the equilibrium levels of the fluorescence (**Figure 19E-F**). The upstream mismatch accelerated the rate of GTP sequestration (**Figure 19G**, purple curve), which was consistent with the fluorescent observations. This suggests that breaking of the base pair at the upstream edge of the RNA:DNA hybrid modulates the rate of CG-TEC17 translocation. In contrast to the matched CG-TEC, the mismatched CG-TEC data could be adequately described using a one-step translocation model, the translocation rates inferred from the kinetics of fluorescence change and the equilibrium levels of fluorescence (**Figure 19H**). The observed inconsistencies between the post-catalytic and pre-catalytic measurements may imply that translocation in the matched CG-TEC17 is limited by an additional transition that is not detected by the 6-MI labels. This further suggests a multi-step translocation mechanism wherein the intermediate state(s) manifest themselves as kinetically distinct steps (see section 5.4.5 below).

Continued figure 19 legend from the previous page. The fraction of extended RNA was fitted into a two-exponential function preceded by a burst phase (non-zero initial level); the rates of the first resolved phase are indicated. The kinetic data are averaged from at least two independent experiments. **(D & H)** The global analysis of the post-catalytic (A&E) and pre-catalytic translocation experiments (C&G) using the minimal kinetic schemes. The values for the mismatched TECs are in purple. The rates of forward and backward translocation were inferred from the apparent translocation rates and the fluorescence levels in B and F. The rates for CTP binding and release were derived by the combined analysis of the 500 µM and 25 µM (not shown) NTP sequestration experiments. The slowly reacting TEC populations are omitted from the schemes for clarity. The complete schemes and parameter confidence intervals are presented in Study VII.

5.4.4 Probing the structure of the intermediate translocation state with photo-crosslinking

The structural transitions in the asynchronous translocation models have centered around the downstream fork junction that must undergo substantial conformational changes during translocation (Brueckner & Cramer, 2008; Silva *et al*, 2014). In order to probe the structure of the downstream DNA and the downstream fork junction we adapted our 8-MP mediated site-specific TA-crosslinking method (described in Figure 10). To this end, transcription templates were designed to harbor a single TA-site in the downstream DNA at different positions allowing the conformational probing of the DNA at multiple sites upon walking the TEC by the addition of subsets of substrate NTPs. In contrast to our expectations and results from crosslinking the upstream DNA, inefficient TA-site crosslinking was observed throughout the downstream DNA between -1 to 13 nucleotides from the RNA 3´ end (**Study VII, Figure S5**). The low crosslinking efficiency was presumably caused by the extensive interactions between the DNA and the RNAP that constrained the conformation of the downstream DNA (Kettenberger *et al*, 2004) (see Figure 7) thus reducing the overall flexibility of the DNA, which is required for the 8-MP photo-binding (Spielmann *et al*, 1995).

Surprisingly, the crosslinking efficiency markedly increased when the last incorporated nucleotide was CMP, the next incoming nucleotide was GTP and the TA-site was positioned two nucleotides downstream of the RNA 3´ end in TEC17 (**Figure 20A**, right). This configuration matched the CG transcribing TEC17 described above. In contrast, crosslinking in the corresponding GC-TEC17 did not differ from the background level (**Figure 20A**, left). Importantly, the CG-TEC17 crosslink did not originate from a pre-translocated complex because *i*) the addition of TGT abolished the crosslinks, *ii*) the β´F773V RNAP variant, which is biased towards the pre-translocated state, produced very little crosslinking (**Figure 20B**). On the other hand, the crosslink did not originate from the post-translocated register either, because the addition of the forward-biasing GMPcPP lowered the crosslinking efficiency. Notably, the crosslinking efficiency almost doubled when the interaction between RNAP and the template DNA was weakened by the β´K334M mutation in the switch 2 region. β´Lys^{334} is absolutely conserved and contacts the phosphate of the kinked template strand between the *i+1* and *i+2* sites (Lane & Darst, 2010b) (see Figure 7 on page 27). These results suggest that the TA-crosslinks are formed in a particular TEC conformation enriched in the CG-TEC17 that involves an altered template DNA conformation, likely reflecting the mobility of the template DNA at the downstream fork junction. Interestingly, pre-melting the base-pairs at the upstream edges of the RNA:DNA hybrid (**Figure 20B**) or the downstream DNA (**Study VII, Figure S6**) increased the crosslinking efficiency. Furthermore, probing the RNA:DNA interaction with 6-TG photocrosslinking indicated that the CG-sequence in the active site inhibited translocation or the separation of the RNA:DNA hybrid (or both) (**Study VII, Figure 5**). As discussed in the next section, these observations give good framework for determining the structures of the kinetically identified intermediate translocation states.

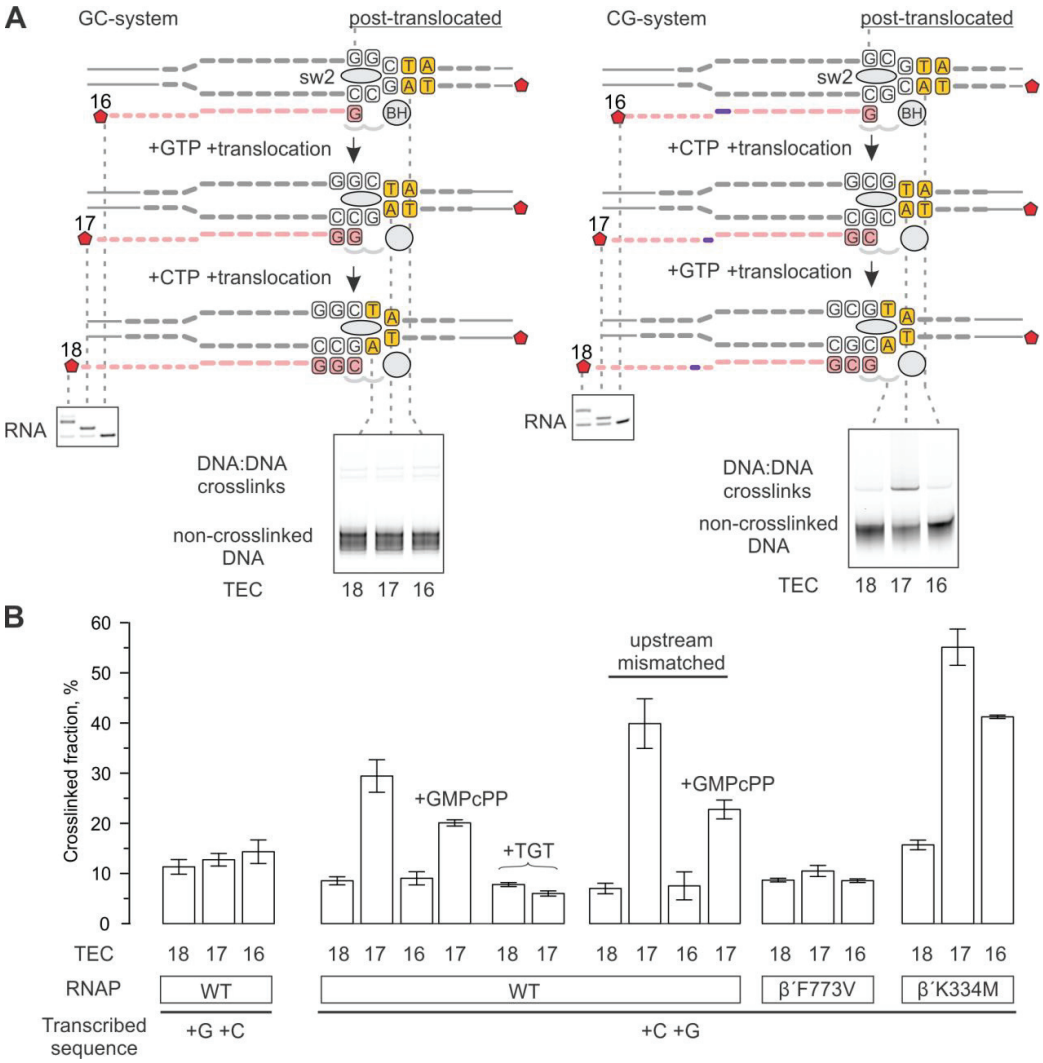


Figure 20. Probing the downstream DNA by photocrosslinking with 8-methoxypsoralen. A) The schematics of the experiments. Photocrosslinking of the downstream DNA was monitored during the transcription of GC (left) and CG (right) sequences. The RNA and template DNA strands were labeled with the Atto-680 infrared fluorophore (red pentagon) for detection in gel after PAGE. The RNA and DNA gel panels were spliced from the same gel. The pixel counts were linearly scaled to span the full 8 bit grayscale range within each gel panel. **B)** DNA:DNA crosslinking in TECs formed by the wild-type and altered RNAPs. TECs containing the unique 8-MP intercalation site were walked by up to two nucleotides, supplemented with 8-MP, and 5 μM TGT (where indicated) or 500 μM GMPcPP (where indicated) and illuminated with UV light. An upstream mismatch shortens the possible RNA:DNA hybrid to 8 bp in TEC16 and to 9 bp in TEC17. The altered RNA nucleotide is colored purple in the CG-TECs in A. Error bars indicate the range of duplicate measurements or the SDs of several measurements. β´Phe^{773} is in the bridge helix (BH in TEC schematics), β´Lys^{334} is in the switch 2 (sw2 in TEC schematics).

5.4.5 Summary: A two-step translocation model

The detailed investigation of nucleotide addition and translocation kinetics as well as structural probing of the upstream and downstream edges of the transcription bubble during the transcription of a CG sequence rule out a simple single-step translocation mechanism and necessitate the development of a more complex multistep mechanistic model. Even kinetically simpler transcription through a GC sequence implicitly suggests the presence of mechanistically but not stoichiometrically significant intermediate species to account for the dependence of the backward translocation rate on the base pair at the upstream edge of the RNA:DNA hybrid. The multistep translocation process may thus be characteristic for all sequence positions, but manifests itself kinetically fully only during the transcription of certain sequences.

The structures of the RNAP elongation complex that cannot be classified as post-translocated or pre-translocated states suggest that while the RNA always translocates by full register steps, the DNA can take various intermediate conformations (discussed in section 2.3.1). As we demonstrated in section 5.2, following the incorporation of a nucleotide into the RNA 3´ end, the opening of the active site unlatches the RNA allowing it to traverse from the substrate binding site (a.k.a. the A-site) to the thermodynamically more favorable nucleophilic site (a.k.a. the P-site). Our biochemical data is thus consistent with the concept that the RNA translocation initiates the rearrangement of all nucleic acid components of the TEC into the post-translocated configuration. Hence, only full-register RNA motions can be kinetically defined.

Next, based on the structural observations it is likely that the movement of the DNA can be partially or fully independent of the movement of the RNA. Two scenarios can be envisaged, and as we describe below, both scenarios possibly take place in our system during transcription of the CG sequence, but with drastically different outcomes. In the first scenario, which is supported by the RNA polymerase II in complex with the inhibitor α-amanitin (Brueckner & Cramer, 2008), the RNA and DNA translocate partially independently. While the RNA translocates a full register, the template DNA translocates only a half-register, which establishes a tilt in the base pairs of the RNA:DNA hybrid. The templating nucleotide occupies a position in between the post-translocated and pre-translocated registers, where it is unable to direct the binding of the incoming substrate NTP and also has to be unpaired from the complementary nucleotide in the non-template DNA. On the other hand, the partial DNA translocation will likely break or at least destabilize the upstream edge of the RNA:DNA hybrid. In the second step of this model, the half-register translocation of the DNA completes the process. In an alternative scenario, inspired by the structure of the RNA polymerase II elongation complex with an RNA backtracked by many nucleotides (Cheung & Cramer, 2011), the RNA and DNA translocate fully asynchronously. Here, the full register translocation of the RNA happens without the movement of the template DNA. As a consequence, the tilt of the RNA:DNA hybrid becomes larger, which should, however, allow maintaining the base pairing at the upstream edges of the RNA:DNA hybrid and the downstream DNA. Consequently, the translocation is completed by the full register movement of the DNA. Importantly, the differences in the effects of the alternative translocational motions on the upstream and the downstream base pairs will be critical in

devising the structural interpretations for the kinetic data because *i*) melting of these base pairs play an important role in lighting up the fluorescent beacons in our fluorescent assays and *ii*) pre-melting of the upstream base pair completely eliminates the slow translocation step inferred from the GTP sequestration experiments in the CG system.

The major observations in favor of the multistep translocation mechanism are as follows. First, with the TEC transcribing the CG sequence we observe disparity in the translocation rates inferred from GTP sequestration and the fluorescence experiments, the former being two to four-fold slower. At the same time, pre-melting of the upstream base pair increases the rate of GTP sequestration to match the rate observed in the fluorescent beacon experiments. Second, probing the conformation of the downstream DNA by photo-crosslinking with 8-MP suggests the existence of an intermediate state that is neither pre- nor post- translocated. Contrary to the situation with the translocation rates disparity, pre-melting of the upstream base pair markedly increases the occupancy of the intermediate state. Third, we observe an uncharacteristically low apparent GTP binding rate. Pre-melting of the upstream base pair increases the GTP binding rate, but the improved rate remains at least twofold lower than that in the other GTP binding TECs.

These observations are accounted for by a translocation model presented in **Figure 21**. The low GTP binding rate can be explained by invoking a rapid equilibrium between a post-translocated state and an incompletely translocated intermediate state that is unable to bind the substrate (**Figure 21**, rightmost equilibrium). Hence, a bias towards the intermediate state tends to lower the apparent GTP binding rate, which equals the product of the true GTP binding rate and the fractional occupancy of the post-translocated state. From a structural standpoint, the α-amanitin stabilized state (see Figure 4 state 7 (Brueckner & Cramer, 2008)) with the fully translocated RNA and half-register translocated DNA accounts best for the properties of such an intermediate state (**Figure 21**, on-pathway intermediate). The template strand acceptor base is positioned outside the active site midway between the pre- and post-translocated registers and is unavailable for binding the incoming substrate NTP. This arrangement of nucleic acids also explains the 8-MP crosslinking results: the downstream pair is melted, thereby allowing for the stretching of the downstream DNA, efficient intercalation of 8-MP and efficient photo-crosslinking of the downstream DNA. Furthermore, the upstream base pair of the RNA:DNA hybrid is expected to be at least fractionally melted because its pre-melting significantly increases the photo-crosslinking of the downstream DNA with 8-MP, and, according the above view, the occupancy of the intermediate state. Accordingly, the intermediate state likely has a similarly bright fluorescence to that of the post-translocated state in both the upstream and downstream 6-MI probe positions. Finally, the intermediate state should not manifest itself in the kinetics of the fluorescence change as its interconversion with the post-translocated state (half-register translocation of the DNA) is postulated to be significantly faster than the preceding translocational motion (full register translocation of RNA and half-register translocation of the DNA) controlled by the opening of the active site.

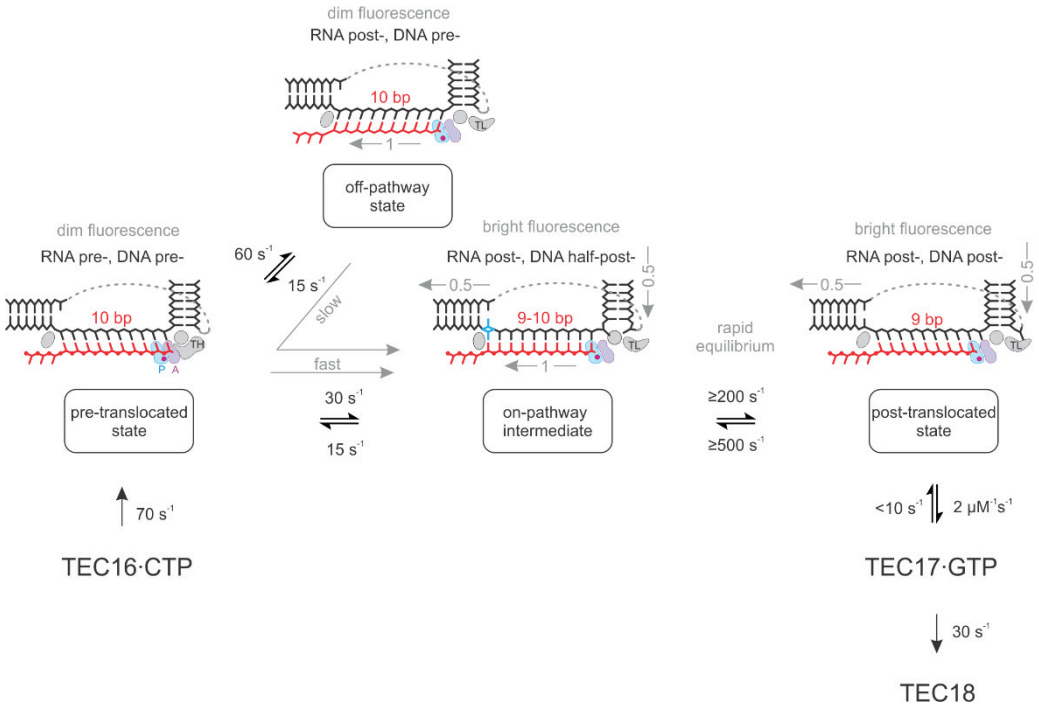


Figure 21. A two-step translocation model. The rate values approximate the kinetics of the CG-system. The forward and backward rates of the rapid equilibrium are not defined by the data but their ratio can be estimated from the difference between the GTP binding rates (k_{on}) observed in the pre-catalytic and post-catalytic experiments. TH, helical trigger loop; TL, unfolded trigger loop; P, P-site; A, A-site. See text for details.

Next, the presence of the slow step that is not detected in the fluorescent translocation measurements calls for a mechanistic explanation. If the slow step were on the on-pathway it should either precede or follow the rapid phase detected by the fluorescence. The slow step does not follow the fast step because *i*) this would be detected as a slow increase in the fluorescence, which is not the case, and *ii*) the fast step is detected as a consequence of the melting of the upstream base pair, yet the pre-melting of the upstream base pair eliminates the slow translocation step. On the other hand, the slow step also cannot obligatorily precede the fast step because then the fast step would not be detected as fast. These arguments suggest that the slow step is an off-pathway state. Therefore, we consider a scenario where the pre-translocation state partitions between an on-pathway state that translocates rapidly and an off-pathway state that does not translocate forward (**Figure 21**, equilibrium between the pre-translocated and the off-pathway state). We also assume that the both the pre-translocated state and the off-pathway state have the similarly dim fluorescence of the upstream and the downstream 6MI fluorophores. Such a scenario can explain the coexistence of the fast and the slow translocation steps: the model attributes the fast step to the on-pathway pre-translocated state and the slow step to the off-pathway state that needs additional time to interconvert back into the pre-translocated state before it can translocate forward (**Figure 21**, grey arrows marked slow step). However, the step that slowly interconverts one state with dim fluorescence into another state with dim fluorescence should also slowly shift the linked equilibrium between the

dim and the bright states and manifest itself as a slow phase in the fluorescence experiments. Our kinetic simulation experiments demonstrate that the above is generally the case but does not hold when the backward rates describing the return of the TEC into the pre-translocated state from the bright state and the dim state (off-pathway) are approximately the same. Under the latter circumstances, the slow interconversion step does not manifest itself in the simulated time traces depicting the accumulation of the brightly fluorescent states. Accordingly, the partitioning of the pre-translocated state into on- and off-pathway states can explain the disparity of translocation rates determined in the fluorescence and the GTP sequestration experiments if the backward rates leading to a pre-translocated state are approximately equal.

We further suggest that the off-pathway transition represents a scenario in which only the RNA translocates (discussed above). First, this would leave the upstream and the downstream base pairs intact resulting in a state with dim fluorescence thereby matching one of the requirements for the off-pathway state. Second, our data are consistent with the forward rate of approximately 60-80 s^{-1} for the transition of the pre-translocated state into the off-pathway state, which is a frequently encountered value for the rate of TL controlled forward translocation in our analyses (Figures 10 and 13). Third, the requirement for the equality of the backward rates leading to the pre-translocation state in a model gains a plausible structural explanation. Assuming the RNA-only scenario above, both isomerizations leading to the pre-translocated state involve backward translocation of the RNA (**Figure 21**, two leftmost equilibriums). As demonstrated in section 5.2.3, the backward translocation is limited by the interaction of the RNA 3' end with the Mg^{2+} (A) in the P-site. Therefore, it is conceivable that the backward translocation of the RNA and the backward translocation of RNA accompanied by a half-register translocation of the DNA occur with similar rates.

In summary, our observations provide evidence for two unorthodox translocation states in addition to the canonical pre- and post-translocation states during the transcription of the CG sequence. The on-pathway intermediate in rapid equilibrium with the post-translocated state is required to account for the uncharacteristically low GTP binding rate and the results of the photocrosslinking experiments. The off-pathway translocational state in equilibrium with the pre-translocation state is required to account for the disparity between the translocation rates inferred from the fluorescence and GTP sequestration experiments. The evaluation of the available TEC structures, kinetic and structural properties of the above mentioned unorthodox translocational states suggest that the on-pathway intermediate has a fully translocated RNA and a half-register translocated DNA, whereas in the off-pathway state the RNA is fully translocated and the DNA is not translocated at all. Accordingly, we argue that the same phenomenon, the asynchronous translocation of the RNA and the DNA is behind all slowdowns during the transcription of the CG sequence. The partially asynchronous translocation directs TEC on-pathway, but leads to a low GTP binding rate after the incorporation of CMP. On the other hand, the fully asynchronous translocation directs the TEC off-pathway and manifest itself in a fraction of the TEC undergoing an apparently slow translocation. Our data suggest that the partially asynchronous translocation only requires a favorable dinucleotide sequence in the

active site and is only lightly modulated by the separation of the RNA:DNA at the upstream edge of the RNA:DNA hybrid. In contrast, the fully asynchronous translocation requires a favorable dinucleotide sequence in the active site and strong base-pairing at the upstream edge of the RNA:DNA hybrid.

We must note that at present it is not clear why the fully asynchronous translocation of the RNA prevents the translocation of the DNA. It likewise remains unclear how the dinucleotide sequence in the active affects the equilibrium between the on-pathway intermediate and post-translocation states (**Figure 21**, rightmost equilibrium), but we can envisage several scenarios that are not mutually exclusive. First, the dinucleotide sequence may affect the inherent structural preference of the RNA:DNA hybrid, downstream DNA and the kink in between them to occupy the intermediate or post-translocated conformations. It has been suggested that the sugar backbone of the dC template base would be flexible and distort the geometry of the active site when combined with the neighboring rC:dG pair (Imashimizu *et al*, 2015; Kireeva & Kashlev, 2009), and could thereby disfavor the post-translocated state. Second, the translocating template strand nucleotide may interact base-specifically with RNAP (Wang *et al*, 2015), which could stabilize the intermediate state. Third, a guanine in the downstream non-template DNA can interact base-specifically with the RNAP "core recognition element" (CRE) which may modulate translocation (Vvedenskaya *et al*, 2014; Petushkov *et al*, 2015). It is likely that multiple sequence encoded interactions work together to affect this equilibrium. These mechanisms can be tested by varying the transcribed sequence, employing abasic DNA oligonucleotides and engineering mutations into the β fork loop, β Dloop II and CRE.

Finally, we should consider whether the asynchronous translocation is specific to certain sequences or whether it is an obligatory part of the translocation mechanism at all sequence positions. The partially asynchronous translocation may possibly represent a universal obligatory step in translocation, but may be undetectable by biochemical and kinetic approaches at some if not most sequence positions. Our model postulates that the last translocation step in the CG system, the half register translocation of the DNA, is considerably faster than the preceding translocation step controlled by the opening of the active site. Such a step is kinetically silent when the equilibrium characterizing it is biased forward, which may be the case at most sequence positions.

In contrast, the fully asynchronous translocation is unlikely an obligatory step in translocation and instead represents an off-pathway transition that may be important for the modulation of the transcription rate by the transcribed sequence and represent the initial state in many types of transcriptional pauses. Indeed, considering the consensus pause sequence $G_{-10}Y_{-1}G_{+1}$ identified by NET-Seq (Larson *et al*, 2014; Vvedenskaya *et al*, 2014; Imashimizu *et al*, 2015) in light with the structure of the proposed off-pathway state suggests that the nucleotides directing pausing are located at sites that modulated the formation of the off-pathway state. In this conformation the G_{-10} would be base-paired and located at the end of the RNA:DNA hybrid, the Y_{-1} (CMP or UMP) would correspond to the full-register translocated RNA 3´ end at the P-

site and the G_{+1} would correspond to the untranslocated dG:dC base pair. This supports the notion that the modulation of base-pairing at the upstream edges of the RNA:DNA hybrid and the downstream DNA affect the partitioning of the TEC between the on-pathway and off-pathway states.

5.5 The mechanism of RNAP regulation by the universal NusG elongation factor (Study **V**)

5.5.1 NusG does not affect the rate of nucleotide addition or forward translocation

NusG/Spt5 transcription elongation factors are universally conserved and accelerate the overall transcription elongation rate *in vitro* and *in vivo* and couple the activity of RNAP to downstream processes, including translation by the ribosome and to termination mediated by the Rho-factor (Burova *et al*, 1995; Burns *et al*, 1998; Bar-Nahum *et al*, 2005; Herbert *et al*, 2010; Werner, 2012). NusG has been proposed to act by stabilizing the clamp domain in the closed state or by interacting with the upstream DNA (Klein *et al*, 2011; Sevostyanova *et al*, 2011; Hirtreiter *et al*, 2010; Martinez-Rucobo *et al*, 2011). It was suggested that NusG biases RNAP translocation forward by selectively stabilizing the post-translocated state (Bar-Nahum *et al*, 2005; Herbert *et al*, 2010). We tested this prediction by measuring the kinetics of the nucleotide addition cycle in the presence of NusG. In contrast to expectation, in these measurements NusG did not measurably affect the rates of nucleotide addition and translocation of the RNA:DNA hybrid in the forward or backward direction nor the translocation equilibrium (**Study V, Figure 2**).

5.5.2 NusG inhibits backtracking measured through GreA-mediated RNA cleavage

An earlier single molecule study revealed that NusG selectively reduced long duration pauses, which were attributed to backtracking events (Herbert *et al*, 2010). In order to study the molecular basis for NusG´s anti-backtracking effect we designed experimental systems for monitoring backtracking. In most sequences, backtracking is reversible and slow compared to forward translocation. Thereby only a minor fraction of TECs is backtracked at a given moment and these do not accumulate, making it difficult to measure the rate of backtracking in an ensemble setting. To overcome this problem, TECs were mixed with a high concentration of the GreA-factor, which accelerates the endonucleolytic RNA cleavage in the backtracked TEC. In effect, the TECs that backtracked were cleaved and irreversibly “trapped” in the cleaved state giving rise to a robust backtracking signal (**Figure 22A**).

In the experiments we employed TEC18s bearing CMP at the RNA 3´ end and 2AP at the penultimate position. In a subset of experiments the 6-MI probe was embedded in the template DNA upstream of the RNA:DNA hybrid. GreA-driven TEC18 backward translocation was monitored as a decrease in the 6-MI signal (6-MI moved 2 registers backwards); the RNA cleavage was followed with PAGE as the shortening of the initial RNA18 and appearance of the 5´ cleavage product RNA16; the release of the dinucleotide p2APpC was followed as an increase in the 2AP fluorescence. All three methods acquired similar kinetics (**Figure 22B**). Importantly, the inactive GreA D41N variant did not bias translocation backwards (**Manuscript V, Figure 3 supplement 1**), indicating that GreA cleaved only the backtracked TECs without affecting their translocation equilibrium. A TEC that was backtracked-biased by a RNA 3´ end mismatch (DNA

sequence was altered) was cleaved 30-fold faster than a matched TEC (**Figure 22C**). This demonstrated that in this experimental system the rate of RNA cleavage in the matched TECs was limited by backtracking.

Importantly, supplementing TEC18 with NusG inhibited ~2.5-fold the rates of GreA mediated backward translocation, RNA cleavage, and dinucleotide release. By inference, NusG inhibited RNAP backtracking, consistent with the observations by (Herbert *et al*, 2010) and the initial proposal by (Pasman & von Hippel, 2000).

5.5.3 NusG inhibits RNAP backtracking by stabilizing the upstream DNA base pairing

The reproducible anti-backtracking effect formed the basis for assaying the structural determinants for NusG´s activity. First, as an objective to abolish the anti-backtracking activity of NusG we modified the structure of RNAP. NusG/Spt5 proteins have been proposed to regulate RNAP activity through the clamp domain (Klein *et al*, 2011; Sevostyanova *et al*, 2011; Hirtreiter *et al*, 2010; Martinez-Rucobo *et al*, 2011). We hypothesized that the loops in the RNAP main cleft connected to the clamp domain – lid (LL) and rudder loops (RL) – could mediate these allosteric effects. Furthermore, the paralogs RfaH and NusG were proposed to interact through a bridging mechanism with the gate loop (GL) in the β-subunit on the opposite side of the main cleft and this interaction was proposed to stabilize the clamp domain in the processive closed state (Sevostyanova *et al*, 2011). The individual deletions of these loops demonstrated that NusG´s anti-backtracking effect was partially dependent on the LL, whereas RL and GL were dispensable (**Figure 22C**).

Next, we disrupted the DNA base pairs expected to melt during entry into the backtracked state (**Figure 22C**, right hand side; blue nucleotides). Strikingly, disruption of the first two upstream DNA base pairs accelerated backtracking five-fold demonstrating that melting of the upstream DNA limits the rate of backtracking (**Figure 22C**, left hand side). Furthermore, the mismatches conferred the TEC insensitive to NusG. This suggested that the anti-backtracking activity was dependent on the interaction of NusG with the upstream DNA and appeared less dependent of the allosteric control through the clamp domain.

To directly probe the interaction between NusG and the upstream DNA we used the RNA:DNA and DNA:DNA crosslinking methodologies together with the 6-MI beacon fluorescence system. NusG affected the upstream DNA conformation as demonstrated by TA-crosslinking and 6-MI fluorescence (**Study V, Figures 4 & 6**), but did not change the length of the RNA:DNA hybrid as was measured by 6-TG crosslinking (**Study V, Figure 5**). More specifically, NusG restored the DNA annealing defects in the ΔLL and ΔRL mutant RNAPs confirming that it stabilized the upstream double stranded DNA (**Study V, Figure 4**). These results suggest that NusG stabilizes the first upstream DNA base pair -10 nucleotides from the RNA 3´ end and effectively keeps the size of the transcription bubble minimal. Because backward translocation entails the melting of the -10 upstream DNA base pair and the strand exchange with RNA, the stabilization of -10 DNA base pair explains how NusG inhibits backtracking.

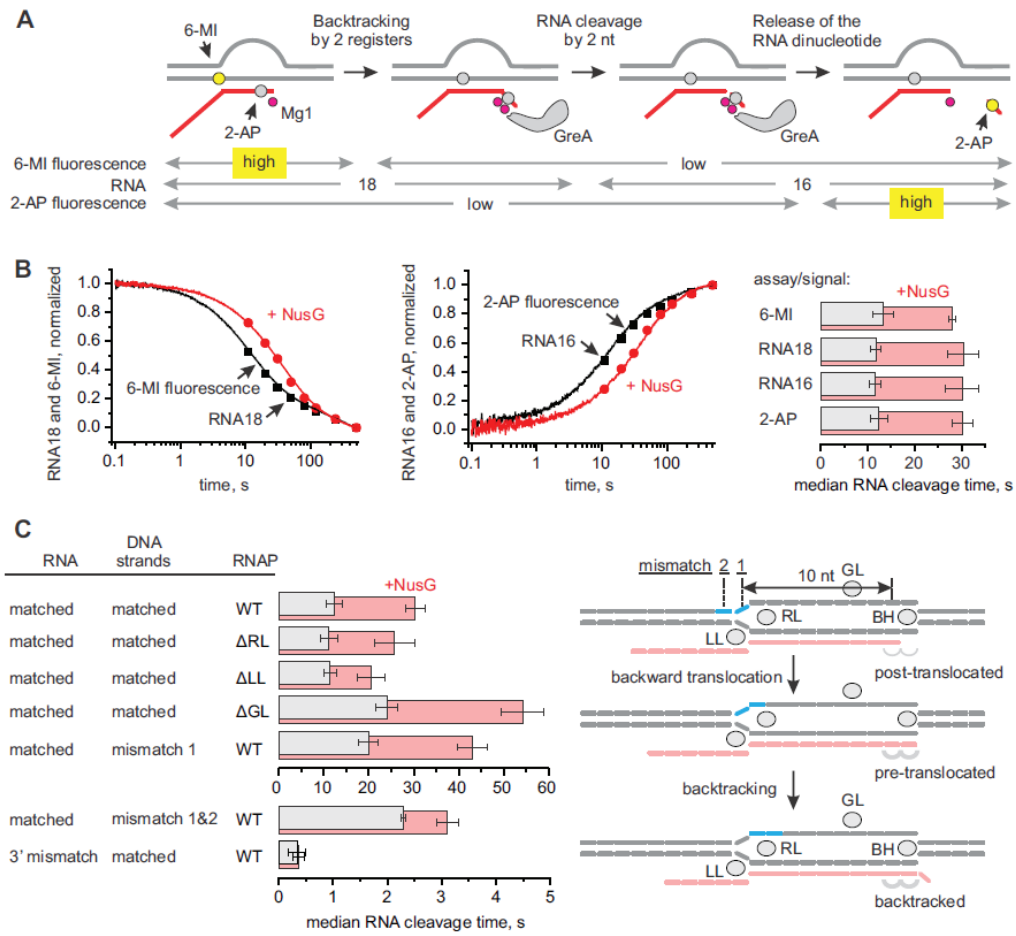


Figure 22. NusG inhibits GreA assisted RNA cleavage by slowing backtracking. A) Three assays for monitoring the GreA-assisted RNA cleavage: TEC backtracking (6-MI fluorescence decrease), RNA cleavage (PAGE), and the dinucleotide release (2-AP fluorescence increase). **B)** Left: the effect of NusG on the TEC backtracking (continuous time-traces) and RNA cleavage (discrete time-points) upon the addition of 8 μM GreA. Center: the effect of NusG on the release of the cleaved dinucleotide (continuous time-traces) and the RNA cleavage (discrete time-points). Right: the median reaction times. **C)** The effect of NusG on RNA cleavage by the TECs with deletions of the RNAP domains, the mismatched upstream DNA and a 3′ rC:dA RNA:DNA mismatch. The median reaction times were determined by monitoring the increase in 2AP fluorescence at 8 μM GreA. The schematic on the right illustrates the location of DNA:DNA mismatches (cyan bars). Error bars indicate the range of the best-fit estimates in duplicate experiments. The figure was adapted from Turtola & Belogurov, 2016.

The results from the 6-MI fluorescence, TA- and 6-TG crosslinking (see section 5.3) and backtracking assays were combined in a structural model of the upstream fork junction together with bound NusG (**Figure 23**). Specifically, a five amino acid loop of the NusG N-terminal domain (NTD) contacts the non-template DNA 11 and 10 nucleotides from the RNA 3´ end. This loop is structurally conserved (**Study V, Figure 8**) and forms one wall of an exit channel for the upstream DNA. The opposite wall of the exit channel is formed by the LL, whereas the RL stacks with the DNA base pair at position -10. Our model is fully consistent with a later cryo-electron microscopy

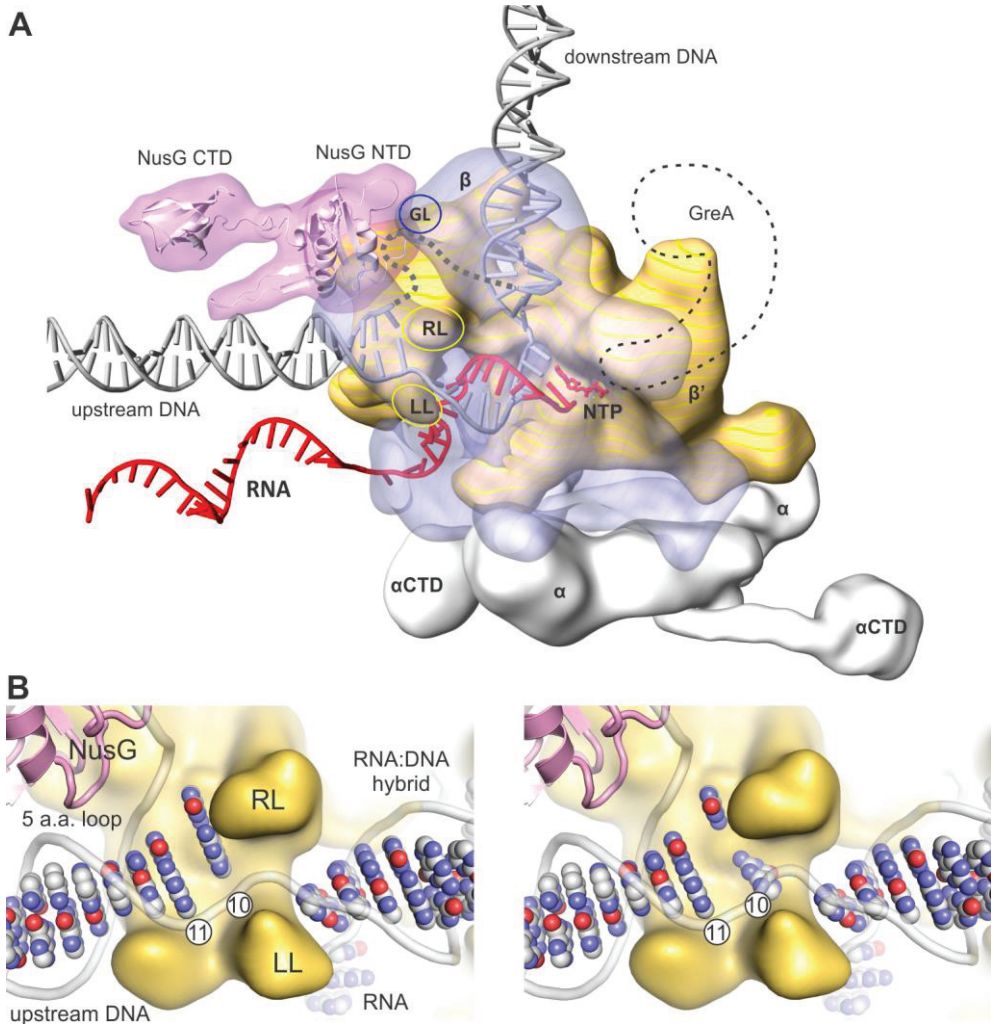


Figure 23. Structure of the TEC with modeled upstream fork junction and NusG. A) RNAP core subunits are depicted by simplified molecular surfaces. β is depicted as transparent to reveal the path of nucleic acids through the enzyme. The positions of NusG CTD and αCTD, connected via flexible linkers, were chosen arbitrarily. The locations of RNAP cleft loops individually deleted in this study, the β Gate Loop (GL), β' Rudder Loop (RL) and β' Lid Loop (LL) are accentuated by ovals. The hypothetical path of the single stranded non-template DNA is depicted by a grey dashed line. The approximate location of the GreA cleavage factor employed in backtracking experiments is depicted as a black dashed contour. The composite model was generated using the *T. thermophilus* TEC (Vassylyev *et al*, 2007b) (Vassylyev *et al*., 2007), NusG NTD from the NusG-RNAP model in (Martinez-Rucobo *et al*, 2011) and the elements from other structures (see Materials and methods in Study V). The duplex DNA immediately upstream of the RNA:DNA hybrid was modeled de novo as based on the results described in the text. **B)** A close view of the upstream fork junction. DNA bases are numbered from the RNA 3' end in the post-translocated TEC. The β subunit is omitted for clarity. β´LL and the structurally conserved five amino acid loop of NusG NTD form a channel accommodating the exiting upstream DNA. The template DNA nucleotide at position ten can be modeled to pair with the non-template DNA in a partially unstacked conformation (left) or to interact with the cleft between the β'RL and β'LL (right). According to the proposed model, NusG stabilizes the base-paired structure and thereby inhibits RNAP backtracking. The figure was adapted from Turtola & Belogurov 2016.

structure of the *E. coli* TEC (Kang *et al*, 2017) but not with the *E. coli* RNAP-NusG complex (Liu & Steitz, 2017) in which the crystal packing prevents NusG binding to the clamp helices in the orientation observed in (Martinez-Rucobo *et al*, 2011). Impressively, the eukaryotic *Komagataella phaffii* Spt4/5-TFIIS-TEC complex showed a similar upstream DNA conformation and the Spt5 NGN-domain interaction surface with the upstream DNA (Ehara *et al*, 2017). The deep evolutionary conservation of these interactions, as well as the anti-arrest activity in Spt4/5 (Crickard *et al*, 2016), suggest that NusG´s anti-backtracking activity is functionally important.

5.6 The structural basis of RNAP backtracking and RNA proofreading (Studies **IV, VI**)

5.6.1 Measuring the RNA 3´ end backtracking dynamics with 2-aminopurine

To study the mechanism of backtracking, we explored an experimental system where RNAP backtracking would produce a direct fluorescence signal. To this end, we extended the RNA 3' end of a TEC by 2-aminopurine (2-AP, TEC-2AP). This has two advantages: First, 2-AP forms a wobble base pair with the template DNA which facilitates backtracking. Second, 2-AP is an environment-sensitive fluorophore which displays quenched fluorescence as a part of a nucleic acid duplex and high fluorescence when unstacked from the neighboring bases (Hardman & Thompson, 2006; Rachofsky *et al*, 2001). These properties make 2-AP an ideal tool for monitoring the nascent RNA dynamics during RNAP backtracking.

Unlike in the previously utilized TEC17s, we designed 11 bp RNA:DNA hybrid in the TEC-2AP, which allowed the optimal 10 bp RNA:DNA hybrid formation in the backtracked state. We first characterized the nucleotide addition properties of the TEC-2AP. The addition of the next nucleotide (CMP) by TEC-2APs displayed biphasic kinetics with fast (18 s^{-1}) and slow phases (~0.4 s^{-1}) of approximately equal amplitudes (**Figure 24A-B)**. The CMP addition reaction became mono-phasic (~18 s^{-1}) in a modified TEC-2AP, which was stabilized to the post-translocated state by a 9 bp RNA:DNA hybrid (**Study VI, Figure 1**). The slow phase also nearly disappeared, when TEC-2AP was incubated without Mg^{2+} before the nucleotide addition reaction. Therefore, TEC-2AP was fractionally and reversibly inactivated, possibly due to backtracking, and the formation of the inactivated state required Mg^{2+}. Upon the addition of Mg^{2+}, TEC-2AP underwent a fractional (~50%) transition into a high fluorescence state (apparent K_D(Mg) ~0.7 mM) (**Study VI, Figures 1 and 6**). On the other hand, upon CMP incorporation the fluorescence decayed close to the Mg^{2+}-free level. Notably, the rate of fluorescence decay closely paralleled the rate of the slow phase of the CMP addition curves, showing that the TEC-2APs occupying the high fluorescent state corresponded to the reversibly inactivated complexes. The inactivated high fluorescent complex was backtracked by 1 nucleotide as shown by the effects of the DNA:DNA and RNA:DNA mismatches that disfavored the backward translocation of the TEC to the 1-nucleotide backtracked state (**Study VI, Figure 2**). Furthermore, the Mg^{2+}-dependent rise in the fluorescence intensity was accompanied by an increase in anisotropy suggesting that formation of the backtracked state was accompanied by the stabilization of 2-AP by a Mg^{2+} ion into a less mobile conformation (**Study VI, Figure 1**). Importantly, the AMP extended TEC (TEC-AMP) was also backtracked albeit less than TEC-2AP and displayed a similar rate of recovery from the

backtracked state. This argues that the results with TEC-2AP could be generalizable to natural TECs.

5.6.2 Helical TL stabilizes the backtracked state

We next used TEC-2AP fluorescence to screen RNAP variants in search for structural elements that control backtracking. Considering the central role of TL in controlling translocation, of particular interest were the perturbations that affected the TL folding equilibrium. Surprisingly, the Mg^{2+}-dependent increase in the 2AP fluorescence was largely abolished in a TEC-2AP assembled with the ΔTL RNAP (**Study VI, Figure 3**), indicating that the TL is a necessary component of the Mg-stabilized backtracked state. Furthermore, addition of STL, which constrained the TL in the unfolded state (Temiakov *et al*, 2005; Vassylyev *et al*, 2007b) reduced the TEC-2AP fluorescence near to the level of ΔTL TEC-2AP, consistent with the earlier observation that STL disfavors the backtracked state (Tuske *et al*, 2005). Furthermore, the amino acid substitutions that were predicted to destabilize the helical TL (β´H936A, β´H936Q, β´Q929A and β´M932L) reduced the TEC-2AP fluorescence, whereas those amino acid substitutions that stabilized the helical TL (β´P750L, β´F773V, and β´G1136S) increased the TEC-2AP fluorescence.

Importantly, the estimates on the occupancy of the backtracked state and its dependence on the closure of the active site by the helical TL obtained in fluorescence measurements were fully consistent with the estimates obtained in time-resolved RNA extension experiments: The TEC-2APs assembled with variant RNAPs and pre-incubated with 1 mM Mg^{2+} showed biphasic CMP incorporation kinetics (**Figure 24B**) with the fraction of a slow phase in good correlation with the amplitude of the Mg^{2+}-dependent increase in fluorescence (**Figure 24C**). Next, the rate of recovery from the backtracked state is the only parameter that directly relates to the stability of the backtracked state. As nucleotide addition is limited by the recovery of backtracking (see Figure 24A), the rate of recovery was determined as the rate of the slow phase or the fluorescence decay in the CMP addition experiments (**Figure 24D**). The RNAP variants that were more backtracked than RNAP had slower recovery rates, whereas those RNAPs (or WT RNAP with 10 bp RNA:DNA hybrid) that were less backtracked had faster recovery rates than WT RNAP (with a 11 bp RNA:DNA hybrid) (**Figure 24E**). The observation that various RNAP mutations had similar effects on the stabilities of the pre-translocated and backtracked TECs suggested that these states share common structural features in the active site: the TL folds into a helical bundle conformation with the BH, closes the active site and interacts with the penultimate (in the backtracked state) or the 3' terminal (in the pre-translocated state) RNA nucleotide.

5.6.3 Structural model of the backtracked TEC

The above conclusion contradicts the published crystal structures of the backtracked *T. thermophilus* and *S. cerevisiae* TECs, where the TL adopts a conformer that neither closes the active site nor approaches the RNA 3' end (Wang *et al*, 2009; Sekine *et al*, 2015). In these structures, the conformation of the RNA 3´ end does not permit the formation of a fully helical TL because the RNA backbone is bent upstream towards the βD-loop II so that the β´His^{936} of

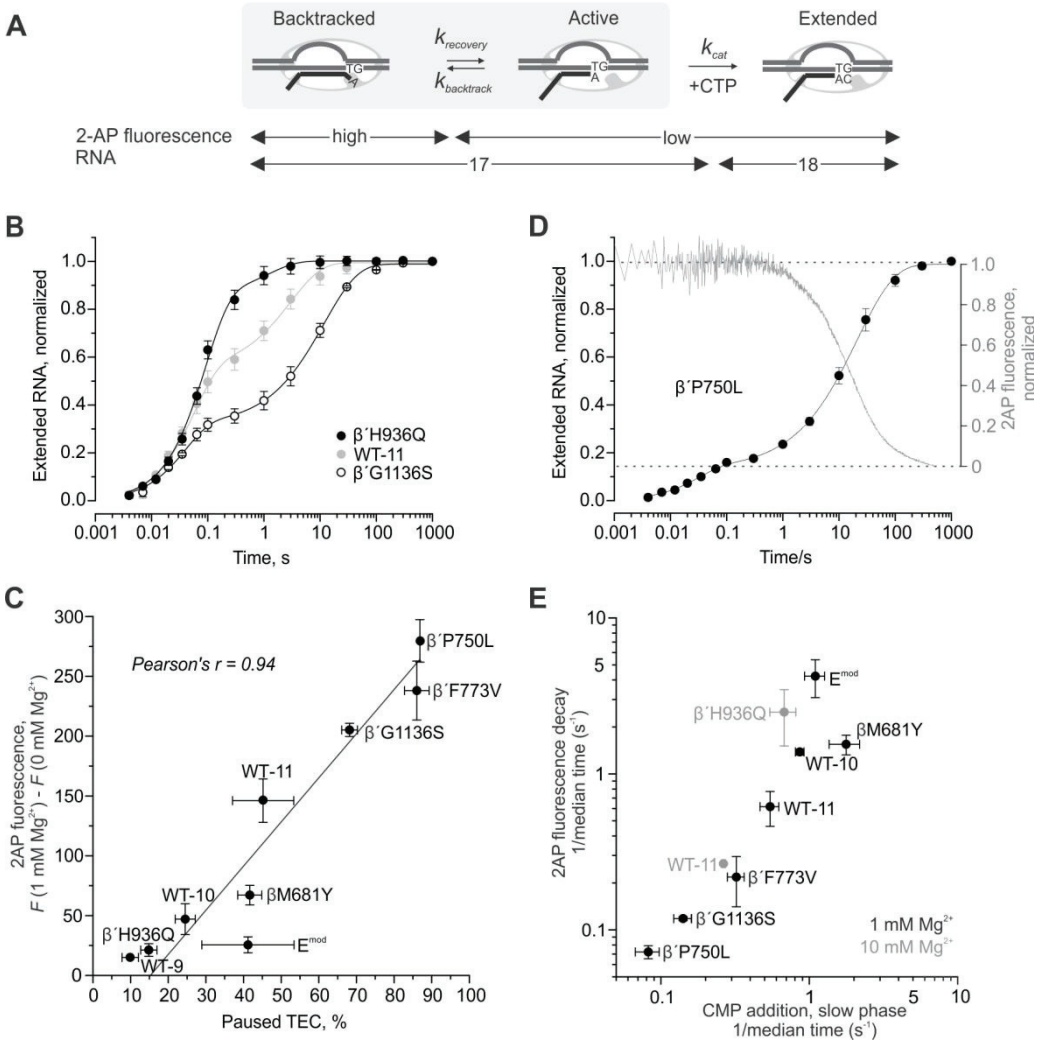


Figure 24. TH stabilizes the backtracked state. A) Experimental approach for measuring the occupancy and escape rate of the backtracked state. **B)** The effect of amino acid substitutions that stabilize (β´G1136S) or destabilize (β´H936Q) the helical TL on the paused state occupancy and recovery rate monitored by CMP incorporation. The data was fitted to a function comprising a sum of a single exponential function describing the rapid phase and a stretched exponential function describing the slower phase. **C)** The correlation of the TEC-2AP fluorescence with the fraction of paused TEC among RNAP variants. The length of the RNA:DNA hybrid in TEC-2APs was 11 bp (RNAP variants and WT-11), 10 bp (WT-10) or 9 bp (WT-9). **D)** Time-resolved measurement of recovery of β´P750L TEC-2AP from the paused state as monitored by RNA extension (filled circles and best-fit line) and 2-AP fluorescence decrease (continuous grey trace) upon the addition of CTP. The area between the dashed horizontal lines indicates the kinetically slow phase denoted as paused in the analyses. **E)** Comparison of pause recovery kinetics followed by the TEC-2AP reaction with CTP as determined from 2-AP fluorescence decay curves or direct measurements of RNA extension. The grey data points indicate that the pre-equilibration and the reaction with CTP were performed at 10 mM Mg^{2+} (all other experiments were performed at 1 mM Mg^{2+}). Error bars represent the range of duplicate measurements or standard deviations from multiple experiments.

the helical TL would clash with the backtracked RNA nucleotide. Therefore, the backtracked state structure was modeled based on the crystal structure of the pre-translocated *E. coli* ITC (Liu *et al*, 2016) and the model of the pre-translocated *T. thermophilus* TEC (refined from an earlier model shown in Figure 13). Both models were *in silico* extended into 1-nucleotide backtracked complexes by attaching a 2AP nucleoside monophosphate to the pre-translocated RNA 3' end. The spatially feasible poses of the backtracked nucleotide were searched for by manipulating the torsion angles around the phosphate and C5' of the backtracked nucleotide as well as C3' of the penultimate nucleotide. The backtracked nucleotide could not be fitted into the active site. At the same time, a narrow passage between the trigger helix 1 (TH1) and the aspartate triad opened towards a larger pocket at the base of the secondary channel, which could fittingly accommodate the backtracked nucleotide in various orientations (**Figure 25**). The upper half of the pocket constitutes the E-site, which would be the most plausible binding site for the backtracked nucleotide. First, the E- site has been observed to bind NTPs (Westover *et al*, 2004b; Wang *et al*, 2006; Sosunov *et al*, 2003) (see Figure 4, state 2), frayed RNA 3' NMP (Sydow *et al*, 2009; Toulokhonov *et al*, 2007) (see Figure 4, state 9), pyrophosphate (Liu *et al*, 2016) and TGT (Vassylyev *et al*, 2005; Artsimovitch *et al*, 2011; Malinen *et al*, 2012) (Figure 13B). Second, as demonstrated in section 5.2.2 and elsewhere (Artsimovitch *et al*, 2011; Yuzenkova *et al*, 2013), TGT stabilizes the closed active site. Therefore, by binding to the same site the backtracked nucleotide could also stabilize the closed active site. Third, TGT coordinates an additional Mg^{2+} ion together with βE813 (Vassylyev *et al*, 2005) and the backtracked nucleotide can be modeled to coordinate an additional Mg^{2+} in a similar position. Binding of an additional Mg^{2+} ion can possibly explain the dependence of the occupancy of the backtracked state on the presence of Mg^{2+} with a K_D~ 1 mM. Fourth, binding of TGT is incompatible with the 1-nucleotide backtracked state (**Study VI, Supplementary figure 5**) suggesting that the backtracked nucleotide and TGT bind in an overlapping volume. Fifth, the E-site pose is incompatible with a 2-nucleotide backtracked state, and consistently, the high-fluorescence backtracked state does not readily form in the 2-nucleotide backtracked TEC (**Study VI, Supplementary figure 5**).

Overall, the results and modeling suggest that the backtracked nucleotide binds to the E-site where it coordinates a Mg^{2+} ion. Binding towards the E-site is supported by the effects of the E^{mod} RNAP variant (β´A455E, V501K, Q504R, these residues are cyan in **Figure 25**), which destabilized the high fluorescent state and accelerated the backtracking recovery rate, but did not affect the fraction of the inactivated state (**Figures 24C, E**). The backtracked RNA stabilizes the helical TL in two ways: First, the phosphate of the backtracked NMP interacts with β´His^{936} and β´Qln^{929}. Accordingly, mutations to these residues destabilized the backtracked state. Second, the penultimate RNA nucleotide in the A-site provides stabilizing interactions with the helical TL β´Met^{932} and the A-site residues. Consistently, destabilization of the RNA in the A-site by the β´N458D and β´M932L substitutions destabilized the backtracked state.

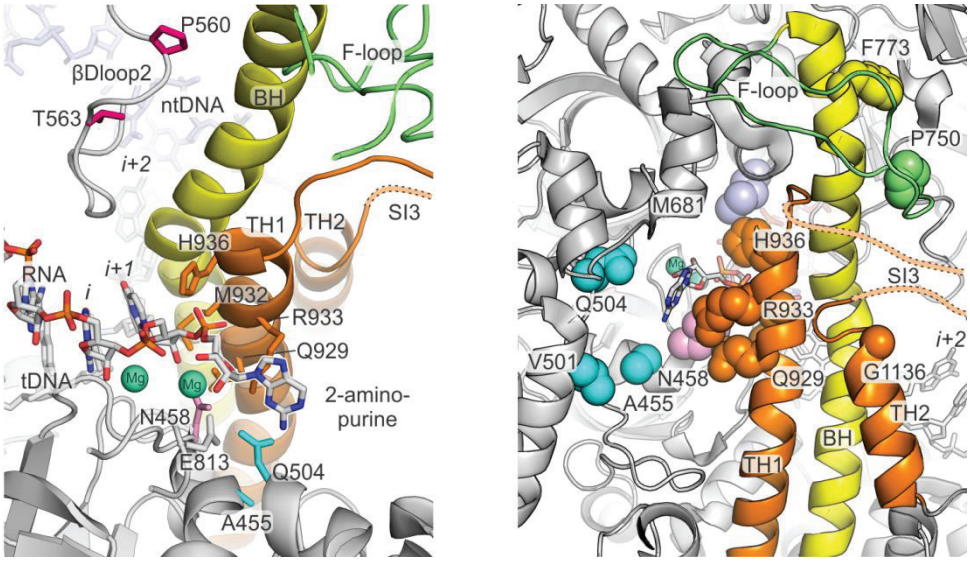


Figure 25. A structural model of the 1 nucleotide backtracked state. The model was built starting from the pre-translocated *E. coli* ITC structure (PDB ID: 5IPL) (Liu *et al*, 2016) by extending the RNA 3´ end with an 2AP nucleoside monophosphate. The spatially feasible poses of the backtracked nucleotide were searched for by manipulating the torsion angles around the phosphate and C5' of the backtracked nucleotide as well as C3' of the penultimate nucleotide. Details of the modelling are described in Study IV. Left panel: a top view of the RNAP active site, most of the β-subunit is omitted for clarity. Right panel: a side view of the active site from the secondary channel. Amino acids relevant to this study are visualized as sticks (left) and spheres (right). The insertion site and the direction of protrusion of the SI3 domain are marked with a dashed line.

5.6.4 TL stabilized backtracked state has low reactivity towards Gre RNA cleavage factors

The recovery from the closed 1 nucleotide backtracked state likely relies on the same conformational transitions that mediate NTP binding and thermally driven translocation during the on-pathway elongation accounting for a relatively fast escape rate (~0.4 s^{-1}). Interestingly, the lifetimes of the backtracking events measured here closely parallel the "ubiquitous pauses" (Neuman *et al*, 2003; Larson *et al*, 2011; Herbert *et al*, 2006; Kireeva & Kashlev, 2009), which have also been argued to be caused by short backtracking (Galburt *et al*, 2007; Depken *et al*, 2009; Galburt *et al*, 2009). These characteristics further suggest that the closed 1-nucleotide backtracked states both play important roles in the control of the sequence specific elongation rate and serve as failsafe checkpoints preventing RNAP from sliding backwards towards the long backtracked states that lead to prolonged inactivation or a permanent arrest.

As part of the transcriptional proofreading, a backtracked TEC is subject to RNA cleavage by Gre factors that excise 2 or more 3'-terminal nucleotides of the nascent RNA. The sensitivity of TECs toward Gre factors is routinely used to evaluate their backtracking propensity. Interestingly, varying the fraction of the pre-formed backtracked state with Mg^{2+} before the reaction with GreA demonstrated that GreA did not prefer the TL stabilized backtracked TEC as a target for RNA cleavage (**Study VI, Figure 5**). As expected, GreA mediated RNA cleavage decreased the

TEC-2AP fluorescence, which coincided with RNA cleavage. Notably, the GreA mediated RNA cleavage and the addition of CMP by the high-fluorescent fraction of TEC-2AP occurred with approximately the same rate suggesting that both processes were limited by the same conformational transition, presumably the unfolding of the helical TL that both speeds up the recovery by forward translocation and allows access to RNA by GreA. The effects of GreA binding on the stability of the high fluorescent state were directly assessed by employing the D41N variant that is incapable of stimulating cleavage (Laptenko *et al*, 2003). Mixing TEC-2AP with GreA D41N reduced fluorescence 20% indicating that the bulk of the TEC maintained the unpaired 2AP in a high fluorescent conformation upon GreA D41N binding. On the other hand, GreB D41N binding reduced fluorescence by >70% suggesting that GreB binding compromised the interactions that stabilized 2AP in a high fluorescent state. Interestingly, GreB was also faster than GreA or CTP in stimulating the TEC-2AP fluorescence decay, perhaps due to its ability to actively disrupt the TL stabilized backtracked state.

The low reactivity of the closed active site towards Gre-factors is not unexpected given the latter were proposed to substitute the TL from the RNAP active site to gain access to RNA (Roghanian *et al*, 2011). However, the results presented in **Supplementary figure 3** (**Study VI**) indicate that this explanation is oversimplified. While GreA was indeed capable of cleaving RNA in ΔTL, STL-bound and β´H936A TEC-2APs, the RNA cleavage rates were 30-fold slower than that of the wild-type TEC-2AP suggesting that GreA requires a certain conformer of TL for optimal activity. Similar results were recently reported by Miropolskaya *et al.* (Miropolskaya *et al*, 2017). Importantly, the TL conformer required for the optimal activity of GreA is apparently distinct from the one that stabilizes the high fluorescent backtracked state in TEC-2AP.

5.6.5 Transcription proofreading by RNA cleavage is modulated by the TL

RNAPs possess an intrinsic RNA cleavage activity and several studies indicate that this activity is dependent on the TL (Yuzenkova & Zenkin, 2010; Bae *et al*, 2015; Mishanina *et al*, 2017). The large variation in intrinsic RNA cleavage activity among bacterial RNAPs is of physiological and mechanistic interest. For example, the *Deinococcus radiodurans* RNAP possesses up to a 200-fold higher RNA cleavage rate compared to the *E. coli* RNAP. The results presented in **IV** demonstrate that at least part of this difference originates from structural variations in the TL. Specifically, swapping the *D. radiodurans* TL or introduction of *D. radiodurans* type TL substitutions (β´G1136M-ΔSI3) to the *E. coli* RNAP substantially increased its RNA cleavage rate (**Study IV, Table 1**). However, our measurements of translocation rates and equilibria with the assays described above did not reveal major differences between the *E. coli* and *D. radiodurans* TECs that could account for their large differences in RNA cleavage activity (**Study IV, Figure 2**).

Because of the central role of TL in RNA cleavage, we examined whether the TH stabilized high fluorescent backtracked state of TEC-2AP reflected the conformation relevant for RNA proofreading by the intrinsic cleavage of the backtracked RNA. TEC-2AP was a good experimental system for this purpose, because it allowed the simultaneous monitoring of backtracking and RNA cleavage. Interestingly, *E. coli* and *D. radiodurans* RNAPs both formed the

backtracked state with a similar rate and amplitude, yet their RNA cleavage rates within TEC-2AP differed by a factor of >200, suggesting that their ability to backtrack could not explain the difference in cleavage activity (**Study VI, Figure 6**). However, we observed a correlation between the TEC-2AP fluorescence and the RNA cleavage rate among RNAP variants, supporting the proposal that the helical TL works as a positional catalyst for RNA proofreading by stabilizing the backtracked state (Mishanina *et al*, 2017). The different apparent K_D for Mg^{2+} in RNA cleavage (13 mM) and backtracked state formation (0.7 mM) argues that the Mg^{2+} ion, which stabilized the 1-nucleotide backtracked state was not coordinated in such a way that it would promote RNA hydrolysis. Furthermore, the rate of formation of the Mg-dependent backtracked state was 6000-fold faster than the rate of RNA cleavage, indicating that backtracking was not the rate-limiting step in intrinsic RNA cleavage. These disparate rates could also imply that the catalytically active conformation represents a minor population among backtracked complexes. A conformational change in the RNA 3´ end or the tip of the helical TL in the backtracked state may present a major rate-limiting step to the intrinsic RNA cleavage reaction. The absence of apparent differences between the *E. coli* and *D. radiodurans* RNAPs in the preceding steps suggests that such a conformational change takes place more readily in the active site of the backtracked *D. radiodurans* RNAP. A second possibility is that the generation of the nucleophile (OH^-) limits the hydrolytic reaction in the *E. coli* RNAP. The reaction center could even be positioned correctly but fail to generate the nucleophile that would attack the phosphodiester bond. These two suggested scenarios need not be mutually exclusive. While these studies do not provide an answer for the mechanism of intrinsic RNA cleavage, they untangle the contributions of backtracking and RNA cleavage in RNA proofreading.

6 CONCLUDING REMARKS

In the experimental part of this thesis we present the development of new methods to study the structure and function of the transcription elongation complex. Specifically, we embedded environment-sensitive fluorophores in DNA and RNA which allowed the time-resolved monitoring of the translocational motions of RNA polymerase (RNAP). The site-specific photo-crosslinking of nucleic acids allowed the structural probing of the front and rear ends of the transcription bubble. In addition, we measured the rate of pyrophosphate (PP_i) release following nucleotide addition with a fluorescent coupled enzyme assay. The employment of these methodologies revealed several important aspects about the RNAP translocation mechanism.

First, besides allowing substrate entry and catalysis, opening and closure of the active site emerged as the central control steps for RNAP translocation. Here, the trigger loop (TL) is the control element that alternates between helical and looped conformations. In the looped conformation the RNAP active site is open biasing the RNA:DNA hybrid towards the post-translocated state. The binding of ligands in the active site, including PP_i and tagetitoxin (TGT) can shift the folding equilibrium of the TL towards the helical conformation and thereby stabilize the pre-translocated state. On the other hand, inhibitors of the CBR-series allosterically destabilize the helical TL and promote forward translocation by RNAP. The measurements of the individual reaction steps along the RNAP nucleotide addition cycle demonstrate that post-catalytic steps contribute to the overall transcription elongation rate and take place at a millisecond timescale. Translocation is weakly coupled to PP_i release, but not in a manner implicated by the power-stroke mechanism. Instead, from a thermodynamic standpoint RNAP translocation can be now specified as operating by a “kinetically controlled thermal ratchet” mechanism. The heights of the kinetic translocation barriers – and thereby the translocation bias – are dictated by the interaction of nucleic acids with RNAP, especially the RNA 3´-terminal end towards the active site aspartate triad and the helical TL.

Second, we found functional evidence that translocation is a multi-step process, as initially proposed in a structural study (Brueckner & Cramer, 2008). Kinetic modeling and DNA:DNA crosslinking indicated the existence of an intermediate translocation state between the canonical pre- and post-translocated states suggesting that translocation takes place via a two-step mechanism: During the first step the template DNA strand migrates only a half register to an intermediate position whereas the second step encompasses the isomerization of the DNA into the post-translocated state. In this case the nucleic acids movements are partially asynchronous. The first step is presumably controlled by the aforementioned active site opening where the unfolding of the helical TL in the pre-translocated state allows the forward migration of the RNA 3´-terminal end by one register. This step can be modulated by small molecule ligands that affect the TL folding dynamics. On the other hand, the translocation rate and equilibrium of the second step are strongly dependent on the DNA dinucleotide sequence in the active site. A backward-bias at this step inhibits the binding of the next incoming nucleotide. Importantly, our data further suggests that the movement of the nucleic acids can also become fully asynchronous. In this scenario following the unfolding of the TL only the RNA translocates

whereas the DNA remains stationary. Because, due to a yet unknown mechanism, this structural organization of the nucleic acids within the TEC does not allow the translocation of DNA, it represents an off-pathway translocation state. This state isomerizes back to the on-pathway through the backward translocation of the RNA. The preference for entry to the off-pathway state is modulated both by the DNA dinucleotide sequence in the active site as well as the RNA:DNA base pairing at the upstream edge of the RNA:DNA hybrid. Therefore, the new translocation model has implications in the sequence dependent control of transcription elongation. How exactly does the DNA sequence affect these structural transitions and how is the impediment in the translocation process connected to other conformational changes that facilitate transcriptional pausing (e.g. bending of the BH and opening of the clamp domain) will be important future question to address. Furthermore, the increasing understanding of the translocation process as well as the wide range of experimental tools available for probing various parameters calls for systematic re-evaluation of the effects of elongation factors on the activity of RNAP.

Third, several observations indicated that the front and rear ends of the transcription bubble are critical sites for regulating RNAP translocation. From this work and others it is also now clear that the DNA sequence at the downstream and upstream ends of the RNA:DNA hybrid can affect the activity of RNAP. Specifically, destabilization of these pairs promote forward translocation whereas their stable association is linked to the formation of the off-pathway translocation state. The DNA sequence around the downstream fork junction modulates the aforementioned second translocation step and the isomerization to the off-pathway state. On the other hand, opening of the DNA helix upstream of the transcription bubble allows RNAP backtracking that reduces the overall transcription elongation rate. Accordingly, we concluded that the universally conserved elongation factor NusG enhances RNAP processivity by stabilizing DNA base pairing at the upstream fork junction. Interestingly, several other transcription factors that regulate RNAP, including RfaH, Sigma^{70}, UvrD and Mfd as well as the viral proteins HK022 Nun and Xp10 p7 bind at or near the upstream fork junction. Furthermore, DNA rewinding at the rear end of the transcription bubble has been implicated as an important factor in intrinsic and Rho-mediated transcription termination. Finally, mis-regulation at the upstream fork junction may lead to persistent RNA:DNA hybrids, or R-loops, that cause genomic instability. Considering their importance to gene regulation, the elucidation of those control mechanisms at the molecular level should be an important future goal. The methods described here, particularly the straightforward DNA conformation specific TA-site photo-crosslinking, are anticipated to be useful in this endeavor.

Finally, we illuminated the structural and thermodynamic basis of backtracking. We developed assays for monitoring the formation of the 1-nucleotide backtracked state and measured the kinetics of the recovery of the backtracked state for a number of structural RNAP variants. The variants with translocation defects, which we identified in our earlier studies, turned out to be very informative also for this investigation. Unexpectedly, we found that closing of the RNAP active site stabilizes the backtracked state and prevents Gre-cleavage factors from accessing the

backtracked RNA. The helical TL stabilizes the pre-translocated and 1-nucleotide backtracked states with similar interactions but, due to additional interactions between RNAP and the backtracked RNA nucleotide, forward translocation from the latter conformation is 100-fold slower. This clarifies the thermodynamics of RNAP translocation and backtracking: the 1-nucleotide backtracked state is in equilibrium with the post- and pre-translocated states, but the thermally driven interconversions to and from the backtracked state are much slower than the other steps in the nucleotide addition cycle. Hence, the relatively high kinetic barrier limits backtracking during transcription elongation. Presumably, the barrier can be overcome if the forward translocation of RNAP is inhibited, by for example a DNA-bound protein, a DNA lesion or a pause sequence. On the other hand, closing of the active site in the 1-nt backtracked state probably restricts any further backtracking that would lead to long pauses or stall transcription. It is interesting that the life-times of the 1-nt backtracked pauses parallel the ubiquitous pauses that occur on average once per every 100 bp (Neuman *et al*, 2003; Larson *et al*, 2011; Herbert *et al*, 2006) and have been argued to be caused by short backtracking (Galburt *et al*, 2007; Depken *et al*, 2009). This could imply that short backtracking happens frequently during transcription elongation, and most backtracking events do not lead to long backtracking. The modulation of backtracking dynamics by targeted RNAP mutations may offer possibilities to investigate the prevalence of backtracking *in vivo*, and to clarify the significance and consequences of transcriptional pauses in gene regulation and cell physiology.

ACKNOWLEDGEMENTS

This work was financially supported by the Academy of Finland, the National Doctoral Programme in Informational and Structural Biology (ISB), the Doctoral Programme in Molecular Life Sciences (DPMLS) and the Emil Aaltonen Foundation. Visits to international research conferences were enabled by funds from ISB, DPMLS, Turku University Foundation, Turku Centre for Systems Biology, Computational and Molecular Methodologies for Life Sciences and the Federation of American Societies for Experimental Biology.

I wish to thank my colleagues in the RNA polymerase group for the successful teamwork throughout these years. First and foremost, I express my deepest gratitude to Dr. Georgiy Belogurov who guided me through the PhD. He trained the nascent scientist by wrapping me up in long detailed discussions and created a sense of freedom while pruning the naivety. I will always be influenced by his rigorous and uncompromised approach to science. I am fortunate to have Dr. Anssi Malinen as a long-term co-worker, instructor and a reliable friend. I am especially thankful for cooperation with Janne Mäkinen and Dr. Thadée Grocholski, who made great contributions to this work. I thank all the past and current members of the RNA polymerase group for helping me and vivifying our lab atmosphere: Salli Keinänen, Mila Vainonen, Pavlina Gregorova, Henri Malmi, Karin Sõstar, Anton Kermanov, Kaisa Palmu, Dr. Ranjit Prajapati, Tatu Porjamo and Armando Alvarez Losada.

I wish to thank my scientific collaborator, Prof. Irina Artsimovitch at the Ohio State University, USA, for always giving the support needed, for being an outspoken mentor and for providing the opportunity to perform experiments in her laboratory. I thank Dr. Andrey Kulbachinskiy and Dr. Daria Esyunina at the Institute of Molecular Genetics, Russian Academy of Sciences, for the opportunity to participate in collaboration as well as Prof. Mark Johnson and his group at Åbo Akademi for cooperation. Mark is also thanked as the director of the ISB for nurturing the communal spirit in the delightful winter schools.

Dr. Roman Tuma at the University of Leeds and Prof. Andre Ribeiro at the Tampere University of Technology are warmly thanked for the review of the thesis and their insightful comments. Language revision of the thesis by Dr. Helen Cooper is sincerely acknowledged.

I am grateful to Dr. Matjaz Barboric at the University of Helsinki for being the member of the thesis committee, for guidance in career-planning and support in realizing the plans. I am also grateful to the other member of the committee, Dr. Anton Zavialov at the University of Turku for fostering insightful scientific discussion wherever we meet.

The experimental work for this thesis was conducted at the Department of Biochemistry, University of Turku. I am grateful to Prof. Jyrki Heino for creating an undisturbed environment for research, for offering advice to the young mind, and for partaking me in the activities of Biocity Turku. I want to thank my colleagues at the unit of Biochemistry for offering their help throughout the years. I am thankful to Dr. Jarmo Käpylä for guidance since the beginning of my

studies. I wish to thank Jani Sointusalo for providing essential technical assistance and building the UV-crosslinking device. I thank Anu Hirvensalo, Heli Kalevo and Teija Luotohaara for laboratory maintenance and Satu Jasu for help in administration. Dr. Jari Nuutila and Dr. Pekka Rappu were very open to providing help with experiments. I warmly remember the discussions, seminars and trips with Dr. Kalle Sipilä, Dr. Maria Salmela, Dr. Pekka Patrikainen and other fellow graduate students.

I wish to thank Prof. Reijo Lahti for sparking my passion into the principles of gene regulation. I remember my previous teachers with gratitude, especially Dr. Juha Vilkki, Markku Tenkanen, Jari Heikkilä, Dr. Simo Veistola, as well as former supervisors Prof. Johanna Vilkki at Luonnonvarakeskus and Prof. Suparna Sanyal at Uppsala University who all helped to build my confidence and inspired the interest towards biology and natural sciences.

The scientific efforts would be meaningless without the people I love. I am grateful to my friends for sharing the excitements and sorrows, and for reflecting the world outside the realm of science. The moments and discussions especially with Kalle and Risto have been great fun but also helped me to define who I am and what I would like to become. I thank my grandparents and other relatives for support and for nurturing my attachment to nature. I am always happy to have Sara, my sister, and Anni with me. From the early years on, Asko as a family friend, and my parents Ari and Eila have instructed my life and have given examples that have inspired my curiosity and demonstrated that being a scientist is not only an exciting profession, but a state of mind.

I am incredibly fortunate to see my children Sofia and Aatos explore the world and to bring great joy and meaning to every day. Lastly, I cannot thank enough my dear wife and companion Enni for her support to my endeavors and for forgiving my weaknesses. Her reservoir of empathy gives me strength and carries me through both good and difficult times.

REFERENCES

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

Abràmoff MD, Magalhães PJ & Ram SJ (2004) Image processing with imageJ. *Biophotonics Int.* **11:** 36–41

Adebali O, Chiou Y-Y, Hu J, Sancar A & Selby CP (2017) Genome-wide transcription-coupled repair in Escherichia coli is mediated by the Mfd translocase. *Proc. Natl. Acad. Sci. USA* **114:** E2116–E2125

Adelman K, Marr MT, Werner J, Saunders A, Ni Z, Andrulis ED & Lis JT (2005) Efficient release from promoter-proximal stall sites requires transcript cleavage factor TFIIS. *Mol. Cell* **17:** 103–112

Artsimovitch I, Chu C, Lynch AS & Landick R (2003) A new class of bacterial RNA polymerase inhibitor affects nucleotide addition. *Science* **302:** 650–654

Artsimovitch I & Landick R (2000) Pausing by bacterial RNA polymerase is mediated by mechanistically distinct classes of signals. *Proc. Natl. Acad. Sci. USA* **97:** 7090–7095

Artsimovitch I & Landick R (2002) The transcriptional regulator RfaH stimulates RNA chain synthesis after recruitment to elongation complexes by the exposed nontemplate DNA strand. *Cell* **109:** 193–203

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:** 40395–40400

Astumian RD (1997) Thermodynamics and kinetics of a Brownian motor. *Science* **276:** 917–922

Bae B, Nayak D, Ray-Soni A, Mustaev A, Landick R & Darst SA (2015) CBR antimicrobials inhibit RNA polymerase via at least two bridge-helix cap-mediated effects on nucleotide addition. *Proc. Natl. Acad. Sci. USA* **112:** E4178-87

Bai L, Fulbright RM & Wang MD (2007) Mechanochemical kinetics of transcription elongation. *Phys. Rev. Lett.* **98:** 68103

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

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:** 183–193

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

Batada NN, Westover KD, Bushnell DA, Levitt M & Kornberg RD (2004) Diffusion of nucleoside triphosphates and role of the entry site to the RNA polymerase II active center. *Proc. Natl. Acad. Sci. USA* **101:** 17361–17364

Belogurov GA & Artsimovitch I (2015) Regulation of Transcript Elongation. *Annu. Rev. Microbiol.* **69:** 49–69

Belogurov GA, Mooney RA, Svetlov V, Landick R & Artsimovitch I (2009) Functional specialization of transcription elongation factors. *EMBO J.* **28:** 112–122

Belogurov GA, Vassylyeva MN, Svetlov V, Klyuyev S, Grishin N V., Vassylyev DG & Artsimovitch I (2007) Structural Basis for Converting a General Transcription Factor into an Operon-Specific Virulence Regulator. *Mol. Cell* **26:** 117–129

Berdygulova Z, Esyunina D, Miropolskaya N, Mukhamedyarov D, Kuznedelov K, Nickels BE, Severinov K, Kulbachinskiy A & Minakhin L (2012) A novel phage-encoded transcription antiterminator acts by suppressing bacterial RNA polymerase pausing. *Nucleic Acids Res.* **40:** 4052–4063

Bintu L, Ishibashi T, Dangkulwanich M, Wu YY, Lubkowska L, Kashlev M & Bustamante C (2012) Nucleosomal elements that control the topography of the barrier to transcription. *Cell* **151:** 738–749

Blank A, Gallant JA, Burgess RR & Loeb LA (1986) An RNA polymerase mutant with reduced accuracy of chain elongation. *Biochemistry* **25:** 5920–5928

Bochkareva A, Yuzenkova Y, Tadigotla VR & Zenkin N (2011) Factor-independent transcription pausing caused by recognition of the RNA–DNA hybrid sequence. *EMBO J.* **31:** 630–639

Borukhov S, Polyakov A, Nikiforov V & Goldfarb A (1992) GreA protein: a transcription elongation factor from Escherichia coli. *Proc. Natl. Acad. Sci. USA* **89:** 8899–8902

Borukhov S, Sagitov V, Goldfarb A. (1993) Transcript cleavage factors from E. coli. *Cell* **72**:459-66

Brueckner F & Cramer P (2008) Structural basis of transcription inhibition by alpha-amanitin and implications for RNA polymerase II translocation. *Nat. Struct. Mol. Biol.* **15:** 811–818

Brune M, Hunter JL, Corrie JET & Webb MR (1994) Direct, real-time measurement of rapid inorganic phosphate release using a novel fluorescent probe and its application to actomyosin subfragment-1 ATPase. *Biochemistry* **33:** 8262–8271

Brune M, Hunter JL, Howell SA, Martin SR, Hazlett TL, Corrie JET & Webb MR (1998) Mechanism of inorganic phosphate interaction with phosphate binding protein from Escherichia coli. *Biochemistry* **37:** 10370–10380

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

206: 179–187

Bubunenko MG, Court DL, Al Refaii A, Saxena S, Korepanov A, Friedman DI, Gottesman ME & Alix J-H (2013) Nus transcription elongation factors and RNase III modulate small ribosome subunit biogenesis in Escherichia coli. *Mol. Microbiol.* **87:** 382–393

Burmann BM, Knauer SH, Sevostyanova A, Schweimer K, Mooney RA, Landick R, Artsimovitch I & Rösch P (2012) An α helix to β barrel domain switch transforms the transcription factor RfaH into a translation factor. *Cell* **150:** 291–303

Burns CM, Richardson L V & Richardson JP (1998) Combinatorial effects of NusA and NusG on transcription elongation and Rho-dependent termination in Escherichia coli. *J. Mol. Biol.* **278:** 307–316

Burova E, Hung SC, Sagitov V, Stitt BL & Gottesman ME (1995) Escherichia coli NusG protein stimulates transcription elongation rates in vivo and in vitro. *J. Bacteriol.* **177:** 1388–1392

Cardinale CJ, Washburn RS, Tadigotla VR, Brown LM, Gottesman ME & Nudler E (2008) Termination Factor Rho and Its Cofactors NusA and NusG Silence Foreign DNA in E. coli. *Science* **320:** 935–938

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. USA* **104:** 4267–4672

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:** 212–218

Chakraborty A, Wang D, Ebright YW, Korlann Y, Kortkhonjia E, Kim T, Chowdhury S, Wigneshweraraj S, Irschik H, Jansen R, Nixon BT, Knight J, Weiss S & Ebright RH (2012) Opening and closing of the bacterial RNA polymerase clamp. *Science* **337:** 591–595

Charlet-Berguerand N, Feuerhahn S, Kong SE, Ziserman H, Conaway JW, Conaway R & Egly JM (2006) RNA polymerase II bypass of oxidative DNA damage is regulated by transcription elongation factors. *EMBO J.* **25:** 5481–5491

Chen VB, Arendall WB, Headd JJ, Keedy DA, Immormino RM, Kapral GJ, Murray LW, Richardson JS & Richardson DC (2010) MolProbity: All-atom structure validation for macromolecular crystallography. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **66:** 12–21

Cheung ACM & Cramer P (2011) Structural basis of RNA polymerase II backtracking, arrest and reactivation. *Nature* **471:** 249–53

Cheung ACM, Sainsbury S & Cramer P (2011) Structural basis of initial RNA polymerase II transcription. *EMBO J.* **30:** 4755–63

Chlenov M, Masuda S, Murakami KS, Nikiforov V, Darst SA & Mustaev A (2005) Structure and function of lineage-specific sequence insertions in the bacterial RNA polymerase β′ subunit. *J. Mol. Biol.* **353:** 138–154

Cimino GD, Gamper HB, Isaacs ST & Hearst JE (1985) Psoralens as photoactive probes of nucleic acid structure and function: organic chemistry, photochemistry, and biochemistry. *Ann. Rev. Biochem.* **54:** 1151–1193

Cramer P, Bushnell DA & Kornberg RD (2001) Structural Basis of Transcription: RNA Polymerase II at 2.8 Ångstrom Resolution. *Science* **292:** 1863–1876

Crawford MA, Tapscott T, Fitzsimmons LF, Liu L, Reyes AM, Libby SJ, Trujillo M, Fang FC, Radi R & Vázquez-Torres A (2016) Redox-active sensing by bacterial DksA transcription factors is determined by cysteine and zinc content. *MBio* **7:** e02161-15

Crickard JB, Fu J & Reese JC (2016) Biochemical Analysis of Yeast Suppressor of Ty 4/5 (Spt4/5) Reveals the importance of Nucleic Acid Interactions in the Prevention of RNA Polymerase II Arrest. *J. Biol. Chem.* **291:** 9853–9870

Da L-T, Pardo-Avila F, Xu L, Silva D-A, Zhang L, Gao X, Wang D & Huang X (2016) Bridge helix bending promotes RNA polymerase II backtracking through a critical and conserved threonine residue. *Nat. Commun.* **7:** 11244

Da L-T, Pardo Avila F, Wang D & Huang X (2013) A Two-State Model for the Dynamics of the Pyrophosphate Ion Release in Bacterial RNA Polymerase. *PLoS Comput. Biol.* **9:** e1003020

Da L-T, Wang D & Huang X (2012) Dynamics of pyrophosphate ion release and its coupled trigger loop motion from closed to open state in RNA polymerase II. *J. Am. Chem. Soc.* **134:** 2399–2406

Dalal R V, Larson MH, Neuman KC, Gelles J, Landick R & Block SM (2006) Pulling on the Nascent RNA during Transcription Does Not Alter Kinetics of Elongation or Ubiquitous Pausing. *Mol. Cell* **23:** 231–239

Dangkulwanich M, Ishibashi T, Liu S, Kireeva ML, Lubkowska L, Kashlev M & Bustamante CJ (2013) Complete dissection of transcription elongation reveals slow translocation of RNA polymerase II in a linear ratchet mechanism. *Elife* **2:** e00971

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

Deighan P, Diez CM, Leibman M, Hochschild A & Nickels BE (2008) The bacteriophage λ Q antiterminator protein contacts the β-flap domain of RNA polymerase. *Proc. Natl. Acad. Sci. USA* **105:** 15305–15310

Deighan P & Hochschild A (2007) The bacteriophage lambdaQ anti-terminator protein regulates late gene expression as a stable component of the

transcription elongation complex. *Mol. Microbiol.* **63:** 911–920

Dennis PP, Ehrenberg M, Fange D & Bremer H (2009) Varying rate of RNA chain elongation during rrn transcription in Escherichia coli. *J. Bacteriol.* **191:** 3740–3746

Depken M, Galburt EA & Grill SW (2009) The origin of short transcriptional pauses. *Biophys. J.* **96:** 2189–2193

Dutta D, Shatalin K, Epshtein V, Gottesman ME & Nudler E (2011) Linking RNA Polymerase Backtracking to Genome Instability in E. coli. *Cell* **146:** 533–543

Ehara H, Yokoyama T, Shigematsu H, Yokoyama S, Shirouzu M & Sekine S (2017) Structure of the complete elongation complex of RNA polymerase II with basal factors. *Science* **357:** 921–924

Epshtein V, Kamarthapu V, McGary K, Svetlov V, Ueberheide B, Proshkin S, Mironov A & Nudler E (2014) UvrD facilitates DNA repair by pulling RNA polymerase backwards. *Nature* **505:** 372–377

Epshtein V, Mustaev A, Markovtsov V, Bereshchenko O, Nikiforov V & Goldfarb A (2002) Swing-gate model of nucleotide entry into the RNA polymerase active center. *Mol. Cell* **10:** 623–634

Epshtein V & Nudler E (2003) Cooperation Between RNA Polymerase Molecules in Transcription Elongation. *Science* **300:** 801–805

Erie DA, Hajiseyedjavadi O, Young MC & von Hippel PH (1993) Multiple RNA polymerase conformations and GreA: control of the fidelity of transcription. *Science* **262:** 867–873

Erie DA, Yager TD & von Hippel PH (1992) The single-nucleotide transcription: a Biophysical and Biochemical Perspective. *Annu. Rev. Biophys. Biomol. Struct.* **21:** 379–415

Esyunina D, Agapov A & Kulbachinskiy A (2016a) Regulation of RNA polymerase through the secondary channel. *Proc. Natl. Acad. Sci. USA* **113:** 8699–8704

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

Favre A, Saintomk C & Clivio P (1998) Thionucleobases as intrinsic photoaffinity probes of nucleic acid structure and nucleic acid-protein interactions. *J. Photochem. Photobiol. B Biol.* **42:** 109–124

Feig M & Burton ZF (2009) RNA polymerase II flexibility during translocation from normal mode analysis. *Proteins Struct. Funct. Bioinforma.* **78:** 434–446

Feig M & Burton ZF (2010) RNA Polymerase II with Open and Closed Trigger Loops: Active Site Dynamics and Nucleic Acid Translocation. *Biophysj* **99:** 2577–2586

Feng Y, Degen D, Wang X, Gigliotti M, Shuang L, Zhang Y, Das D, Trevor M, Ebright YW, Talaue M, Connell N & Ebright RH (2015) Structural Basis of Transcription Inhibition by CBR Hydroxamidines and CBR Pyrazoles. *Structure* **23:** 1470–1481

Forde NR, Izhaky D, Woodcock GR, Wuite GJL & Bustamante C (2002) Using mechanical force to probe the mechanism of pausing and arrest during continuous elongation by Escherichia coli RNA polymerase. *Proc. Natl. Acad. Sci. USA* **99:** 11682–11687

Freudenthal BD, Beard WA, Perera L, Shock DD, Kim T, Schlick T & Wilson SH (2015) Uncovering the polymerase-induced cytotoxicity of an oxidized nucleotide. *Nature* **517:** 635–639

Freudenthal BD, Beard WA, Shock DD & Wilson SH (2013) Observing a DNA polymerase choose right from wrong. *Cell* **154:** 1571–68

Furman R, Danhart EM, NandyMazumdar M, Yuan C, Foster MP & Artsimovitch I (2015) pH dependence of the stress regulator DksA. *PLoS One* **10:** e0120746

Furman R, Sevostyanova A & Artsimovitch I (2012) Transcription initiation factor DksA has diverse effects on RNA chain elongation. *Nucleic Acids Res.* **40:** 3392–3402

Galburt EA, Grill SW & Bustamante C (2009) Single molecule transcription elongation. *Methods* **48:** 323–332

Galburt EA, Grill SW, Wiedmann A, Lubkowska L, Choy J, Nogales E, Kashlev M & Bustamante C (2007) Backtracking determines the force sensitivity of RNAP II in a factor-dependent manner. *Nature* **446:** 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 Å Resolution. *Science* **292:** 1876–1882

Gordon AJE, Halliday JA, Blankschien MD, Burns PA, Yatagai F & Herman C (2009) Transcriptional infidelity promotes heritable phenotypic change in a bistable gene network. *PLoS Biol.* **7:** e1000044

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

Greive SJ & von Hippel PH (2005) Thinking quantitatively about transcriptional regulation. *Nat. Rev. Mol. Cell Biol.* **6:** 221–232

Guajardo R & Sousa R (1997) A model for the mechanism of polymerase translocation. *J. Mol. Biol.* **265:** 8–19

Hardman SJO & Thompson KC (2006) Influence of base stacking and hydrogen bonding on the fluorescence of 2-aminopurine and pyrrolocytosine in nucleic acids. *Biochemistry* **45:** 9145–9155

Hawkins ME (2008) Fluorescent Pteridine Probes for Nucleic Acid Analysis. *Methods Enzymol.* **450:** 201–231

Hawkins ME, Pfleiderer W, Balis FM, Porter D & Knutson JR (1997) Fluorescence properties of pteridine nucleoside analogs as monomers and incorporated into oligonucleotides. *Anal. Biochem.* **244:** 86–95

Hein PP, Kolb KE, Windgassen TA, Bellecourt MJ, Darst SA, Mooney RA & Landick R (2014) RNA polymerase pausing and nascent-RNA structure formation are linked through clamp-domain movement. *Nat. Struct. Mol. Biol.* **21:** 794–802

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

Heinonen JK (2001) Biological Role of Inorganic Pyrophosphate. Kluwer Academic Publishers

Herbert KM, Greenleaf WJ & Block SM (2008) Single-molecule studies of RNA polymerase: motoring along. *Annu. Rev. Biochem.* **77:** 149–176

Herbert KM, La Porta A, Wong BJ, Mooney RA, Neuman KC, Landick R & Block SM (2006) Sequence-Resolved Detection of Pausing by Single RNA Polymerase Molecules. *Cell* **125:** 1083–1094

Herbert KM, Zhou J, Mooney RA, Porta AL, Landick R & Block SM (2010) E. coli NusG inhibits backtracking and accelerates pause-free transcription by promoting forward translocation of RNA polymerase. *J. Mol. Biol.* **399:** 17–30

von Hippel PH & Pasman Z (2002) Reaction pathways in transcript elongation. *Biophys. Chem.* **101:** 401–423

Hirtreiter A, Damsma GE, Cheung ACM, Klose D, Grohmann D, Vojnic E, Martin ACR, Cramer P & Werner F (2010) Spt4/5 stimulates transcription elongation through the RNA polymerase clamp coiled-coil motif. *Nucleic Acids Res.* **38:** 4040–4051

Howan K, Monnet J, Fan J & Strick TR (2014) Stopped in its tracks: The RNA polymerase molecular motor as a robust sensor of DNA damage. *DNA Repair (Amst).* **20:** 49–57

Huang X, Wang D, Weiss DR, Bushnell DA, Kornberg RD & Levitt M (2010) RNA polymerase II trigger loop residues stabilize and position the incoming nucleotide triphosphate in transcription. *Proc. Natl. Acad. Sci. USA* **107:** 15745–15750

Hurwitz J (2005) The discovery of RNA polymerase. *J. Biol. Chem.* **280:** 42477–42485

Hwang CS, Xu L, Wang W, Ulrich S, Zhang L, Chong J, Shin JH, Huang X, Kool ET, McKenna CE & Wang D (2016) Functional interplay between NTP leaving group and base pair recognition during RNA polymerase II nucleotide incorporation revealed by methylene substitution. *Nucleic Acids Res.* **44:** 3820–3828

Imashimizu M, Kireeva ML, Lubkowska L, Gotte D, Parks AR, Strathern JN & Kashlev M (2013a) Intrinsic translocation barrier as an initial step in pausing by RNA polymerase II. *J. Mol. Biol.* **425:** 697–712

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

Imashimizu M, Shimamoto N, Oshima T & Kashlev M (2014) Heterogeneous tracking of RNA polymerase and its biological implications. *Transcription* **5:** e28285

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

Irvin JD, Kireeva ML, Gotte DR, Shafer BK, Huang I, Kashlev M & Strathern JN (2014) A Genetic Assay for Transcription Errors Reveals Multilayer Control of RNA Polymerase II Fidelity. *PLoS Genet.* **10:** e1004532

James K, Gamba P, Cockell SJ & Zenkin N (2016) Misincorporation by RNA polymerase is a major source of transcription pausing in vivo. *Nucleic Acids Res.* **45:** 1105–1113

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

Jovanovic M, Burrows PC, Bose D, Camara B, Wiesler S, Zhang X, Wigneshweraraj S, Weinzierl ROJ & Buck M (2011) Activity map of the Escherichia coli RNA polymerase bridge helix. *J. Biol. Chem.* **286:** 14469–14479

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

Kaplan CD, Larsson KM & Kornberg RD (2008) The RNA Polymerase II Trigger Loop Functions in Substrate Selection and Is Directly Targeted by alpha-Amanitin. *Mol. Cell* **30:** 547–556

Kashkina E, Anikin M, Brueckner F, Lehmann E, Kochetkov SN, McAllister WT, Cramer P & Temiakov D (2007) Multisubunit RNA polymerases melt only a single DNA base pair downstream of the active site. *J. Biol. Chem.* **282:** 21578–21582

Kashkina E, Anikin M, Tahirov TH, Kochetkov SN, Vassylyev DG & Temiakov D (2006) Elongation complexes of thermus thermophilus RNA polymerase that possess distinct translocation conformations. *Nucleic Acids Res.* **34:** 4036–4045

Kellinger MW, Ulrich S, Chong J, Kool ET & Wang D (2012) Dissecting chemical interactions governing RNA polymerase II transcriptional fidelity. *J. Am. Chem. Soc.* **134:** 8231–8240

Kent T, Kashkina E, Anikin M & Temiakov D (2009) Maintenance of RNA-DNA hybrid length in bacterial RNA polymerases. *J. Biol. Chem.* **284:** 13497–13504

Kettenberger H, Armache KJ & Cramer P (2004) Complete RNA polymerase II elongation complex structure and its interactions with NTP and TFIIS. *Mol. Cell* **16:** 955–965

Kingston RE, Nierman WC & Chamberlin MJ (1981) A direct effect of guanosine tetraphosphate on pausing of Escherichia coli RNA polymerase during RNA chain elongation. *J. Biol. Chem.* **256:** 2787–2797

Kireeva ML, Domecq C, Coulombe B, Burton ZF & Kashlev M (2011) Interaction of RNA polymerase

II fork loop 2 with downstream non-template DNA regulates transcription elongation. *J. Biol. Chem.* **286:** 30898–30910

Kireeva ML, Hancock B, Cremona GH, Walter W, Studitsky VM & Kashlev M (2005) Nature of the nucleosomal barrier to RNA polymerase II. *Mol. Cell* **18:** 97–108

Kireeva ML & Kashlev M (2009) Mechanism of sequence-specific pausing of bacterial RNA polymerase. *Proc. Natl. Acad. Sci. USA* **106:** 8900–8905

Kireeva ML, Kashlev M & Burton ZF (2010) Translocation by multi-subunit RNA polymerases. *Biochim. Biophys. Acta - Gene Regul. Mech.* **1799:** 389–401

Kireeva ML, Komissarova N & Kashlev M (2000a) Overextended RNA:DNA hybrid as a negative regulator of RNA polymerase II processivity. *J. Mol. Biol.* **299:** 325–335

Kireeva ML, Komissarova N, Waugh DS & Kashlev M (2000b) The 8-nucleotide-long RNA:DNA hybrid is a primary stability determinant of the RNA polymerase II elongation complex. *J. Biol. Chem.* **275:** 6530–6536

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:** 557–566

Kireeva ML, Nedialkov YA, Gong XQ, Zhang C, Xiong Y, Moon W, Burton ZF & Kashlev M (2009) Millisecond phase kinetic analysis of elongation catalyzed by human, yeast, and Escherichia coli RNA polymerase. *Methods* **48:** 333–345

Kireeva ML, Opron K, Seibold SA, Domecq C, Cukier RI, Coulombe B, Kashlev M & Burton ZF (2012) Molecular dynamics and mutational analysis of the catalytic and translocation cycle of RNA polymerase. *BMC Biophys.* **5:** 11

Klein BJ, Bose D, Baker KJ, Yusoff ZM, Zhang X & Murakami KS (2011) RNA polymerase and transcription elongation factor Spt4/5 complex structure. *Proc. Natl. Acad. Sci. USA* **108:** 546–550

Kolb KE, Hein PP & Landick R (2014) Antisense oligonucleotide-stimulated transcriptional pausing reveals RNA exit channel specificity of RNA polymerase and mechanistic contributions of NusA and RfaH. *J. Biol. Chem.* **289:** 1151–1163

Komissarova N, Becker J, Solter S, Kireeva ML & Kashlev M (2002) Shortening of RNA:DNA hybrid in the elongation complex of RNA polymerase is a prerequisite for transcription termination. *Mol. Cell* **10:** 1151–1162

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:** 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. USA* **94:** 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

Korzheva N, Mustaev A, Kozlov M, Malhotra A, Nikiforov V, Goldfarb A & Darst SA (2000) A Structural Model of Transcription Elongation. *Science* **289:** 619–625

Kotlajich M V., Hron DR, Boudreau BA, Sun Z, Lyubchenko YL & Landick R (2015) Bridged filaments of histone-like nucleoid structuring protein pause RNA polymerase and aid termination in bacteria. *Elife* **4:** e04970

Krummel B & Chamberlin MJ (1992) Structural analysis of ternary complexes of Escherichia coli RNA polymerase. Deoxyribonuclease I footprinting of defined complexes. *J. Mol. Biol.* **225:** 239–250

Kusuya Y, Kurokawa K, Ishikawa S, Ogasawara N & Oshima T (2011) Transcription factor GreA contributes to resolving promoter-proximal pausing of RNA polymerase in Bacillus subtilis cells. *J. Bacteriol.* **193:** 3090–3099

Kuznedelov K, Korzheva N, Mustaev A & Severinov K (2002) Structure-based analysis of RNA polymerase function: the largest subunit´s rudder contributes critically to elongation complex stability and is not involved in the maintenance of RNA-DNA hybrid length. *EMBO J.* **21:** 1369–1378

Kyzer S, Kook SH, Landick R & Palangat M (2007) Direct versus limited-step reconstitution reveals key features of an RNA hairpin-stabilized paused transcription complex. *J. Biol. Chem.* **282:** 19020–19028

Landick R (2001) RNA Polymerase Clamps Down. **105:** 567–570

Landick R (2005) NTP-entry routes in multi-subunit RNA polymerases. *Trends Biochem. Sci.* **30:** 651–654

Lane WJ & Darst SA (2010a) Molecular Evolution of Multisubunit RNA Polymerases: Sequence Analysis. *J. Mol. Biol.* **395:** 671–685

Lane WJ & Darst SA (2010b) Molecular Evolution of Multisubunit RNA Polymerases: Structural Analysis. *J. Mol. Biol.* **395:** 686–704

Laptenko O, Kim S-S, Lee J, Starodubtseva M, Cava F, Berenguer J, Kong X-P & Borukhov S (2006) pH-dependent conformational switch activates the inhibitor of transcription elongation. *EMBO J.* **25:** 2131–2141

Laptenko O, Lee J, Lomakin I & Borukhov S (2003) Transcript cleavage factors GreA and GreB act as transient catalytic components of RNA polymerase. *EMBO J.* **22:** 6322–6334

Larson MH, Greenleaf WJ, Landick R & Block SM (2008) Applied Force Reveals Mechanistic and Energetic Details of Transcription Termination. *Cell* **132:** 971–982

Larson MH, Landick R & Block SM (2011) Single-Molecule Studies of RNA Polymerase: One Singular Sensation, Every Little Step It Takes. *Mol. Cell* **41:** 249–262

Larson MH, Nayak D, Gross CA, Block SM & Greenleaf WJ (2014) A pause sequence enriched at translation start sites drives transcription dynamics in vivo. *Science* **344:** 1042–1047

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. USA* **109:** 6555–6560

Liu B & Steitz TA (2017) Structural insights into NusG regulating transcription elongation. *Nucleic Acids Res.* **45:** 968–974

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

Liu C & Martin CT (2001) Fluorescence characterization of the transcription bubble in elongation complexes of T7 RNA polymerase. *J. Mol. Biol.* **308:** 465–475

Liu X, Bushnell DA, Silva D-A, Huang X & Kornberg RD (2011) Initiation Complex Structure and Promoter Proofreading. *Science* **333:** 633–638

Ma C, Yang X & Lewis PJ (2016) Bacterial Transcription as a Target for Antibacterial Drug Development. *Microbiol. Mol. Biol. Rev.* **80:** 139–60

Ma J & Wang MD (2014) Interplay between DNA supercoiling and transcription elongation. *Transcription* **5:** e28636

de Maddalena L, Niederholtmeyer H, Turtola M, Swank ZN, Belogurov GA & Maerkl SJ (2016) GreA and GreB Enhance Expression of Escherichia coli RNA Polymerase Promoters in a Reconstituted Transcription–Translation System. *ACS Synth. Biol.* **5:** 929–935

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

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

Marchand B, Tchesnokov EP & Götte M (2007) The pyrophosphate analogue foscarnet traps the pre-translocational state of HIV-1 reverse transcriptase in a Brownian ratchet model of polymerase translocation. *J. Biol. Chem.* **282:** 3337–3346

Marr MT & Roberts JW (2000) Function of transcription cleavage factors GreA and GreB at a regulatory pause site. *Mol. Cell* **6:** 1275–1285

Martinez-Rucobo FW, Sainsbury S, Cheung ACM & Cramer P (2011) Architecture of the RNA polymerase-Spt4/5 complex and basis of universal transcription processivity. *EMBO J.* **30:** 1302–1310

Mathews E & Durbin D (1989) Tagetitoxin inhibits RNA synthesis directed by RNA polymerases from chloroplasts and Escherichia coli. *J. Biol. Chem.* **265:** 493–498

McGary K & Nudler E (2013) RNA polymerase and the ribosome: the close relationship. *Curr. Opin. Microbiol.* **16:** 112–117

Mejia YX, Nudler E & Bustamante C (2015) Trigger loop folding determines transcription rate of Escherichia coli's RNA polymerase. *Proc. Natl. Acad. Sci. USA* **112:** 743–748

Mellon I & Hanawalt PC (1989) Induction of the *Escherichia coli* lactose operon selectively increases repair of its transcribed DNA strand. *Nature* **342:** 95–98

Miropolskaya N, Artsimovitch I, Klimašauskas S, Nikiforov V & Kulbachinskiy A (2009) Allosteric control of catalysis by the F loop of RNA polymerase. *Proc. Natl. Acad. Sci. USA* **106:** 18942–18947

Miropolskaya N, Esyunina D, Klimašauskas S, Nikiforov V, Artsimovitch I & Kulbachinskiy A (2014) Interplay between the trigger loop and the F loop during RNA polymerase catalysis. *Nucleic Acids Res.* **42:** 544–552

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

Miropolskaya N, Nikiforov V, Klimašauskas S, Artsimovitch I & Kulbachinskiy A (2010) Modulation of RNA polymerase activity through trigger loop folding. *Transcription* **1:** 89–94

Mishanina T V, Palo MZ, Nayak D, Mooney RA, Landick R & Roberts JW (2017) Trigger loop of RNA polymerase is a positional, not acid–base, catalyst for both transcription and proofreading. *Proc. Natl. Acad. Sci. USA* **114:** E5103–E5112

Mooney RA, Artsimovitch I & Landick R (1998) Information Processing by RNA Polymerase: Recognition of Regulatory Signals during RNA Chain Elongation. *J. Bacteriol.* **180:** 3265–3275

Mooney RA, Davis SE, Peters JM, Rowland JL, Ansari AZ & Landick R (2009) Regulator Trafficking on Bacterial Transcription Units In Vivo. *Mol. Cell* **33:** 97–108

Murakami K (2015) Structural Biology of Bacterial RNA Polymerase. *Biomolecules* **5:** 848–864

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. USA* **94:** 6641–6645

Mustaev A, Roberts J & Gottesman M (2017) Transcription elongation. *Transcription* **8:** 150–161

Naryshkina T, Kuznedelov K & Severinov K (2006) The

Role of the Largest RNA Polymerase Subunit Lid Element in Preventing the Formation of Extended RNA-DNA Hybrid. *J. Mol. Biol.* **361:** 634–643

Nayak D, Voss M, Windgassen TA, Mooney RA & Landick R (2013) Cys-Pair Reporters Detect a Constrained Trigger Loop in a Paused RNA Polymerase. *Mol. Cell* **50:** 882–893

Nedialkov YA, Gong XQ, Hovde SL, Yamaguchi Y, Handa H, Geiger JH, Yan H & Burton ZF (2003) NTP-driven Translocation by Human RNA Polymerase II. *J. Biol. Chem.* **278:** 18303–18312

Nedialkov YA, Nudler E & Burton ZF (2012) RNA polymerase stalls in a post-translocated register and can hyper-translocate. *Transcription* **3:** 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 - Gene Regul. Mech.* **1829:** 187–198

Neuman KC, Abbondanzieri EA, Landick R, Gelles J & Block SM (2003) Ubiquitous Transcriptional Pausing Is Independent of RNA Polymerase Backtracking. *Cell* **115:** 437–447

Nudler E (2012) RNA Polymerase Backtracking in Gene Regulation and Genome Instability. *Cell* **149:** 1438–1445

Nudler E, Goldfarb A & Kashlev M (1994) Discontinuous Mechanism of Transcription Elongation. *Science* **265:** 793–796

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:** 33–41

Opalka N, Chlenov M, Chacon P, Rice WJ, Wriggers W & Darst SA (2003) Structure and Function of the Transcription Elongation Factor GreB Bound to Bacterial RNA Polymerase. *Cell* **114:** 335–345

Orlova M, Newlandst J, Dast A, Goldfarb A & Borukhov S (1995) Intrinsic transcript cleavage activity of RNA polymerase. *Proc. Natl. Acad. Sci. USA* **92:** 4596–4600

Pais JE, Bowers KE, Stoddard AK & Fierke CA (2005) A continuous fluorescent assay for protein prenyltransferases measuring diphosphate release. *Anal. Biochem.* **345:** 302–311

Palangat M, Hittinger CT & Landick R (2004) Downstream DNA Selectively Affects a Paused Conformation of Human RNA Polymerase II. *J. Mol. Biol.* **341:** 429–442

Pan T, Artsimovitch I, Fang XW, Landick R & Sosnick TR (1999) Folding of a large ribozyme during transcription and the effect of the elongation factor NusA. *Proc. Natl. Acad. Sci. USA* **96:** 9545–9550

Pan T & Sosnick T (2006) RNA Folding During Transcription. *Annu. Rev. Biophys. Biomol. Struct.* **35:** 161–175

Park J-S, Marr MT & Roberts JW (2002) E. coli transcription repair coupling factor (Mfd protein) rescues arrested complexes by promoting forward translocation. *Cell* **109:** 757–767

Park J-S & Roberts JW (2006) Role of DNA bubble rewinding in enzymatic transcription termination. *Proc. Natl. Acad. Sci. USA* **103:** 4870–4875

Pasman Z & von Hippel PH (2000) Regulation of Rho-dependent transcription termination by NusG is specific to the Escherichia coli elongation complex. *Biochemistry* **39:** 5573–5585

Pasman Z & von Hippel PH (2002) Active Escherichia coli transcription elongation complexes are functionally homogeneous. *J. Mol. Biol.* **322:** 505–519

Perdue SA & Roberts JW (2010) A backtrack-inducing sequence is an essential component of Escherichia coli Sigma70-dependent promoter-proximal pausing. *Mol. Microbiol.* **78:** 636–650

Perdue SA & Roberts JW (2011) Sigma70-dependent transcription pausing in Escherichia coli. *J. Mol. Biol.* **412:** 782–792

Peters JM, Mooney RA, Grass JA, Jessen ED, Tran F & Landick R (2012) Rho and NusG suppress pervasive antisense transcription in Escherichia coli. *Genes Dev.* **26:** 2621–2633

Peters JM, Mooney RA, Kuan PF, Rowland JL, Keles S & Landick R (2009) Rho directs widespread termination of intragenic and stable RNA transcription. *Proc. Natl. Acad. Sci. USA* **106:** 15406–15411

Petushkov I, Pupov D, Bass I & Kulbachinskiy A (2015) Mutations in the CRE pocket of bacterial RNA polymerase affect multiple steps of transcription. *Nucleic Acids Res.* **43:** 5798–5809

Proshkin S, Rahmouni AR, Mironov A & Nudler E (2010) Cooperation Between Translating Ribosomes and RNA Polymerase in Transcription Elongation. *Science* **328:** 504–508

Purich DL (2010) Enzyme kinetics: Catalysis & Control. Elsevier.

Rachofsky EL, Osman R & Ross JB (2001) Probing structure and dynamics of DNA with 2-aminopurine: effects of local environment on fluorescence. *Biochemistry* **40:** 946–956

Ray-Soni A, Bellecourt MJ & Landick R (2016) Mechanisms of Bacterial Transcription Termination: All Good Things Must End. *Annu. Rev. Biochem.* **85:** 319–347

Reeder TC & Hawley DK (1996) Promoter proximal sequences modulate RNA polymerase II elongation by a novel mechanism. *Cell* **87:** 767–777

Roberts J & Park J-S (2004) Mfd, the bacterial transcription repair coupling factor: Translocation, repair and termination. *Curr. Opin. Microbiol.* **7:** 120–125

Roghanian M, Yuzenkova Y & Zenkin N (2011) Controlled interplay between trigger loop and Gre factor in the RNA polymerase active centre. *Nucleic Acids Res.* **39:** 4352–4359

Ryder AM & Roberts JW (2003) Role of the non-template strand of the elongation bubble in intrinsic transcription termination. *J. Mol. Biol.* **334:** 205–213

Said N, Krupp F, Anedchenko E, Santos KF, Dybkov O, Huang YH, Lee CT, Loll B, Behrmann E, Bürger J, Mielke T, Loerke J, Urlaub H, Spahn CMT, Weber G, Wahl MC (2017) Structural basis for λN-dependent processive transcription antitermination. *Nat. Microbiol.* **2**:17062

Sankar TS, Wastuwidyaningtyas BD, Dong Y, Lewis SA & Wang JD (2016) The nature of mutations induced by replication–transcription collisions. *Nature* **535:** 178–181

Schulz S, Gietl A, Smollett K, Tinnefeld P, Werner F & Grohmann D (2016) TFE and Spt4/5 open and close the RNA polymerase clamp during the transcription cycle. *Proc. Natl. Acad. Sci. USA* **113:** E1816–E1825

Seibold SA, Singh BN, Zhang C, Kireeva ML, Domecq C, Bouchard A, Nazione AM, Feig M, Cukier RI, Coulombe B, Kashlev M, Hampsey M & Burton ZF (2010) Conformational coupling, bridge helix dynamics and active site dehydration in catalysis by RNA polymerase. *Biochim. Biophys. Acta - Gene Regul. Mech.* **1799:** 575–587

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:** 408–421

Sevostyanova A, Belogurov GA, Mooney RA, Landick R & Artsimovitch I (2011) Article The b Subunit Gate Loop Is Required for RNA Polymerase Modification by RfaH and NusG. *Mol. Cell* **43:** 253–262

Shaevitz JW, Abbondanzieri EA, Landick R & Block SM (2003) Backtracking by single RNA polymerase molecules observed at near-base-pair resolution. *Nature* **426:** 684–687

Shimamoto N (2013) Nanobiology of RNA polymerase: biological consequence of inhomogeneity in reactant. *Chem. Rev.* **113:** 8400–8422

Sidorenkov I, Komissarova N & Kashlev M (1998) Crucial role of the RNA:DNA hybrid in the processivity of transcription. *Mol. Cell* **2:** 55–64

Sigurdsson S, Dirac-Svejstrup AB & Svejstrup JQ (2010) Evidence that Transcript Cleavage Is Essential for RNA Polymerase II Transcription and Cell Viability. *Mol. Cell* **38:** 202–210

Silva D-A, Weiss DR, Pardo Avila F, Da L-T, Levitt M, Wang D & Huang X (2014) Millisecond dynamics of RNA polymerase II translocation at atomic resolution. *Proc. Natl. Acad. Sci. USA* **111:** 7665–7670

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:** 2234–44

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:** 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. USA* **100:** 15469–15474

Spielmann HP, Dwyer TJ, Hearst JE & Wemmer DE (1995) Solution structures of psoralen monoadducted and cross-linked DNA oligomers by NMR spectroscopy and restrained molecular dynamics. *Biochemistry* **34:** 12937–12953

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

Strathern JN, Jin DJ, Court DL & Kashlev M (2012) Isolation and characterization of transcription fidelity mutants. *Biochim. Biophys. Acta* **1819:** 694–699

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:** 5694–5705

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

Sydow JF, Brueckner F, Cheung ACM, Damsma GE, Dengl S, Lehmann E, Vassylyev DG & Cramer P (2009) Article Structural Basis of Transcription: Mismatch-Specific Fidelity Mechanisms and Paused RNA Polymerase II with Frayed RNA. *Mol. Cell* **34:** 710–721

Sydow JF & Cramer P (2009) RNA polymerase fidelity and transcriptional proofreading. *Curr. Opin. Struct. Biol.* **19:** 732–739

Tadigotla VR, O Maoiléidigh D, Sengupta AM, Epshtein V, Ebright RH, Nudler E & Ruckenstein AE (2006) Thermodynamic and kinetic modeling of transcriptional pausing. *Proc. Natl. Acad. Sci. USA* **103:** 4439–4444

Tagami S, Sekine S, Kumarevel T, Hino N, Murayama Y, Kamegamori S, Yamamoto M, Sakamoto K & Yokoyama S (2010) Crystal structure of bacterial RNA polymerase bound with a transcription inhibitor protein. *Nature* **468:** 978–982

Tan L, Wiesler S, Trzaska D, Carney HC & Weinzierl ROJ (2008) Bridge helix and trigger loop perturbations generate superactive RNA polymerases. *J. Biol.* **7:** 40

Tehranchi AK, Blankschien MD, Zhang Y, Halliday JA, Srivatsan A, Peng J, Herman C & Wang JD (2010) The Transcription Factor DksA Prevents Conflicts between DNA Replication and Transcription Machinery. *Cell* **141:** 595–605

Temiakov D, Anikin M, Ma K, Jiang M & McAllister William T. (2003) Probing the Organization of Transcription Complexes Using Photoreactive 4-Thio-Substituted Analogs of Uracil and Thymidine. *Methods Enzymol.* **371:** 133–143

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:** 655–666

Tetone LE, Friedman LJ, Osborne ML, Ravi H, Kyzer S, Stumper SK, Mooney RA, Landick R & Gelles J (2017) Dynamics of GreB-RNA polymerase interaction allow a proofreading accessory protein to patrol for transcription complexes needing rescue. *Proc. Natl. Acad. Sci. USA* **114:** E1081–E1090

Toulmé F, Mosrin-Huaman C, Sparkowski J, Das A, Leng M & Rahmouni AR (2000) GreA and GreB proteins revive backtracked RNA polymerase in vivo by promoting transcript trimming. *EMBO J.* **19:** 6853–6859

Toulokhonov I & Landick R (2006) The Role of the Lid Element in Transcription by E. coli RNA Polymerase. *J. Mol. Biol.* **361:** 644–658

Toulokhonov I, Zhang J, Palangat M & Landick R (2007) Article A Central Role of the RNA Polymerase Trigger Loop in Active-Site Rearrangement during Transcriptional Pausing. *Mol. Cell* **27:** 406–419

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

Tuske S, Sarafianos SG, Wang X, Hudson B, Sineva E, Mukhopadhyay J, Birktoft JJ, Leroy O, Ismail S, Clark AD, Dharia C, Napoli A, Laptenko O, Lee J, Borukhov S, Ebright RH & Arnold E (2005) Inhibition of Bacterial RNA Polymerase by Streptolydigin: Stabilization of a Straight-Bridge-Helix Active-Center Conformation. *Cell* **122:** 541–552

Vassylyev DG, Sekine S, Laptenko O, Lee J, Vassylyeva MN & Borukhov S (2002) Crystal structure of a bacterial RNA polymerase holoenzyme at 2.6 Å resolution. *Nature* **417:** 712–719

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:** 1086–1093

Vassylyev DG, Vassylyeva MN, Perederina A, Tahirov TH & Artsimovitch I (2007a) Structural basis for transcription elongation by bacterial RNA polymerase. *Nature* **448:** 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:** 163–168

Vitiello CL, Kireeva ML, Lubkowska L, Kashlev M & Gottesman M (2014) Coliphage HK022 Nun protein inhibits RNA polymerase translocation. *Proc. Natl. Acad. Sci. USA* **111:** E2368-2375

Vogel U & Jensen KF (1994a) Effects of guanosine 3´,5´-bisdiphosphate (ppGpp) on rate of transcription elongation in isoleucine-starved Escherichia coli. *J. Biol. Chem.* **269:** 16236–16241

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

Vvedenskaya IO, Vahedian-Movahed H, Bird JG, Knoblauch JG, Goldman SR, Zhang Y, Ebright RH & Nickels BE (2014) Interactions between RNA polymerase and the core recognition element'' counteract pausing. *Science* **344:** 1285–1289

Vyas R, Reed AJ, Tokarsky EJ & Suo Z (2015) Viewing human DNA polymerase β faithfully and unfaithfully bypass an oxidative lesion by time-dependent crystallography. *J. Am. Chem. Soc.* **137:** 5225–5230

Wang B, Predeus A V., Burton ZF & Feig M (2013) Energetic and structural details of the trigger-loop closing transition in RNA polymerase II. *Biophys. J.* **105:** 767–775

Wang B, Sexton RE & Feig M (2017) Kinetics of nucleotide entry into RNA polymerase active site provides mechanism for efficiency and fidelity. *BBA - Gene Regul. Mech.* **1860:** 482–490

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:** 941–954

Wang D, Levitt M & Kornberg RD (2009) Structural Basis of Transcription: Backtracked RNA Polymerase II at 3.4 Angstrom Resolution. *Science* **324:** 1203–1206

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

Wang L, Zhou Y, Xu L, Xiao R, Lu X, Chen L, Chong J, Li H, He C, Fu X-D & Wang D (2015) Molecular basis for 5-carboxycytosine recognition by RNA polymerase II elongation complex. *Nature* **523:** 621–625

Weixlbaumer A, Leon K, Landick R & Darst SA (2013) Structural Basis of Transcriptional Pausing in Bacteria. *Cell* **152:** 431–441

Werner F (2012) A Nexus for Gene Expression — Molecular Mechanisms of Spt5 and NusG in the Three Domains of Life. *J. Mol. Biol.* **417:** 13–27

Werner F & Grohmann D (2011) Evolution of multisubunit RNA polymerases in the three domains of life. *Nat. Rev. Microbiol.* **9:** 85–98

Westover KD, Bushnell DA & Kornberg RD (2004a) Structural Basis of Transcription: Separation of RNA from DNA by RNA Polymerase II. *Science* **303:** 1014–1016

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

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:** 12707–12721

Xu L, Wang W, Wu J, Shin JH, Wang P, Unarta IC, Chong J, Wang Y & Wang D (2017) Mechanism of DNA alkylation-induced transcriptional stalling, lesion bypass, and mutagenesis. *Proc. Natl. Acad. Sci.* **114:** E7082–E7091

Yager TD & von Hippel PH (1991) A thermodynamic analysis of RNA transcript elongation and termination in Escherichia coli. *Biochemistry* **30:** 1097–1118

Yakhnin A V, Murakami KS & Babitzke P (2016) NusG Is a Sequence-specific RNA Polymerase Pause Factor That Binds to the Non-template DNA within the Paused Transcription Bubble. *J. Biol. Chem.* **291:** 5299–5308

Yanofsky C (1981) Attenuation in the control of expression of bacterial operons. *Nature* **289:** 751–758

Yin YW & Steitz TA (2004) The Structural Mechanism of Translocation and Helicase Activity in T7 RNA Polymerase. *Cell* **116:** 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, Gamba P, Herber M, Attaiech L, Shafeeq S, Kuipers OP, Klumpp S, Zenkin N & Veening JW (2014) Control of transcription elongation by GreA determines rate of gene expression in Streptococcus pneumoniae. *Nucleic Acids Res.* **42:** 10987–10999

Yuzenkova Y, Roghanian M, Bochkareva A & Zenkin N (2013) Tagetitoxin inhibits transcription by stabilizing pre-translocated state of the elongation complex. *Nucleic Acids Res.* **41:** 9257–9265

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

Zahn KE, Tchesnokov EP, Götte M & Doublié S (2011) Phosphonoformic acid inhibits viral replication by trapping the closed form of the DNA polymerase. *J. Biol. Chem.* **286:** 25246–25255

Zamft B, Bintu L, Ishibashi T & Bustamante C (2012) Nascent RNA structure modulates the transcriptional dynamics of RNA polymerases. *Proc. Natl. Acad. Sci. USA* **109:** 8948–8953

Zaychikov E, Denissova L & Heumann H (1995) Translocation of the Escherichia coli transcription complex observed in the registers 11 to 20: 'jumping' of RNA polymerase and asymmetric expansion and contraction of the 'transcription bubble'. *Proc. Natl. Acad. Sci. USA* **92:** 1739–1743

Zaychikov E, Emil M, Denissova L, Maxim K, Markovtsov V, Kashlev M, Heumann H, Nikiforov V & Goldfarb A (1996) Mapping of Catalytic Residues in the RNA Polymerase Active Center. *Science* **273:** 107–110

Zellars M & Squires CL (1999) Antiterminator-dependent modulation of transcription elongation rates by NusB and NusG. *Mol. Microbiol.* **32:** 1296–1304

Zenkin N, Severinov K & Yuzenkova Y (2015) Bacteriophage Xp10 anti-termination factor p7 induces forward translocation by host RNA polymerase. *Nucleic Acids Res.* **43:** 6299–6308

Zenkin N, Yuzenkova Y & Severinov K (2006) Transcript-Assisted Transcriptional Proofreading. *Science* **313:** 518–520

Zhang G, Campbell EA, Minakhin L, Richter C, Severinov K & Darst SA (1999) Crystal structure of Thermus aquaticus core RNA polymerase at 3.3 Å resolution. *Cell* **98:** 811–824

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

Zhang L, Pardo-Avila F, Unarta IC, Cheung PP-H, Wang G, Wang D & Huang X (2016) Elucidation of the Dynamics of Transcription Elongation by RNA Polymerase II using Kinetic Network Models. *Acc. Chem. Res.* **49:** 687–694

Zhang Y, Mooney RA, Grass J, Sivaramakrishnan P, Herman C, Landick R & Wang JD (2014) DksA Guards Elongating RNA Polymerase against Ribosome-Stalling-Induced Arrest. *Mol. Cell* **53:** 766–778

Zhilina E, Esyunina D, Brodolin K & Kulbachinskiy A (2012) Structural transitions in the transcription elongation complexes of bacterial RNA polymerase during sigma-dependent pausing. *Nucleic Acids Res.* **40:** 3078–3091

Zhou J, Ha KS, La Porta A, Landick R & Block SM (2011) Applied Force Provides Insight into Transcriptional Pausing and Its Modulation by Transcription Factor NusA. *Mol. Cell* **44:** 635–646

Zhou YN, Lubkowska L, Hui M, Court C, Chen S, Court DL, Strathern JN, Jin DJ & Kashlev M (2013) Isolation and characterization of RNA polymerase rpoB mutations that alter transcription slippage during elongation in Escherichia coli. *J. Biol. Chem.* **288:** 2700–2710

ORIGINAL STUDIES I-VII

Annales Universitatis Turkuensis

Turun yliopisto
University of Turku

Juvenes Print, Helsinki, Finland 2018

ISBN 978-951-29-7274-6 (print)
ISBN 978-951-29-7275-3 (pdf)
ISSN 0082-7002