

**The interplay between polymerase organization and nucleosome occupancy along DNA
How dynamic roadblocks on the DNA induce the formation of RNA polymerase pelotons**

van den Berg, Aafke

DOI

[10.4233/uuid:993e98ca-3c91-4591-9fbf-26bd6eea2354](https://doi.org/10.4233/uuid:993e98ca-3c91-4591-9fbf-26bd6eea2354)

Publication date

2017

Document Version

Final published version

Citation (APA)

van den Berg, A. (2017). *The interplay between polymerase organization and nucleosome occupancy along DNA: How dynamic roadblocks on the DNA induce the formation of RNA polymerase pelotons*. [Dissertation (TU Delft), Delft University of Technology]. <https://doi.org/10.4233/uuid:993e98ca-3c91-4591-9fbf-26bd6eea2354>

Important note

To cite this publication, please use the final published version (if applicable).
Please check the document version above.

Copyright

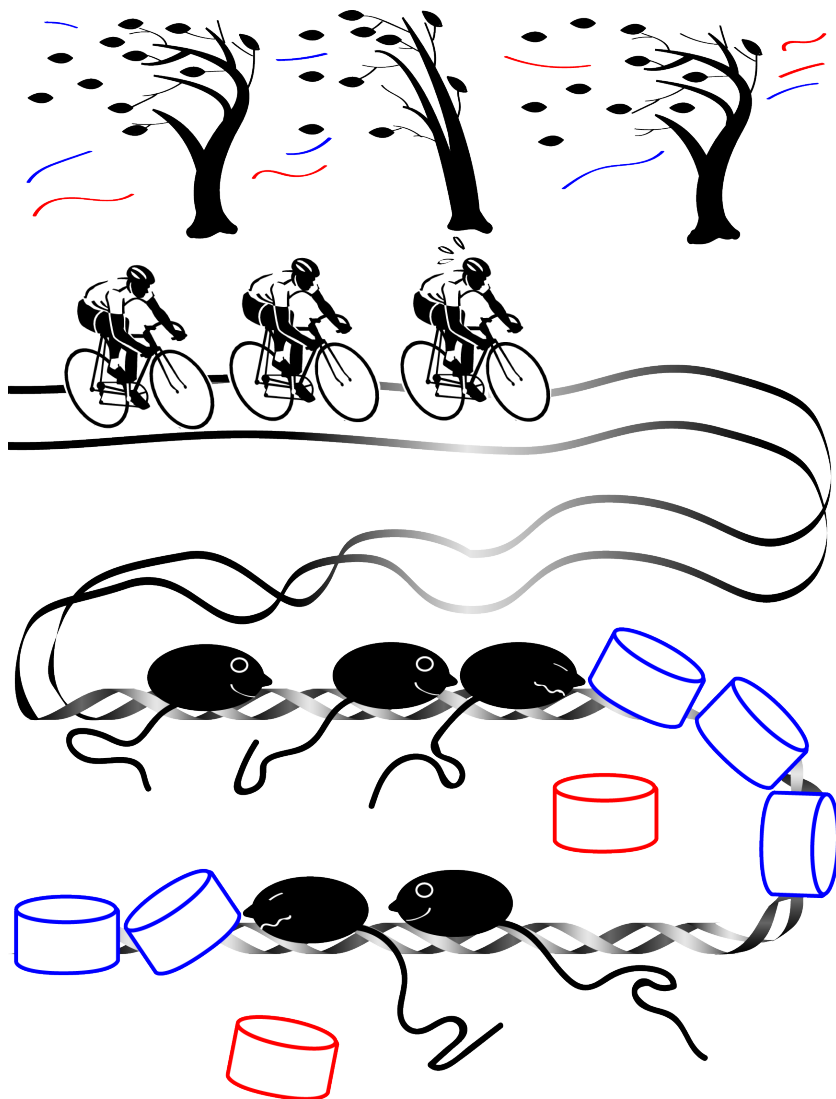
Other than for strictly personal use, it is not permitted to download, forward or distribute the text or part of it, without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license such as Creative Commons.

Takedown policy

Please contact us and provide details if you believe this document breaches copyrights.
We will remove access to the work immediately and investigate your claim.

THE INTERPLAY BETWEEN POLYMERASE ORGANIZATION AND NUCLEOSOME OCCUPANCY ALONG DNA

HOW DYNAMIC ROADBLOCKS ON THE DNA INDUCE THE FORMATION OF RNA POLYMERASE PELOTONS



THE INTERPLAY BETWEEN POLYMERASE ORGANIZATION AND NUCLEOSOME OCCUPANCY ALONG DNA

HOW DYNAMIC ROADBLOCKS ON THE DNA INDUCE THE
FORMATION OF RNA POLYMERASE PELOTONS

Proefschrift

ter verkrijging van de graad van doctor
aan de Technische Universiteit Delft,
op gezag van de Rector Magnificus prof. ir. K.C.A.M. Luyben,
voorzitter van het College voor Promoties,
in het openbaar te verdedigen op vrijdag 29 september 2017 om 12:30 uur

door

Aafke Andrée VAN DEN BERG

Master of Science in Applied Physics
Technische Universiteit Delft
geboren te Rotterdam, Nederland.

Dit proefschrift is goedgekeurd door de

promotor: Prof. dr. N.H. Dekker

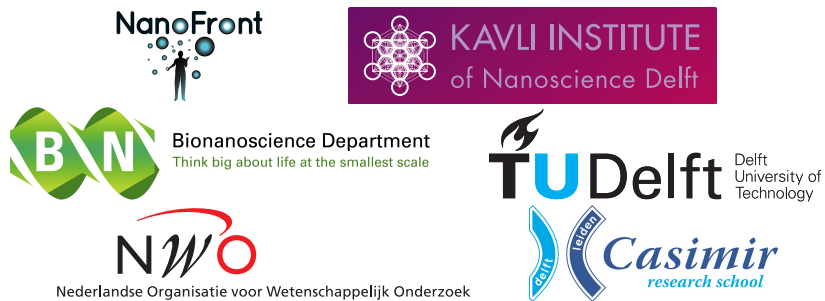
copromotor: Dr. S.M. Depken

Samenstelling promotiecommissie:

Rector Magnificus,	voorzitter
Prof. dr. N.H. Dekker	Technische Universiteit Delft
Dr. S.M. Depken	Technische Universiteit Delft

Onafhankelijke leden:

Prof. dr. M. Dogterom	Technische Universiteit Delft
Prof. dr. H. Schiessel	Universiteit Leiden
Prof. dr. P.R. ten Wolde	AMOLF
Prof. dr. A. Parmeggiani	Université de Montpellier
Dr. T. Idema	Technische Universiteit Delft
Prof. dr. ir. S. Tans	Technische Universiteit Delft en AMOLF reservelid



Keywords: Transcription, Nucleosomes, Crowding, TASEP,
Bus route model, Transcriptional bursts

Printed by: Gildeprint

Front and back: 'Peloton formation', by A.A. van den Berg

Copyright © 2017 by A.A. van den Berg

Casimir PhD series, Delft-Leiden 2017-24

ISBN 978-90-8593-308-3

An electronic version of this dissertation is available at

<http://repository.tudelft.nl/>.

*The ultimate aim of the modern movement in biology
is to explain all biology in terms of physics and chemistry*

Francis Crick

CONTENTS

Preface	xi
Summary	xiii
Samenvatting	xv
1 Introduction to transcription	1
1.1 The central dogma of molecular biology	2
1.2 Transcription through nucleosomes	3
1.3 Backtracking and RNA cleavage	4
1.4 Conclusion	5
References	6
2 Modeling transcription	9
2.1 Thermodynamic equilibrium	10
2.1.1 Local equilibrium	11
2.1.2 Far out of equilibrium and steady state	11
2.2 Transcription as a stochastic process	11
2.2.1 Master equation	13
2.2.2 The effect of force	14
2.3 Continuum time random walks	14
2.4 Monte Carlo simulations	14
2.5 The Totally Asymmetric Simple Exclusion Process	15
2.6 Conclusion	16
2.7 Thesis outline	17
References	18
3 Crowding-induced transcriptional bursts dictate polymerase and nucleosome density profiles along genes	19
3.1 Introduction	20
3.2 Results	21
3.2.1 A minimal model of motors interacting with roadblocks	21
3.2.2 Spontaneous formation of stable pelotons	22
3.2.3 A hierarchy of TASEPs control motor organization	23
3.2.4 A heuristic solution	24
3.2.5 Motor and roadblock reorganization on finite genes	25
3.2.6 From pelotons to bursts	27
3.2.7 Transcription on highly induced genes	28

3.3	Discussion	31
3.3.1	Peloton formation has been observed <i>in vivo</i>	31
3.3.2	Predicted density profiles agree with observations in yeast	31
3.3.3	Burst characteristics agree with <i>in vivo</i> observations.	32
3.3.4	DNA supercoils as a source of bursts.	33
3.3.5	Experimental testing and alternative models.	33
3.3.6	Conclusion and outlook	34
3.4	Supplementary material	35
3.4.1	Heuristic solution for of the hierarchical TASEP model.	35
3.4.2	Relation between heuristic arguments and mean-field solution of BRM	37
3.4.3	Observable bulk quantities.	37
3.4.4	Asymptotic behavior in the SPR	38
3.4.5	The bulk state is never reached	40
3.4.6	Initiation limited dynamics	40
3.4.7	Bursts from terminating pelotons	44
3.4.8	Monte Carlo Simulations.	46
	References	47
4	Principles of histone replacement in the wake of a transcribing RNAP	53
4.1	Introduction	54
4.2	Results	55
4.2.1	A general model for transcription through two competing DNA binding proteins	55
4.2.2	Rebinding dynamics of two competing proteins after pol passage	57
4.2.3	Physiological limits of the model.	57
4.2.4	RNAP peloton formation.	58
4.2.5	General solution for protein coverage during active transcription in the presence of one protein species	58
4.2.6	General solution for protein coverage during active transcription in the presence of two DNA binding proteins.	59
4.2.7	Transcription in the presence of competing DNA binders of equal size.	60
4.2.8	Transcription in the presence of proteins binding in multiple steps	61
4.2.9	Transcription in the presence of competing DNA binders of unequal size.	64
4.3	Discussion	66
4.3.1	Chromatin changes due to polymerase eviction are fast and specific	67
4.3.2	Increasing accessibility of DNA to CRISPR-Cas using transcription	67
4.3.3	Chromatin changes of ribosomal genes	68
4.3.4	Conclusion and outlook	68
4.4	Supplementary material	69
4.4.1	Modeling protein binding	69
4.4.2	Peloton formation	69
4.4.3	The average speed	70
4.4.4	Equilibration of two equally sized DNA binding proteins	71

4.4.5	Equilibrium coverage of competing proteins with different sizes. . .	73
4.4.6	Scaling of dissociation constant with protein size	73
4.4.7	Competition of proteins with unequal size.	74
	References	76
5	Transcription elongation factors modify nucleosome density and transcriptional bursts	81
5.1	Introduction	82
5.2	Results	83
5.2.1	A model for transcription with elongation factors	83
5.2.2	Velocity and pause density and duration for a single RNAP	83
5.2.3	Macroscopic effects of RNAP cooperation	86
5.2.4	The effect of elongation factors on transcription on nucleosome coverage and output dynamics.	86
5.3	Discussion	91
5.3.1	Elongation factors modify transcriptional bursts.	91
5.3.2	Pioneering polymerases	91
5.3.3	Heavily transcribed genes	92
5.3.4	Conclusion and outlook	92
5.4	Supplementary material	93
5.4.1	Average velocity of single RNAP with transcription factors and nucleosome remodelers	93
5.4.2	Pause duration	95
5.4.3	Monte Carlo simulations.	95
	References	98
6	Discussion and perspectives	103
6.1	Transcriptional bursts.	103
6.2	Specific gene targeting using transcription	104
6.3	The robustness of spontaneous processes	104
6.4	Optimizing bus traffic.	105
6.5	Phase transitions	105
	References	107
	Acknowledgements	109
	Curriculum Vitae	111
	List of Publications	113

PREFACE

How does life work? A simple, but intriguing question that triggers the imagination. Life has to follow the same physical laws as everything else in the universe. Yet, even a description of the smallest units of life, cells, is difficult, because the interior of the cell is complex and interconnected and cellular processes are noisy. Experimental studies of cellular processes in idealized environments have been fruitful approaches to tackle these problems. In this work, we use theoretical modeling to synthesize the experimentally obtained facts into predictions for a more complex environment that is closer to the situation in the cell, thereby guiding future experiments. Through many iterations between theoretical predictions and experiments we can bring the pieces of the puzzle called life together.

*Aafke Andrée van den Berg
Delft, March 2017*

SUMMARY

During transcription RNA polymerase (RNAP) moves along a DNA molecule to copy the information on the DNA to an RNA molecule. Many textbook pictures show an RNAP sliding along empty DNA. In reality however, it is crowded on the DNA and RNAP competes for space with many proteins such as other RNAPs and histones. Coverage of DNA by histones is essential for DNA protection and signaling. Yet, RNAP evicts histones during transcription, which then rebind quickly or are replaced by other proteins. How does crowding of RNAP and histones on the DNA affect transcription dynamics on the one hand, and how does transcription activity change the density and exchange of histones along the DNA on the other hand? Those are the central questions of this thesis.

The transfer of information from DNA to other molecules is central to every living system. The central dogma of molecular biology describes how information transfer is carried out in three processes: replication, transcription and translation (Chapter 1). During transcription, the focus of this thesis, RNAP binds to the start of a gene (initiation) and then slides along the gene, copying the information to an RNA molecule (elongation). Once the RNAP reaches the end of the gene, it unbinds and the RNA molecule is released (termination). During transcription elongation, RNAP meets many proteins such as histones that affect the transcription process. Single-molecule experiments have shown that histones along the DNA slow down RNAP and that RNAP can evict histones. Though it is known how a single or a pair of RNAPs interact with histones, it is not clear what the effect of these interactions are in a living cell where multiple RNAPs can transcribe a gene at the same time and a gene is occupied by many histones. Here, we try to understand these interactions in more detail using theoretical modeling and Monte Carlo simulations.

We theoretically conceptualized transcription as a Totally Asymmetric Simple Exclusion Process (TASEP) which describes motors that stochastically initiate to the beginning of a lattice (RNAP binding to the promoter), hop along the lattice (elongation) and terminate once they reach the end of the track (RNAP unbinds from DNA) (Chapter 2). We model transcription on crowded DNA by extending the TASEP to include roadblocks (histones) of arbitrary size that dynamically bind to DNA and are evicted by passing motors (RNAPs).

In Chapter 3 we find that multiple RNAPs spontaneously group into pelotons as they transcribe a gene and interact with obstacles like histones. This process is comparable to peloton formation during cycling races, where the cyclists (RNAPs) form pelotons to reduce the air resistance (the obstacle). The RNAP pelotons are stable as they move along a gene and result in bursty transcriptional output, even if initiation is not bursty. We analytically predict the peloton size, the RNAP and histone density profiles along a gene and the burst characteristics. The predictions are verified with Monte Carlo sim-

ulations. The predicted density profiles and transcription output dynamics agree with multiple *in vivo* observations, thereby unifying many experimental observations into a single mechanism of peloton formation.

Many studies have shown that histones are often replaced by more dynamic histone variants or other proteins on transcribed DNA. Understanding the underlying mechanisms for histone exchange is important as it relates to epigenetics and disease and it ensures better accessibility for RNAP on active genes. Multiple experiments have indicated that replacement of histones takes place in the wake of a transcribing RNAP that just evicted a histone. We therefore extended the model to include two roadblock species (Chapter 4). We derive the roadblock densities as a function of transcription levels and discuss three physiologically relevant examples: competing roadblocks of equal size (two histone variants), roadblocks binding in steps (histones) and roadblocks of different sizes (histones competing with HMG box protein). We find that transcription activity can lead to spontaneous replacement of a stable binding protein by a faster binding or smaller protein. The duration of such a memory of transcription activity depends on the unbinding rates and the size difference of the competing proteins. These predictions agree with the experimental observations that the initial replacement of histones is fast and tightly coupled to transcription, while the memory of transcription activity can last longer than the cell cycle. We give testable predictions and suggest experiments to further test the applicability of this model to transcription.

In Chapter 3, where we discuss the TASEP with one type of roadblock, and in 4, where we add a second type, we model transcription elongation with a single rate, but the elongation phase is actually interspersed with pauses: RNAP backtracks regularly to remove errors and also in response to obstacles along a gene. To establish whether or not RNAPs form pelotons under physiological conditions we need to test whether the pelotons are stable in the presence of backtracks and transcription factors that can modify the frequency and duration of these backtracks. Using Monte Carlo simulations we show that peloton formation is also expected when RNAPs backtrack (Chapter 5). Further we point to novel large scale effects of elongation factors: elongation factors and nucleosome remodelers cannot only change the nucleosome density on a gene by modifying the RNAP density (through the velocity), but also by reshaping the pelotons. Our work takes an important step along the road towards integrating what we know about transcription with the dynamic environment of the cell.

We conclude this thesis by discussing wider implications for molecular biophysics (transcriptional noise and epigenetics), traffic studies (optimizing traffic flows) and medicine (Chapter 6).

SAMENVATTING

Tijdens transcriptie kopieert RNA polymerase (RNAP) de genetische informatie van een DNA naar een RNA molecuul. Vaak wordt transcriptie afgebeeld als een enkele RNAP op een leeg DNA molecuul, maar in werkelijkheid is het druk op het DNA, bijvoorbeeld met andere RNAPs en andere eiwitten, zoals histonen. Tijdens transcriptie verwijdert RNAP de histonen van het DNA, waarna ze snel terugbinden, of vervangen worden door andere eiwitten. Hoe beïnvloedt de drukte op het DNA transcriptie en hoe beïnvloedt transcriptie de dichtheid van histonen langs het DNA? Dat zijn de centrale vragen van dit proefschrift.

De overdracht van genetische informatie van DNA naar andere moleculen is een centraal proces in het leven zoals we dat kennen. Het centrale dogma van de moleculaire biologie beschrijft hoe genetische informatie wordt overgedragen in drie processen: replicatie (het kopiëren van DNA), transcriptie en translatie (het maken van eiwitten met behulp van RNA) (Hoofdstuk 1). Tijdens transcriptie, waar dit proefschrift over gaat, bindt het enzym RNAP aan het DNA (initiatie) en beweegt vervolgens langs het DNA terwijl het genetische informatie naar een RNA molecuul kopieert (elongatie). RNAP en het RNA molecuul dissociëren van het DNA zodra ze het einde van het gen hebben bereikt (terminatie). Tijdens elongatie komt RNAP eiwitten tegen die het transcriptie proces beïnvloeden. Experimenten met enkele moleculen hebben bijvoorbeeld aangetoond dat histonen obstakels vormen langs het DNA en RNAP vertragen tijdens elongatie. Wanneer RNAP een histoon passeert, verwijdert deze een deel van het histoon. We hebben dus een goed beeld van hoe een enkele RNAP een enkel histoon van het DNA verwijdert, maar het is niet duidelijk wat het effect van deze interactie is op de situatie in de cel, waar meerdere RNAPs een gen aflezen dat bedekt is met vele histonen. In dit proefschrift proberen wij deze interacties beter te begrijpen met behulp van wiskundige modellen en Monte Carlo simulaties.

Om transcriptie op druk DNA te modelleren gebruiken we als basis het Totally Asymmetric Simple Exclusion Process (TASEP). Dit model wordt vaak gebruikt om transcriptie te modelleren en beschrijft motoren (RNAPs) die stochastisch aan het begin van een een-dimensionaal rooster binden (initiatie), die van roosterpunt naar roosterpunt springen (elongatie) en dissociëren zodra ze het einde van het rooster hebben bereikt (terminatie, Hoofdstuk 2). We modelleren transcriptie op DNA dat bedekt is met histonen door obstakels toe te voegen aan de TASEP, die stochastisch aan lege roosterpunten kunnen binden en verwijderd worden door passerende motoren.

In Hoofdstuk 3 tonen we aan dat RNAPs zich spontaan organiseren in pelotons wanneer ze DNA aflezen dat bedekt is met histonen. Dit proces is vergelijkbaar met peloton formatie tijdens de tour de France, waar wielrenners (de RNAPs) pelotons vormen om het effect van luchtweerstand (het obstakel) te minimaliseren. De RNAP pelotons zijn stabiel wanneer ze zich over het DNA bewegen en door de peloton formatie wordt het

RNA niet gelijkmatig, maar in schokken geproduceerd. We doen voorspellingen over de grootte van een peloton en de verdeling van RNAP en histonen langs het DNA. De voorspellingen komen overeen met meerdere *in vivo* observaties, die dus allemaal verklaard kunnen worden met hetzelfde mechanisme van peloton formatie.

Meerdere studies hebben aangetoond dat histonen tijdens transcriptie vaak vervangen worden door andere soorten histonen of door andere eiwitten. Het vervangen van histonen is belangrijk om genen beter toegankelijk te maken voor RNAP en als signaal aan andere eiwitten dat een gen geactiveerd is, een soort geheugen van transcriptie. Verder zijn problemen met het vervangen van histonen gerelateerd aan ziektes, zodat het belangrijk is dit proces beter te begrijpen. Experimentele studies hebben aangetoond dat de vervanging plaatsvindt in de kielzog van een RNAP die net een histoon van het DNA heeft verwijderd. Om dit proces te modelleren hebben we het model verder uitgebreid tot een TASEP met twee soorten obstakels (Hoofdstuk 4). We berekenen hoe de dichtheid van obstakels langs het DNA afhangt van transcriptie activiteit en bespreken drie fysiologische situaties: twee soorten obstakels die even groot zijn (bijvoorbeeld twee histoon varianten), obstakels die in twee stappen binden (histonen) en obstakels met verschillende groottes (bijvoorbeeld histonen en HMG box eiwitten). We tonen aan dat histonen tijdens transcriptie vervangen worden door eiwitten die sneller binden, of kleiner zijn en dat het geheugen van transcriptie afhangt van de relatieve grootte van de twee obstakels. Deze voorspellingen komen overeen met de observaties dat vervanging van histonen vaak heel snel gebeurt en direct gekoppeld is aan transcriptie, terwijl het geheugen van transcriptie veel langer duurt. We geven voorspellingen om verder te testen hoe goed dit model toepasbaar is op transcriptie.

In Hoofdstuk 3 and 4 is de verdeling van tijden om een stap te maken voor RNAP een simpele exponentiële functie. In werkelijkheid pauzeert RNAP regelmatig langs het DNA, zodat de verdeling een complexere vorm heeft. RNAP stapt bijvoorbeeld vaak achteruit om transcriptie fouten te verwijderen, of wanneer RNAP een obstakel tegenkomt. Dit type pauze heet een 'backtrack'. Om te voorspellen of RNAPs in de cel pelotons vormen moeten we deze backtracks meenemen in ons model, evenals transcriptie factoren en histoon modificaties die de frequentie en de duur van een backtrack veranderen. Met Monte Carlo simulaties tonen we aan dat RNAPs inderdaad pelotons vormen wanneer backtrack pauzes worden meegenomen (Hoofdstuk 5). Bovendien laten de simulaties nieuwe effecten zien van transcriptie factoren die tot nu toe genegeerd zijn. Transcriptie factoren en histoon modificaties kunnen namelijk de dichtheid van histonen op het DNA verhogen, niet alleen door de RNAP dichtheid te verlagen (door middel van een verhoogde RNAP snelheid), maar ook door pelotons compacter te maken. Deze resultaten laten zien, dat als we de effecten van transcriptie factoren en histoon modificaties goed willen begrijpen, ze in een fysiologische omgeving bestudeerd moeten worden.

Doordat het model algemeen is opgezet, zijn de voorspellingen toepasbaar op veel verschillende systemen. We eindigen dit proefschrift met een discussie van de implicaties van de resultaten voor de moleculaire biologie, het verkeer en de geneeskunde (Hoofdstuk 6). Pelotons zijn overal...

1

INTRODUCTION TO TRANSCRIPTION

Here we introduce the relevant biological concepts for this thesis. We start with the central dogma of molecular biology, which describes how information stored by a DNA molecule is transferred to DNA and RNA, for heredity and protein production respectively. Then we focus on one of the processes covered by the central dogma, transcription. Transcription is catalyzed by a protein called RNA polymerase (RNAP) that slides along the DNA, copying information from the DNA to an RNA molecule. We discuss the detailed dynamics of RNAP during transcription and how RNAP is hindered by proteins called histones, which structure and organize DNA. We conclude the chapter by discussing the highly complex dynamics of the transcription process in a living cell where multiple RNAPs can interact while encountering many histones.

1.1. THE CENTRAL DOGMA OF MOLECULAR BIOLOGY

A DNA molecule contains the blueprint for building a cell and the instructions for all cellular processes. In this section we discuss the structure of DNA and how the information it stores is transferred to daughter cells or to proteins, the workers and building blocks of the cell. A DNA molecule is a double helix formed by two interwoven strands with a sugar-phosphate backbone and four types of nucleobases attached to the backbone: Adenine (A), Cytosine (C), Thymine (T) and Guanine (G) (Figure 1.1). The backbone is asymmetric and one end of the strands is called the 3' end and the other the 5' end. The nucleobases represent the alphabet in which the blueprint and instructions for living systems are written, and they bind the two helices together by forming hydrogen bonds with nucleobases from the other strand, thereby forming base pairs. Thymine is always paired with Adenine and Guanine is always paired with Cytosine. The double stranded form of the DNA molecule allows for an elegant mechanism for copying the information on the DNA, namely by separating the two strands and forming base pairs with the exposed nucleobases. The information on the DNA can be transferred in two different processes: replication and transcription.

During replication, the DNA double helices are separated and a protein called DNA Polymerase catalyzes the binding of free nucleotides (= nucleobase + sugar-phosphate) to the exposed nucleotides, thereby forming copies of the original strand. Transcription is catalyzed by a protein called RNA polymerase (RNAP). During transcription the information on the DNA is copied to an RNA molecule, which has a chemical structure similar to DNA, but normally exist in a single-stranded form and the nucleobase T is replaced by Uracil (U) (Figure 1.1).

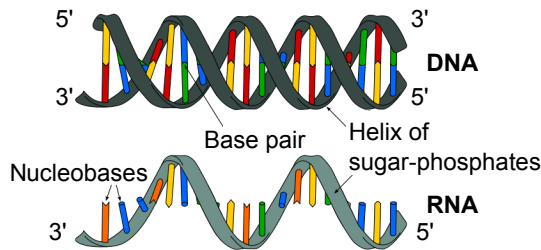


Figure 1.1: A schematic of a DNA and an RNA molecule showing the DNA and RNA backbone in gray and the nucleobases T (red), A (yellow), C (blue), G (green) and U (orange). As indicated in the figure, T always pairs with A and C always with G. The DNA strands are bound together in an anti parallel fashion, one strand is oriented from the 3' end to the 5' end, while the other strand is oriented from the 5' end to the 3' end. Adapted from [1].

During a third process of information transfer, translation, the information on RNA is used as a template to construct proteins. The replication of DNA and the transference of genetic information from DNA to RNA to proteins plays such a fundamental role in living systems that Francis Crick called it the Central Dogma of Molecular Biology [2] (Figure 1.2 A).

This thesis concentrates on one of the processes from the central dogma, transcription. During transcription, RNAP first binds to a particular sequence of DNA, called the

promoter, see Figure 1.2 B, often accompanied by other regulatory proteins called transcription factors, that help to start transcription. When RNAP binds to the promoter, the two helices of DNA are separated and RNAP binds to a single strand of DNA to form a transcription bubble (Figure 1.2 B middle panel). Then RNAP proceeds to the elongation phase where RNAP makes single-base pair steps along the DNA and selects nucleotides complementary to one of the two DNA strands. In the catalytic site of RNAP the complementary nucleotides are attached to an RNA molecule (Figure 1.2 B, middle panel). Just as for the initiation phase, there are many transcription factors that regulate the elongation phase [3], for example to modulate RNAP dynamics or correct transcriptional errors.

When RNAP reaches the end of the gene, the RNA molecule is released and RNAP unbinds from the DNA. This process is called the termination phase (lower panel of Figure 1.2 B). Though the above description is correct, Figure 1.2 only gives a highly simplified version of the transcription process, as will be explained in the rest of this chapter.

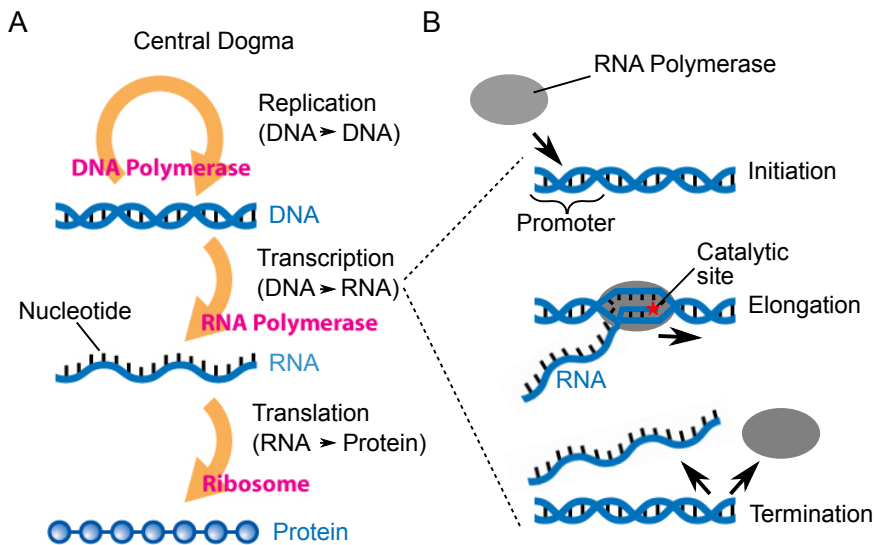


Figure 1.2: In formation flow in the cell. A) The central dogma of molecular biology: genetic information goes from DNA to RNA to proteins or between DNA molecules of mother and daughter cells, adapted from [4]. B) Schematic of the transcription process, adapted from [4].

1.2. TRANSCRIPTION THROUGH NUCLEOSOMES

After the discovery of DNA, many scientists assumed that the genetic information on the DNA is a complete blueprint for living systems. This idea nicely agrees with the evolution theory of Charles Darwin in 'The Origin of Species', that natural selection is the driving force of evolution: the DNA with its genetic material is passed on to the next generation, and partners with a better chance of survival, will more likely pass on their DNA. Now we know that not only the DNA, but also the spatial organization of the DNA is part of the blueprint for living systems and can be inherited. The organization of the DNA is therefore part of the 'epigenetics' of a cell, which is the heritable information other than the

genetic information [5]. Interestingly, the epigenetic state of a cell can be changed during its lifetime. As a consequence, not only natural selection, but also the environment drives evolution, such that Darwin's evolution theory was not entirely complete (and his colleague Lamarck was partially right, but that is a different story [6]). The spatial organization of DNA is regulated differently in prokaryotes (bacteria and archaea) and eukaryotes (for example yeast and human cells). In this thesis, we mainly concentrate on eukaryotes.

In eukaryotes, DNA is organized around cylindrical protein complexes called histones that compactify, protect and structure the DNA. DNA wrapped around a single histone is called a nucleosome and nucleosomes again fold into higher order structures (Figure 1.3 A). A histone consists of an H3/H4 tetramer and two H2A/H2B dimers that together form a histone octamer (Figure 1.3 B). There are many different ways in which histones can be modified. The histone tails attached to the octamer can for example be changed, parts of the histone can be exchanged [7], and there are many more histone modifications [8]. The wide diversity in histone modifications and their function in gene regulation led to the term 'histone code' additional to the genetic code [9]. Interestingly, the density of histones on a gene depends on the transcription intensity [10] and some histone modifications are tightly coupled to transcription [11, 12], which suggests that RNAP interferes with the organization of histones.

Indeed, as much as 80 % of the DNA is covered by histones [13], such that a transcribing RNAP will encounter many histones while transcribing. Histones form obstacles to RNAP: when encountering a histone along the DNA, RNAP slows down or stops completely, depending on the properties of the histone [14]. When RNAP passes through the histone, part of the histone is evicted. A single passing RNAP only evicts a dimer [15], while a second passing RNAP evicts the remaining hexamer (=tetramer+dimer) [16] (Figure 1.3 C) leaving behind bare DNA. Complete histones are only evicted on heavily transcribed genes where RNAPs are closely spaced [10]. Before and during passage through a histone, RNAP pauses for a long time [17] and the nature of these pauses (backtracks) is discussed in the next section.

1.3. BACKTRACKING AND RNA CLEAVAGE

Transcription elongation is interspersed by pauses, such that the time RNAP needs for adding a nucleotide is highly variable. Transcriptional pauses have many different functions such as regulating co-transcriptional processes and ensuring genome stability [19]. In this thesis, we are interested in pauses that are induced by the presence of obstacles that oppose forward motion of RNAP, such as histones. A common pause that is highly sensitive to opposing force is a backtrack, which is important for a host of regulatory processes [20]. A backtrack starts with backward motion of RNAP which then performs a random walk along the DNA [21]. As the RNAP steps backwards, the 3' end of the RNA protrudes from a channel in the front of RNAP and blocks the catalytic site such that the RNA molecule cannot be elongated, (Figure 1.4 B). When RNAP returns to the starting point of the backtrack and the catalytic site is accessible again, RNAP can continue transcription (Figure 1.4 A). Both the frequency and the duration of backtracks increase significantly when passing through a histone [17].

Instead of returning to the starting point of the backtrack, there is another mecha-

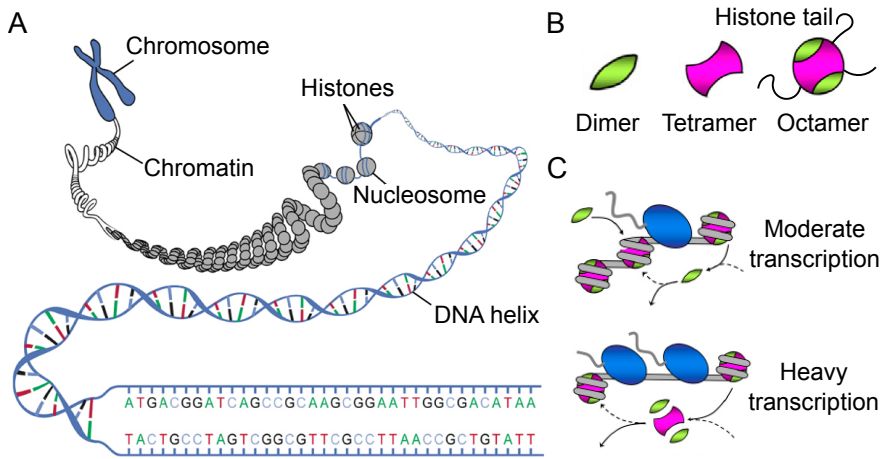


Figure 1.3: Higher order DNA structures and how they affect transcription A) DNA is wrapped around histones, which in turn form higher order structures resulting in a chromatin, adapted from [18]. B) The elements of a histone: two H2A/H2B dimers, and an H3/H4 tetramer together form a histone octamer, adapted from [16]. C) Eviction of histones by RNAP occurs in two steps. The first passing polymerase only evicts a histone dimer, while a second, closely spaced RNAP evicts the remaining hexamer, adapted from [16].

nism for RNAP to resume transcription: the transcription factor TFIIS catalyzes cleavage of the RNA that sticks out of the front channel (Figure 1.4 C), such that the catalytic site is exposed again and RNA elongation can continue [22]. TFIIS has multiple regulatory functions: TFIIS can remove transcriptional errors [23] and increases the ability of RNAP to resist an opposing force (Figure 1.4 D).

1.4. CONCLUSION

Over the years, much has been learned from *in vitro* studies in highly idealized, experimental settings. Single-molecule experiments have taught us about the effect of transcription factors on a single RNAP and interactions between one or two RNAPs and a nucleosome. However, for a complete model of transcription we need to take into account the presence of multiple histones and multiple RNAPs on a gene.

In this thesis, we combine what has been learned from different single-molecule experiments into a single model to study transcription on crowded DNA, taking into account the presence of obstacles like histones, the interaction between multiple polymerases and transcription factors. In the next chapter we discuss the theoretical framework that we used to model transcription.

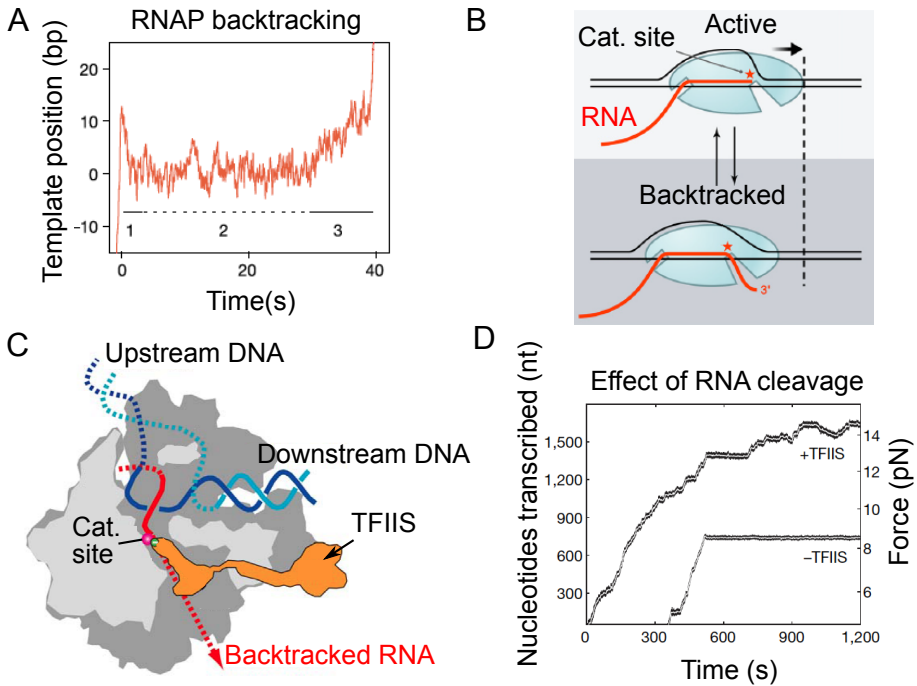


Figure 1.4: Backtracking and RNA cleavage. A) A trace of an RNAP in a backtrack, recorded using an optical tweezer [24]. Initially, the RNAP actively transcribes DNA, then the RNAP reverses direction and enters a backtrack, from which it escapes after about 40 seconds. B) Schematic of a backtracked RNAP, indicating an active RNAP with an accessible catalytic site and RNA and a backtracked RNAP where the catalytic site is blocked, adapted from [20]. C) TFIIS (orange) cleaving RNA inside the pore of a backtracked RNAP (gray), adapted from [25]. D) Single molecule traces of RNAP transcribing against a force. In the presence of TFIIS, RNAP can resist higher forces than without TFIIS [22].

REFERENCES

- [1] Wikipedia, The Free Encyclopedia, *Nucleic acid*, (2016).
- [2] F. H. Crick, *On protein synthesis*, Symposia of the Society for Experimental Biology **12**, 138 (1958).
- [3] R. J. Sims, R. Belotserkovskaya, and D. Reinberg, *Elongation by RNA polymerase II: the short and long of it*. *Genes & development* **18**, 2437 (2004).
- [4] Wikipedia, The Free Encyclopedia, *Central dogma of molecular biology*, (2016).
- [5] A. D. Goldberg, C. D. Allis, and E. Bernstein, *Epigenetics: a landscape takes shape*. *Cell* **128**, 635 (2007).
- [6] J. Parrington, *The deeper genome* (Oxford University Press, 2015).
- [7] C. Thiriet and J. J. Hayes, *Histone Dynamics During Transcription : Exchange of H2A / H2B Dimers and H3 / H4 Tetramers During Pol II Elongation*, (2006).

- [8] T. Kouzarides, *Chromatin modifications and their function*. *Cell* **128**, 693 (2007).
- [9] S. B. Rothbart and B. D. Strahl, *Interpreting the language of histone and DNA modifications*, *Biochimica et Biophysica Acta - Gene Regulatory Mechanisms* **1839**, 627 (2014), arXiv:NIHMS150003 .
- [10] H. A. Cole, J. Ocampo, J. R. Iben, R. V. Chereji, and D. J. Clark, *Heavy transcription of yeast genes correlates with differential loss of histone H2B relative to H4 and queued RNA polymerases*. *Nucleic acids research* **42**, 12512 (2014).
- [11] D. Ray-Gallet, A. Woolfe, I. Vassias, C. Pellentz, N. Lacoste, A. Puri, D. C. Schultz, N. A. Pchelintsev, P. D. Adams, L. E. T. Jansen, and G. Almouzni, *Dynamics of Histone H3 Deposition In Vivo Reveal a Nucleosome Gap-Filling Mechanism for H3.3 to Maintain Chromatin Integrity*, *Molecular Cell* **44**, 928 (2011).
- [12] B. E. Schwartz and K. Ahmad, *Transcriptional activation triggers deposition and removal of the histone variant H3.3*, *Genes and Development* **19**, 804 (2005).
- [13] W. Lee, D. Tillo, N. Bray, R. H. Morse, R. W. Davis, T. R. Hughes, and C. Nislow, *A high-resolution atlas of nucleosome occupancy in yeast*. *Nature genetics* **39**, 1235 (2007).
- [14] B. Li, M. Carey, and J. L. Workman, *The role of chromatin during transcription*. *Cell* **128**, 707 (2007).
- [15] M. L. Kireeva, W. Walter, V. Tchernajenko, V. Bondarenko, M. Kashlev, and V. M. Studitsky, *Nucleosome remodeling induced by RNA polymerase II: Loss of the H2A/H2B dimer during transcription*, *Molecular Cell* **9**, 541 (2002).
- [16] O. I. Kulaeva, F.-K. Hsieh, and V. M. Studitsky, *RNA polymerase complexes cooperate to relieve the nucleosomal barrier and evict histones*. *Proceedings of the National Academy of Sciences* **107**, 11325 (2010).
- [17] L. Bintu, T. Ishibashi, M. Dangkulwanich, Y.-Y. Wu, L. Lubkowska, M. Kashlev, and C. Bustamante, *Nucleosomal elements that control the topography of the barrier to transcription*. *Cell* **151**, 738 (2012).
- [18] Textbook Equity College Edition, *Anatomy and Physiology Volume 1 of 3* (Rice University, 2013).
- [19] I. Jonkers and J. T. Lis, *Getting up to speed with transcription elongation by RNA polymerase II*. *Nature reviews. Molecular cell biology* **16**, 167 (2015).
- [20] E. Nudler, *RNA polymerase backtracking in gene regulation and genome instability*. *Cell* **149**, 1438 (2012).
- [21] M. Depken, J. M. R. Parrondo, and S. W. Grill, *Intermittent transcription dynamics for the rapid production of long transcripts of high fidelity*. *Cell reports* **5**, 521 (2013).

- [22] E. A. Galburt, S. W. Grill, A. Wiedmann, L. Lubkowska, J. Choy, E. Nogales, M. Kashlev, and C. Bustamante, *Backtracking determines the force sensitivity of RNAP II in a factor-dependent manner*. [Nature](#) **446**, 820 (2007).
- [23] D. a. Erie, O. Hajiseyedjavadi, M. C. Young, and P. H. von Hippel, *Multiple RNA polymerase conformations and GreA: control of the fidelity of transcription*. [Science \(New York, N.Y.\)](#) **262**, 867 (1993).
- [24] J. W. Shaevitz, E. a. Abbondanzieri, R. Landick, and S. M. Block, *Backtracking by single RNA polymerase molecules observed at near-base-pair resolution*. [Nature](#) **426**, 684 (2003).
- [25] H. Kettenberger, *Structure of the Complete RNA Polymerase II Elongation Complex and its Interaction with the Elongation Factor TFIIS*, Ph.D. thesis (2005).

2

MODELING TRANSCRIPTION

Modeling cellular processes is challenging: proteins are constantly wiggling and jiggling due to thermal motion and many processes are out of equilibrium. This chapter explains how we still can model cellular processes such as transcription using concepts from thermodynamics, non-equilibrium physics and mathematics.

2.1. THERMODYNAMIC EQUILIBRIUM

My desk evolved from empty, at the start of my PhD, to a pile of mess in the final months. In general, spontaneous processes around us seem to lead to more disorder. How is it possible that living systems stay ordered? One way of looking at this question is using a quantity from thermodynamics, the entropy. The entropy is a measure for the number of microscopic configurations of a system (for example the number of configurations of the paper work on my desk) that is consistent with some observable macroscopic state of the system (for example the macroscopic state 'tidy' or 'messy'). That disorder usually increases over time is implied by the second law of thermodynamics, which states that an isolated system evolves through spontaneous processes to a state where the entropy is maximized. Since there are many more paper configurations in which my desk looks messy than that my desk looks tidy, the state of maximal entropy for my desk is a messy one¹.

When the system is not isolated because of exchange of energy with the surroundings, the Gibbs free energy G is used instead of the entropy. The Gibbs free energy is also referred to as useful energy and is defined as

$$G = E + W - TS \quad (2.1)$$

where E is the internal energy, W is work, S the entropy and T the temperature. The Gibbs free energy is always minimized by spontaneous processes [1]. When the total energy and the temperature are constant, and there is no work done by the system, a minimal free energy implies a maximal entropy (Equation 2.1). A system is in equilibrium when its free energy is minimal and the probability currents into and out of a microstate balance each other. When probability currents are pairwise balanced, the system is said to obey detailed balance. Systems in thermodynamic equilibrium are nice to study, as we always know that their probability distribution follows from the Boltzmann distribution and the probability P_i for the system to be in the microscopic state i with energy E_i is given by the Boltzmann weight

$$P_i \propto e^{-E_i/k_B T}, \quad (2.2)$$

where k_B is Boltzmann's constant. As an example, Figure 2.1 A shows a transcription process where RNAPs bind from solution to a promoter, then transcribe to the end of the gene and return to the solution, or take the same path backward along the gene. The system in this example obeys detailed balance, because all the probability currents are exactly balanced by the opposite reaction, as is illustrated by the equal sizes of the arrows. However, in the cell detailed balance would be a rather unfavorable state as there would be no net production of RNA. In a living cell, the transcription process is therefore driven out of equilibrium.

Living systems as a whole are also not in equilibrium. They are not in thermal equilibrium with their environment (humans have a higher temperature than the surroundings) nor in chemical equilibrium (not every chemical reaction is cancelled by its reverse reaction) [1], because living systems are continuously changing useful energy from sunlight and nutrients into heat and entropy in the surroundings. However, some cellular

¹The example of the messy desk is highly simplified and should not be used as an excuse for not cleaning up.

processes are in local equilibrium, which allows us to use concepts from equilibrium physics.

2.1.1. LOCAL EQUILIBRIUM

Though cellular processes never fully equilibrate, some can be approximated as equilibrated when they relax much faster than other processes. A system is then said to be in a local equilibrium, or a metastable state. For example, when the papers on my desk move around much quicker than the books, it could be that the papers have reached a very likely configuration given the configuration of books, while the configuration of books has not. Similarly, the position of RNAP at a certain nucleotide along the DNA is a metastable state, as small thermal fluctuations around the position are much faster than that RNAP hops to the next site.

The free energy of (meta)stable states and the energetic barriers between them can be visualized in a free energy diagram. Figure 2.1 B shows the energetic states of RNAP for catalyzation and backtracking. Each minimum in the landscape is a metastable state and corresponds to RNAP occupying a certain position along the DNA. The rightmost minimum corresponds to the active state of RNAP and the other two minima correspond to one or two steps into the backtrack. The barriers between the metastable states are energetic barriers that RNAP has to overcome to hop from one nucleotide to the next. In other words, the movement of RNAP during elongation can be described as discrete steps from one locally equilibrated state to another.

2.1.2. FAR OUT OF EQUILIBRIUM AND STEADY STATE

The position of RNAP on DNA is a metastable state, but transcription as a whole is far out of equilibrium, because there constantly is a flow of RNAPs from the promoter, to the end of the gene, to the solution and back (Figure 2.1 C). If the probability to be in each of the three states (promoter, end gene and solution) is constant over time while there is a net flow (no detailed balance), transcription is said to be in non-equilibrium steady state. The constant properties of steady state provide an opportunity for a theoretical description.

Transcription can thus be described as a mix of local equilibrium and non-equilibrium processes. For equilibrium processes, thermodynamics and (equilibrium) statistical physics provide a wide range of tools. Non-equilibrium processes in biology can often be modeled as a stochastic process [2] where events are not deterministic, but random. In Section 2.2 we explain how transcription can be described as a stochastic process.

2.2. TRANSCRIPTION AS A STOCHASTIC PROCESS

One process that makes the microscopic world in the cell so different from the macroscopic world that we know is Brownian motion. RNAP is bombarded from all directions by smaller molecules whose average kinetic energy is related to temperature as described by the equipartition theorem. These bombardments make that everything in the cell is constantly wiggling and jiggling, undergoing Brownian motion. The mean square displacement of an object undergoing Brownian motion in one dimension is given by

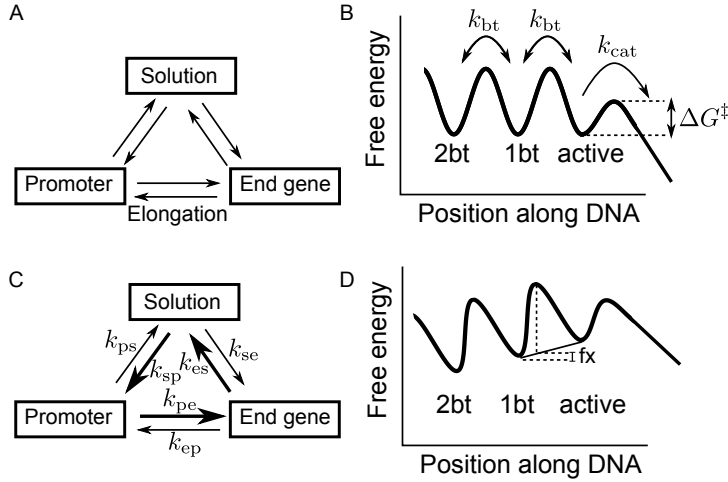


Figure 2.1: A) A schematic diagram of transcription. The state 'Promoter' corresponds to RNAP bound to a promoter, the state 'end gene' corresponds to RNAP at the end of a gene and the state 'solution' to RNAP that is unbound from DNA. This system obeys detailed balance, which is indicated by that the probability currents (the arrows) are all pair wise balanced. B) Free energy landscape for transcription elongation. The three metastable states represent an active RNAP (active), an RNAP that is one step into the backtrack (1bt) and an RNAP that is two steps into the backtrack (2bt). The rates with which RNAP jumps over the energy barriers are also indicated and they are related to the height of the energy barrier according to Equation 2.4. C) This diagram has the same states as in A), but the probability currents do not balance each other, such that the system does not obey detailed balance. D) This energy landscape contains the same states as in B), but is tilted due to an external force, f . The variable x is the distance of the minimum in the metastable state to the peak of the energy barrier

$$\langle \Delta x^2 \rangle = 2Dt \quad (2.3)$$

where $\langle \Delta x^2 \rangle$ is the mean square displacement, D the diffusion coefficient and t the time. The diffusion coefficient can be related to the drag coefficient γ through the Einstein relation $D = k_B T / \gamma$. The drag coefficient for a spherical object with radius r moving through a fluid with viscosity η is $\gamma = 6\pi\eta r$. For macroscopic objects, such as a human, the mean square displacement that follows from Equation 2.3 is less than the radius of an atom per second ($k_B T \approx 4.1$ pN nm, $\eta_{\text{air}} \approx 2 \cdot 10^{-5}$ Pa s, $a_{\text{human}} \approx 0.5$ m $\rightarrow \langle \Delta x^2 \rangle / t \approx 4 \cdot 10^{-11}$ m²/s). Therefore thermal noise is negligible in the macroscopic world.

In the microscopic world inside the cell, displacements by thermal noise are not negligible. In fact, many cellular processes, for example transcription, would not happen without thermal noise. RNAP relies on thermal motion to diffuse to the promoter and hop along the DNA during the elongation phase. In order to move from one nucleotide position to the next, RNAP has to cross a free energy barrier with a certain height, ΔG^\ddagger . Thermal fluctuations bring RNAP to an excited state with a certain energy, eventually pushing it over the barrier. When RNAP fluctuates many times before making a step, we can assume the RNAP to be equilibrated between excited states. From the Maxwell Boltzmann distribution we know the probability for an equilibrated particle to have at least an energy ΔG^\ddagger is proportional to

$$P \propto e^{-\Delta G^\ddagger/k_B T}, \quad (2.4)$$

where the Gibbs free energy ΔG^\ddagger is given in units of $k_B T$. Equation 2.4 is called the Arrhenius factor [2].

The overall rate forward of RNAP depends on the frequency of attempts to cross the energy barrier. An RNAP is bombarded from all directions by surrounding molecules, sometimes pushing the polymerase forward, resulting in an attempt to cross the barrier. Let p be the probability that RNAP has not crossed the barrier. In a well mixed environment and when attempts in different time intervals are independent, the change in probability dp in a time dt is proportional to the probability p , $dp = kpdt$, where k is a constant and dt is small. The evolution of the probability p can be described by the following differential equation

$$\frac{dp}{dt} = kp \quad (2.5)$$

Equation 2.5 is an example of a simple master equation, which is further explained below. The solution of Equation 2.5 is an exponential function, $p \propto e^{-kt}$ and the constant k is called the attempt rate k . In this case, the durations of the steps that RNAP takes are exponentially distributed with a rate given by the product of the probability to have a high enough energy (Equation 2.4) and the attempt rate k

$$k_{\text{tot}} = ke^{-\Delta G^\ddagger/k_B T}. \quad (2.6)$$

The attempt rate depends on the diffusion constant and the details of the process [2]. In this thesis we are mostly interested in the total rates k_{tot} , Equation 2.6.

To summarize this section, due to thermal noise, the time for making a step can be described as an exponentially distributed stochastic variable with a characteristic rate k_{tot} . The study of stochastic processes is a field on itself and there is a wide range of mathematical theories available to study them, such as the master equation.

2.2.1. MASTER EQUATION

The master equation is a differential equation that describes the evolution of a probability $P_n(t)$ that the system is in a state n at time t . Take for example the kinetic scheme in Figure 2.1 C. The evolution of the probability (or fraction) of RNAP at the promoter is given by

$$\frac{dP_{\text{Promoter}}}{dt} = k_{\text{sp}}P_{\text{Solution}} + k_{\text{ep}}P_{\text{Endgene}} - (k_{\text{ps}} + k_{\text{pe}})P_{\text{Promoter}}, \quad (2.7)$$

where k_{nm} is the rate (transition probability per unit time) at which the system changes from state n to state m . The master equations for the probability for RNAP to be in solution, P_{Solution} , or at the end of a gene, P_{Endgene} have the same form as in Equation 2.7, giving a system of coupled differential equations. When the system is in steady state the probabilities are constant over time, $dP_n/dt = 0$, and can be solved for in terms of the transition rates.

2.2.2. THE EFFECT OF FORCE

The presence of a force opposing transcription, for example an obstacle like a histone on the DNA, increases the height of the energy barrier for forward motion, because RNAP has to do extra work. The work done by RNAP when exerting a force f over a distance x is $W = x \cdot f$. The work done by RNAP results in an increase of the free energy (Equation 2.1), which can be visualized as a tilt of the free energy diagram (Figure 2.1 D). The decrease of the hopping rate over an energy barrier due to the work that has to be performed depends on the distance from the the minimum in the metastable state to the peak of the energy barrier. If this distance is x then the hopping rate over the barrier changes to $k_{\text{tot}} = ke^{-(\Delta G^\ddagger + fx)}$. In Figure 2.1 D the applied force results in a decreased forward rate and an increased backward rate.

2.3. CONTINUUM TIME RANDOM WALKS

The average speed for transcription can be calculated using the theory for continuum time random walks (CTRW) [3]. This theory considers walkers, such as RNAP, that make discrete steps, but the dwell time (the time to make a step) is a continuous, stochastic variable. We here make the simplifying assumption that dwell times for RNAP are distributed exponentially: $P_1(t) = (\sum k_i) e^{-t(\sum k_i)}$, where t is the waiting time and $\sum k_i$ the sum of the rates leaving a state.

Often, one needs to know the dwell time distribution for making multiple steps. This is given by the convolution of single steps, for example, the dwell time distribution for making three steps is given by $P_3(t) = (P_1 * P_1 * P_1)(t)$. Rather than calculating the convolution, it is simpler to go to Laplace space where convolutions become products. Let Ψ_1 be the Laplace transform of the dwell-time distribution to make one step, $\Psi_1(s) = \int_{t=0}^{\infty} P_1(t) e^{-st} dt$. The Laplace transform of the dwell time distribution to make three steps is simply given by $\Psi_3(s) = \Psi_1(s)^3$. The average dwell time can also be easily obtained from the Laplace transform. The average waiting time is given by $\tau = -1/\Psi \partial \Psi / \partial s |_{s=0}$ and the probability for a transition to happen at any time is $\int_{t=0}^{\infty} P(t) dt = \Psi(s=0)$.

The CTRW allows us to calculate dwell time distributions and average dwell times for complex kinetic diagrams. In Chapter 5 we elaborate on the CTRW to calculate the average dwell time for RNAP including backtracking, histones and transcription factors.

2.4. MONTE CARLO SIMULATIONS

As a numerical approach to studying stochastic processes, Monte Carlo simulations are often used to simulate biological systems. We used Monte Carlo simulations with fixed time steps for every iteration. Though this approach does not give exact dwell time distributions it is accurate enough for our purposes and it has the advantage that time is a simple linear function of the number of iterations, such that the evolution of the process can be plotted. Dwell time distributions for fixed time step Monte Carlo become more accurate for smaller time steps. When waiting times are exponentially distributed, and the time step per iteration is small enough, the probability that an event is happens in time interval can be approximated by

$$P(\Delta t) \approx k\Delta t, \quad k\Delta t \ll 1. \quad (2.8)$$

The time steps for the simulations in this thesis are chosen such that $k_{\max}\Delta t \approx 0.1$, with k_{\max} the maximum rate in the system. Sometimes, when multiple events can happen (for example backtracking and elongation of a polymerase), k_{\max} is a sum of all the rates leaving a node. During every iteration step of time Δt the lattice is updated random sequentially: all lattice sites are updated once in random order. For every lattice site, a random number a is drawn from a uniform distribution between 0 and 1, and the event happens if $a < k\Delta t$.

A more elaborate discussion on simulating transcription using fixed time step Monte Carlo is given in Chapter 5.

2.5. THE TOTALLY ASYMMETRIC SIMPLE EXCLUSION PROCESS

Transcription can be described as RNAPs that stochastically bind to a promoter and then hop with a certain rate from one nucleotide to the next, until they reach the end of the gene and unbind from the gene (Figure 2.2 A). This process can be modeled as an Asymmetric Simple Exclusion Process (ASEP).

The ASEP was introduced half a century ago to describe translation [4] of mRNA by ribosomes. In physics, the ASEP is widely studied, because it is one of the few non-equilibrium models that has been solved exactly [5] and shows rich dynamics with boundary induced phase transitions. The ASEP describes motors that initiate with rate k_{in} to the first site of a one-dimensional lattice, hop forward with rate k_{f} , backward with rate $k_{\text{b}} < k_{\text{f}}$ and terminate from the last site of the lattice with rate k_{ter} . The word 'Exclusion' indicates that motors cannot overlap. Since a polymerase transcribes DNA in one direction, we only consider forward hopping ($k_{\text{b}} = 0$), such that the ASEP reduces to the Totally Asymmetric Simple Exclusion Process (TASEP), see Figure 2.2 A.

A complete solution for the TASEP gives the probability for every configuration of motors along the lattice. In a mean-field approximation, gaps between motors $\{g_i\}$ are uncorrelated, such that the probability for a certain configuration motors $P(\{g_i\})$ can be written as a product of probabilities

$$P(\{g_i\}) = \frac{1}{Z} P(g_1)P(g_2)\dots \quad (2.9)$$

Here Z is a normalization factor and $P(g_i)$ is the probability for a gap of size g_i between motors. When ρ is the density of motors one empty site occurs with probability $1 - \rho$ and the occurrence of g neighbouring empty sites has probability $(1 - \rho)^g$, such that the distribution of gaps between motors is geometrically distributed as [5]

$$P(g) = \rho(1 - \rho)^g \quad (2.10)$$

A motor can only hop if the next site is empty, such that the average velocity is given by $v = k_{\text{f}}(1 - \rho)$ and the flux of motors, $J = v\rho$, is given by

$$J = k_{\text{f}}\rho(1 - \rho), \quad (2.11)$$

which is plotted in Figure 2.2 B. For the TASEP, the mean-field current-density relation in Equation 2.11 turns out to be exact [5]. However, the mean-field solution does not capture all the different behaviors of the exact solution, such as the width and dynamics of shocks, for further details we refer to [5, 6].

In open systems, the density of motors is set by the initiation and termination rate. The TASEP with open boundary conditions has three different phases, each corresponding to a different part of the flux density relation. In the initiation-limited phase (low density phase), the initiation rate limits the flux into the system, such that the flux is given by $J = k_{\text{in}}(1 - \rho)$. This flux should match the flux in the bulk (Equation 2.11). The density in the system in the initiation limited regime is therefore given by $\rho = k_{\text{in}}/k_{\text{f}}$. In the termination limited regime (high density phase), the density is only a function of the termination rate and can be determined from the flux leaving the system: $J = k_{\text{ter}}\rho$ giving $\rho = 1 - k_{\text{ter}}/k_{\text{f}}$. In the maximal current phase (bulk limited phase), the flux reaches its maximal value and the corresponding density, $\rho = 0.5$ (Figure 2.2 B), is independent of the initiation or termination rate. Every phase corresponds to a distinct part of the flux density curve, which is indicated in Figure 2.2 B and the phase diagram is shown in Figure 2.2 C. RNA production rate is usually a function of the initiation rate only [7], indicating that transcription is in the initiation-limited regime.

As a model for directed transport along a one-dimensional track, extended versions of the TASEP has been applied to many (biological) systems: translation, transcription [8], kinesin along microtubules and even ant traffic [9]. One of these extensions considers buses picking up passengers [10], the Bus Route Model (BRM), see Figure 2.2 D. The bus drives from one bus stop to the next with rate k_{f} . Passengers arrive at the bus stop with rate k_{b} and the bus has to stop for waiting passengers, which slows the bus down to a rate $k_{\text{f}}^* < k_{\text{f}}$. In a mean field approximation, this model can be mapped onto a zero-range process [10]. The BRM explains why buses have a tendency to cluster along a bus route and even has a jamming transition for $k_{\text{b}} \rightarrow 0$ where all buses form one stable jam moving with rate k_{f} . Though bus traffic and transcription have little to do with each other, the BRM can, with some modifications, describe polymerases (the buses) transcribing DNA while interacting with histones (the passengers), which is discussed in the next chapter.

2.6. CONCLUSION

This chapter summarized the framework and the techniques used to study transcription theoretically. We started by recognizing that each nucleotide position along the DNA is a metastable state for RNAP and that RNAP dynamics can be modeled as hopping from one nucleotide to the next with an exponential rate. Transcription initiation, elongation and termination can be coupled together and modeled as a TASEP, a widely studied model from non-equilibrium physics. The Bus Route Model is an extension of the TASEP that, after some modifications, can be used to model transcription in the presence of obstacles like histones.

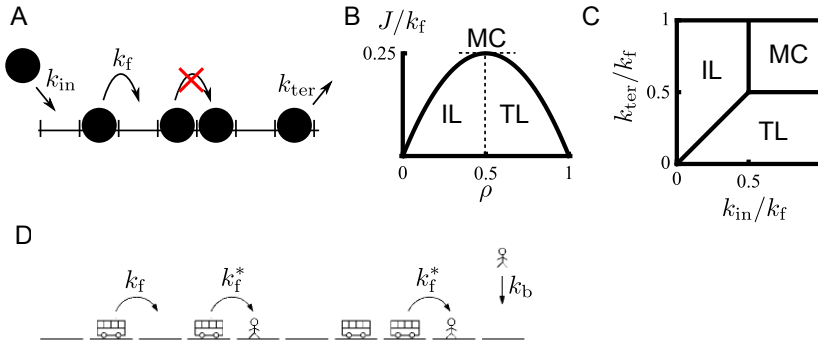


Figure 2.2: The TASEP. A) The TASEP with all the parameters as defined in the text. B) The flux of motors in the TASEP as a function of the motor density, Equation 2.11. The figure also indicates which part of the flux-density curve corresponds to which phase: the initiation-limited phase (IL), termination-limited phase (TL) or the maximal current phase (MC). C) The phase diagram for the TASEP. D) BRM adapted from [10].

2.7. THESIS OUTLINE

We now established both the biological and theoretical background necessary to read this thesis. In the following chapters we take a bottom-up approach in trying to understand transcription on crowded DNA taking into account interactions between RNAPs and between RNAP and histones.

In Chapter 3 we study a model for transcription of multiple polymerases on a gene covered with nucleosomes. So far the BRM was only studied for periodic boundary conditions and for buses and passengers of size 1, while both RNAP and histones occupy multiple nucleotides. We therefore study the initiation-limited regime and allow motors and roadblocks to occupy more than one lattice site. We predict that polymerases spontaneously self organize into pelotons due to non-specific interactions with nucleosomes. The formation of pelotons along a gene results in non-homogeneous RNAP and nucleosome density profiles and transcriptional bursts, agreeing with multiple experimental observations.

Many *in vivo* studies have shown that the composition of histone species on a gene changes after transcription activation. In Chapter 4 we extend the model to include two roadblock species. We predict that histones are replaced by faster binding or smaller proteins during transcription. The results agree with multiple *in vivo* observations on transcription-coupled histone replacements.

Backtracking and transcription factors were not considered in Chapters 3 and 4. In Chapter 5 we show that peloton formation is expected, even in the presence of backtracking. Further we show that elongation factors can modify the density of nucleosomes on a gene, both by changing the speed of RNAP and by reorganizing pelotons along a gene.

REFERENCES

- [1] P. Nelson, *Biological Physics*, Vol. 3rd (W.H. Freeman and Company, 2004).
- [2] N. G. Kampen, *Stochastic Processes in Physics and Chemistry*, 3rd ed. (Elsevier, 2007).
- [3] M. Depken, J. M. R. Parrondo, and S. W. Grill, *Intermittent transcription dynamics for the rapid production of long transcripts of high fidelity*. *Cell reports* **5**, 521 (2013).
- [4] C. T. MacDonald, J. H. Gibbs, and A. C. Pipkin, *Kinetics of Biopolymerization on Nucleic Acid Templates*, *Biopolymers* **6**, 1 (1968).
- [5] B. Derrida, M. R. Evans, V. Hakim, and V. Pasquier, *Exact solution of a 1D asymmetric exclusion model using a matrix formulation*, *Journal of Physics A: Mathematical and General* **26**, 1493 (1993).
- [6] R. A. Blythe and M. R. Evans, *Nonequilibrium steady states of matrix-product form: a solver's guide*, *Journal of Physics A: Mathematical and Theoretical* **40**, R333 (2007).
- [7] G. M. Cooper, *The Cell: A Molecular Approach. 2nd edition* (Sinauer Associates, 2000).
- [8] S. Klumpp, *Pausing and Backtracking in Transcription Under Dense Traffic Conditions*, *Journal of Statistical Physics* **142**, 1252 (2011).
- [9] A. Kunwar, A. John, K. Nishinari, A. Schadschneider, and D. Chowdhury, *Collective traffic-like movement of ants on a trail: dynamical phases and phase transitions*, *Journal of the Physical Society of Japan* **73**, 2979 (2004).
- [10] O. J. O. Loan, M. R. Evans, and M. E. Cates, *Jamming transition in a homogeneous one-dimensional system: The bus route model*, *Physical Review E* **58**, 1404 (1998).

3

CROWDING-INDUCED TRANSCRIPTIONAL BURSTS DICTATE POLYMERASE AND NUCLEOSOME DENSITY PROFILES ALONG GENES

During transcription, RNA polymerase competes for space with other DNA binding proteins and higher order DNA structures acting as roadblocks. The effect of such roadblocks on the transcriptional time series and the nucleosome/polymerase organization has not been investigated. Without understanding the local organization of polymerases and nucleosomes it will remain a challenge to understand the action of transcription factors regulating the elongation phase. Based on quantitative theoretical modeling, we here show that interactions with roadblocks induce a strong kinetic attraction between polymerases, causing them to self-organize into stable and moving pelotons. This peloton formation explains observed nucleosome and polymerase density profiles close to the initiation site on highly transcribed genes, and how these densities depend on induction levels. At termination, pelotons translate into transcriptional bursts, with the same characteristics as those observed in vivo. The generality of our model suggests that peloton formation might be ubiquitous in systems where molecular motors interact with dynamic roadblocks.

The work described in this chapter has been published as
AA van den Berg and M Depken, *Nucleic Acids Res* **45** 7623–7632 (2017)

3.1. INTRODUCTION

On every scale, motility is a hallmark of life [1, 2]. On the smallest scales, directed motion through the densely packed interior of cells is crucial for biogenesis, morphogenesis, and the timely delivery of vital cargo to distant parts [3]. The motion is often induced by large molecular complexes, powered along tracks by internal chemical reactions: polymerase and helicases move along DNA and RNA, ribosome along RNA, myosin along actin filaments, and dynein and kinesin along microtubules, to name but a few.

The intracellular environment is crowded [4]. Crowding of molecular motors can result in emergent behavior that is not present for single motors [5] and motors often have to bypass large amounts of other proteins bound to their track [6]. This is particularly true for the eukaryotic RNA polymerases, as over 80% of eukaryotic DNA is organized into nucleosomes [7] that consists of 147 base pairs (bps) of DNA wrapped tightly around an octameric core of histone proteins. Maintaining this dense nucleosome coverage is important since it organizes genomic DNA into compact, higher order structures that can fit within the limited space of the cell nucleus, but it also creates a formidable barrier to transcription [8]. Importantly, the local degree of nucleosome coverage correlates with gene-expression levels [7, 9–13] showing that transcription activity has important implications for nucleosome coverage and vice versa.

To shed light on the mechano-chemistry of transcription in the presence of nucleosomes (Figure 3.1 A), single-molecule experiments have been used to show that polymerases slow down at positions where nucleosomes are formed [14]. It is also known that multiple polymerases can cooperate to increase the transcription rate through nucleosomes [15] showing that the spatial organization of polymerases along a gene could be of crucial importance for understanding transcription in crowded environments.

Even though it is experimentally established that polymerase organization and nucleosome coverage affect the transcriptional output, it remains unclear how this is actualized on a mechanistic level [6]. With the aim to understand the basic implications of molecular crowding in eukaryotic transcription, we here construct a theoretical model that quantitatively describes the motion of polymerases interacting with dynamic nucleosomes. Taking into account that polymerases are slowed down by such roadblocks, we show that polymerases attract each other through a physical mechanism analogous to drafting in racing sports [16]. At physiological conditions, the attraction is so strong that two polymerases that meet along a gene remain together until termination, thus ensuring a progressive clustering of polymerases into stable pelotons as they move along the gene.

Our calculations show that peloton formation should be expected as soon as transcription initiation rates exceed the nucleosome exchange rate. Local polymerase clustering into pelotons could thus function to increase polymerase cooperation on highly transcribed genes, and it is interesting to note that clustering has been directly observed in Miller spreads of ribosomal genes [17–20], and for polymerases moving along heavily transcribed genes in live-cell experiments [21]. The model further explains how both nucleosome and polymerase densities can increase along heavily transcribed genes, even though polymerases and nucleosomes compete for space [13, 22, 23]. Lastly, the peloton formation predicted by our model results in bursts of mRNA production when the pelotons arrive at the termination site, pointing to a so-far unrecognized type of tran-

scriptional bursts[24–26].

To facilitate future experimental testing, we analyze our model analytically and present quantitative relationships that capture how nucleosome and polymerase densities, peloton sizes and separation, and transcriptional burst parameters depend on polymerase initiation and translocation rates, as well as nucleosome turnover times. As our model is based on general principles, it has the potential to describe motor and obstacle interactions in many other biological systems, suggesting that peloton formation should be expected as soon as motors interact with dynamical roadblocks.

3.2. RESULTS

The theoretical modeling of stochastic and driven molecular traffic on one-dimensional tracks has a long history in biology, starting almost half a century ago with the introduction of the Totally Asymmetric Simple Exclusion Process (TASEP) [27]. The TASEP consists of motors hopping stochastically in one direction along a one-dimensional lattice, moving only if the track just ahead is empty. Coupling this simple bulk rule to injection and extraction of motors at the boundaries gives rise to rich dynamical behavior, and the model has been extended to describe a wide range of physical and biological systems [28–34]. Here we extend the TASEP to include the interaction with roadblocks by building on earlier studies that considered a single roadblock [35, 36], as well as multiple dynamic roadblocks in the so-called Bus-Route Model (BRM) [37].

3.2.1. A MINIMAL MODEL OF MOTORS INTERACTING WITH ROADBLOCKS

To capture motor and roadblock interactions, we consider a system (Figure 3.1 B) for which: (i) motors move stochastically in one direction along a track, (ii) motors cannot overtake each other, (iii) roadblocks dynamically appear on empty sites of the track, (iv) roadblocks immediately ahead of a motor impede the motion of the motor, and (v) a passing motor temporarily removes a roadblock. The BRM is a specifically simple realization of the above criteria on a circular track, and with motor and roadblock sizes equal to the motor step size. As both nucleosome and polymerase are orders of magnitude larger than the basic polymerase step size, we here extend this model to the physiologically more relevant situation with larger motor and roadblock sizes (δ_m and δ_{rb} respectively, measured in units of the motor step size). To allow for transcription initiation and termination, we further allow motors to enter and leave the track at specific initiation and termination sites. The above rules are captured in the microscopic model illustrated in Figure 3.2 A.

Though we are not able to solve our model exactly, it is readily analyzed by computer simulations. Still, simulations only yield results for the particular parameter values tested, and will not give the general relationship between input and output parameters needed for easy comparison to future experimental results. Therefore, we here opt for a heuristic approach that yields approximate analytical relations between input and output parameters. Monte-Carlo simulations are then used to check validity of our approximations, showing that we lose little precision by taking a heuristic approach. Instead, this approach allows us to capture the dominant behavior of the wide class of motor systems satisfying condition (i)-(v).

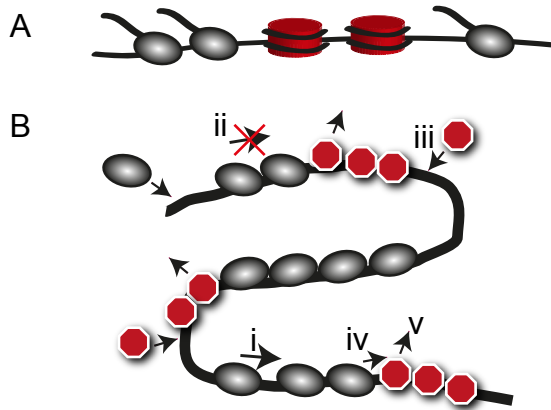


Figure 3.1: A model for motors interacting with dynamic roadblocks. A) The system studied in this Chapter, RNA polymerase interacting with nucleosomes. B) Schematic illustration of model features i)-v) (see text) for motors (ovals) interacting with dynamic roadblocks (octagons) along a one-dimensional track.

3.2.2. SPONTANEOUS FORMATION OF STABLE PELOTONS

To build intuition for the phenomenology of motor-roadblock-track interactions, we first investigate the dynamics in the bulk of the track, far away from initiation and termination sites. The roadblock occupancy should vary depending on the roadblock binding dynamics and motor-roadblock interactions. We start by consider the two limits of rare and ubiquitous roadblocks. The former limit is reached when roadblocks bind slowly, or the motor density is high such that roadblocks are excluded from the track and the dynamics should approach that of the TASEP with the motor hopping rate set by the rate of hopping into empty sites. The limit of ubiquitous roadblocks is reached when roadblocks rebind quickly behind a motor or when the motor density is low enough for roadblocks to bind between every pair of motors. The dynamics in that limit approaches that of a TASEP with a motor hopping rate set by the rate of hopping into a site occupied by a roadblock. In either limit, the exact solution of the TASEP [38] gives a geometric distribution of gap sizes between adjacent motors (see 2 and the Supplementary Material). In Figure 3.2 B and C we show kymographs and gap-size distributions generated by Monte Carlo simulations (see Supplementary Material) of the BRM [37]. As expected, both ubiquitous (left panels Figure 3.2 B and C) and sparse (right panels Figure 3.2 B and C) roadblocks result in gap-sizes distributions that are well described as geometrical.

For intermediate roadblock densities, the situation is subtler. Motors that are slowed down by roadblocks induce trailing traffic jams, while the gap to the motor ahead increases. As a gap opens up ahead of the motor causing the jam, it grows more likely to be slowed down by further roadblocks deposited in the gap, and the jam stabilizes. The jams should not grow indefinitely though, but organize into finite moving pelotons, as can be seen by the following argument: Defining a peloton as a group of motors with no interspersing roadblocks, a peloton can split at any position through the binding of a roadblock between two motors in the peloton. The rate of this splitting should be roughly proportional to the number of internal gaps in (i.e. the size of) the peloton. Pelotons can

also merge, but with a rate that is independent of peloton size. In the steady state we expect pelotons to have a well-defined typical size, such that the average peloton merging and splitting rates balance.

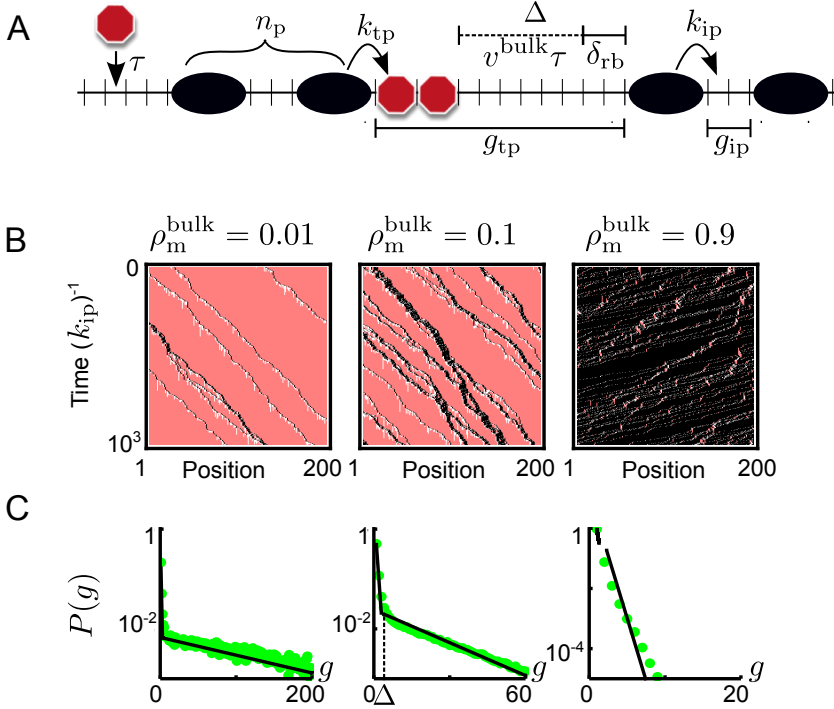


Figure 3.2: Roadblocks induce a hierarchy of TASEPs. A) Schematic illustration of the rules of the BRM. Microscopic rates are indicated, as well as the roadblock-DNA binding equilibration time τ , the roadblock shadow Δ , instances of the peloton size n_p , and the gap size for both trans- and intra-peloton gaps, g_{ip} and g_{tp} respectively. In this example, the motors occupy four lattice sites, $\delta_m = 4$, and roadblocks two, $\delta_{rb} = 2$, but the heuristic solution of the model is valid for arbitrary sizes. B) Kymographs generated through Monte-Carlo simulations of the BRM for systems with low roadblock density (left), intermediate roadblock density (middle), and high roadblock densities (right) for $k_{ip}\tau = 10$. Motors are shown in black, roadblocks in pink and the roadblock shadow is visible as a roadblock depleted region (white) behind the motors. C) The simulated gap-size distributions (green dots) corresponding to B), together with our analytical results (black lines). The left and right panel show a dominating single exponential (note the log-scales on the y-axes), which corresponds to a single TASEP. The gap size distribution in the middle panel shows two exponentials, which suggests that the system can be described as a combination of two TASEPs.

3.2.3. A HIERARCHY OF TASEPs CONTROL MOTOR ORGANIZATION

To understand the interactions between pelotons, we note that when two pelotons of a typical size merge, the new peloton is larger than the typical stable peloton. The new peloton is therefore typically unstable, and will split in two. This merging and subsequent splitting can be seen as an effective steric repulsion between pelotons, much like the interaction between motors in the original TASEP. The steady-state system can

therefore be seen as a superposition of two steady-state TASEP models: the intra-peloton TASEP (ipTASEP) originating from motor dynamics within pelotons, and the trans-peloton TASEP (tpTASEP) originating in the dynamics of the pelotons themselves acting as basic units of a TASEP. This heuristic hierarchical picture is confirmed in the middle panels of Figure 3.2 B and C, where we show a kymograph and a double-geometric gap distribution for intermediate roadblock coverage in the BRM.

3.2.4. A HEURISTIC SOLUTION

It is now important to establish how large the typical bulk pelotons are, as this will give an indication of the effective interaction strength between motors. Here we describe the important features of our heuristic solution, but refer the interested reader to the Supplementary Material for further details. Due to the finite size and equilibration time of roadblocks, roadblocks cannot rebind as soon as they have been evicted: the motor must both have moved away from the site of binding, and have allowed for enough time for the stochastic rebinding of the roadblock. Consequently, there is a region behind every moving motor that is depleted of roadblocks. We will refer to this region as the roadblock shadow, and estimate its size to be $\Delta \approx v^{\text{bulk}}\tau + \delta_{\text{rb}}$ (Figure 3.2 A). With v^{bulk} being the average motor velocity in the bulk, the term $v^{\text{bulk}}\tau$ captures the average distance traveled by a motor during the equilibration time τ of roadblocks, and δ_{rb} accounts for the fact that the motor must clear the whole footprint of the roadblock before it can rebind. Note that as long as the roadblock has a substantial size as compared to the basic step of the motor, the roadblock shadow will here remain extensive even for very fast roadblock rebinding, contrary to the situation in the BRM.

We have defined gaps between pelotons as those gaps that have roadblocks in them, meaning that they are typically larger than the roadblock shadow. Conversely, gaps within pelotons are devoid of roadblocks, and thus they are typically smaller than the roadblock shadow. We denote the effective motor hopping rate into gaps without roadblocks as the intra-peloton hopping rate k_{ip} , and the rate of hopping into gaps with roadblocks as the trans-peloton hopping rate k_{tp} . In the Supplementary Material we show that knowing the density of motors along the track we can analytically predict the dynamic state of the system, and that the average velocity in the system is $v^{\text{bulk}} = k_{\text{tp}}$ as soon as pelotons form.

In the Supplementary Material we further show that the typical size of a bulk peloton is proportional to $(k_{\text{ip}}/k_{\text{tp}})^{\Delta/2}$, which can be seen to combine the strength of the interaction between motors and roadblocks ($k_{\text{ip}}/k_{\text{tp}}$ is large if roadblocks substantially slow down motors) with the range of the interaction (Δ is the maximum typical distance over which two motors dynamically interact through roadblock depletion). As the size of a peloton depends linearly on the roadblock shadow size Δ (Equation 3.1), and the roadblock size is generally substantially larger than the motor step for any physiological system (c.f. the fact that the nucleosome covers 147 bp of DNA, while the polymerase step is 1 bp), we expect the system to strive towards extremely large steady-state pelotons in the bulk. Indeed, for transcription with realistic parameter values (see Table 3.1), the steady-state peloton size is so enormous that it will never fit on any gene (see Supplementary Material). Though the steady-state peloton size is thus never realized, its magnitude shows that two motors that meet will typically stay together until termination. We

now use this observation to derive the peloton size reached over finite tracks, such as genes.

Table 3.1: Parameter values as estimated from the literature and implemented in the simulations.

Microscopic parameter	Value	Citation
δ_{Pol} : RNAP footprint	35 bp	[39]
δ_{rb} : Footprint Nucleosome + linker DNA	167 bp	[40, 41]
k_{ip} : Average RNAP elongation rate on bare DNA	10 bp/s	[42–44]
k_{tp} : RNAP elongation rate through nucleosome	3 bp/s	[43]
k_{in} : Initiation rate on highly transcribed genes	0.6–3/min	[45]
k_{b} : Nucleosome binding rate	0.02 s^{-1}	[46, 47]
τ : Hexamer binding time	$\tau k_{\text{tp}} \ll \delta_{\text{rb}} \rightarrow \Delta \approx \delta_{\text{rb}}$	[46–48]

3.2.5. MOTOR AND ROADBLOCK REORGANIZATION ON FINITE GENES

Initiation of transcription generally controls transcription levels [49], which means that transcription is in the initiation-limited regime where initiation rates are low (2). To capture this situation, we consider open systems where initiation rates are low enough that a motor does not generally block the initiation of subsequent motors. During eukaryotic transcription, the initiation site is kept free of nucleosomes [7, 50], and consequently we will assume the initiation site in our model to also be devoid of roadblocks. Taking motors to initiate with a finite rate k_{in} , some of the motors will have a roadblock just ahead of them, while some will not. Motors unhindered by roadblocks catch up with motors slowed down by roadblocks (see schematic kymograph in Figure 3.3 A upper panel), and motors start to collect into pelotons.

As more and more motors are absorbed into pelotons, the average motor velocity goes down. To maintain a constant steady-state motor flux, the motor density then goes up as we move away from initiation (Figure 3.3 A lower panel). Simultaneously, as motors organize into pelotons, they leave more room available for roadblocks to bind. Spontaneous organization into pelotons thus allow both motor and roadblock densities to increase along genes.

After the initial pelotons are formed, these will continue to evolve towards the bulk peloton size through a merging process described by diffusion-limited coagulation [51]. Relaxation in such systems (referred to as aging in the physics literature) is exceedingly slow, and we do not expect to see any appreciable evolution of the initially formed pelotons over a finite track (such as a gene).

In the Supplementary Material we give the general expressions relating the microscopic parameters to the average size n_{p} of pelotons, and the distance x_{p} over which they form (Equation 3.33). For simplicity we here give the physiologically relevant limit where motors typically clear the initiation site between attempted initiation events,

$$n_{\text{p}} \approx 1 + k_{\text{in}} \tau_{\Delta}, \quad x_{\text{p}} = \frac{1/k_{\text{in}}}{1/k_{\text{tp}} - 1/k_{\text{ip}}} \frac{1}{\ln(1 + 1/k_{\text{in}} \tau_{\Delta})}, \quad \tau_{\Delta} = \Delta/k_{\text{tp}} = \tau + \delta_{\text{rb}}/k_{\text{tp}}. \quad (3.1)$$

Here we recognize the timescale τ_{Δ} as the time needed for a motor to clear the road-

block shadow and allow a new roadblock to bind. We will refer to this time as the effective roadblock-rebinding time. The average peloton size beyond the leading motor, n_p , can be understood as a ratio between the effective roadblock-rebinding time and the time to initiate a new motor ($1/k_{in}$), giving the number of motors initiated between roadblock binding events at the start of the track. The distance over which pelotons are formed, x_p , is given by the ratio of effective roadblock-rebinding time and the time difference between taking a step for slow and fast motors ($1/k_{tp} - 1/k_{ip}$) (see the Supplementary Material for details), giving the number of steps needed for the last motor to catch up with the rest of a forming peloton.

As it is often easier to experimentally measure relative rather than absolute changes in densities and velocities, we here report the evolution of the motor density/roadblock density/motor velocity ($\rho_m(x)/\rho_{rb}(x)/v_m(x)$) relative to their final value once pelotons are formed ($\rho_m/\rho_{rb}/v_m$) (for details see the Supplementary Material)

$$\begin{aligned}\rho_m(x)/\rho_m &\approx 1 - \frac{k_{in}\tau_\Delta}{1+k_{in}\tau_\Delta} \left(1 - \frac{k_{tp}}{k_{ip}}\right) e^{-x/x_p}, \\ v_m(x)/v_m &\approx 1 + \frac{k_{in}\tau_\Delta}{1+k_{in}\tau_\Delta} \left(\frac{k_{ip}}{k_{tp}} - 1\right) e^{-x/x_p}, \\ \rho_{rb}(x)/\rho_{rb} &\approx 1 - \frac{k_{in}\tau_\Delta}{e^{\delta_{rb}/x_p} + k_{in}\tau_\Delta} e^{-x/x_p}.\end{aligned}\tag{3.2}$$

From Equation 3.2 we see that all measures approach their final value with an exponential decay over the region where pelotons form. We see that relative changes along the track grow in magnitude with the initiation rate, but that the effect saturates around $k_{in} \sim 1/\tau_\Delta$ for motor-density and velocity changes, while the roadblock density saturates later, around $k_{in} \sim e^{\delta_{rb}/x_p}/\tau_\Delta$. The relative increase of the velocity and the densities saturate because when the initiation rate is larger than the effective roadblock binding time $k_{in} > 1/\tau_\Delta$ most initiating motors have no roadblock in front of them and the initial density and velocity become independent of the initiation rate. The evolution of the density and velocity further depends on the strength of interactions between motors and roadblocks (k_{tp}/k_{ip}) since the change in velocity (density) when a motor catches up with a peloton is larger when k_{tp}/k_{ip} is smaller.

In Figure 3.3 B we illustrate how the relative change in motor density is affected by the motor initiation rate, comparing the full expressions derived in the Supplementary Material to simulations. For slow initiation ($k_{in}\tau_\Delta < 1$, green arrow in Figure 3.3 B), a roadblock typically binds between every two initiating motors. Therefore, pelotons are typically of size one, giving only a marginal motor density change along the track (top density profile of Figure 3.3 D). For faster initiation ($k_{in}\tau_\Delta > 1$, red arrow Figure 3.3 B), multiple motors bind before a roadblock rebinds to the start of the track, pelotons are larger than one, and we have a substantial increase in motor density as we move away from the initiation site (lower density profile of Figure 3.3 D). In Figure 3.3 C we compare our prediction for the distance over which pelotons form to estimates extracted by fitting an exponential relaxation distance to the density profiles generated by simulations. It is quite remarkable that our crude approximations capture the simulated data without any adjustable parameters.

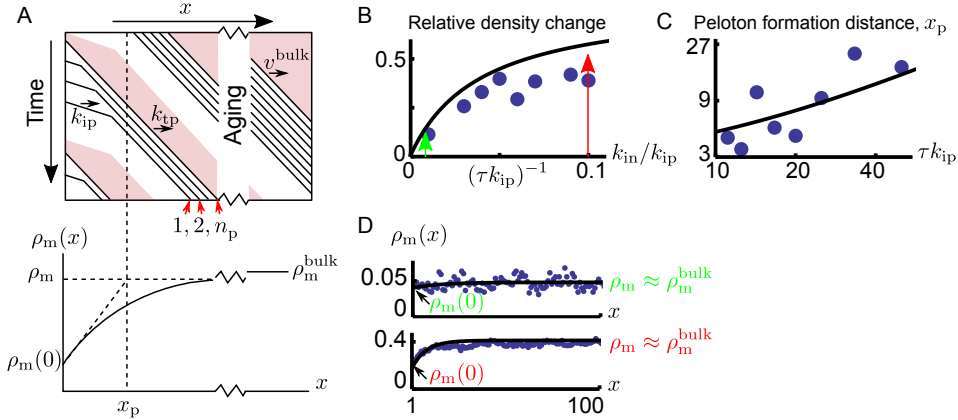


Figure 3.3: Peloton formation close to the initiation site. A) A schematic kymograph showing motors (black lines) initially not interacting with each other, until they reach the growing peloton. Motors initiate from the left and then travel into the system, moving through roadblock depleted regions (white) and roadblock filled regions (pink). If a roadblock is deposited between two motor initiation events, the last motor propagates with rate k_{tp} and otherwise with rate k_{ip} . After a typical distance x_p , a peloton of size n_p is formed. Below the kymograph we sketch the corresponding motor density. B) The relative change in motor density, $(\rho_m - \rho_m(0))/\rho_m$, from the start of the lattice to the point where all the initial pelotons have formed. The line represents our theoretical predictions (Equation 3.2) and the green and red arrows indicate the initiation rates used for Figure D. C) The distance x_p over which pelotons form as a function of the roadblock equilibration time for $k_{in}/k_{ip} = 0.1$. The dots are values for x_p obtained by fitting an exponential distribution to the peloton forming region (estimated as the first $4x_p$ lattice points) of simulated data, while the line represents our theoretical predictions (Equation 3.1). D) Motor density profiles for $\tau k_{ip} = 20$, and $k_{in}/k_{ip} = 0.01 < (\tau k_{ip})^{-1}$ in the top panel, and $k_{in}/k_{ip} = 0.1 > (\tau k_{ip})^{-1}$ for the lower panel. Blue dots are the result of Monte Carlo simulations and black lines are our analytical predictions. Note, there are no free parameters in any of the analytical predictions in B)-D).

3.2.6. FROM PELOTONS TO BURSTS

Transcriptional bursts have been observed in both eukaryotic and prokaryotic systems [24, 52], and are often ascribed to a promoter that can be turned on and off [53]. In the presence of roadblocks along the gene, our model shows that we should expect the same type of bursts even for promoters that are constantly turned on (Figure 3.4 B). To facilitate future experimental testing through the many known downstream effects of a bursty promoter [53], we here relate the bursts of motor activity in our model to those arising from the standard assumption of a promoter that turns on and off as described by a two-state model [54], (Figure 3.4 C). This should prove especially useful when characterizing the level of noise in mRNA production by using the Fano factor (the ratio between the variance and the mean in mRNA copy numbers), which has been widely used as a measure to classify transcriptional noise experimentally [55, 56]. The fano factor for the two-state model is known [53, 54], such that a mapping of our model to the two-state model allows for a direct comparison with experiments.

In the two-state model (Figure 3.4 C), the system switches between an on-state with production rate k_{tr} , and an off-state where nothing is produced. The off-state switches to the on-state with rate k_{on} , and back again with rate k_{off} . Though we present the full form

of how the effective burst parameters depend on microscopic parameters in the Supplementary Material, we here again limit ourselves to the physiologically relevant case where initiation rates are low such that motors typically clear the initiation site between attempted initiation events,

$$k_{\text{tr}} = \frac{k_{\text{in}}\tau_{\Delta}}{1 + k_{\text{in}}\tau_{\Delta}} \frac{k_{\text{tp}}}{\delta_{\text{m}}}, \quad k_{\text{off}} = \frac{1}{1 + k_{\text{in}}\tau_{\Delta}} \frac{k_{\text{tp}}}{\delta_{\text{m}}}, \quad \text{and} \quad k_{\text{on}} \approx \frac{1}{\tau_{\Delta}}. \quad (3.3)$$

For significant peloton formation $k_{\text{in}}\tau_{\Delta} > 1$ (Equation 3.1), the production rate k_{tr} becomes insensitive to the initiation rate, and is simply set by the rate at which motors pass termination. The off-rate k_{off} depends strongly on the initiation rate, and is given by the rate at which a typical peloton passes termination. The on-rate k_{on} is simply given by the inverse roadblock-rebinding time. It should also be noted that as long as the track extends further than the peloton forming distance x_{p} (which is the case for transcription, see below), the burst characteristics do not depend on the length of the track. The analytical relationships given in Equation 3.1-3.3 are the main results of this chapter. As these relationships dictate the precise dependence of a number of observables on microscopic parameters they should be well suited for falsification through comparison to future experiments (see Discussion). Here we show that the predictions are in line with the results from a number of recent studies, using as input parameters values from the literature (see Table 3.1).

3.2.7. TRANSCRIPTION ON HIGHLY INDUCED GENES

Now that we have a qualitative understanding of how the non-specific interactions between motors and roadblocks give rise to peloton formation, we consider transcription on inducible genes in eukaryotes. These considerations are complicated by the fact that nucleosome assembly and disassembly are not single step processes, as a tetramer and two dimers come together to make the full histone octamer contained in the nucleosome. *In vitro* studies have shown that a single polymerase only removes the histone dimer [57–59], while a second polymerase can dislodge the remaining hexamer [60]. These *in vitro* results broadly agree with the *in vivo* observations that the density of histone dimers decreases strongly with transcription intensity genome wide, while an increased exchange and depletion of all core histones is only observed on highly transcribed genes [9–13, 61, 62].

When polymerases form pelotons, they are expected to cooperate in dislodging the full nucleosome [60]. Therefore, we assume that the roadblocks consist of histone hexamers that are dislodged by a passing polymerase. Further, we only consider genes where initiation is both active and non-paused, excluding situations where transcription is stalled by promoter-proximally paused polymerases [63]. As Equation 3.1 shows that pelotons form when initiation is high ($k_{\text{in}}\tau_{\Delta} > 1$), we compare our model to experiments tracking highly expressed genes (see Table 3.1).

We compare our analytical predictions of our heuristic approach to simulations with physiologically relevant parameters (Figure 3.5). In our simulations we assume that the motors are only impeded at the nucleosome dyad, since this forms the largest obstacle for RNA polymerase II translocation [14]. We consider the initiation rates $k_{\text{in}} = 0.6$ pol/min and $k_{\text{in}} = 3.0$ pol/min, where both rates correspond to highly induced genes,

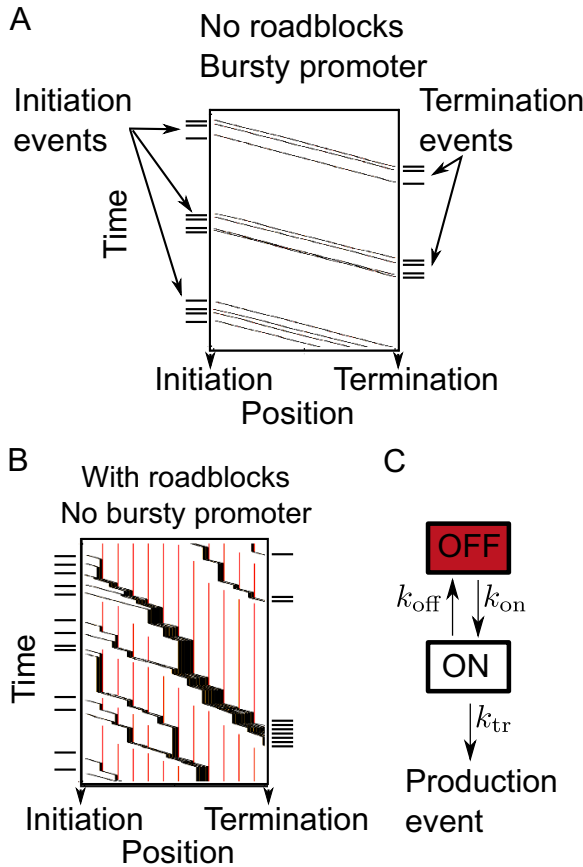


Figure 3.4: Burst generation from the promoter and during elongation. A) A kymograph of a Monte Carlo simulation for transcription with a bursty promoter and no roadblocks. The distribution of initiation events is similar to the time distribution at termination. B) A kymograph for transcription without a bursty promoter, but with roadblocks. Though the initiation events are exponentially distributed over time, the events at termination are more clustered, resulting in bursts in RNA production. C) The phenomenological two-state model normally used to describe bursty transcription. In Equation 3.3 we report the parameters that would result from fitting the bursts generated by our model to the two-state model.

and the highest rate is chosen to match the maximal estimate of initiation rates on yeast genes [45]. It is known that histones rebind on a sub-minute time scale [48], while it takes about a minute to clear space for a roadblock (see Table 3.1). Consequently, the nucleosome shadow is dominated by the roadblock size, and we assume $\Delta \approx \delta_{\text{rb}}$ for simplicity in the analytical theory.

With only a small set of known microscopic input parameters (Table 3.1), our theory quantitatively captures the dynamics without free parameters. As predicted by Equation 3.1, we see that pelotons grow over the first few hundred base pairs after initiation (Figure 3.5 A and C). The peloton growth in turn means that the density of polymerases and nucleosomes near the initiation site is lower than further into the gene, while the veloc-

ity decreases as we move away from initiation (Figure 3.5 B and D). After this point, the polymerases and nucleosome densities, as well as polymerase velocities, remain virtually constant throughout the bulk of the gene. In Table 3.2 we give an overview of the predicted values of several observables, including burst parameters.

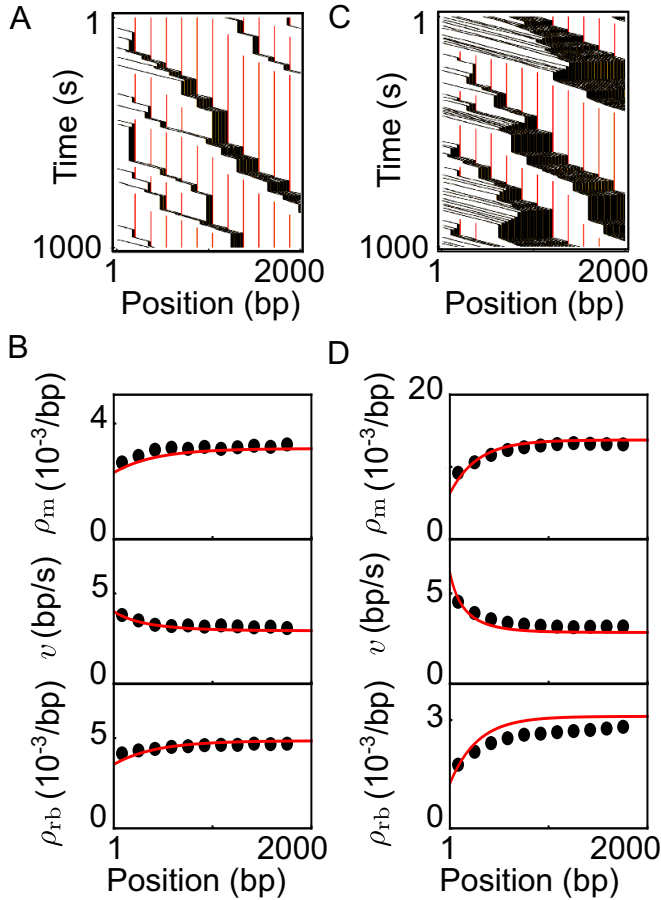


Figure 3.5: Bursts and density evolution for eukaryotic transcription on inducible genes. The parameter values used are shown in Table 3.1. A) Kymograph for relatively moderate initiation rates. A polymerase (shown in black) evicts a nucleosome when it passes its center (the dyads, indicated by red lines). As the polymerases enter the gene, pelotons form over a distance of a few hundred base pairs. B) The polymerase density, polymerase velocity, and nucleosome density corresponding to the kymograph in A). Simulations were averaged over the size of a nucleosome and are shown as black dots, with our analytical predictions as red lines. C) Kymograph for relatively high initiation rates, resulting in larger pelotons as compared to A). D) The polymerase density, polymerase velocity, and nucleosome density corresponding to the kymograph in C). Comparing B) and D), we see that larger pelotons give a visibly stronger density and velocity evolution.

Table 3.2: Calculated observables for the physiological parameters in Table 3.1.

Initiation rate	0.6 pol/min	3.0 pol/min
$v_m(0)$: RNAP velocity at initiation	5.5 bp/s	8.1 bp/s
v_m : RNAP velocity once pelotons have formed	3 bp/s	3 bp/s
J : Transcriptional output	0.6 RNA/min	2.3 RNA/min
$\rho_m(0)$: RNAP density at initiation	0.002 pol/bp	0.006 pol/bp
ρ_m : RNAP density once pelotons formed	0.003 pol/bp	0.013 pol/bp
n_p : Peloton size	1.6 pol	3.8 pol
x_p : Distance over which pelotons form	420 bp	280 bp
k_{on} : Apparent on rate in two-state model	1.1/min	1.1/min
k_{tr} : Apparent production rate in two-state model	1.8/min	3.8/min
k_{off} : Apparent off rate in two-state model	3.3/min	1.4/min

3.3. DISCUSSION

With the aim of describing transcription in the crowded environment of the cell, we have introduced a general model that captures a large class of systems where molecular motors interact with dynamic roadblocks (Figure 3.1 B). Assuming that roadblocks have a finite size and/or rebinding time, and that motors slow down when evicting nucleosomes, we show that a physical mechanism reminiscent of drafting in racing sports gives rise to a strong kinetic attraction between motors. This attraction induces the motors to spontaneously reorganize into pelotons, and motors arrive to the terminus in bursts. Our analysis shows that one should always expect bursts in the presence of roadblocks if the motor initiation rate exceeds the inverse effective roadblock-rebinding time (τ_Δ in Equation 3.1), independently of if the promoter itself is bursty.

3.3.1. PELOTON FORMATION HAS BEEN OBSERVED *in vivo*

Already 40 years ago there was evidence from Miller spreads suggesting that polymerases cluster on heavily transcribed genes [17–20]. Recently, direct real-time evidence of polymerase ‘convoys’ on HIV-1 and POLR2A genes in HeLa cells has been found [21]. In accordance with our predictions, the typical distance between polymerases within convoys is too small for a nucleosome to bind, the distances between convoys are geometrically distributed, and a typical convoy includes several polymerases (see Table 2). Though our model cannot rule out alternative explanations relying on specific polymerase-nucleosome interactions, the agreement between our predictions and the experimental observations without any adjustable fit parameters suggests that peloton formation through non-specific interactions with nucleosomes should be seriously considered, and could drive the formation of the observed polymerase convoys. This hypothesis can readily be tested by correlating the average peloton size (Equation 3.1) with changes in induction levels.

3.3.2. PREDICTED DENSITY PROFILES AGREE WITH OBSERVATIONS IN YEAST

Our model also gives parsimonious explanations for several recent *in vivo* experimental observations pertaining to density profiles of polymerases and nucleosomes along in-

ducible genes. In both yeast and human cells, highly transcribed genes without promoter-proximally paused polymerases show low nucleosome and polymerase densities for the first few hundred base pairs after initiation [13, 64]. This distance agrees with our predicted distance over which pelotons are formed, x_p (Table 3.2). Though there are many specific interactions that could give rise to nucleosome and polymerase depletion [65], the fact that this is a general property of heavily transcribed genes [13, 66] suggests a non-specific mechanism. Indeed, our model accurately predicts the occurrence and extent of such depletion without invoking any specific interactions (see Equation 3.1 and 3.2, as well as Figure 3.5 and Table 3.2). Our model correctly predicts a pausing-index (relative polymerase density within the promoter-proximal region compared to the bulk) that is below one for highly transcribed genes [64], and can be further tested by correlating changes in the pausing-index with for example histone modifications [43] that modify the transcription rate through nucleosomes.

The predicted increase of the polymerase density along the gene coincides with a decrease in the elongation rate (see Figure 3.5 B and D). Global run-on sequencing (GRO-seq) [64] experiments on active genes, on the contrary, have shown that the elongation rate *increases* over the first 15 kbp [67]. The velocity increase in these GRO-seq experiments was likely caused by the gradual maturation of the transcription machinery [65] with mechanisms such as interactions with elongation factors not presently included in our model. Though it would be interesting to see how such mechanisms modulate the formation of pelotons, the observed increase of the elongation rate takes place over distances much longer than the few hundred base pairs over which pelotons form (see Table 3.2), and we do not expect our quantitative results to change due to these moderate velocity changes close to the promoter.

Another mechanism suggested to explain the observed density profiles is that transcription becomes termination-limited for high transcription rates [13]. However, on a termination-limited gene, cued polymerases typically block each other's movement sterically [27] and leave no place for nucleosomes in between, which is inconsistent with experimental observations [13].

It is interesting to note that a nucleosome-free region at the start of genes [13, 22, 23] has been suggested to increase the accessibility of transcription factor binding sites close to the initiation site, thereby increasing the potential for transcriptional regulation [22, 23]. Our model thus suggests that nucleosome depletion close to the initiation site could be a transcriptional epiphenomenon that has been coopted/adapted to allow for a greater regulatory response.

3.3.3. BURST CHARACTERISTICS AGREE WITH *in vivo* OBSERVATIONS

Though bursts in RNA production have been observed in both prokaryotes [24, 68] and eukaryotes [55, 69], their origin is unclear, and usually modeled phenomenologically as arising from a promoter that turns on and off [70, 71]. Though not developed with bursts in mind, our model predicts that transcriptions should be expected to be bursty as soon as the transcription initiation rate is comparable to the nucleosome binding rate, Equation 3.3, even if the promoter is constantly turned on. Several properties of the predicted bursts agree quantitatively with experimental observations. First, the pelotons are completed over a few hundred base pairs, which is shorter than the most genes. Therefore

the predicted burst size is independent of gene length, agreeing with observations in yeast [72]. Second, the predicted time between production events is on a sub-minute time scale (see Table 3.2), which falls within the experimentally observed range [72]. Third, our model predicts that when bursts are significant, only the apparent burst duration should be sensitive to transcription intensity, and that it decreases with increasing induction (see Equation 3.3). This behavior is broadly agreeing with the behavior reported for transcriptional bursting in *Escherichia coli* (*E. coli*) [26], where many other DNA binding proteins might act as the necessary roadblocks [73].

It should be noted that bursts generated during elongation do not rule out a bursty promoter. Instead, multi-scale bursting was recently reported [21], which could very well originate in a promoter turning on and off on long timescales, while pelotons still form during elongation, giving rise to bursting on shorter timescales.

3.3.4. DNA SUPERCOILS AS A SOURCE OF BURSTS

Though there are many DNA binding proteins in *E. coli* [73], another interesting candidate for producing bursts is DNA supercoiling. Due to the helicity of DNA, transcribing polymerases are known to induce positive supercoils ahead and negative supercoils behind [74]. Such supercoils slow down the polymerase [68], and will in the steady state extend some finite distance in front and behind. As negative supercoils spontaneously annihilate with positive supercoils, any DNA between two polymerases will have a lower net supercoiling density the closer together they are. With a lower supercoiling density ahead, a trailing polymerase will move faster than a leading polymerase, and all the conditions for peloton formation as described by our model (Figure 3.1 A) are fulfilled.

Our general mechanism of burst generation is connected to the mechanism suggested as a source of transcriptional bursting observed in bacteria, where a buildup of supercoils in torsional constrained plasmids was shown to suppress transcription until the supercoils were released [68]. Importantly though, our model does not require the DNA to be torsionally constrained as the supercoiling density around polymerases is set by the supercoil's diffusivity [75] and a balance between supercoil creation and release.

3.3.5. EXPERIMENTAL TESTING AND ALTERNATIVE MODELS

Many mechanisms have been suggested for promoter induced bursting. As indicated in Figure 3.4, our model can be differentiated from such models by comparing the input and output dynamics. Our model could also be refuted by using existing techniques reporting on polymerase "convoys" [21] or transcriptional bursting [24] by manipulating or screening the limited set of effective parameters that controls the spatial and temporal evolution of the system (Equation 3.1-3.3). For example, the typical peloton size could be manipulated by changing the transcription initiation rate, or through histone modifications [43] that change the transcription rate through nucleosomes or the nucleosome rebinding time.

If the initiation dynamics cannot explain the bursts of RNA production, the elongation phase is likely the source of the bursts. To our knowledge, there are only two previous theoretical studies suggesting that bursts are created during transcription elongation [29, 76]. In both cases, intrinsic polymerase pausing through backtracking [77] was suggested as the source. However, backtracking is unlikely to produce bursts at termi-

nation, as it does not induce an effective attraction between polymerases, but rather an effective repulsion: interaction with a trailing polymerase is known to help terminate backtracks of a leading polymerase, and so speeds it up; interaction with a leading polymerase increases the chance of pausing in a trailing polymerase [60, 78], and so slows it down. Polymerases thus kinetically repel each other, and jams induced by backtracks are unstable and should typically dissolve before termination. Instead, we have shown that the interaction with roadblocks induces a persistent effective attraction between polymerases, resulting in a fast buildup of stable pelotons as polymerases move through the gene to terminate in bursts.

3.3.6. CONCLUSION AND OUTLOOK

Our model points to a single source for a wide range of observed phenomena, from burst characteristics to the spatial organization of polymerases and nucleosomes. Further experiments are needed to determine the degree to which the observed phenomena can be explained through non-specific polymerase and nucleosome interactions as we suggest. Surprisingly though, the model agrees quantitatively with multiple experimental observations without adjustable parameters. Only by first understanding the organization of polymerases and nucleosomes along the DNA, and how it can be modulated to effect things like polymerase cooperation, will it be possible to fully understand the action of transcription factors and other important cellular responses acting on the elongation phase of transcription.

ACKNOWLEDGEMENTS

We thank Joachim Griesenbeck for insightful discussions, Misha Klein, Orkide Ordu, Behrouz Eslami, John van Noort, and Stephan Grill for comments on the manuscript, and Ruben van Drongelen for help with questions about programming. MD and AAvdB acknowledge financial support from a TU Delft startup grant to MD. AAvdB further acknowledges financial support by the Netherlands Organization for Scientific Research (NWO/OCW), as part of the Frontiers of Nanoscience program.

3.4. SUPPLEMENTARY MATERIAL

We here give the derivations and arguments that were left out of the main text for clarity. For brevity we will here only keep the 'b' of 'bulk' in the subscript, for example writing v^b instead of v^{bulk} .

3.4.1. HEURISTIC SOLUTION FOR OF THE HIERARCHICAL TASEP MODEL

The average TASEP motor velocity

For the TASEP, gap sizes are geometrically distributed as (21),

$$P_{\text{TASEP}}(g; a) = (1 - a) a^g, \quad (3.4)$$

for some constant $a < 1$. Unless a motor is blocked by another motor ($g = 0$), it will hop forward with rate v , and the average velocity of motors can be calculated as

$$v^b = k \sum_{g=1}^{\infty} P_{\text{TASEP}}(g; a) = ka. \quad (3.5)$$

From this it follows that $a = v^b/k$, and we can write

$$P_{\text{TASEP}}(g; v^b/k) = \left(1 - v^b/k\right) \left(v^b/k\right)^g. \quad (3.6)$$

Intra- and trans-peloton gap sizes

The inclusion of dynamic roadblocks will split the motor dynamics into an intra-peloton and a trans-peloton TASEP describing gap sizes below and above the roadblock shadow Δ . In line with our heuristic argument, we assume there to be no roadblocks within a peloton. Apart from the leading motor, all motors within a peloton thus attempt to hop forward with rate k_{ip} . The pelotons themselves are controlled by the leading motor, which faces a trans-peloton gap filled with roadblocks and thus attempts to hop forward with rate k_{tp} . Assuming that the gap-size distribution is geometric both below and above the roadblock shadow, we can now write our normalized heuristic gap-size distribution for the intra- and trans-peloton regimes as

$$P_{\text{ip}}(g; v^b/k_{\text{ip}}) = \frac{(v^b/k_{\text{ip}})^g}{\sum_{g=0}^{\Delta-1} (v^b/k_{\text{ip}})^g} = \left(1 - v^b/k_{\text{ip}}\right) \left(v^b/k_{\text{ip}}\right)^g + O\left[\left(v^b/k_{\text{ip}}\right)^\Delta\right], \quad g < \Delta$$

$$P_{\text{tp}}(g; v^b/k_{\text{tp}}) = \frac{(v^b/k_{\text{tp}})^g}{\sum_{g=\Delta}^{\infty} (v^b/k_{\text{tp}})^g} = \left(1 - v^b/k_{\text{tp}}\right) \left(v^b/k_{\text{tp}}\right)^{g-\Delta}, \quad g \geq \Delta. \quad (3.7)$$

The above equations are valid when $(k_{\text{ip}}/k_{\text{tp}})^\Delta$ is large, which we refer to as the stable peloton regime (SPR). The condition for the SPR can intuitively be seen as combining the strength of the attraction between motors and roadblocks ($k_{\text{ip}}/k_{\text{tp}}$) and its range (Δ). Due to the SPR conditions exponential dependence on the roadblock shadow size, we

expect physiological systems where the roadblock size is substantially larger than the motor step to always be in the SPR. In this limit, we have

$$\begin{aligned} P_{\text{ip}}(g; v^b/k_{\text{ip}}) &= P_{\text{TASEP}}(g; v^b/k_{\text{ip}}), \quad g < \Delta, \\ P_{\text{tp}}(g; v^b/k_{\text{tp}}) &= P_{\text{TASEP}}(g - \Delta; v^b/k_{\text{tp}}), \quad g \geq \Delta \end{aligned} \quad (3.8)$$

With these conditional distributions, we can now calculate the average gap sizes for both regimes

$$\begin{aligned} \langle g \rangle_{\text{ip}}^b &= \sum_{g=0}^{\infty} g P_{\text{TASEP}}(g; v^b/k_{\text{ip}}) = \frac{v^b}{k_{\text{ip}} - v^b}, \\ \langle g \rangle_{\text{tp}}^b &= \sum_{g=\Delta}^{\infty} g P_{\text{TASEP}}(g - \Delta; v^b/k_{\text{tp}}) = \Delta + \frac{v^b}{k_{\text{tp}} - v^b}. \end{aligned} \quad (3.9)$$

Defining p^b as the probability of a gap in the bulk being a trans-peloton gap, we can write the average gap between motors as $\langle g \rangle^b = (1 - p^b) \langle g \rangle_{\text{ip}}^b + p^b \langle g \rangle_{\text{tp}}^b$. Taking the average motor size into account, the average motor density is the inverse of the typical distance between the fronts of neighboring motors,

$$\rho_m^b = \frac{1}{\langle g \rangle^b + \delta_m} = \frac{1}{(1 - p^b) \langle g \rangle_{\text{ip}}^b + p^b \langle g \rangle_{\text{tp}}^b + \delta_m}. \quad (3.10)$$

The relative fraction of trans peloton gaps

Combining the intra- and trans-peloton distributions we can write the complete gap-size distribution of the hierarchical TASEP as

$$P_{\text{hTASEP}}(g; v^b/k_{\text{ip}}, v^b/k_{\text{tp}}, \Delta) = \begin{cases} (1 - p^b) P_{\text{TASEP}}(g; v^b/k_{\text{ip}}), & g < \Delta \\ p^b P_{\text{TASEP}}(g - \Delta; v^b/k_{\text{tp}}), & g \geq \Delta \end{cases}. \quad (3.11)$$

In steady state, the probabilistic flow from intra- to trans-peloton gaps (a peloton is split in two) and from trans- to intra-peloton gaps (two pelotons merge) should balance. The motor ahead of a gap of size $\Delta - 1$ hops with an average rate v and extends the gap to size Δ , inducing the mean-field probabilistic flow $v^b P_{\text{hTASEP}}(\Delta - 1; v^b/k_{\text{ip}}, v^b/k_{\text{tp}}, \Delta) = v^b (1 - p^b) P_{\text{TASEP}}(\Delta - 1; v^b/k_{\text{ip}})$. In turn, a motor behind a gap of size hops forward with an average rate and decreases the gap to size $\Delta - 1$, inducing the probabilistic flow $k_{\text{tp}} P_{\text{hTASEP}}(\Delta; v^b/k_{\text{ip}}, v^b/k_{\text{tp}}, \Delta) = k_{\text{tp}} p^b P_{\text{TASEP}}(0; v^b/k_{\text{tp}})$. In the steady state these two flows should balance, and equating these rates gives

$$p^b = \frac{[\langle g \rangle_{\text{tp}}^b - \Delta]}{[\langle g \rangle_{\text{tp}}^b - \Delta] + \langle g \rangle_{\text{ip}}^b \left(\frac{k_{\text{ip}}}{v^b} \right) \Delta}. \quad (3.12)$$

Taken together, Equations 3.9, 3.10, and 3.12 relate the average density in the system to the velocity and the microscopic model parameters. In Figure 3.2 C in the main text we compare the gap-size distribution resulting from our heuristic arguments (solid lines) with ones generated through simulations of the BRM (dots) at different motor densities.

3.4.2. RELATION BETWEEN HEURISTIC ARGUMENTS AND MEAN-FIELD SOLUTION OF BRM

For periodic boundary conditions and irreversible roadblock binding, there exists a mean-field solution for the special case $\delta_m = \delta_{rb} = 1$ [37]. It is informative to consider our more general but less precise heuristic approach in the light of this mean-field solution. We first present a short summary of the mean-field solution, and then show how a few approximations give our heuristic results. The gap-size distribution for the BRM in the mean-field is given by

$$P_{\text{BRM}}(g) = \frac{\prod_{y=1}^g \frac{v^b}{k(y)}}{\sum_{g=0}^{\infty} \prod_{y=1}^g \frac{v^b}{k(y)}}. \quad (3.13)$$

Here $k(g)$ is average stepping rate of a motor into a gap of size g . This average stepping rate can be determined by calculating the probability of a roadblock with binding rate k_b being bound to a lattice site after a time t since the last motor passed, $P_{rb}(t) = 1 - e^{-k_b t}$. If we make the mean-field assumption that the motor moves with the average velocity v of the system, the probability of a roadblock being bound can be written in terms of the gap size to the motor ahead $P_{rb}(g) = 1 - e^{-gk_b/v^b}$. The effective stepping rate of a motor as a function of the gap ahead is now given by

$$k(g) = P_{rb}(g)k_{tp} + (1 - P_{rb}(g))k_{ip} = k_{tp} + (k_{ip} - k_{tp})e^{-gk_b/v^b} = k_{tp} + (k_{ip} - k_{tp})e^{-g/(\Delta-1)} \quad (3.14)$$

where we in the last step introduced the roadblock shadow $\Delta = 1 + v^b/k_b$ for the BRM, and in direct analogy with our heuristic approach. If we now approximate $k(g)$ by a step function

$$k(g) = \begin{cases} k_{ip}, & g < \Delta \\ k_{tp}, & g \geq \Delta \end{cases} \quad (3.15)$$

the gap-size distribution is given by

$$P_{\text{BRM}}(g) \propto \begin{cases} \left(v^b/k_{ip}\right)^g, & g < \Delta \\ \left(k_{tp}/k_{ip}\right)^{\Delta-1} \left(v^b/k_{tp}\right)^g, & g \geq \Delta \end{cases} \quad (3.16)$$

Enforcing normalization, Equation 3.16 implies Equation 3.7.

3.4.3. OBSERVABLE BULK QUANTITIES

In addition to the motor density, there are other interesting observables that can be calculated if we know the average velocity in the bulk. Among them are the current of motors, $J = \rho_m^b v^b$, and the average number of motors in a peloton, $\langle n_p \rangle^b = 1/p^b$. Further, only gaps between pelotons are filled with roadblocks, and then typically only beyond the roadblock shadow. If we let ρ_{rb}^{eq} be the equilibrium roadblock occupancy in the absence of motor activity, we can estimate the average roadblock occupancy as

$$\rho_{rb}^b = \rho_{rb}^{eq} \frac{p^b [\langle g \rangle_{tp}^b - \Delta]}{\langle g \rangle^b + \delta_m}. \quad (3.17)$$

In Figure 3.6 we illustrate the relationships derived for various observables and check our arguments against simulations of the BRM. In Figure 3.6 A-D we show the effect of varying the trans-peloton hopping rates for long roadblock equilibration times. As predicted by our analytical arguments, up to a critical motor density ρ_1 the velocity remains approximately constant (Figure 3.6 A), and the total current of motors grows linearly with the motor density (Figure 3.6 B), while the roadblock occupancy decreases linearly with motor density (Figure 3.6 C). At the critical density, and long before the track is completely covered by motors, all roadblocks are evicted. For motor densities above the critical density, the velocity and motor current follows the relationship for the TASEP without roadblocks (the ipTASEP) (Figure 3.6 A and B). In Figure 3.6 D we plot the typical peloton size up to the critical density, after which whole system acts as one large roadblock-excluding peloton. In Figure 3.6 E-H we vary the roadblock equilibration times. For rapid roadblock equilibration (red curves Figure 3.6 E-H) the roadblock shadow is small, gaps are largely filled with roadblocks, and the system is well described by a single TASEP with roadblocks in every gap (the tpTASEP) (dashed line in Figure 3.6 E). In this regime, the total density of roadblocks decreases weakly with motor density (red curve Figure 3.6 G), as roadblock shadows are small and the motor footprints are all that excludes the roadblocks. For intermediate roadblock equilibration times (the blue curves in Figure 3.6 E-H) the roadblock shadow is larger, resulting in peloton formation, a velocity that is less sensitive to motor density, and a system that is better at evicting roadblocks. The breakdown of our predictions for roadblock densities and peloton size in the case of fast roadblock binding (red and blue curves in Figure 3.6 G and H) is not surprising given that we here have small enough roadblock shadows to push the system outside the SPR.

3

3.4.4. ASYMPTOTIC BEHAVIOR IN THE SPR

We here show how Equations 3.9, 3.10, and 3.12 can be solved explicitly in terms of the motor density rather than velocity in the limit SPR. For notational convenience we introduce the average excess trans-peloton gap $\langle \tilde{g} \rangle_{tp}^b$ and the critical density ρ_0 , to write Equation (9) as

$$p^b = \frac{\langle \tilde{g} \rangle_{tp}^b \rho_0}{1 + \langle \tilde{g} \rangle_{tp}^b \rho_0}, \quad \langle \tilde{g} \rangle_{tp}^b = \langle g \rangle_{tp}^b - \Delta = \frac{\nu^b}{k_{ip} - \nu^b}, \quad \rho_0 = \frac{1}{\langle g \rangle_{ip}^b} \left(\frac{\nu^b}{k_{ip}} \right)^\Delta. \quad (3.18)$$

In the SPR, the transition density ρ_0 is very small by definition. In the limit $\rho_0 \langle \tilde{g} \rangle_{tp}^b \gg 1$, pelotons are small and p^b is close to one, and Equation 3.6 can be written as

$$1/\rho_m^b - \delta_m \approx \langle g \rangle_{ip}^b \quad (3.19)$$

Here we have a system controlled by the tpTASEP. In the limit $\rho_0 \langle \tilde{g} \rangle_{tp}^b \ll 1$, pelotons are large, p^b is small, and Equation 3.10 can be written as

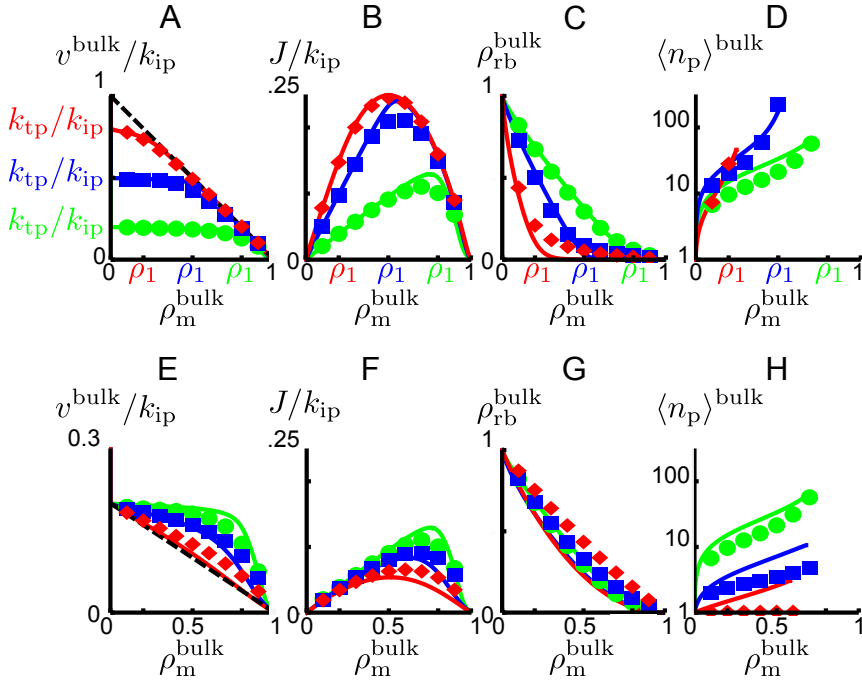


Figure 3.6: General model captures bulk dynamics of the BRM. Solid lines are analytical predictions and symbols are results from Monte Carlo simulations. Within the rows, each color represents the same parameter values, while green represents the same parameter values across all panels. A-D) Systems in to the SPR: the parameter values are $k_{ip}\tau = 20$ for all curves, with $k_{tp}/k_{ip} = 0.8$ (red), $k_{tp}/k_{ip} = 0.5$ (blue), and $k_{tp}/k_{ip} = 0.2$ (green). The dashed line in A) is the velocity relation for the ipTASEP. E-H) Sweep from stable to non-stable pelotons: the parameter values are $k_{tp}/k_{ip} = 0.2$ for all curves, and $k_{ip}\tau = 20$ (green), $k_{ip}\tau = 10$ (blue), $k_{ip}\tau = 2$ (red). The dashed line in E) is the velocity relation for the tpTASEP. In A) and E) we show the velocity, in B) and F) the motor current, in C) and G) the roadblock occupancy and in D) and H) the peloton size, all as a function of motor density. In the SPR, it only makes sense to talk about pelotons when there are roadblocks in the system ($\rho_m^{\text{bulk}} < \rho_1$), and we only show the burst size as a function of motor activity for these densities. For small enough roadblock shadows (red and blue line in Figure E-H) our description breaks down as the system falls outside the SPR.

$$1/\rho_m^b - \delta_m \approx \langle g \rangle_{ip}^b + \langle \tilde{g} \rangle_{ip}^b (\langle \tilde{g} \rangle_{ip}^b + \Delta) \rho_0. \quad (3.20)$$

If the first two terms on the right hand side dominate, we have the standard ipTASEP. If the last term dominates, we have a composite system. Taken together, there are three limits given by

$$\begin{aligned} \langle g \rangle_{ip}^b &\approx 1/\rho_m^b - \delta_m, & \rho_m^b &\ll \rho_0 \\ \langle \tilde{g} \rangle_{ip}^b &\approx \frac{1}{\sqrt{\rho_0 \rho_m^b}}, & \rho_0 &\ll \rho_m^b \ll \rho_1 = \frac{1}{\langle g \rangle_{ip}^b + \delta_m} \\ \langle g \rangle_{ip}^b &\approx 1/\rho_m^b - \delta_m, & \rho_1 &\ll \rho_m^b. \end{aligned} \quad (3.21)$$

In the middle regime is large, and the average velocity (see Equation 3.9) is close to k_{tp} ,

$$v^b = k_{tp} \left(1 - \sqrt{\rho_0 \rho_m^b} \right). \quad (3.22)$$

Using Equations 3.21 and 3.22, all observables can be written to leading order in $1/\rho_0$ as

$$\begin{aligned} v^b &= k_{tp}, \quad \langle g \rangle_{ip}^b = \frac{k_{tp}}{k_{ip} - k_{tp}}, \quad \langle g \rangle_{tp}^b = \frac{1}{\sqrt{\rho_0 \rho_m^b}}, \\ \langle n_p \rangle^b &= \sqrt{\frac{\rho_m^b}{\rho_0}}, \quad \frac{\rho_{rb}^b}{\rho_{rb}^{eq}} = 1 - \frac{\rho_m^b}{\rho_1}, \quad \rho_0 = \frac{k_{ip} - k_{tp}}{k_{tp}} \left(\frac{k_{tp}}{k_{ip}} \right)^{k_{ip}\tau + \delta_{rb}} \end{aligned} \quad (3.23)$$

In this regime, ($\rho_0 \ll \rho_m \ll \rho_1$), the average bulk peloton size is large, and that is why we refer to this as the stable peloton regime.

3.4.5. THE BULK STATE IS NEVER REACHED

With the parameter values in Table 3.1 in the main text, the bulk state is given by

$$\langle g \rangle_{ip}^b \approx 0.4, \quad \langle g \rangle_{tp}^b \approx 1 \frac{5 \cdot 10^{82}}{\sqrt{\rho_m}} \gg 1, \quad v^b \approx 3/s, \quad \langle n_p \rangle^b \approx 5 \cdot 10^{82} \sqrt{\rho_m} \gg 1. \quad (3.24)$$

From the above it is clear that the average steady-state peloton size $\langle n_p \rangle^b$ is in general enormous throughout the experimentally accessible range, and that the true bulk-dynamics will never be reached over a finite gene. Judging by the size of bulk pelotons, polymerases that meet along the gene stay together until termination, invariably producing bursts of RNA production.

3.4.6. INITIATION LIMITED DYNAMICS

Without roadblocks, and if the initiation rate k_{in} is limiting, the density and the distribution of motors at the start of the lattice are the same as in the bulk [27]. With the inclusion of roadblocks, the microscopic organization among motors and roadblocks can change from the start of the lattice to the bulk. Here we assume that the density of motors is low enough, such that once a motor is initiated it is only slowed down by other motors when it encounters a peloton. This condition means that once the initiation site is cleared, motors step away much faster at the beginning of the lattice than a new motor typically initiates. We will refer to this regime as the slow initiation regime, and we detail its extent below.

The formation of pelotons

A schematic kymograph for the TASEP with roadblocks and open boundary conditions is shown in Figure 6. At the start of the lattice, the time gaps between newly initiated motors are exponentially distributed with average time τ . As the motors move into the system, those motors that happened to have a roadblock bound ahead will start inducing peloton-forming traffic jams. For convenience we here call these jams proto-pelotons, and they will grow until all motors between roadblocks are absorbed into one

peloton. Once all motors are collected into pelotons, we will refer to these as the fully formed pelotons. We here set out to determine the nature of this peloton formation, and what effects it has on both motor and roadblock density profiles.

In all the expressions below, the superscript ‘in’ refers to the first site after the initiation site for which $x = 0$. For a roadblock to bind to the first site after a motor just left, the motor first has to take δ_{rb} steps and then a roadblock has to bind, all before another motor initiates. Considering the splitting probabilities for each step, we can write the probability of a roadblock binding between two motor initiation events as

$$p^{\text{in}} = \left(\frac{1}{1 + k_{\text{in}} \tau_{\text{m}}^{\text{in}}} \right)^{\delta_{\text{rb}}} \frac{1}{1 + k_{\text{in}} \tau}. \quad (3.25)$$

Here $\tau_{\text{m}}^{\text{in}}$ is the average time it takes a motor to take a step at the start of the track,

$$\tau_{\text{m}}^{\text{in}} = \frac{1 - p^{\text{in}}}{k_{\text{ip}}} + \frac{p^{\text{in}}}{k_{\text{tp}}}. \quad (3.26)$$

The definition of the slow initiation regime implies $\tau_{\text{m}}^{\text{in}} \ll 1/k_{\text{in}}$, and Equation 3.25 can be simplified as

$$p^{\text{in}} = \frac{e^{-\delta_{\text{rb}} \tau_{\text{m}}^{\text{in}} k_{\text{in}}}}{1 + k_{\text{in}} \tau}. \quad (3.27)$$

Equation 3.26 and 3.27 can be used to solve for p^{in} explicitly in the steady state,

$$p^{\text{in}} = \Lambda^{-1} W \left(\frac{k_{\text{in}} \tau}{1 + k_{\text{in}} \tau} \delta_{\text{rb}} \Lambda e^{-k_{\text{in}} \delta_{\text{rb}} / k_{\text{ip}}} \right), \quad \Lambda = \left(\frac{1}{k_{\text{tp}}} - \frac{1}{k_{\text{ip}}} \right) \frac{1}{\tau}, \quad (3.28)$$

where W is the Lambert W function. In the limit of low initiation rates we can write

$$p^{\text{in}} = \frac{1}{1 + k_{\text{in}} \tilde{\tau}}, \quad \tilde{\tau} = \tau + \delta_{\text{rb}} / k_{\text{tp}}. \quad (3.29)$$

Next we calculate the typical time it takes for n_{p}^* motors to aggregate into a peloton. Let Δx be the typical distance that the proto-peloton back end moves between two motors joining (Figure 3.7). From the geometry of typical times and distances sketched in the kymograph of Figure 3.7, we can write

$$\left. \begin{aligned} \Delta t k_{\text{ip}} &= \Delta x \\ \langle g \rangle_{\text{ip}}^{\text{in}} &= \frac{k_{\text{ip}}}{k_{\text{ip}} - k_{\text{tp}}} \\ (1/k_{\text{in}} + \delta_{\text{m}}/k_{\text{ip}} + \Delta t) k_{\text{tp}} &= \Delta x + \delta_{\text{m}} + \langle g \rangle_{\text{ip}}^{\text{in}} \end{aligned} \right\} \Rightarrow \Delta x = \frac{1}{k_{\text{in}} \tau \Lambda} - \delta_{\text{rb}}, \quad (3.30)$$

where Λ is given in Equation (25). In the last step above, we used the fact that we are considering the slow initiation regime, $k_{\text{in}} \ll k_{\text{ip}}, k_{\text{tp}}$. Proto-pelotons grow as more and more motor catch up, and again ignoring correlations, the final size n_{p}^* of a proto peloton is geometrically distributed as

$$P(n_p^*) = p^{\text{in}}(1 - p^{\text{in}})^{n_p^* - 1} \Rightarrow \langle n_p^* \rangle = 1/p^{\text{in}}. \quad (3.31)$$

Here the first factor in the probability function accounts for the probability of having a roadblock in a gap, and the following factors accounts for the probability of having no roadblocks in the preceding $n_p^* - 1$ gaps. The probability that a proto-peloton is still growing at position x , or equivalently the probability that $n_p^* > x/\Delta x$, is then given by

$$P_{\text{grow}}(x) = \sum_{n_p^* > x/\Delta x} p^{\text{in}}(1 - p^{\text{in}})^{n_p^* - 1} = (1 - p^{\text{in}})^{x/\Delta x}. \quad (3.32)$$

Letting $\langle n_p(x) \rangle$ be the size of the average proto-peloton at a distance from the initiation site, we can now write down the discrete evolution equation $\langle n_p(x + \Delta x) \rangle = \langle n_p(x) \rangle + 1 \cdot P_{\text{grow}}$, with $\langle n_p(0) \rangle = 1$, giving

$$\langle n_p(x) \rangle = \sum_{n=1}^{x/\Delta x} (1 - p^{\text{in}})^n + \langle n_p(0) \rangle = \langle n_p^* \rangle - \left[\langle n_p^* \rangle - 1 \right] e^{-x/x_p}, \quad x_p = -\Delta x / \ln(1 - p^{\text{in}}). \quad (3.33)$$

Though strictly only true for x in multiples of Δx , we take the above equation to be valid for any position $x \geq 0$.

The macroscopic effects of peloton formation

We now turn to calculate how the gradual growth of the proto pelotons impacts the motor density and velocity. The motor number density $\rho_m(x)$ at any position x is defined as the fraction of time that a site is occupied by (say) the front of a motor. The proto-peloton size at position x tells us that the fraction $\langle n_p(x) \rangle p^{\text{in}}$ motors take an average time $1/k_{\text{tp}}$ to step, while the rest take $1/k_{\text{ip}}$, giving the average stepping time

$$\tau_m(x) = \frac{\langle n_p(x) \rangle p^{\text{in}}}{k_{\text{tp}}} + \frac{1 - \langle n_p(x) \rangle p^{\text{in}}}{k_{\text{ip}}}, \quad \tau_m(0) = \tau_m^{\text{in}}. \quad (3.34)$$

The total time T between the seeding of two pelotons (Figure 3.7) averages over peloton sizes to

$$\langle T \rangle = \langle n_p^* \rangle (1/k_{\text{in}} + \delta_m \tau_m^{\text{in}}). \quad (3.35)$$

The fraction of time that the track is occupied by motors at position x is then

$$\rho_m = \frac{\langle n_p \rangle \tau_m(x)}{\langle T \rangle} = \frac{\tau_m(x)}{1/k_{\text{in}} + \delta_m \tau_m^{\text{in}}} = \rho_m^* + (\rho_m^{\text{in}} - \rho_m^*) e^{-x/x_p} \quad (3.36)$$

where the ρ_m^* denotes the motor density where the proto pelotons have fully formed. The average motor hopping rate can similarly be written as

$$k_m(x) = p^{\text{in}} \langle n_p(x) \rangle k_{\text{tp}} + (1 - p^{\text{in}} \langle n_p(x) \rangle) k_{\text{ip}} = k^* + (k^{\text{in}} - k^*) e^{-x/x_p}. \quad (3.37)$$

Though true for the TASEP, note that here the average motor hopping rate (Equation 3.37) is typically not the same as the average velocity $v(x) = \tau_m^{-1}(x)$ in our system. With

the velocity defined this way, it satisfies the standard relation $J = \rho_m(x)v(x)$ where $J = \langle n_p \rangle^* / \langle T \rangle$ is the flux of motors.

In Figure 3.7 we indicate the times and positions where there typically are roadblocks in pink. Moving away from the initiation site, the fraction of time a site is occupied by roadblocks grows because motors that have not yet caught up with a proto peloton move faster than the proto pelotons themselves. The moving front of equilibrating roadblocks (intersection of white and pink region in Figure 3.7) is typically offset with respect to the last motor of the peloton by a distance δ_{rb} (see dashed square in Figure 3.7). Taking the offset δ_{rb} into account, Equation 3.32 implies that a fraction $P_{grow}(x + \delta_{rb})$ of the proto pelotons is still evolving when the front of equilibrating roadblocks is at position x . The increase of the average time a site at position x (see Figure 3.7) is occupied by a roadblock then grows with distance as

$$t_{rb}(x + \Delta x) = t_{rb}(x) + (1/k_{tp} - 1/k_{ip}) \Delta x P_{grow}(x + \delta_{rb}), \quad t_{rb}(0) = 1/k_{in}. \quad (3.38)$$

This expression can be summed in the same manner as we previously summed to solve for $\langle n_p(x) \rangle$ in Equation 3.33, yielding

$$t_{rb}(x) = [1/k_{tp} - 1/k_{ip}] \Delta x e^{-\delta_{rb}/x_p} (\langle n_p^* \rangle - 1) (1 - e^{-x/x_p}) + 1/k_{in}. \quad (3.39)$$

The total fraction of time a site is covered by roadblocks is now

$$\rho_{rb}(x) = \rho_{rb}^{eq} \frac{t_{rb}(x)}{\langle T \rangle} = \rho_{rb}^* - (\rho_{rb}^{in} - \rho_{rb}^*) e^{-x/x_p}. \quad (3.40)$$

Relative changes are often easier to measure experimentally than absolute changes, therefore we here also give the relative changes in velocity and density at the beginning, and compared to well after the pelotons are formed

$$\begin{aligned} \frac{\rho_m^* - \rho_m^{in}}{\rho_m^*} &= (1 - p^{in}) \left(1 - \frac{k_{tp}}{k_{ip}} \right), & \frac{v^* - v^{in}}{v^*} &= -(1 - p^{in}) \left(\frac{k_{ip}}{k_{tp}} - 1 \right), \\ \frac{\rho_{rb}^* - \rho_{rb}^{in}}{\rho_{rb}^*} &= \frac{1}{1 + \frac{p^{in}}{1-p^{in}} \frac{\Delta x + \delta_m}{\Delta x} \cdot e^{\delta_{rb}/x_p}}. \end{aligned} \quad (3.41)$$

To have an appreciable motor density and velocity evolution, we need only the typical peloton to have a size of a few motors. In the limit of low initiation rates, the above can be written as

$$\begin{aligned} \frac{\rho_m^* - \rho_m^{in}}{\rho_m^*} &= \frac{k_{in} \bar{\tau}}{1 + k_{in} \bar{\tau}} \left(1 - \frac{k_{tp}}{k_{ip}} \right), & \frac{v^* - v^{in}}{v^*} &= -\frac{k_{in} \bar{\tau}}{1 + k_{in} \bar{\tau}} \left(\frac{k_{ip}}{k_{tp}} - 1 \right), \\ \frac{\rho_{rb}^* - \rho_{rb}^{in}}{\rho_{rb}^*} &= \frac{k_{in} \bar{\tau}}{e^{\delta_{rb}/x_p} + k_{in} \bar{\tau}}. \end{aligned} \quad (3.42)$$

We see that relative changes along the track grow in magnitude with the initiation rate, but that the effect saturates around $k_{in} = 1/\bar{\tau}$ for motor-density and velocity changes, while the roadblock density saturates later, around $k_{in} = e^{\delta_{rb}/x_p} / \bar{\tau}$. Interestingly, we see

that the total shift in motor density and velocity along the track is set solely by the ratio of motor stepping rates with and without roadblocks ahead.

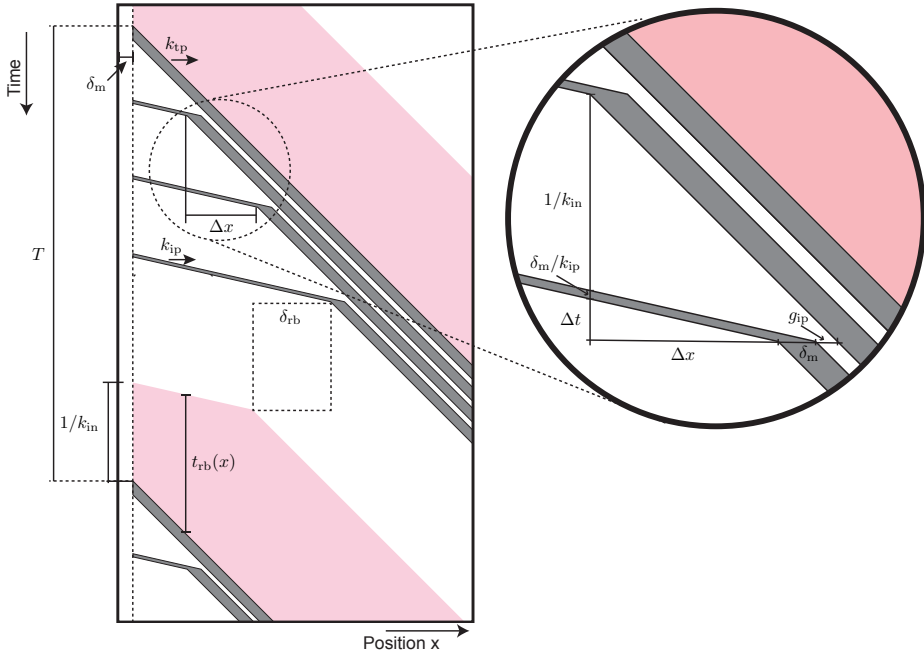


Figure 3.7: A schematic kymograph of the system with open boundary conditions. Motors (grey) initiate on the left side of the lattice, with an initiation site free of roadblocks. As motors travel into the system, their speed depends on if there is a roadblock (pink) ahead of them. With a roadblock ahead, the motor speed is k_{tp} , while without roadblock ahead it is k_{ip} . The typical time that a roadblock at position x is bound before being evicted by the next peloton is given by $t_{rb}(x)$. The right figure is a magnification of the region defining Δx , which is the typical distance a motor travels to catch up with the proto peloton since the last motor caught up.

3.4.7. BURSTS FROM TERMINATING PELOTONS

The peloton-forming dynamics of our model will manifest as burst of motor activity if viewed from a specific position (for example the transcription-termination site). Using the average velocity of the system, we can translate the average gap-sizes to average time gaps between motors arrivals

$$\tau_{ip} = \frac{\langle g \rangle_{ip} + \delta_m}{v}, \quad \tau_{tp} = \frac{\langle g \rangle_{tp} + \delta_m}{v}. \quad (3.43)$$

Letting k_{tr} be the rate of reaction when the process is in the on state, k_{off} be the rate at which the system transitions to the off state, and k_{on} be the rate at which the system transitions back to the on state, we can relate our first-principles model to the phenomenological two-state model traditionally used to describe transcriptional bursts (Figure 3.4 C) [53]. As both models generate double exponentials, we relate them to each other by equating the time constant and the corresponding relative probabilistic weight. Since

we are interested in describing transcriptional bursts, we here consider the limit where the two-state model gives clearly separated bursts, $k_{\text{on}} \ll k_{\text{off}} + k_{\text{tr}}$. In this limit we have, to leading order in k_{on} ,

$$p = \frac{k_{\text{off}}}{k_{\text{off}} + k_{\text{tr}}}, \quad \tau_{\text{ip}} = \frac{1}{k_{\text{off}} + k_{\text{tr}}}, \quad \text{and} \quad \tau_{\text{tp}} = \frac{1}{k_{\text{on}}} \left(1 + \frac{k_{\text{off}}}{k_{\text{tr}}} \right). \quad (3.44)$$

The relations in Equation 3.44 can be inverted to give

$$k_{\text{tr}} = \frac{1-p}{\tau_{\text{ip}}}, \quad k_{\text{off}} = \frac{p}{\tau_{\text{ip}}}, \quad \text{and} \quad k_{\text{on}} = \frac{1}{(1-p)\tau_{\text{tp}}}. \quad (3.45)$$

For the bulk we can in principle calculate the corresponding two-state model using Equation 3.9, though this state is likely never reached. The physiologically more interesting situation is just after the initial pelotons have fully formed. The relative probability of a gap being between pelotons is then (see Equation 3.29)

$$p = p^{\text{in}} = \frac{1}{1 + k_{\text{in}}\tilde{\tau}}. \quad (3.46)$$

Once the initial pelotons have formed beyond x_p , we know that the average velocity is k_{tp} , and we have

$$\langle g \rangle_{\text{ip}}^* = \frac{k_{\text{tp}}}{k_{\text{ip}} - k_{\text{tp}}} \Rightarrow \tau_{\text{ip}}^* = \frac{1}{k_{\text{ip}} - k_{\text{tp}}} + \frac{\delta_m}{k_{\text{tp}}}. \quad (3.47)$$

The trans-peloton gaps can be written as (see Figure 3.7)

$$\begin{aligned} \langle T \rangle &= \tau_{\text{tp}}^* + \tau_{\text{ip}}^* (\langle n_p \rangle^* - 1) \\ \Rightarrow \tau_{\text{tp}}^* &= \langle T \rangle - \tau_{\text{ip}}^* k_{\text{in}}\tilde{\tau} = \left(\frac{1}{k_{\text{in}}} + \frac{\delta_m}{k_{\text{tp}}} \right) + \left(\frac{1}{k_{\text{in}}} - \frac{1}{k_{\text{ip}} - k_{\text{tp}}} \right) k_{\text{in}}\tilde{\tau} \end{aligned} \quad (3.48)$$

Combining Equations 3.45-3.48 we have

$$k_{\text{tr}} = \frac{k_{\text{in}}\tilde{\tau}}{1 + k_{\text{in}}\tilde{\tau}} \frac{1}{\tau_{\text{ip}}^*}, \quad k_{\text{off}} = \frac{1}{1 + k_{\text{in}}\tilde{\tau}} \frac{1}{\tau_{\text{ip}}^*}, \quad \text{and} \quad k_{\text{on}} = \frac{1 + k_{\text{in}}\tilde{\tau}}{k_{\text{in}}\tilde{\tau}} \frac{1}{\tau_{\text{tp}}^*}. \quad (3.49)$$

In the main text we are interested in the case where motors are large and the initiation rate is low compared to motor stepping rates. In this limit we have

$$k_{\text{tr}} = \frac{k_{\text{in}}\tilde{\tau}}{1 + k_{\text{in}}\tilde{\tau}} \frac{k_{\text{tp}}}{\delta_m}, \quad k_{\text{off}} = \frac{1}{1 + k_{\text{in}}\tilde{\tau}} \frac{k_{\text{tp}}}{\delta_m}, \quad \text{and} \quad k_{\text{on}} = \frac{1}{\tilde{\tau} + \frac{k_{\text{in}}\tilde{\tau}}{1 + k_{\text{in}}\tilde{\tau}} \frac{\delta_m}{k_{\text{tp}}}} \approx \frac{1}{\tilde{\tau}}. \quad (3.50)$$

Here the approximate relation should be valid in the case of transcription through nucleosomes as here the roadblock are substantially larger than the motors, and $\tilde{\tau}$ is consequently substantially larger than δ_m/k_{tp} .

3.4.8. MONTE CARLO SIMULATIONS

We validate our heuristic arguments using a random-sequential-update Monte Carlo scheme with fixed time step dt to simulate our model. During a Monte Carlo step on a lattice of size $L+1$ there are $L+1$ possible events: all the motors on the lattice can make a step forward, bound roadblocks can unbind, roadblocks in solution can bind to an empty lattice site, and a motor can bind at the start of the lattice. The time step is chosen small enough that the probability of any event occurring with rate k in a time dt can be approximated as kdt . In our simulation k_{ip} is the fastest rate, and we choose $k_{ip}dt = 0.1$. We verified that our results are robust towards changes in dt . Without roadblocks, the time to equilibration for periodic systems scales with the system size as $L^{3/2}$ [28]. With roadblocks, the time to equilibration is expected to be larger due to the slow peloton dynamics. For the simulations with periodic boundary conditions we waited L^2/dt iterations for the system to reach steady state and let simulations run a total amount of $10L^2/dt$ iterations, and checked that the peloton size did not change for longer equilibration times. The velocities presented in Figure 3.5 and 3.6 are calculated by averaging over the instantaneous hopping rates of the motors.

REFERENCES

- [1] D. Chowdhury, A. Schadschneider, and K. Nishinari, *Physics of transport and traffic phenomena in biology: From molecular motors and cells to organisms*, *Physics of Life Reviews* **2**, 318 (2005), 0509025 .
- [2] M. A. Hoyt, A. A. Hyman, and M. Bähler, *Motor proteins of the eukaryotic cytoskeleton*. *Proceedings of the National Academy of Sciences of the United States of America* **94**, 12747 (1997).
- [3] J. Howard, *Mechanics of Motor Proteins and the Cytoskeleton* (Sinauer Associates, Sunderland, Massachusetts, 2001) p. 367.
- [4] D. S. Goodsell, *The Machinery of Life* (Springer Sciences & Business Media, 2009).
- [5] L. Ciandrini, M. C. Romano, and A. Parmeggiani, *Stepping and crowding of molecular motors: Statistical kinetics from an exclusion process perspective*, *Biophysical Journal* **107**, 1176 (2014), arXiv:1312.1911 .
- [6] I. J. Finkelstein and E. C. Greene, *Molecular traffic jams on DNA*. *Annual review of biophysics* **42**, 241 (2013).
- [7] W. Lee, D. Tillo, N. Bray, R. H. Morse, R. W. Davis, T. R. Hughes, and C. Nislow, *A high-resolution atlas of nucleosome occupancy in yeast*. *Nature genetics* **39**, 1235 (2007).
- [8] S. S. Teves, C. M. Weber, and S. Henikoff, *Transcribing through the nucleosome*, *Trends in Biochemical Sciences* **39**, 577 (2014).
- [9] C. Thiriet and J. J. Hayes, *Replication-independent core histone dynamics at transcriptionally active loci in vivo*, *Genes and Development* **19**, 677 (2005).
- [10] M. F. Dion, T. Kaplan, M. Kim, S. Buratowski, N. Friedman, and O. J. Rando, *Dynamics of Replication-Independent Histone Turnover in Budding Yeast*, *Science* **315**, 1405 (2007).
- [11] A. Jamai, R. M. Imoberdorf, and M. Strubin, *Continuous Histone H2B and Transcription-Dependent Histone H3 Exchange in Yeast Cells outside of Replication*, *Molecular Cell* **25**, 345 (2007).
- [12] A. Rufiange, P.-E. Jacques, W. Bhat, F. Robert, and A. Nourani, *Genome-Wide Replication-Independent Histone H3 Exchange Occurs Predominantly at Promoters and Implicates H3 K56 Acetylation and Asf1*, *Molecular Cell* **27**, 393 (2007).
- [13] H. A. Cole, J. Ocampo, J. R. Iben, R. V. Chereji, and D. J. Clark, *Heavy transcription of yeast genes correlates with differential loss of histone H2B relative to H4 and queued RNA polymerases*. *Nucleic acids research* **42**, 12512 (2014).
- [14] C. Hodges, L. Bintu, L. Lubkowska, M. Kashlev, and C. Bustamante, *Nucleosomal fluctuations govern the transcription dynamics of RNA polymerase II*. *Science (New York, N.Y.)* **325**, 626 (2009).

- [15] V. Epshtein and E. Nudler, *Cooperation Between RNA Polymerase Molecules in Transcription Elongation*, *Science* **300**, 801 (2003).
- [16] H. Trenchard, *The peloton superorganism and protocooperative behavior*, *Applied Mathematics and Computation* **270**, 179 (2015).
- [17] B. Albert, I. Léger-Silvestre, C. Normand, M. K. Ostermaier, J. Pérez-Fernández, K. I. Panov, J. C. B. M. Zomerdijk, P. Schultz, and O. Gadal, *RNA polymerase I-specific subunits promote polymerase clustering to enhance the rRNA gene transcription cycle*. *The Journal of cell biology* **192**, 277 (2011).
- [18] F. Harper and F. Puvion-Dutilleul, *Non-Nucleolar Transcription Complexes of Rat Liver as Revealed by Spreading Isolated Nuclei*, *J Cell Sci* **40**, 181 (1979).
- [19] C. D. Laird and W. Y. Chooi, *Morphology of transcription units in Drosophila melanogaster*, *Chromosoma* **58**, 193 (1976).
- [20] S. L. Mcknight and O. L. Miller, *Post-replicative D. melanogaster Nonribosomal Embryos Transcription Units in*, *Cell* **17**, 551 (1979).
- [21] K. Tantale, F. Mueller, A. Kozulic-Pirher, A. Lesne, J.-M. Victor, M.-C. Robert, S. Capozzi, R. Chouaib, V. Bäcker, J. Mateos-Langerak, X. Darzacq, C. Zimmer, E. Basyuk, and E. Bertrand, *A single-molecule view of transcription reveals convoys of RNA polymerases and multi-scale bursting*, *Nature Communications* **7**, 12248 (2016).
- [22] S. Shivaswamy, A. Bhinge, Y. Zhao, S. Jones, M. Hirst, and V. R. Iyer, *Dynamic remodeling of individual nucleosomes across a eukaryotic genome in response to transcriptional perturbation*, *PLoS Biology* **6**, 0618 (2008).
- [23] A. Weiner, A. Hughes, M. Yassour, O. J. Rando, and N. Friedman, *High-Resolution Nucleosome Mapping Reveals Transcription-Dependent Promoter Packaging*, *Genome Research* , 90 (2010).
- [24] I. Golding, J. Paulsson, S. M. Zawilski, and E. C. Cox, *Real-time kinetics of gene activity in individual bacteria*. *Cell* **123**, 1025 (2005).
- [25] M. Kaern, T. C. Elston, W. J. Blake, and J. J. Collins, *Stochasticity in gene expression: from theories to phenotypes*. *Nature reviews. Genetics* **6**, 451 (2005).
- [26] L.-H. So, A. Ghosh, C. Zong, L. a. Sepúlveda, R. Segev, and I. Golding, *General properties of transcriptional time series in Escherichia coli*. *Nature genetics* **43**, 554 (2011).
- [27] C. T. MacDonald, J. H. Gibbs, and A. C. Pipkin, *Kinetics of Biopolymerization on Nucleic Acid Templates*, *Biopolymers* **6**, 1 (1968).
- [28] R. A. Blythe and M. R. Evans, *Nonequilibrium steady states of matrix-product form: a solver's guide*, *Journal of Physics A: Mathematical and Theoretical* **40**, R333 (2007).

- [29] M. Dobrzynski and F. J. Bruggeman, *Elongation dynamics shape bursty transcription*, *PNAS* **106**, 2583 (2009).
- [30] S. Klumpp, *Pausing and Backtracking in Transcription Under Dense Traffic Conditions*, *Journal of Statistical Physics* **142**, 1252 (2011).
- [31] S. Klumpp and T. Hwa, *Stochasticity and traffic jams in the transcription of ribosomal RNA: Intriguing role of termination and antitermination*. *Proceedings of the National Academy of Sciences of the United States of America* **105**, 18159 (2008).
- [32] A. Kunwar, A. John, K. Nishinari, A. Schadschneider, and D. Chowdhury, *Collective traffic-like movement of ants on a trail: dynamical phases and phase transitions*, *Journal of the Physical Society of Japan* **73**, 2979 (2004).
- [33] A. Parmeggiani, T. Franosch, and E. Frey, *Totally asymmetric simple exclusion process with Langmuir kinetics*, *Physical Review E* **70**, 046101 (2004), [arXiv:0408034 \[cond-mat\]](https://arxiv.org/abs/0408034).
- [34] A. Schadschneider, D. Chowdhury, and K. Nishinari, *Stochastic Transport in Complex Systems* (Elsevier B.V., 2011).
- [35] M. Sahoo, J. Dong, and S. Klumpp, *Dynamic blockage in an exclusion process*, *Journal of Physics A: Mathematical and Theoretical* **48**, 015007 (2015).
- [36] F. Turci, A. Parmeggiani, E. Pitard, M. C. Romano, and L. Ciandrini, *Transport on a lattice with dynamical defects*, *Phy* **012705**, 1 (2013).
- [37] O. J. O. Loan, M. R. Evans, and M. E. Cates, *Jamming transition in a homogeneous one-dimensional system : The bus route model*, *Physical Review E* **58**, 1404 (1998).
- [38] B. Derrida, M. R. Evans, V. Hakim, and V. Pasquier, *Exact solution of a 1D asymmetric exclusion model using a matrix formulation*, *Journal of Physics A: Mathematical and General* **26**, 1493 (1993).
- [39] S. J. Greive and P. H. V. Hippel, *Thinking quantitatively about transcriptional regulation*. *Nature reviews. Molecular cell biology* **6**, 221 (2005).
- [40] K. Luger, a. W. Mäder, R. K. Richmond, D. F. Sargent, and T. J. Richmond, *Crystal structure of the nucleosome core particle at 2.8 Å resolution*. *Nature* **389**, 251 (1997).
- [41] K. Brogaard, L. Xi, J.-P. Wang, and J. Widom, *A map of nucleosome positions in yeast at base-pair resolution*. *Nature* **486**, 496 (2012).
- [42] K. C. Neuman, E. A. Abbondanzieri, R. Landick, J. Gelles, and S. M. Block, *Ubiquitous Transcriptional Pausing Is Independent of RNA Polymerase Backtracking*, *Cell* **115**, 437 (2003).
- [43] L. Bintu, T. Ishibashi, M. Dangkulwanich, Y.-Y. Wu, L. Lubkowska, M. Kashlev, and C. Bustamante, *Nucleosomal elements that control the topography of the barrier to transcription*. *Cell* **151**, 738 (2012).

- [44] X. Darzacq, Y. Shav-Tal, V. de Turris, Y. Brody, S. M. Shenoy, R. D. Phair, and R. H. Singer, *In vivo dynamics of RNA polymerase II transcription*. *Nature structural & molecular biology* **14**, 796 (2007).
- [45] V. Pelechano, S. Chavez, and J. E. Perez-Ortin, *A Complete Set of Nascent Transcription Rates for Yeast Genes*, *PloS one* **5**, e15442 (2010), arXiv:1203.2655 .
- [46] J. Mazurkiewicz, J. F. Kepert, and K. Rippe, *On the mechanism of nucleosome assembly by histone chaperone NAPI*. *The Journal of biological chemistry* **281**, 16462 (2006).
- [47] A. Worcel, S. Han, and M. L. Wong, *Assembly of newly replicated chromatin*. *Cell* **15**, 969 (1978).
- [48] M. a. Schwabish and K. Struhl, *Evidence for eviction and rapid deposition of histones upon transcriptional elongation by RNA polymerase II*. *Molecular and cellular biology* **24**, 10111 (2004).
- [49] G. M. Cooper, *The Cell: A Molecular Approach*. 2nd edition (Sinauer Associates, 2000).
- [50] G.-C. Yuan, L. Y-J, M. F. Dion, M. D. Slack, L. F. Wu, A. J. Altshuler, and O. J. Rando, *Genome-Scale Identification of Nucleosome Positions in *S. cerevisiae**, *Science* **309**, 626 (2005).
- [51] K. P. N. Murthy and G. M. Schutz, *Aging in two- and three-particle annihilation processes*, *Physical Review E* **57**, 1388 (1998).
- [52] T. Lenstra, J. Rodriguez, H. Chen, and D. R. Larson, *Transcription Dynamics in Living Cells **, *Annu. Rev. Biophys* **45**, 25 (2016).
- [53] B. Munsky, G. Neuert, and A. van Oudenaarden, *Using gene expression noise to understand gene regulation*. *Science* **336**, 183 (2012).
- [54] J. Peccoud and B. Ycart, *Markovian modelling of gene product synthesis*, *Theoretical Population Biology* **48**, 222 (1995).
- [55] A. Raj, C. S. Peskin, D. Tranchina, D. Y. Vargas, and S. Tyagi, *Stochastic mRNA Synthesis in Mammalian Cells*, *PLoS biology* **4** (2006), 10.1371/journal.pbio.0040309.
- [56] J. M. Raser and E. K. O. Shea, *Control of Stochasticity in Eukaryotic Gene Expression Jonathan*, *Science* **304**, 1811 (2006), arXiv:NIHMS150003 .
- [57] D. Angelov, V. a. Bondarenko, S. Almagro, H. Menoni, F. Mongélard, F. Hans, F. Mietton, V. M. Studitsky, A. Hamiche, S. Dimitrov, and P. Bouvet, *Nucleolin is a histone chaperone with FACT-like activity and assists remodeling of nucleosomes*. *The EMBO journal* **25**, 1669 (2006).
- [58] R. Belotserkovskaya, S. Oh, V. a. Bondarenko, G. Orphanides, V. M. Studitsky, and D. Reinberg, *FACT facilitates transcription-dependent nucleosome alteration*. *Science* **301**, 1090 (2003).

- [59] M. L. Kireeva, W. Walter, V. Tchernajenko, V. Bondarenko, M. Kashlev, and V. M. Studitsky, *Nucleosome remodeling induced by RNA polymerase II: Loss of the H2A/H2B dimer during transcription*, *Molecular Cell* **9**, 541 (2002).
- [60] O. I. Kulaeva, F.-K. Hsieh, and V. M. Studitsky, *RNA polymerase complexes cooperate to relieve the nucleosomal barrier and evict histones*. *Proceedings of the National Academy of Sciences* **107**, 11325 (2010).
- [61] A. Kristjuhan and J. Q. Svejstrup, *Evidence for distinct mechanisms facilitating transcript elongation through chromatin in vivo*. *The EMBO journal* **23**, 4243 (2004).
- [62] C.-K. Lee, Y. Shibata, B. Rao, B. D. Strahl, and J. D. Lieb, *Evidence for nucleosome depletion at active regulatory regions genome-wide*. *Nature genetics* **36**, 900 (2004).
- [63] L. J. Core and J. T. Lis, *Transcription regulation through promoter-proximal pausing of RNA polymerase II*. *Science (New York, N.Y.)* **319**, 1791 (2008).
- [64] L. J. Core, J. Waterfall, and J. T. Lis, *Nascent RNA Sequencing Reveals Widespread Pausing and Divergent Initiation at Human Promoters*, *Science* **322**, 1845 (2008).
- [65] I. Jonkers and J. T. Lis, *Getting up to speed with transcription elongation by RNA polymerase II*. *Nature reviews. Molecular cell biology* **16**, 167 (2015).
- [66] P. P. Dennis, M. Ehrenberg, D. Fange, and H. Bremer, *Varying rate of RNA chain elongation during rrn transcription in Escherichia coli*, *Journal of Bacteriology* **191**, 3740 (2009).
- [67] C. G. Danko, N. Hah, X. Luo, A. L. Martins, L. Core, J. T. Lis, A. Siepel, and W. L. Kraus, *Signaling Pathways Differentially Affect RNA Polymerase II Initiation, Pausing, and Elongation Rate in Cells*, *Molecular Cell* **50**, 212 (2013).
- [68] S. Chong, C. Chen, H. Ge, and X. S. Xie, *Mechanism of Transcriptional Bursting in Bacteria*, *Cell* **158**, 314 (2014).
- [69] J. R. Chubb, T. Trcek, S. M. Shenoy, and R. H. Singer, *Transcriptional Pulsing of a Developmental Gene*, *Current Biology* **16**, 1018 (2006).
- [70] G.-w. Li, X. S. Xie, and T. Hirschfeld, *Central dogma at the single-molecule level in living cells*, *Nature* **475**, 308 (2011).
- [71] A. Raj and A. van Oudenaarden, *Nature, nurture, or chance: stochastic gene expression and its consequences*. *Cell* **135**, 216 (2008).
- [72] D. Zenklusen, D. R. Larson, and R. H. Singer, *Single-RNA counting reveals alternative modes of gene expression in yeast*. *Nature structural & molecular biology* **15**, 1263 (2008).
- [73] M. S. Luijsterburg, M. F. White, R. van Driel, and R. T. Dame, *The Major Architects of Chromatin: Architectural Proteins in Bacteria, Archaea and Eukaryotes*, *Critical Reviews in Biochemistry and Molecular Biology* **43**, 393 (2008).

- [74] L. F. Liu and J. C. Wang, *Supercoiling of the DNA template during transcription*, *Proceedings of the National Academy of Sciences* **84**, 7024 (1987).
- [75] M. T. J. van Loenhout, M. V. de Grunt, and C. Dekker, *Dynamics of DNA Supercoils*, *Science* **338**, 94 (2012).
- [76] M. Voliotis, N. Cohen, C. Molina-París, and T. B. Liverpool, *Fluctuations, pauses, and backtracking in DNA transcription*. *Biophysical journal* **94**, 334 (2008).
- [77] J. W. Shaevitz, E. a. Abbondanzieri, R. Landick, and S. M. Block, *Backtracking by single RNA polymerase molecules observed at near-base-pair resolution*. *Nature* **426**, 684 (2003).
- [78] J. Jin, L. Bai, D. S. Johnson, R. M. Fulbright, M. L. Kireeva, M. Kashlev, and M. D. Wang, *Synergistic action of RNA polymerases in overcoming the nucleosomal barrier*. *Nature structural & molecular biology* **17**, 745 (2010).

4

PRINCIPLES OF HISTONE REPLACEMENT IN THE WAKE OF A TRANSCRIBING RNAP

DNA is covered by histone proteins and other DNA binding proteins that ensure genome integrity, but also have to provide access to transcription and replication machinery. To fulfill both tasks, histones are dynamic proteins that can be modified or replaced, depending on regulatory processes in the cell. Transcription is often tightly coupled to replacement of histones by other histone variants or DNA binding proteins, but the underlying mechanism has never been examined in detail. Using a theoretical model of RNA polymerases transcribing a gene with two competing proteins, we show that transcription activity can spontaneously change the relative density of DNA binding proteins. The change in protein occupancy forms a memory of transcription activity that depends both on the equilibration time and the size difference of the competing DNA binding proteins. Our model gives testable predictions for a mechanism of transcription-coupled chromatin changes on many types of genes and in many different organisms.

4.1. INTRODUCTION

Coverage of DNA by histones is essential for protection of DNA against radiation, as well as for structuring DNA for gene regulation [1]. Other than protecting DNA, histones are essential for transcription regulation and there is a variety of processes that regulate accessibility to replication and transcription machinery. Over the coding region of active genes there is an increased histone turnover [2–5] and histones are reorganized [6, 7]. Furthermore there are many examples where the relative occupancy of DNA binding proteins changes and (part of) the histone is replaced by more dynamic proteins. The histone variant H3.3 for example accumulates on active chromatin [8, 9] while histones are replaced by the smaller and more dynamic HMG box proteins on the highly transcribed ribosomal genes [10]. Though initial replacement of histones is fast, the equilibrium coverage is not directly restored after transcription stops. Instead both H3.3 and HMG box proteins are only removed during cell division [11, 12], thereby constituting a long memory of transcription activity that distinguishes genes that have been activated from silent loci [13]. The exchange of nucleosomal histones by new variants lowers chromatin stability, making it more permissive to transcription, while ensuring chromatin integrity on actively transcribed genes [9].

In higher organisms, transcription elongation is associated with the replacement of the histone variant H3 with H3.3 [13–15] and HMG box protein does not bind without transcription [12]. These observations suggest that histone replacement and turnover takes place in the wake of passing polymerase that evicts DNA binding proteins, thereby freeing up space for other protein species to bind [8, 9, 16] (Figure 4.1 A and B). Though this mechanism can explain why histone replacements are tightly coupled to transcription, it has not been studied whether additional regulation is necessary for the observed protein exchange and the memory of transcription activity and what the requirements are on protein binding dynamics.

To elucidate the conditions for histone exchange in the wake of a transcribing polymerase we studied a minimal theoretical model of transcribing polymerases evicting two competing DNA binding proteins. We focus our analysis on three different physiologically relevant scenarios: two competing DNA binding proteins with an equal footprint (for example H3 competing with H3.3), protein complexes that bind in steps (histone octamers) and two proteins with different footprints (for example a histone and HMG box protein).

We show that the relative occupancy of DNA binding proteins on the DNA is controlled both by the density and organization of polymerases along a gene. As described in Chapter 3, polymerases spontaneously organize into pelotons during transcription. In this chapter we show that histones, which are evicted in steps [17], are efficiently evicted on heavily transcribed genes in the presence of peloton formation. Further, histones are replaced in the wake of a passing polymerase by histone variants that bind faster or by smaller proteins such as HMG box protein. Interestingly, the memory of transcription activity depends on the relative size of competing DNA binding proteins, giving an unforeseen function for the footprint of small proteins like HMG box protein as compared to a histone. The dynamic replacement of histones during transcription is a robust mechanism, independent of specific regulatory factors, that can explain both the fast replacement of histones during transcription and the long memory of transcription

activity, making it a plausible mechanism for transcription-coupled chromatin changes.

4.2. RESULTS

4.2.1. A GENERAL MODEL FOR TRANSCRIPTION THROUGH TWO COMPETING DNA BINDING PROTEINS

Transcription and the interaction of polymerases with nucleosomes and other obstacles can be modeled by extending the Totally Asymmetric Simple Exclusion Process (TASEP) [18] to include one [19] or multiple ([20],[21]) obstacles along the track. To model transcription and two competing DNA proteins, we extended the TASEP further to include two types of obstacles (Figure 4.1 C), A and B, that compete for space with polymerases. On bare DNA, polymerases elongate with rate k_t and they initiate with rate k_{in} when the promoter is empty. The binding and unbinding rates for the DNA binding proteins are $k_b^{A,B}$ and $k_u^{A,B}$ respectively and a polymerase is slowed down to a rate $k_t^{A,B}$ when it faces a region of equilibrated protein coverage. The proteins occupy $\delta_{A,B}$ base pairs and the polymerase δ_m base pairs. Since the initiation site, the promoter, is typically free of nucleosomes on active genes [22, 23], the proteins in our model cannot bind to the first δ_m lattice sites.

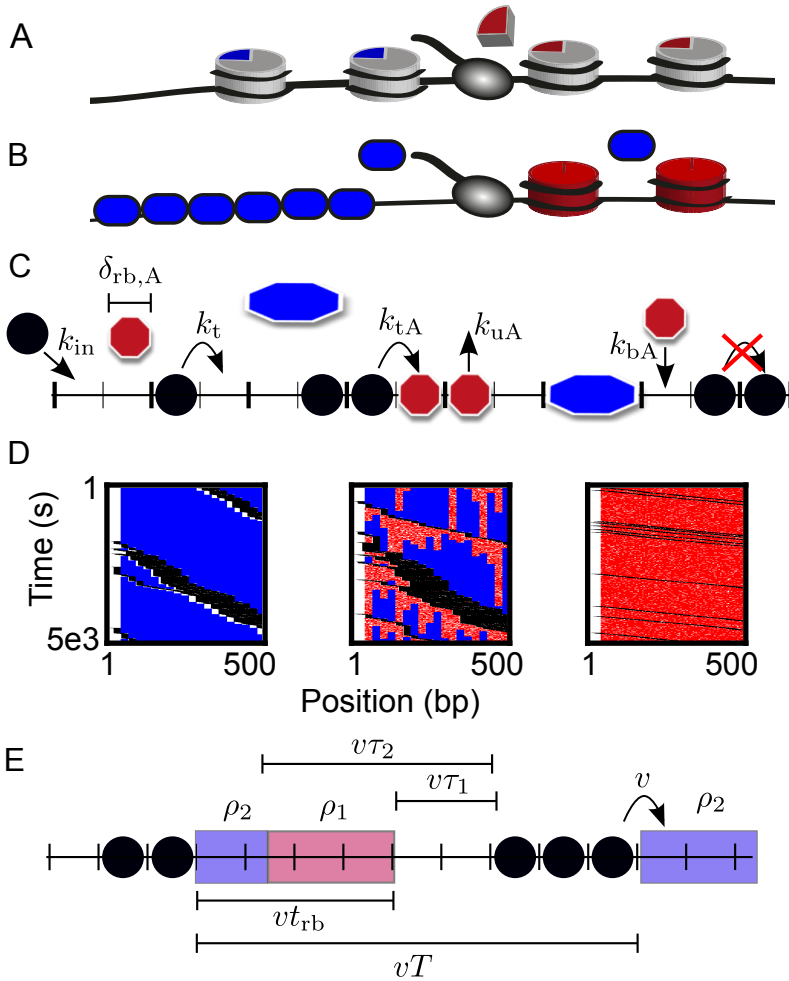


Figure 4.1: A model for competing DNA binding proteins interacting with RNA polymerase on a gene. A) An example of a situation where part of a protein is evicted by RNAP: the replacement of H3.3 histones by H3 histones during transcription. B) An example of a situation where large proteins are replaced by smaller proteins during transcription: HMG box protein replaces histones on ribosomal genes [10]. C) The complete model with all the parameters. For clarity we only labeled the rates of roadblock species A. D) Kymographs of polymerases (black) interacting with one or two DNA binding proteins. Motors initiate on the left and roadblock shadows can be seen as the absence of proteins behind the motors (white). In the left panel, there is only one type of protein (blue) and motors spontaneously form pelotons due to non-specific interactions with roadblocks, as described in Chapter 3. The results were obtained using Monte Carlo simulations (see the Supplementary Material) with parameters $k_{ip} = 1/s$, $k_b = 1/s$, $k_u = 10^{-6}$, $k_{dyad} = 0.12$, $k_{in} = 0.001$, $\delta_{rb} = 25$ and $\delta_m = 35$. In the middle panel, we added a second smaller (red) protein which delays binding of the larger (blue) protein, the chosen parameters are: $\delta_{rb} = 5$, $k_b^s = 1$ and $k_u^s = 0.4$. The small red proteins only hinder binding of the large ones, but do not slow down the polymerase: $k_{tp}^s = k_{ip}$. In the right panel, the unloading rate of the small protein has been decreased, such that the larger blue one does not have time to bind between passing motors. Since the small red proteins form no obstacle to the polymerase, the pelotons have disappeared. E) Illustration of the relevant effective parameters in the system. After proteins are evicted by a passing motor, the protein density relaxes over two time scales, after a time τ_1 the density is ρ_1 and after a time τ_2 the density is equilibrated to ρ_2 , Equation 4.1. The average density of proteins depends on the distance between pelotons, T , and the fraction of time between pelotons occupied by proteins, t_{rb} , Equation 4.3. The times between passing polymerases are translated to distance through the polymerase speed, v .

4.2.2. REBINDING DYNAMICS OF TWO COMPETING PROTEINS AFTER POLYMERASE PASSAGE

A qualitative discussion of protein binding already reveals that the model in Figure 4.1 C can lead to a change in relative occupancy of the two competing proteins. After the proteins are evicted by a polymerase, their binding to the DNA equilibrates with one or multiple relaxation time scales. If the proteins were to relax over a single time scale, the relative density would not change and such a process could not explain the shift in occupancy as seen after transcription activation. Consider therefore two competing proteins for which the joined relaxation can be described by two well-separated relaxation time scales, τ_1 and τ_2 . After eviction of both proteins at $t = 0$ by a polymerase, the proteins will compete for space on the DNA and the protein density over time can approximately be written as

$$\boldsymbol{\rho}(t) = \begin{cases} \mathbf{0}, & 0 \ll t \ll \tau_1 \\ \boldsymbol{\rho}_1, & \tau_1 \ll t \ll \tau_2 \\ \boldsymbol{\rho}_2, & \tau_2 \ll t. \end{cases} \quad (4.1)$$

Here $\boldsymbol{\rho}_{1,2} = (\rho_{1,2}^A, \rho_{1,2}^B)$ is a vector containing the respective densities of the two proteins, τ_1 is the relaxation time into the meta-stable state with density $\boldsymbol{\rho}_1$ and τ_2 is the total equilibration time, setting the duration of memory of transcriptional activity on the gene. The protein densities and the relaxation times are functions of the dissociation constants, binding rates and the sizes of the proteins as will be specified later in this chapter.

During transcription, the time between two passing polymerases sets what density the roadblocks have time to relax to before being evicted again. If this time is smaller than τ_2 , the proteins do not have time to equilibrate between two passing polymerases, resulting in a steady-state occupancy of DNA binding proteins that depends on transcription intensity. When transcription levels are even higher, such that the average time between passing polymerases is smaller than τ_1 , the gene would be cleared of proteins. Before quantifying how the protein densities depend on the initiation rate, we first make several simplifications to the model.

4.2.3. PHYSIOLOGICAL LIMITS OF THE MODEL

The competitive binding of proteins with different sizes to DNA is a complicated problem with a long history [24]. Adsorption of two proteins with different sizes has been studied [25] as well as adsorption and desorption of a single protein [26], but the full dynamics of proteins with different sizes that associate and dissociate from a lattice has yet to be solved. Including transcription complicates the problem even further. To narrow down the problem, we here explore the physiologically relevant limits.

First, since a change in relative occupancy will only occur if the proteins equilibrate over multiple time scales, we only consider equilibration over two well-separated time scales.

Second, since the halflife of histones and HMG box protein is longer than cell cycle (Table 4.1) we only consider small unbinding rates, $k_{\text{ub}} \ll k_{\text{b}}$, such that the equilibration time τ of a single protein is effectively set by the binding time, $\tau \approx 1/k_{\text{b}}$.

Third, nucleosomes and other DNA binding proteins rebind quickly after eviction by a passing RNAP [5], which results in a regular packing with high packing densities behind a polymerase [27]. Other mechanisms such as sliding of nucleosomes, or cooperative binding of HMG box proteins [28] also ensure regular spacing and high packing densities [29]. We model the fast rebinding and dense packing by dividing the lattice into sections with the size of the DNA binding proteins, and let proteins equilibrate to each section. For a comparison of our model to other models of proteins binding to DNA, see the Supplementary Material.

Fourth, the distribution of RNAPs along a gene depend on the strength of the interaction between RNAP and the proteins (chapter 3). If proteins form weak obstacles, they can be ignored. We assume that at least one of the DNA binding proteins slows down a polymerase significantly, $k_{tp}/k_t < 0.5$, which applies to nucleosomes (Table 4.1).

Lastly, we only consider initiation-limited systems. The TASEP has multiple phases that depend on the initiation and termination rate, but transcription is typically in the initiation-limited phase [30], which means that the initiation rate controls the output. Within the initiation-limited phase, we consider the limit where initiation is low enough that polymerases do not hinder each other during initiation. As discussed in the following sections, the solution in this limit gives good predictions, even for initiation rates that are considered high for physiological conditions, while strongly simplifying the solution.

Further note that all the binding rates in the main text are effective binding rates per binding site with units (s^{-1}), accounting for protein concentration and remodeling factors. The dissociation constants are unitless and defined as the ratio of the unbinding rate and the effective binding rate.

4.2.4. RNAP PELOTON FORMATION

To calculate the occupancies for both proteins on the DNA we need to know the typical time between two passing polymerases, which is determined by the speed of a polymerase and the gap sizes between polymerases. For typical transcription rates, at most one polymerase occupies a gene at a time [31] such that the time between passing polymerases is set by the initiation rate. For high transcription levels where multiple polymerases transcribe the same gene at a time, one would naively expect distances between polymerases to be geometrically distributed when initiation is exponentially distributed [32]. However, polymerases self-organize into pelotons as they move along a gene and are slowed down by and evict proteins through non-specific interactions [20, 21] (Figure 4.1 D left panel). For two competing species of DNA binding proteins, polymerases can also form pelotons, see the kymograph in Figure 4.1 D middle panel. We refer to [21] and chapter 3 for an extensive discussion on peloton formation and here only outline the argument, starting with the solution for a single protein species.

4.2.5. GENERAL SOLUTION FOR PROTEIN COVERAGE DURING ACTIVE TRANSCRIPTION IN THE PRESENCE OF ONE PROTEIN SPECIES

Consider RNAPs that initiate with a rate k_{in} . Some RNAPs will have an obstacle right in front of them when initiating, such that they elongate with a slow rate k_t^A and evict DNA binding proteins. When the next RNAP initiates before a protein can rebind, it hops with a faster rate k_t . The fast motor will catch up with the slow one, forming a peloton

(Figure 4.1 D left panel). As pelotons form along the gene, the average RNAP speed goes down, and therefore the RNAP density goes up. Since the distance over which pelotons are formed is small compared to a gene length (Figure 4.1 D) for physiological rates we ignore this regions effect on the overall density.

Polymerase pelotons are separated by DNA binding proteins and the average protein density on the lattice depends on the typical time between two passing pelotons that a site is occupied by proteins, t_{rb} , and the total time, T , between passage of two pelotons (Figure 4.1 E),

$$\langle \rho \rangle \approx \rho_{\text{eq}} t_{\text{rb}} / T. \quad (4.2)$$

Here T and t_{rb} depend on the parameters of the model (see Equation 4.15 and 4.18 respectively in the Supplementary Material). The average protein density in Equation 4.2 decreases with an increasing initiation rate or decreasing polymerase speed, since more polymerases exclude proteins. More interestingly, the protein coverage goes up when the polymerase density remains the same, but the peloton size increases, because there is more space between pelotons for proteins to bind. The peloton size changes independently of the polymerase density when the protein binding rate k_b decreases or when the protein size δ_{rb} increases (see the Supplementary Material).

4.2.6. GENERAL SOLUTION FOR PROTEIN COVERAGE DURING ACTIVE TRANSCRIPTION IN THE PRESENCE OF TWO DNA BINDING PROTEINS

For two protein species, the solution is very similar as for one protein species, except that the composition of proteins between pelotons changes over time. For high enough initiation rates, when t_{rb} is less than $\tau_2 - \tau_1$, protein densities only relax to the intermediate density ρ_1 while for lower initiation rates, there is a mix of ρ_1 and ρ_2 . Hence, the average density of DNA binding proteins is given by

$$\langle \rho \rangle \approx \begin{cases} \rho_1 t_{\text{rb}} / T, & t_{\text{rb}} + \tau_1 < \tau_2 \\ \left(\rho_1 (\tau_2 - \tau_1) + \rho_2 (t_{\text{rb}} + \tau_1 - \tau_2) \right) / T, & t_{\text{rb}} + \tau_1 \geq \tau_2 \end{cases} \quad (4.3)$$

where $\langle \rho \rangle = (\langle \rho^A \rangle, \langle \rho^B \rangle)$ is a vector containing the average densities of the two protein species.

The average density in Equation 4.3 depends on the polymerase speed. In the presence of obstacles the velocity of pelotons is limited by the front motor [20, 21], such that the average velocity of a peloton is the same as the velocity of a single polymerase that is hindered by roadblocks. Unequal eviction rates for the two proteins gives highly complex dynamics as the speed of a peloton depends on the distance to the next peloton. However, when one of the proteins barely hinders the polymerase, or when the protein binding is much slower than the time for a polymerase to clear enough space for a protein to bind, we can still give an accurate description (see the Supplementary Material). When the eviction rates are equal, $k_t^A = k_t^B$, and unbinding rates are low, such that equilibrated proteins have a density close to one, we can write [20] (see the Supplementary Material)

$$v \approx k_t^A \quad (4.4)$$

Apart from the velocity, the average protein densities in Equation 4.3 also depend on the densities and relaxation time scales that were introduced in Equation 4.1. In the following sections we give those quantities in three physiologically relevant situations: proteins that bind in steps, transcription in the presence of competing equally sized proteins and competing proteins with different sizes.

4.2.7. TRANSCRIPTION IN THE PRESENCE OF COMPETING DNA BINDERS OF EQUAL SIZE.

Consider two equally sized competing DNA binding proteins, Figure 4.2 A, which could correspond to, for example, two nucleosome variants with different binding properties or a nucleosome dimer competing with a transcription factor [33, 34]. The relaxation dynamics of protein binding behind the polymerase is given by

$$\rho(t) = -\rho_1 e^{-(t-\delta_{rb}/v)/(\tau_1-\delta_{rb}/v)} + (\rho_1 - \rho_2) e^{-(t-\delta_{rb}/v)/(\tau_2-\delta_{rb}/v)} + \rho_2. \quad (4.5)$$

An example of Equation 4.5 is plotted in Figure 4.2 B. For short times after eviction, the faster binding red species dominates, while in equilibrium the blue protein species, which has a smaller dissociation constant, dominates. For well-separated relaxation time scales τ_1 and τ_2 and the other approximation described in Section 4.2.3, Equation 4.5 can be approximated as in Equation 4.3.

The densities of the two proteins in equilibrium are given by

$$\rho_2^{A,B} = \frac{1/K_D^{A,B}}{1/K_D^B + 1/K_D^A + 1}, \quad (4.6)$$

where $K_D^{A,B}$ are the dissociation constants of protein A and B respectively. The intermediate densities and relaxation times are reduced in the limit of large time scale separation and for slow unbinding to

$$\rho_1^{A,B} \approx \frac{k_b^{A,B}}{k_b^A + k_b^B} \quad (4.7)$$

$$\tau_1 \approx (k_b^A + k_b^B)^{-1} + \delta_{rb}/v, \quad (4.8)$$

$$\tau_2 \approx \left(\frac{k_b^B}{k_b^B + k_b^A} k_u^A + \frac{k_b^A}{k_b^B + k_b^A} k_u^B \right)^{-1} + \delta_{rb}/v, \quad (4.9)$$

where the velocity, v , is given in Equation 4.4 and the full derivations are given in the Supplementary Material. Since we here consider equally sized roadblocks, the roadblock size δ_{rb} drops out when plugging Equations 4.8 and 4.9 into Equation 4.3, making the average density independent of the protein size. Equation 4.7 shows that the intermediate densities and the shortest time scale τ_1 are a function of the binding rates only, such that the *fastest* rather than the *strongest* binder dominates for high transcription levels. This

change in protein occupancy with transcription is illustrated in Figure 4.2 C. Initially proteins are equilibrated, but after transcription starts, the occupancy quickly switches. The protein occupancy then remains constant for a constant initiation rate, and after transcription stops, the protein occupancy slowly relaxes to the equilibrated densities, over a timescale τ_2 .

The two components of the density in Equation 4.3 are plotted in Figure 4.2 D as a function of initiation rate together with Monte Carlo simulations. Since the relaxation time scales are well separated, the intermediate densities are constant over several orders of magnitude of the initiation rate. Note that, despite the complexity of the model and our simplifying assumptions, we have a quantitative agreement between simulations (dots) and calculations (solid lines).

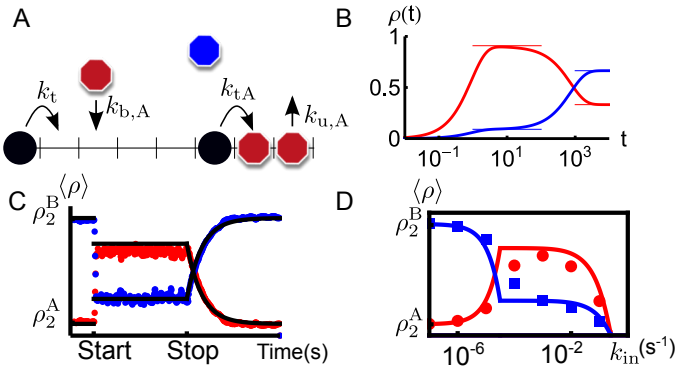


Figure 4.2: DNA coverage for transcription through two equally sized DNA binding proteins. A) Schematic of the model with parameters. For clarity we only indicate the rates relating to the red roadblock species, A . B) The evolution of the protein densities over time after eviction at $t = 0$, Equation 4.5. The horizontal bars are the intermediate and equilibrium protein densities given by Equation 4.6 and 4.7 respectively. C) In equilibrium, the blue species dominates, while during active transcription, between 'start' and 'stop' the red species dominates. When transcription ends, the equilibrium densities are restored. Blue and red dots are simulations and the solid black line are analytical predictions from Equations 4.6, 4.7 and 4.9. The parameter values are: $k_{b,A} = 0.01 k_t$, $K_{D,A} = 0.01$ (red), $k_{b,B} = 0.004 k_t$, $K_{D,B} = 0.001$ (blue), $k_{tA} = k_{tB} = 0.2 k_t$, $k_{in} = 0.001 k_t$, $\delta_{rA,B} = 1$. D) Protein densities for proteins that occupy one lattice site with $k_b^A = 0.01 k_t$, $K_D^A = 0.01$ (red) and $k_b^B = 0.004 k_t$, $K_D^B = 0.001$ (blue). The roadblocks are evicted by a passing motor with rate $k_{tA} = k_{tB} = 0.2 k_t$.

4.2.8. TRANSCRIPTION IN THE PRESENCE OF PROTEINS BINDING IN MULTIPLE STEPS

Histones assemble in steps: first two H3-H4 dimers form a tetrasome on the DNA, then two H2A-H2B dimers attach to form the complete histone octamer [36]. RNAP also evicts histones in steps: a single RNAP evicts an H2A-H2B dimer [43] while a trailing RNAP can evict the remaining hexamer [17]. The model needed to describe proteins that bind in steps is different from the one needed for two competing proteins. For the latter, a passing polymerase always leaves behind an empty lattice site, while the type of protein behind an RNAP for two-step binding and eviction depends on the distance to the RNAP ahead. When the distance is large enough for a dimer to have time to rebind, the trail-

Table 4.1: Properties of DNA binding proteins and transcription

Microscopic parameter	Value	Citation
Binding H2A/H2B dimer	2-10 minutes	[35, 36]
Residence time H2B	>60 min	[37]
Nucleosome with linker DNA	167 bp	[38]
Memory ribosomal genes	> 6 hours	[12]
k_b H3 histone	< 1 minute	[36]
halflife chromatin bound H3.3 histone (flies)	24 hours	[8]
halflife chromatin bound H3 histone (mice)	> 100 days	[39]
halflife HMG chromatin bound HMG	> cell cycle	[12]
Footprint HMG	26 bp	[40]
Footprint RNAP	35 bp	[41]
Footprint nucleosome + linker DNA	167 bp	[38, 42]
Transcription intensity (yeast)	10^{-4} -0.05/s	[31]

ing RNAP encounters a full histone and once again only evicts a dimer, but when two RNAPs follow closely, the trailing RNAP encounters a hexamer and leaves behind bare DNA (Figure 4.3 A).

However, under physiological conditions, transcription in the presence of histones effectively reduces to the model with only one obstacle. Hexamers rebind fast behind a transcribing polymerase, such that we can neglect the time that DNA is bare between pelotons. The transcription dynamics for histones that bind in steps is also similar as for only one obstacle. Since hexamers form less strong obstacles than full histones [17], the RNAP in front of a peloton, which faces full histones, always limits the speed of the rest of the peloton and pelotons are stable, see the kymograph in Figure 4.3 B where we used experimentally observed dimer and hexamer binding rates, Table 4.1.

The velocity of a peloton is $v = k_{tp}^{\text{dimer}}$, with k_{tp}^{dimer} the eviction rate of a dimer. In a nucleosomal region, the nucleosome dyad forms the largest obstacle to transcription [44], therefore polymerases in our model are only slowed down at the center of a protein and evict that protein with rate k_{dyad} , in such a way that the average rate in a nucleosomal region equals k_{tp}^{dimer} , $k_{tp}^{\text{dimer}} = \delta_{rb} / ((\delta_{rb} - 1) / k_t + 1 / k_{dyad})$.

The equilibration time and intermediate densities for hexamers can be approximated with

$$\tau = \delta_{rb} / v + 1 / k_b^{\text{dimer}} + 1 / k_b^{\text{hexamer}}, \quad (4.10)$$

$$(\rho_1^{\text{hexamer}}, \rho_1^{\text{histone}}) = (1, 0), \quad (\rho_2^{\text{hexamer}}, \rho_2^{\text{histone}}) = (0, 1), \quad (4.11)$$

where τ is the time in which full histones are formed after eviction and k_b^{dimer} and k_b^{hexamer} are the binding rate for dimers and hexamers respectively. Since histones have long lifetimes (Table 4.1), we ignore unbinding. The densities of histones and hexamers are ρ_i^{histone} and ρ_i^{hexamer} respectively. Using Equation 4.2 and 4.11 we can calculate the density of hexamers, dimers and empty lattice sites (Figure 4.3 C).

For moderate transcription rates, Table 4.1, there is a reduced occupancy of dimers,

while for high transcription rates both dimers and hexamers are depleted, agreeing with *in vivo* observations [45]. Due to the formation of pelotons, histones are efficiently evicted as soon as two or more polymerases occupy a gene. Further, the formation of pelotons, leaves more space on the gene for histones to reassemble as compared to the situation where RNAPs would be randomly spread over the gene. For high transcription rates, the gap between pelotons decreases and the theoretical predictions and simulations start to deviate (Figure 4.3 C). The time for a hexamer to rebind between two pelotons is not negligible compared to the time that hexamers are bound, such that we can no longer ignore the contribution of bare DNA to the overall density. Further, by only taking into account the average gap between pelotons, we underestimate the number of full histones for high transcription. For high transcription rates histone dynamics can no longer be approximated as that of a single binding protein.

We can also calculate the dimer and hexamer density over time for a gene that is switched on and off (Figure 4.3 D). When transcription stops, RNAP first has to reach the end of the gene before histones can fully relax. Depending on the gene length, this time can be comparable to the equilibrium time of dimers. For long genes, the time for RNAP to reach the end of the gene has to be taken into account to determine the histone relaxation time.

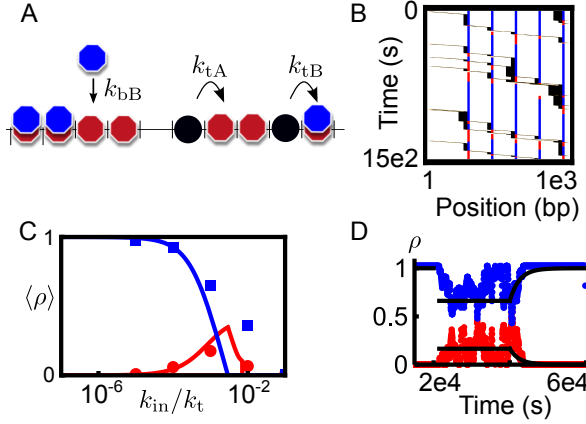


Figure 4.3: Relative protein densities and motor velocity for proteins that bind in two steps. A) A blue protein (eg a dimer) binds on top of a red one (eg a hexamer). A polymerase encountering a fully formed protein only evicts the blue protein, leaving behind a red one. A second passing polymerase evicts the red protein leaving behind an empty lattice site. B) Kymographs of transcription through histones using parameters from Table 4.1, generated using Monte Carlo simulations. Hexamers are shown in red and full histones in blue. In the left panel, the eviction rate for hexamers and full histones is the same such that the RNAP in front of a peloton often runs away from the peloton. In the right panel, full histones form stronger obstacles to transcription than hexamers such that the speed of the RNAP in front of a peloton is rate-limiting and pelotons are stable. The input parameters are $k_b^{\text{dimer}} = 0.02/s$, $k_u^{\text{dimer}} = 10^{-6}/s$, $k_b^{\text{hexamer}} = 10/s$, $k_u^{\text{hexamer}} = 10^{-6}/s$, $k_{ip} = 10/s$, $k_{in} = 0.01/s$, $k_t^{\text{dimer}} = 2/s$, $k_t^{\text{hexamer}} = 2.5/s$, $\delta_m = 35bp$, $\delta_{rb} = 167bp$ C) Average density of full histones (blue), hexamers (red) and bare DNA (black) as a function of the initiation rate. All rates are normalized by the bare motor stepping rate, k_t . Dots are results of Monte Carlo simulations and the solid lines are the theoretical predictions in Equation 4.11 and [21]. The input parameters are $k_b^{\text{dimer}} = 0.02/s$, $k_u^{\text{dimer}} = 10^{-6}/s$, $k_b^{\text{hexamer}} = 10/s$, $k_u^{\text{hexamer}} = 10^{-6}/s$, $k_{ip} = 10/s$, $k_t^{\text{dimer}} = 2/s$, $k_t^{\text{hexamer}} = 4/s$, $\delta_m = 35bp$, $\delta_{rb} = 167bp$ D) Hexamer (red) and full histone (blue) densities over time. The gene is turned on at $t = 2e4$ s and turned off at $t = 4e4$ s. The input parameters are the same as in C) with an initiation rate of $k_{in} = 0.005/s$.

4.2.9. TRANSCRIPTION IN THE PRESENCE OF COMPETING DNA BINDERS OF UNEQUAL SIZE.

Histones do not only compete with other histone variants, but also with other protein species with different sizes. HMG box protein, for example, has a footprint that is six times smaller than that of a histone, Table 4.1. Consider therefore two roadblocks with different sizes where the smallest roadblock is N times smaller than the large roadblock, $N \geq 1$ (Figure 4.4 A). To investigate the effect of size difference, we keep the binding energy per base pair constant while scaling the dissociation constant of the small roadblock with N , $\epsilon_s = \epsilon_l/N$, where ϵ_s and ϵ_l are the binding energy for the small and the large protein respectively. Equilibrium coverages are given by (see the Supplementary Material)

$$\frac{\rho_2^s}{\rho_2^l} = \frac{1/K_D^s(1 + 1/K_D^s)^{N-1}}{1/K_D^l}. \quad (4.12)$$

For equally sized protein, Equation 4.12 reduces to Equation 4.6. Interestingly, Equation 4.12 shows that even when the binding energies per basepair are the same for the

large and the small protein (see the Supplementary Material), the small protein will dominate, because it has more ways to bind.

In the limit where the small protein species binds fast behind a passing polymerase ($1/k_b^s \ll \delta_{rb}/v$) and equilibrates between pelotons before a large one binds, the intermediate density and the smallest relaxation time scale (Equation 4.1) are respectively given by

$$\rho_1 = \frac{1}{1 + K_D^s}, \quad \tau_1 = \delta_{rb}^s/v. \quad (4.13)$$

Full relaxation of two proteins with different sizes can be described by a birth death process [46] (see the Supplementary Material). Though we could solve for the transcription memory for arbitrary protein dynamics, we here are interested in the case in which small proteins bind and equilibrate before the large protein attempts to bind. The time until a large roadblock binds can then be approximated by

$$\tau_2 = (P_0(N)k_b^l)^{-1}, \quad (4.14)$$

where $P_0(N)$ is the equilibrium probability to have zero small proteins bound at the binding site of a large protein in the absence of polymerases (Supplementary Material) and k_b^l is the binding rate of a large protein. By combining Equation 4.3, 4.12 and 4.13 we can plot the density of the small and large protein as a function of initiation rate, Figure 4.4 B, showing a good agreement between simulations and theory.

Interestingly the transcription memory for competing proteins with different sizes increases with the size difference N (Figure 4.4 D), because N small proteins have to unbind, before a large protein can bind (compare the two panels in Figure 4.4 C).

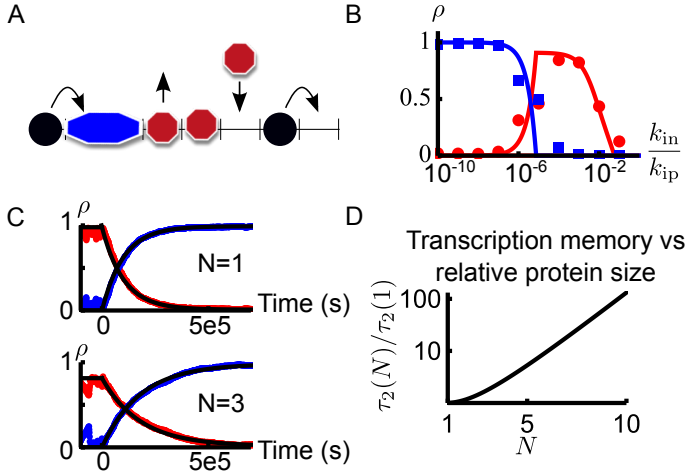


Figure 4.4: DNA coverage and memory for transcription through DNA binding proteins with different sizes. A) Schematic of the model, the blue protein is n times larger than the red. We here only consider the situation where the smaller protein binds first after eviction by a polymerase and the larger protein dominates in equilibrium. B) The protein density for a protein occupying 1 lattice site (red) and a protein occupying 5 lattice sites. Our theoretical predictions given in Equations 4.12-4.14 (solid lines) agree well with MC simulations (dots). The parameter values for the simulation are: $k_t = 1/s$, $k_b^s = 1/s$, $k_u^s = 10^{-1}/s$, $k_b^l = 10^{-2}/s$, $k_u^l = 10^{-7}/s$, $k_{\text{dyad}}^s = 0.3$, $k_{\text{dyad}}^l = 0.3$, $k_{\text{dyad}}^1 = 0.125$, $\delta_b^s = 1$, $\delta_l = 3$, $\delta_m = 35$. Motors are only slowed down in the middle of a protein (for the nucleosome that corresponds to the dyad), such that the average time to hop through a protein is $(\delta_{\text{rb}} - 1)/k_t + 1/k_{\text{dyad}}$ and the average speed through large and small proteins is the same in this simulation. C) The protein density over time for proteins that are equal in size ($N = 1$), or for proteins that differ a factor three in size ($N = 3$). Initially, there is active transcription. Then, at Time=0, transcription stops and proteins levels start to equilibrate. When the size difference between the two competing proteins is larger, equilibration takes longer. D) The memory of active transcription for proteins that differ a factor N in size ($\tau_2(N)$ in Equation 4.14) divided by the memory for two proteins of equal size ($\tau_2(N = 1)$ in Equation 4.14). Note that the y-axis has a log scale. The binding rate and the energy per base pair are kept constant (Supplementary Material).

4.3. DISCUSSION

To understand the process of histone replacement in the wake of a polymerase, we introduced a model of transcription through two competing DNA binding proteins and proteins that bind in steps. We studied three types of transcription-coupled chromatin changes: histone exchange, two-step histone eviction and assembly in steps and competition of histones with smaller proteins. A central assumption in our model is that relaxation of proteins after eviction happens over two well-separated time scales (Equation 4.1). Since initial rebinding of proteins after eviction by RNAP is typically on the order of minutes, while spontaneous unbinding of proteins is on the order of hours or even days (Table 4.1), the time scales τ_1 and τ_2 are indeed well-separated for competing protein species (Equation 4.7 and Equation 4.13). As a consequence, it is the fastest or smallest binder that dominates first after histone eviction, rather than the strongest binder. The relative occupancy of many DNA binding proteins can therefore be controlled by changing the transcription rate. The binding time of hexamers and dimers is both on the order of minutes, such that our predictions for step-wise binding of histones

is less accurate. This is likely a cause for the discrepancy between theoretical predictions and simulations for high transcription rates in Figure 4.3 C.

Interestingly, addition of a protein that binds faster, delays binding of the first protein, giving a longer memory of polymerase passage (compare Figure 4.1 D left and middle panel). A memory of transcription activity could also have been achieved by having only one protein species with very slow equilibration. However, adding a second protein has the advantage that it allows for both a long memory of transcription, large τ_2 , and a high coverage and integrity of the DNA, small τ_1 , (Equation 4.1). Furthermore, the memory for two competing proteins in the limits discussed in this chapter is long as long as unbinding is slow, independent of the protein concentration, which makes the memory robust against protein fluctuations in the cytosol.

Below we discuss the implications of our results.

4.3.1. CHROMATIN CHANGES DUE TO POLYMERASE EVICTION ARE FAST AND SPECIFIC

Changing the relative occupancy of DNA binding proteins can be achieved through several mechanisms: changes in protein concentration or protein binding properties result in a different (equilibrium) stoichiometry, while exchange in the wake of the polymerase results in a temporary switch in relative occupancy, while the equilibrium coverage remains the same.

Concentration changes depend on the slow reorganization of proteins (Equations 4.9, 4.11 and 4.14) as the equilibration time for DNA binding proteins is often longer than the cell cycle, Table 4.1. In contrast, for exchange in the wake of a polymerase, the elongation and initiation rate of the polymerase set the reorganization time. Therefore changes in protein occupation due to eviction by a polymerase is more specific and faster than changes induced by varying the protein concentrations or nucleosome remodelers in the cytosol. Further transcription induced protein exchange changes allow for gene specific changes as transcription factors can target one gene at a time, while changes of concentrations or remodeling factors in the cytosol could easily target neighboring genes.

There are multiple examples where the cell exploits the fast and specific exchange in the wake of a polymerase. Eviction of a histone dimer leaves enough space for transcription factors to bind and transcription triggers site specific binding of transcription factors, like RSS proteins [34] essential for the adaptive immune system, or Rad26p, involved in transcription-coupled DNA repair [33]. Interestingly, transcription coupled binding of Rad26p involves transcription on non-coding genes, suggesting that constitutive transcription of non-coding genes [47] could play a role in transcription factor binding.

4.3.2. INCREASING ACCESSIBILITY OF DNA TO CRISPR-CAS USING TRANSCRIPTION

The tight control of chromatin remodeling to transcription has a potential use in drug targeting. The CRISPR-Cas system discovered in 2012 is part of the bacterial immune system and is widely studied now as it allows for fast and specific gene editing [48]. However, in eukaryotes nucleosomes constrain CRISPR-Cas function [49, 50]. The access of

DNA to CRISPR-Cas has been increased experimentally using chromatin remodeling factors [49]. Targeted transcription activity could also swiftly and specifically increase the accessibility of a gene by evicting dimers resulting in accessibility for several minutes, Table 4.1, giving a possibly more specific gene targeting than chromatin changes due to nucleosome remodelers only.

4.3.3. CHROMATIN CHANGES OF RIBOSOMAL GENES

The total production of ribosomal RNA accounts for more than half of the total amount of transcripts produced. The high initiation rate and speed of Pol I and the structure of ribosomal genes allow for such high transcription rates. Ribosomal genes exist in multiple chromatin states, open and closed, that correlate with transcription activity: open genes are covered by High Mobility Group (HMG) box proteins, while closed genes form a more dense structure with nucleosomes (reviewed in [10]). After transcription ends, active genes do not immediately return to a closed chromatin state [12]. Our predictions agree with these observations: proteins are only exchanged during active transcription and there is a long memory of transcription activity (Figures 4.2 C, 4.3 D and 4.4 C). High transcription results in competition between histones and HMG box protein. Even when histones and HMG box protein assemble at a similar rate, HMG box protein dominates on the DNA, because this protein has a smaller footprint, Table 4.1. The smaller size of HMG box protein, allows for this protein to bind weakly, while preserving a long memory of transcription. A more weakly binding protein likely hinders the polymerase less than a nucleosome allowing higher elongation rates.

4.3.4. CONCLUSION AND OUTLOOK

By studying a theoretical model for transcription in the presence of two competing proteins, we have shown that protein replacement in the wake of a transcribing RNAP is a mechanism for fast and specific protein exchange that is tightly coupled to transcription. Despite the complexity of the model, we are able to accurately predict protein densities that we obtained from Monte Carlo simulations. Our predictions agree with the observations on transcription-induced chromatin changes as described in the introduction of this chapter.

The model can be extended to m competing proteins and n relaxation time scales, giving more than two possible protein compositions on the DNA, depending on the initiation rate. Such an extension of the model would make it possible to describe a wider range of (biological) systems or give more accurate predictions for the systems discussed in this chapter. For modeling the dynamics of protein exchange on ribosomal genes, for example, we ignored that histones are evicted in steps. Full modeling of ribosomal genes, would require at least three protein levels in Equation 4.3. For low transcription the gene is occupied with complete nucleosomes, for intermediate transcription nucleosomes without dimers and for high transcription nucleosomes are evicted and replaced by HMG box proteins.

Before extending the model, however, more detailed experiments are necessary to further test the applicability of the model to transcription. A system with DNA curtains would be ideal for testing the model as it provides a lot of control [51]. The protein densities on DNA could be measured as a function of transcription intensity to test whether

the predictions of our model agree quantitatively with experimental observations. Our calculations show that, in theory, chromatin changes can be controlled by transcription dynamics only, without changes in protein concentrations. Future studies can reveal to what extent these predictions capture the *in vivo* dynamics for transcription-coupled chromatin changes.

ACKNOWLEDGMENTS

We thank Joachim Griesenbeck and Helmut Schiessel for helpful discussions on chromatin structure. MD and AAvdB acknowledge financial support from a TU Delft startup grant to MD. AAvdB further acknowledges financial support by the Netherlands Organisation for Scientific Research (NWO/OCW), as part of the Frontiers of Nanoscience program.

4.4. SUPPLEMENTARY MATERIAL

We here give the derivations and arguments that were left out of the main text for clarity.

4.4.1. MODELING PROTEIN BINDING

Nucleosome formation is a complicated process as histones form in multiple steps [36] and DNA is tightly bend. Yet, experiments have shown a single dominant time scale in nucleosome formation [52, 53]. Models for protein binding to a lattice (DNA) often assume that protein binding is reaction-limited and that the effective rate of protein binding to an empty lattice increases linearly with every added lattice site [24, 54]. For large lattices, the effective binding rate for such models only weakly depends on the lattice size, but for the small lattices, the scaling becomes unrealistic. If a lattice for example has the size of the footprint of a histone and increases by only one basepair, the binding rate doubles. Since the pieces of lattice between polymerases in our model can be small, we here choose a scaling that depends more weakly on lattice size. By dividing the lattice into binding sites with the size of the proteins, the effective binding rate naturally scales with the lattice size as $k_{b,eff} = k_b N/L$ with N the size of the lattice (in base pair) and L the footprint of a protein. This simple exponential relaxation to separate binding sites will not capture the details of histone binding, but will result in high histone coverage as observed in experiments.

4.4.2. PELOTON FORMATION

When RNAP evicts a DNA binding protein it takes time for the RNAP to clear enough space for the protein to rebind and the protein has a finite binding time. Therefore there is a region behind every RNAP that is depleted of proteins. Following [21] we refer to this region as the roadblock shadow. A trailing RNAP at a distance less than the roadblock shadow experiences less hinder from DNA binding proteins and catches up with the first RNAP to form a peloton. The size and stability of a peloton depends on the details of the system. Since histones have a large footprint, RNAPs attract each other from a long distance resulting in large pelotons. Furthermore histones slow down RNAPs significantly, such that RNAPs form stable pelotons during transcription [21]. We here give the expressions in the limit of strong peloton formation and refer to [21] for the full deriva-

tions

To calculate the average density on a gene after pelotons have formed (Equation 4.3) we need the typical time between two passing pelotons T and the typical time t_{rb} that a DNA site is occupied by proteins between the passage of pelotons. When initiation is slow enough such that motors do not hinder each other during initiation, $\tau_{\text{m}}^{\text{in}} \ll 1/k_{\text{in}}$, with $\tau_{\text{m}}^{\text{in}}$ the motor stepping time near the initiation site, we can write the total distance between pelotons as:

$$T = 1/p_{\text{in}}(1/k_{\text{in}} + \delta_{\text{m}}\tau_{\text{m}}^{\text{in}}) \quad (4.15)$$

$$\tau_{\text{m}}^{\text{in}} = \frac{1-p^{\text{in}}}{k_{\text{t}}} + \frac{p^{\text{in}}}{k_{\text{tP}}} \quad (4.16)$$

$$p_{\text{in}} = \left(1 + k_{\text{in}}\tau_1\right)^{-1} \quad (4.17)$$

The typical peloton size is $\langle n_{\text{p}} \rangle = p_{\text{in}}^{-1}$. The average time a site is occupied by a road-block between two passing polymerases is given by

$$t_{\text{rb}} = (1/k_{\text{in}} - \delta_{\text{m}}(1/k_{\text{tP}} - 1/k_{\text{t}}))e^{-\delta_{\text{rb}}/x^*} (1/p_{\text{in}} - 1) + 1/k_{\text{in}} \quad (4.18)$$

$$x^* = -\frac{1}{\ln(1-p_{\text{in}})} \left(\frac{1}{k_{\text{in}}(1/k_{\text{tP}} - 1/k_{\text{t}})} \right), \quad (4.19)$$

where x^* is the typical distance over which pelotons are formed and k_{t}^{P} is the eviction rate of the protein species that forms the strongest obstacle. For transcription, emptying of a promoter by a previous polymerase is typically not rate limiting, $k_{\text{in}}^{-1} \gg \delta_{\text{m}}/k_{\text{tP}}$ such that we can simplify Equation 4.18 further as

$$t_{\text{rb}} = (1/k_{\text{in}})(1/p_{\text{in}} - 1)e^{-\delta_{\text{rb}}/x^*} + 1/k_{\text{in}}. \quad (4.20)$$

Equation 4.20 and 4.15 can be plugged into Equation 4.3 in the main text to calculate the protein densities.

4.4.3. THE AVERAGE SPEED

The average velocity can be obtained by averaging the RNAP speed over the protein densities:

$$v = \begin{cases} (\langle \rho_1^{\text{A}} \rangle / k_{\text{t}}^{\text{A}} + \langle \rho_1^{\text{B}} \rangle / k_{\text{t}}^{\text{B}} + (1 - \langle \rho_1^{\text{A}} \rangle - \langle \rho_1^{\text{B}} \rangle) / k_{\text{t}})^{-1}, & t_{\text{rb}} < \tau_2 - \tau_1 \\ (\langle \rho^{\text{A}} \rangle / k_{\text{t}}^{\text{A}} + \langle \rho^{\text{B}} \rangle / k_{\text{t}}^{\text{B}} + (1 - \langle \rho^{\text{A}} \rangle - \langle \rho^{\text{B}} \rangle) / k_{\text{t}})^{-1}, & t_{\text{rb}} \geq \tau_2 - \tau_1. \end{cases} \quad (4.21)$$

Since we assume low unbinding rates, the intermediate and equilibrium densities are close to unity and Equation 4.21 reduces to Equation 4.4 in the main text for equal eviction rates for both proteins, $k_{\text{t}}^{\text{A}} = k_{\text{t}}^{\text{B}}$.

There are two limits in which the solution in Equation 4.3 is accurate for two unequal eviction rates, $k_{\text{t}}^{\text{A}} \neq k_{\text{t}}^{\text{B}}$. The first limit is when the fastest binding protein does not hinder the polymerase significantly, $k_{\text{t}}^{\text{A}} \approx k_{\text{t}}$. Then the system can be described by the equations for one competing protein [21], with a protein equilibration time τ_2 . The second limit

is when protein binding is much slower than the time needed for a polymerase to clear space for a protein, $\tau^{A,B} \gg \delta_{rb}/\nu$, and the evictions rates for both proteins are similar such that pelotons do not interact over the length of a gene. By combining Equation 4.8, 4.20 and 4.17 one can see that t_{rb} becomes independent of $k_t^{A,B}$ when the equilibration time behind a polymerase, τ_1 is dominated by the binding rates, $k_b^A + k_b^B$, rather than the time needed for a polymerase to clear enough space for a protein, δ_{rb}/ν .

4.4.4. EQUILIBRATION OF TWO EQUALLY SIZED DNA BINDING PROTEINS

Equilibration of two equally-sized proteins to a DNA binding site after eviction at $t = 0$ can be described by the master equation

$$\frac{d}{dt} \begin{pmatrix} \rho^A \\ \rho^B \end{pmatrix} = \begin{pmatrix} -k_b^A(1+K_D^A) & -k_b^A \\ -k_b^B(1+K_D^B) & -k_b^B \end{pmatrix} \begin{pmatrix} \rho^A \\ \rho^B \end{pmatrix} + \begin{pmatrix} k_b^A \\ k_b^B \end{pmatrix}, \quad \begin{pmatrix} \rho^A \\ \rho^B \end{pmatrix} (t=0) = 0 \quad (4.22)$$

Which after solving gives the following solutions for the intermediate densities $\rho_1^{A,B}$

$$\rho_1^A = \rho_2^A \frac{\sqrt{(\tau_A^{-1} + \tau_B^{-1})^2 - 4k_b^A k_b^B (K_D^A K_D^B + K_D^A + K_D^B)} + \tau_A^{-1} - \tau_B^{-1} + 2k_b^A K_D^A / K_D^B}{2\sqrt{(\tau_A^{-1} + \tau_B^{-1})^2 - 4k_b^A k_b^B (K_D^A K_D^B + K_D^A + K_D^B)}} \quad (4.23)$$

$$\rho_1^B = \rho_2^B \frac{\sqrt{(\tau_A^{-1} + \tau_B^{-1})^2 - 4k_b^A k_b^B (K_D^A K_D^B + K_D^A + K_D^B)} + \tau_B^{-1} - \tau_A^{-1} + 2k_b^B K_D^B / K_D^A}{2\sqrt{(\tau_A^{-1} + \tau_B^{-1})^2 - 4k_b^A k_b^B (K_D^A K_D^B + K_D^A + K_D^B)}}, \quad (4.24)$$

where ρ_2^A and ρ_2^B are the equilibrium densities, Equation 4.6, and the equilibration times of the individual proteins are $\tau_{A,B} = k_b^{A,B}(1 + K_D^{A,B})$. The relaxation time scales are given by

$$\tau_{\pm} = \left(\frac{1}{2}(\tau_A^{-1} + \tau_B^{-1}) \pm \frac{1}{2}\sqrt{(\tau_A^{-1} + \tau_B^{-1})^2 - 4k_b^A k_b^B (K_D^A K_D^B + K_D^A + K_D^B)} \right)^{-1}, \quad (4.25)$$

where $\tau_+ = \tau_1 - \delta_{rb}/\nu$ and $\tau_- = \tau_2 - \delta_{rb}/\nu$. The relaxation time scales τ_{\pm} are well separated when

$$(\tau_A^{-1} + \tau_B^{-1})^2 \gg 4k_b^A k_b^B (K_D^A K_D^B + K_D^A + K_D^B) \quad (4.26)$$

The intermediate densities in the limit of a large separation of equilibration time scales (Equation 4.26) are given by

$$\rho_1^A / \rho_1^B \quad (4.27)$$

$$= \rho_2^A \left(\sqrt{(\tau_A^{-1} + \tau_B^{-1})^2 - 4k_b^A k_b^B (K_D^A K_D^B + K_D^A + K_D^B)} + \tau_A^{-1} - \tau_B^{-1} + 2k_b^A K_D^A / K_D^B \right) / \quad (4.28)$$

$$\rho_2^B \left(\sqrt{(\tau_A^{-1} + \tau_B^{-1})^2 - 4k_b^A k_b^B (K_D^A K_D^B + K_D^A + K_D^B)} + \tau_B^{-1} - \tau_A^{-1} + 2k_b^B K_D^B / K_D^A \right) \quad (4.29)$$

$$\approx \left(\tau_A^{-1} K_D^B + k_b^A K_D^A \right) / \left(\tau_B^{-1} K_D^A + k_b^B K_D^B \right) \quad (4.30)$$

$$\approx k_b^A / k_b^B, \quad (4.31)$$

where in the third line we assume a large separation of time scales (Equation 4.26) and in the last line we took a zeroth order expansion in the two unbinding rates. The equilibration times in Equation 4.25 in the limit of a large time scale separation and slow unbinding reduce to

$$\tau_+ \approx (k_b^B + k_b^A)^{-1} \quad (4.32)$$

$$\tau_- \approx \left(\frac{k_b^B}{k_b^B + k_b^A} k_u^A + \frac{k_b^A}{k_b^B + k_b^A} k_u^B \right)^{-1}, \quad (4.33)$$

where we take Taylor expansions to leading order in the unbinding rates. The short equilibration time τ_+ only depends on the binding rates, while full the full equilibration time, τ_- , also depends on rare unbinding events.

We can be more precise in how large the separation of time scales should be. Let n be defined as $n \equiv \tau_-/\tau_+$ and rearrange terms using Equation 4.25, giving

$$\tau_- = n\tau_+ \rightarrow \frac{(\tau_A^{-1} - \tau_B^{-1})^2 + 4k_b^A k_b^B}{(\tau_A^{-1} + \tau_B^{-1})^2} = \left(\frac{n-1}{n+1} \right)^2. \quad (4.34)$$

A large separation of the time scales τ_{\pm} implies that $n \gg 1$, such that $((n-1)/(n+1))^2 \approx 1$. Equation 4.34 shows that two limits that give a large separation of time scales are $\tau_{A,B} \gg \tau_{B,A}$ and $k_b^A, k_b^B \gg k_u^A, k_u^B$.

A more precise requirement for n in Equation 4.34 is that there must be a time t_{int} , at which the densities are set by the intermediate densities. t_{int} must be much larger than τ_+ , but much smaller than τ_- and we write $t_{\text{int}} = i\tau_+$. We can now write a condition for i and n using Equation 4.5:

$$\rho(i\tau_+) \approx e^{-i/n}(\rho_1 - \rho_2) + \rho_2 \approx \rho_1 \quad (4.35)$$

$$\Rightarrow \rho_2^A(1 - e^{-i/n}) \ll \rho_1^A, \rho_2^B(1 - e^{-i/n}) \ll \rho_1^B \quad (4.36)$$

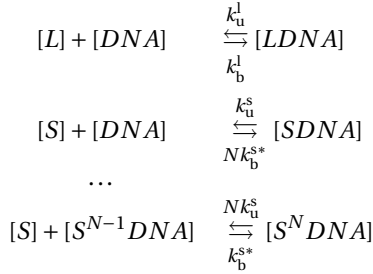
The condition in Equation 4.36 is satisfied when i is much smaller than n , but how much smaller depends on the ratios ρ_1^A/ρ_2^A and ρ_1^B/ρ_2^B :

$$\rho(i\tau_+) \approx \rho_1 \Rightarrow \begin{cases} 10i \leq n, & \rho_2^A \leq \rho_1^A, \rho_2^B \leq \rho_1^B \\ \frac{10i}{-\text{Log}(1 - \rho_1^A/\rho_2^A)} \leq n, & \rho_2^A < \rho_1^A \\ \frac{10i}{-\text{Log}(1 - \rho_1^B/\rho_2^B)} \leq n, & \rho_2^B < \rho_1^B \end{cases} \quad (4.37)$$

Equation 4.37 gives sufficient conditions for the binding rates of the proteins, for there to be a time $\tau_+ \ll t_{\text{int}} \ll \tau_-$ at which the density of proteins is given by the intermediate densities, ρ_1 . In physiological conditions, both the intermediate and the equilibrium densities are typically close to 1, such that the requirement $10i \leq n$ is enough.

4.4.5. EQUILIBRIUM COVERAGE OF COMPETING PROTEINS WITH DIFFERENT SIZES

We here calculate the equilibrium coverage of small and large roadblocks to a lattice with binding sites as described in the main text. Consider pieces of DNA with the size of a large protein and N binding sites for a small protein and define the binding rate per binding site k_b . In steady state, the chemical reactions describing binding of protein P to DNA are given by



The dimensionless dissociation constant in the text is the actual dissociation constant divided by protein concentration $[P]$, $K_D = K_D^*/[P]$. Detailed balance dictates $[P^N DNA] = [DNA]/K_D^N$, such that the average fraction of DNA occupied by small or large proteins is given by

$$\rho^s = \frac{1}{NZ} \sum_{m=1}^N \binom{N}{m} m \left(\frac{1}{K_D^s} \right)^m \quad (4.38)$$

$$\rho^l = \frac{1}{Z} \frac{1}{K_D^l} \quad (4.39)$$

$$Z = \frac{1}{K_D^l} + \sum_{m=0}^N \binom{N}{m} \left(\frac{1}{K_D^s} \right)^m, \quad (4.40)$$

giving Equation 4.12 in the main text. For $N = 1$, Equation 4.38 and 4.39 reduce to Equation 4.6 in the main text.

4.4.6. SCALING OF DISSOCIATION CONSTANT WITH PROTEIN SIZE

In this section we show that the memory of transcription increases when the size difference of competing DNA binding proteins increases, even if the binding energy per base pair is the same for both proteins. The binding rates can be written as

$$k_b = k_b^{\text{at}} e^{-\Delta G^\ddagger}, \quad k_u = k_u^{\text{at}} e^{-(\Delta G^\ddagger + \epsilon)}. \quad (4.41)$$

Here k_b^{at} and k_u^{at} are the attempt rates for binding and unbinding respectively, ΔG^\ddagger is the activation energy, ϵ the binding energy and the exponential terms are the Arrhenius factors [55]. From Equation 4.41 the dissociation constant can be derived as $K_D =$

$k_u^{\text{at}}/k_b^{\text{at}} e^{-\epsilon}$. Let the subscripts s and l indicate the small and large protein respectively. We keep the binding energy per base pair constant by setting $\epsilon_l = N\epsilon_s$. The resulting dissociation constants are $K_{D,s} = k_{u,s}^{\text{at}}/k_{b,s}^{\text{at}} e^{-\epsilon_l/N}$ and $K_{D,l} = k_{u,l}^{\text{at}}/k_{b,l}^{\text{at}} e^{-\epsilon_l}$. In Figure 5.4 in the main text we set the attempt rates for the small and large protein equal,

$$k_{b,s}^{\text{at}} = k_{u,s}^{\text{at}}, \quad k_{b,l}^{\text{at}} = k_{u,l}^{\text{at}}, \quad (4.42)$$

while varying the binding energy with protein size. The dissociation constants in that limit are related as

$$K_D^s = \sqrt[N]{K_D^l}. \quad (4.43)$$

4.4.7. COMPETITION OF PROTEINS WITH UNEQUAL SIZE

The dynamics of roadblocks binding and unbinding to a lattice site is described by a birth death process [46] (Figure 4.5). The states in the birth death process are the number of proteins bound and the rates are the effective binding and unbinding rates as indicated in Figure 4.5. We consider the limit where unbinding of the large roadblocks can be neglected such that the state where a large roadblock is bound is an absorbing state and the equilibration time (memory of transcription), T_N , is the typical time for a large roadblock to bind. After transcription stops, new binding sites open up that were previously occupied by a polymerase. For very slow binding of the large roadblock, $k_b^l \ll k_b^s, k_u^s$ binding and unbinding of the small roadblock equilibrates to those sites, before the large roadblock binds, such that we can write for $P_0(N)$, the probability of having 0 small proteins bound when small protein binding is equilibrated

$$P_0(N) = \left(\frac{k_u^s}{k_b^s + k_u^s} \right)^N = K_D^l \left(1 + \sqrt[N]{K_D^l} \right)^{-N}, \quad (4.44)$$

where we set the attempt rates equal for simplicity, Equation 4.42. To show when the memory of transcription increases with N we calculate the derivative of $P_0(N)$

$$\frac{dP_0(N)}{dN} = (1 + \sqrt[N]{K_D^l})^{-N} \left(\frac{\sqrt[N]{K_D^l} \ln(K_D^l)}{N(1 + \sqrt[N]{K_D^l})} - \ln(1 + \sqrt[N]{K_D^l}) \right) \quad (4.45)$$

Since we consider slow binding of the large protein, $K_D^l < 1$, Equation 4.45 is always negative. Hence, the memory of transcription activity τ_2 increases with the size difference N (Equation 4.14 in the main text). For large N , $\sqrt[N]{K_D^l}$ approaches unity, such that the memory of transcription depends exponentially on N , (Figure 4.4 D main text).

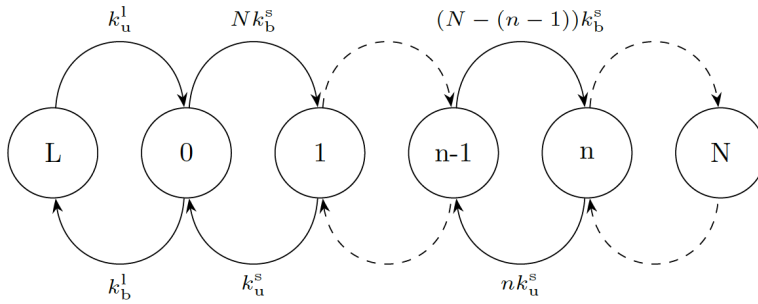


Figure 4.5: Birth death process for binding of a large roadblock to a lattice with small roadblocks. The numbers $0 \dots N$ represent the number of small roadblocks bound and the left node, L represents a bound large roadblocks. The arrows correspond to binding or unbinding of roadblocks and are labeled with the effective binding rates. If the unbinding rate of a large roadblock is zero, the left node becomes an absorbing state.

REFERENCES

- [1] B. Alberts, D. Bray, K. Hopkin, A. Johnson, J. Lewis, M. Raff, K. Roberts, and P. Walter, *Essential cell biology*, 2nd ed. (Garland Science, Taylor and Francis Group, 2004).
- [2] M. F. Dion, T. Kaplan, M. Kim, S. Buratowski, N. Friedman, and O. J. Rando, *Dynamics of Replication-Independent Histone Turnover in Budding Yeast*, *Science* **315**, 1405 (2007).
- [3] A. Rufiange, P.-E. Jacques, W. Bhat, F. Robert, and A. Nourani, *Genome-Wide Replication-Independent Histone H3 Exchange Occurs Predominantly at Promoters and Implicates H3 K56 Acetylation and Asf1*, *Molecular Cell* **27**, 393 (2007).
- [4] A. Jamaï, R. M. Imoberdorf, and M. Strubin, *Continuous Histone H2B and Transcription-Dependent Histone H3 Exchange in Yeast Cells outside of Replication*, *Molecular Cell* **25**, 345 (2007).
- [5] M. a. Schwabish and K. Struhl, *Evidence for eviction and rapid deposition of histones upon transcriptional elongation by RNA polymerase II*. *Molecular and cellular biology* **24**, 10111 (2004).
- [6] D. E. Schones, K. Cui, S. Cuddapah, T.-Y. Roh, A. Barski, Z. Wang, G. Wei, and K. Zhao, *Dynamic regulation of nucleosome positioning in the human genome*. *Cell* **132**, 887 (2008).
- [7] F. Cui, H. A. Cole, D. J. Clark, and V. B. Zhurkin, *Transcriptional activation of yeast genes disrupts intragenic nucleosome phasing*, *Nucleic Acids Research* **40**, 10753 (2012).
- [8] B. E. Schwartz and K. Ahmad, *Transcriptional activation triggers deposition and removal of the histone variant H3.3*, *Genes and Development* **19**, 804 (2005).
- [9] D. Ray-Gallet, A. Woolfe, I. Vassias, C. Pellentz, N. Lacoste, A. Puri, D. C. Schultz, N. A. Pchelintsev, P. D. Adams, L. E. T. Jansen, and G. Almouzni, *Dynamics of Histone H3 Deposition In Vivo Reveal a Nucleosome Gap-Filling Mechanism for H3.3 to Maintain Chromatin Integrity*, *Molecular Cell* **44**, 928 (2011).
- [10] S. Hamperl, M. Wittner, V. Babl, J. Perez-Fernandez, H. Tschochner, and J. Griesenbeck, *Chromatin states at ribosomal DNA loci*, *Biochimica et Biophysica Acta - Gene Regulatory Mechanisms* **1829**, 405 (2013).
- [11] T. Tamura, M. Smith, T. Kanno, H. Dasenbrock, A. Nishiyama, and K. Ozato, *Inducible deposition of the histone variant H3.3 in interferon-stimulated genes*, *Journal of Biological Chemistry* **284**, 12217 (2009).
- [12] M. Wittner, S. Hamperl, U. Stöckl, W. Seufert, H. Tschochner, P. Milkereit, and J. Griesenbeck, *Establishment and maintenance of alternative chromatin states at a multicopy gene locus*, *Cell* **145**, 543 (2011).
- [13] K. Ahmad and S. Henikoff, *The Histone Variant H3.3 Marks Active Chromatin by Replication-Independent Nucleosome Assembly*, *Molecular Cell* **9**, 1191 (2002).

- [14] C. Wirbelauer and O. Bell, *Variant histone H3.3 is deposited at sites of nucleosomal displacement throughout transcribed genes while active histone modifications show a promoter-proximal bias*, *Genes & Development* **4**, 1761 (2005).
- [15] Y. Mito, J. G. Henikoff, and S. Henikoff, *Genome-scale profiling of histone H3.3 replacement patterns*. *Nature genetics* **37**, 1090 (2005).
- [16] R. B. Deal, J. G. Henikoff, and S. Henikoff, *Genome-Wide Kinetics of Nucleosome*, *Science* **328**, 1161 (2010).
- [17] O. I. Kulaeva, E.-K. Hsieh, and V. M. Studitsky, *RNA polymerase complexes cooperate to relieve the nucleosomal barrier and evict histones*. *Proceedings of the National Academy of Sciences* **107**, 11325 (2010).
- [18] C. T. MacDonald, J. H. Gibbs, and A. C. Pipkin, *Kinetics of Biopolymerization on Nucleic Acid Templates*, *Biopolymers* **6**, 1 (1968).
- [19] M. Sahoo, J. Dong, and S. Klumpp, *Dynamic blockage in an exclusion process*, *Journal of Physics A: Mathematical and Theoretical* **48**, 015007 (2015).
- [20] O. J. O. Loan, M. R. Evans, and M. E. Cates, *Jamming transition in a homogeneous one-dimensional system : The bus route model*, *Physical Review E* **58**, 1404 (1998).
- [21] A. A. van den Berg and M. Depken, *Crowding-induced transcriptional bursts dictate nucleosome and polymerase density profiles along genes*, *Nucleic Acids Res* **45**, 7623–7632 (2017).
- [22] G.-C. Yuan, L. Y-J, M. F. Dion, M. D. Slack, L. F. Wu, A. J. Altshuler, and O. J. Rando, *Genome-Scale Identification of Nucleosome Positions in *S. cerevisiae**, *Science* **309**, 626 (2005).
- [23] W. Lee, D. Tillo, N. Bray, R. H. Morse, R. W. Davis, T. R. Hughes, and C. Nislow, *A high-resolution atlas of nucleosome occupancy in yeast*. *Nature genetics* **39**, 1235 (2007).
- [24] J. D. McGhee and P. H. von Hippel, *Theoretical aspects of DNA-protein interactions: Co-operative and non-co-operative binding of large ligands to a one-dimensional homogeneous lattice*, *Journal of Molecular Biology* **86**, 469 (1974).
- [25] L. Reeve and J. a. D. Wattis, *Random sequential adsorption with two components: asymptotic analysis and finite size effects*, *Journal of Physics A: Mathematical and Theoretical* **48**, 235001 (2015).
- [26] P. L. Krapivsky and E. Ben-Naim, *Collective properties of adsorption – desorption processes* *Collective properties of adsorption-desorption processes*, *J Chem Phys* **100**, 6778 (1994).
- [27] B. Osberg, J. Nuebler, P. Korber, and U. Gerland, *Replication-guided nucleosome packing and nucleosome breathing expedite the formation of dense arrays*, *Nucleic Acids Research* **42**, 13633 (2014).

- [28] D. Murugesapillai, M. J. Mccauley, R. Huo, M. H. N. Holte, A. Stepanyants, L. J. Mather, N. E. Israeloff, and M. C. Williams, *DNA bridging and looping by HMO1 provides a mechanism for stabilizing nucleosome-free chromatin*, *Nucleic Acids Research*, **1** (2014).
- [29] T. van der Heijden and C. Dekker, *Monte carlo simulations of protein assembly, disassembly, and linear motion on DNA*. *Biophysical journal* **95**, 4560 (2008).
- [30] G. M. Cooper, *The Cell: A Molecular Approach. 2nd edition* (Sinauer Associates, 2000).
- [31] V. Pelechano, S. Chavez, and J. E. Perez-Ortin, *A Complete Set of Nascent Transcription Rates for Yeast Genes*, *PLoS one* **5**, e15442 (2010), arXiv:1203.2655 .
- [32] B. Derrida, M. R. Evans, V. Hakim, and V. Pasquier, *Exact solution of a 1D asymmetric exclusion model using a matrix formulation*, *Journal of Physics A: Mathematical and General* **26**, 1493 (1993).
- [33] S. Malik and S. R. Bhaumik, *Rad26p, a transcription-coupled repair factor, promotes the eviction and prevents the reassociation of histone H2A-H2B dimer during transcriptional elongation in vivo*, *Biochemistry* **51**, 5873 (2012).
- [34] S. Bevington and J. Boyes, *Transcription-coupled eviction of histones H2A/H2B governs V(D)J recombination*. *The EMBO journal* **32**, 1381 (2013).
- [35] A. Worcel, S. Han, and M. L. Wong, *Assembly of newly replicated chromatin*. *Cell* **15**, 969 (1978).
- [36] J. Mazurkiewicz, J. F. Kepert, and K. Rippe, *On the mechanism of nucleosome assembly by histone chaperone NAP1*. *The Journal of biological chemistry* **281**, 16462 (2006).
- [37] R. Phair, P. Scaffidi, C. Elbi, J. Vecerová, A. Dey, K. Ozato, D. Brown, G. Hager, M. Bustin, and T. Misteli, *Global Nature of Dynamic Protein-Chromatin Interactions In Vivo: Three-Dimensional Genome Scanning and Dynamic Interaction Networks of Chromatin Proteins*, *Mol. Cell. Biol.* **24**, 6393 (2004).
- [38] K. Luger, a. W. Mäder, R. K. Richmond, D. F. Sargent, and T. J. Richmond, *Crystal structure of the nucleosome core particle at 2.8 Å resolution*. *Nature* **389**, 251 (1997).
- [39] B. Piña and P. Suau, *Changes in histones H2A and H3 variant composition in differentiating and mature rat brain cortical neurons*, *Developmental Biology* **123**, 51 (1987).
- [40] E. Kamau, K. T. Bauerle, and A. Grove, *The Saccharomyces cerevisiae high mobility group box protein HMO1 contains two functional DNA binding domains*. *The Journal of biological chemistry* **279**, 55234 (2004).
- [41] S. J. Greive and P. H. V. Hippel, *Thinking quantitatively about transcriptional regulation*. *Nature reviews. Molecular cell biology* **6**, 221 (2005).

- [42] K. Brogaard, L. Xi, J.-P. Wang, and J. Widom, *A map of nucleosome positions in yeast at base-pair resolution*. *Nature* **486**, 496 (2012).
- [43] M. L. Kireeva, W. Walter, V. Tchernajenko, V. Bondarenko, M. Kashlev, and V. M. Studitsky, *Nucleosome remodeling induced by RNA polymerase II: Loss of the H2A/H2B dimer during transcription*, *Molecular Cell* **9**, 541 (2002).
- [44] C. Hodges, L. Bintu, L. Lubkowska, M. Kashlev, and C. Bustamante, *Nucleosomal fluctuations govern the transcription dynamics of RNA polymerase II*. *Science (New York, N.Y.)* **325**, 626 (2009).
- [45] H. A. Cole, J. Ocampo, J. R. Iben, R. V. Chereji, and D. J. Clark, *Heavy transcription of yeast genes correlates with differential loss of histone H2B relative to H4 and queued RNA polymerases*. *Nucleic acids research* **42**, 12512 (2014).
- [46] C. W. Gardiner, *Handbook of Stochastic Methods*, third edit ed., edited by H. Haken (Springer-Verlag, Berlin Heidelberg New York, 2004).
- [47] H. L. Ashe, J. Monks, M. Wijgerde, P. Fraser, and N. J. Proudfoot, *Intergenic transcription and transinduction of the human β -globin locus*, *Genes and Development* **11**, 2494 (1997).
- [48] M. Jinek, K. Chylinski, I. Fonfara, M. Hauer, J. A. Doudna, and E. Charpentier, *A Programmable Dual-RNA – Guided DNA Endonuclease in Adaptive Bacterial Immunity*, *Science* **337**, 816 (2012).
- [49] M. A. Horlbeck, L. B. Witkowsky, B. Guglielmi, J. M. Replogle, L. A. Gilbert, J. E. Villalta, S. E. Torigoe, R. Tijan, and J. S. Weissman, *Nucleosomes impede Cas9 access to DNA in vivo and in vitro*, *eLife* **5**, e12677 (2016).
- [50] R. S. Isaac, F. Jiang, J. A. Doudna, W. A. Lim, G. J. Narlikar, and R. A. Almeida, *Nucleosome breathing and remodeling constrain CRISPR-Cas9 function*, *eLife* **5**, e13450 (2016).
- [51] A. Granéli, C. C. Yeykal, and E. C. Greene, *Organized Arrays of Individual DNA Molecules Tethered to Supported Lipid Bilayers*, *Langmuir* **22**, 292 (2006).
- [52] P. Ranjith, J. Yan, and J. F. Marko, *Nucleosome hopping and sliding kinetics determined from dynamics of single chromatin fibers in Xenopus egg extracts*. *Proceedings of the National Academy of Sciences of the United States of America* **104**, 13649 (2007).
- [53] N. Filigheddu, V. F. Gnocchi, M. Coscia, M. Cappelli, P. E. Porporato, and R. Taulli, *Ghrelin and Des-Acyl Ghrelin Promote Differentiation and Fusion of C2C12 Skeletal Muscle Cells*, *Molecular biology of the cell* **18**, 986 (2007).
- [54] J. J. Parmar, D. Das, and R. Padinhateeri, *Theoretical estimates of exposure timescales of protein binding sites on DNA regulated by nucleosome kinetics*. *Nucleic acids research* **44**, 1630 (2015).
- [55] N. G. Kampen, *Stochastic Processes in Physics and Chemistry*, 3rd ed. (Elsevier, 2007).

5

TRANSCRIPTION ELONGATION FACTORS MODIFY NUCLEOSOME DENSITY AND TRANSCRIPTIONAL BURSTS

Transcription elongation is regulated by elongation factors like TFIIS and Spt4/5 and nucleosome remodelers that modify the pause or elongation dynamics of RNAP. Single-molecule experiments have taught us how TFIIS, Spt4/5 or one nucleosome interacts with RNAP, but in vivo multiple factors and nucleosome remodelers simultaneously orchestrate the transcription process. Using theoretical modeling we here study the combined effects of Spt4/5, TFIIS and nucleosomes on transcription on heavily transcribed genes. Surprisingly, elongation factors and nucleosome remodelers have two completely overlooked effects on transcription. First, they can increase the nucleosome density on a gene, not only by increasing the RNAP speed, but also by changing the RNAP distribution along a gene. Second, the modified RNAP distribution results in a modified transcription output dynamics. The two overlooked effects of transcription factors and nucleosome remodelers are only apparent on crowded DNA and show that for a full understanding of the biological function of transcription factors, they have to be studied in physiologically relevant environments.

Manuscript in preparation: A.A. van den Berg, R. Molenaar and M. Depken, *Transcription elongation factors modify nucleosome density and transcriptional bursts.*

5.1. INTRODUCTION

Transcription is typically initiation-limited [1, 2], such that an increase of the elongation rate does not increase the overall rate of mRNA production. Instead, transcription regulation during the transcription elongation phase has many other functions. The transcription factor TFIIS is for example important for error correction [3] and enhancing the chance of passage of RNAP through nucleosomes [4–6], while other factors are important for coordination of transcription-coupled processes and signaling [2, 7]. Nucleosomes, which are essential for compactification and protection of DNA [8], also regulate transcription elongation as they form obstacles for RNAP translocation [9], and RNAP can evict (part of) a nucleosome [10]. Nucleosome remodeling, such as acetylation or removal of the histone tails reduces the nucleosomal barrier to transcription elongation and nucleosome eviction [11–13].

Through experimental studies in idealized *in vitro* environments, it is fairly well understood how RNAP functions while interacting with one nucleosome or one elongation factor at a time [14, 15]. RNAP frequently backtracks during transcription and moves upstream. As a consequence, the 3' end of RNA sticks out from the front channel and blocks the catalytic site such that RNA cannot be elongated [16–18]. Some elongation factors, like the the widely conserved elongation factor Spt4/5, reduce entry into a backtrack [19], while others, like TFIIS, reduce the duration of a backtrack through RNA cleavage [3, 20].

It often remains an experimental challenge to translate results from single-molecule studies into an understanding of biological function in the crowded and actively regulated internal environment of cells. Theoretical modeling can extrapolate these experimental results to more complex settings and make testable predictions to guide experiments. Elongation factors are for example expected to change the nucleosome density, because an increase in speed results in a decrease of RNAP density, leaving more space for nucleosomes to bind.

In Chapter 3 and 4, we used theoretical modeling to show that actively transcribing RNAPs organize into pelotons when interacting with dynamic roadblocks like nucleosomes [21], but RNAP dynamics was modeled as a single hopping rate, ignoring the effect of backtracking and elongation factors. Taking into account RNAP traffic [22], internal motor pause dynamics [23], as well as the interaction with multiple nucleosomes along the track (Figure 5.1 A), we here analytically solve for the dynamics of a single RNAP and simulate the motion of multiple RNAP along a gene.

We show how elongation factors change the nucleosome coverage on a gene, not only by changing the density of RNAP, but also by changing the organization of RNAPs along genes. The changed RNAP organization also results in a different temporal transcription output dynamics. Nucleosome remodelers have a larger effect on reorganization and temporal output than elongation factors that decrease the internal pausing dynamics of RNAP. The influence of elongation factors on RNAP organization and transcription output dynamics has been overlooked so far in studies on the elongation phase and our study sheds new light on the possible biological functions for transcription regulation during the elongation phase.

5.2. RESULTS

5.2.1. A MODEL FOR TRANSCRIPTION WITH ELONGATION FACTORS

The full model for transcription on crowded DNA, Figure 5.1 B, is based on the the Totally Asymmetric Exclusion Process (TASEP) in which motors initiate, elongate and terminate stochastically along a 1D lattice and motors cannot overtake nor overlap [24]. We add nucleosomes to the TASEP that can bind to unoccupied lattice sites [21] and RNAPII backtracking and transcript cleavage [23] (Figure 5.1 B). Following observations from single-molecule experiments, the transcription factors and nucleosome modifications are included as variations in the pausing and elongation dynamics of RNAP. To study the whole spectrum of force-dependent RNAP dynamics we consider elongation factors that modify the backtrack entry rate (Spt4/5), the backtrack duration (TFIIS) and the translocation rate (nucleosome modifications).

The parameters of the model are shown in Figure 5.1 B. RNAP initiates at the start of a gene with a rate k_{in} if the promoter is empty. Since transcription is typically initiation-limited, the termination rate does not affect the transcription dynamics nor the density of polymerases along the gene [1, 2]. After initiation, RNAP catalyzes the production of RNA and pauses frequently along the DNA. Some pauses, like backtracking, are force-dependent [4], while others, such as the short-lived elemental pause [25], are not. We include the catalysis rate and force-independent pauses in a single effective RNA production rate k_{cat} .

At each site, an RNAP can enter a backtrack with rate k_{bt}^* and an increase of the concentration of Spt4/5 is modeled as a decrease in k_{bt}^* . Within the backtrack, a polymerase moves back and forth diffusively with rate k_{bt} . The transcription factor TFIIS reduces the duration of a backtrack by cleaving RNA protruding from the secondary channel of RNAPII with rate k_{clv} .

The rate of transcription elongation is reduced in a region with a nucleosome of size δ_{nuc1} (Figure 5.1 B), such that the frequency and the time spent in backtracks increases [9]. Since the nucleosome dyad is the strongest roadblock to transcription [9] we model passage through a nucleosome as one slow node with rate k_{nuc1} and $\delta_{\text{nuc1}} - 1$ nodes where the rate is unaltered (k_{cat}). We choose k_{nuc1} such that the average RNAP speed through a nucleosomal region matches single molecule results [13]. After eviction, nucleosomes can rebind with rate k_{b} if there are δ_{nuc1} adjacent available lattice sites. We model modifications of the histone tails as changes in the elongation rate through a nucleosome dyad k_{nuc1} [13].

In the following sections we increase the complexity step by step using the experimentally observed rates in Table 5.1. First we calculate the velocity and pause dynamics of a single RNAP, then we discuss different forms of cooperation between RNAPs and in the last section, we show the effect of elongation factors on nucleosome density on heavily transcribed genes.

5.2.2. VELOCITY AND PAUSE DENSITY AND DURATION FOR A SINGLE RNAP

The kymograph in Figure 5.2 A illustrates how elongation factors modify transcription dynamics. The black trace has a typical backtrack entry rate and cleavage rate. When the concentration of Spt4/5 increases, the backtrack entry rate is reduced and RNAP speeds

Table 5.1: Parameters of the model

Microscopic parameter	Symbol	Value	Citation
DNA Footprint RNAP	δ_{Pol}	35 bp	[26]
DNA Footprint Nucleosome + linker DNA	δ_{nucl}	167 bp	[27, 28]
Catalyzation rate	k_{cat}	10 s^{-1}	[13, 29, 30]
Catalyzation rate through nucleosome dyad	k_{nucl}	$0.01\text{-}0.1 \text{ s}^{-1}$	[13]
Initiation rate	k_{in}	$0.01\text{-}0.03 \text{ s}^{-1}$	[31]
Nucleosome binding rate	k_{b}	0.02 s^{-1}	[32, 33]
Backtracking rate	k_{bt}	1 s^{-1}	[34]
Cleavage rate	k_{clv}	$0.01\text{-}0.1 \text{ s}^{-1}$	[4, 35]
Backtrack entry rate w/o Spt4/5	k_{bt}^*	$0.1\text{-}1.0 \text{ s}^{-1}$	[36]

up (red trace) and when the concentration of TFIIS is reduced, the duration of backtracks increases and RNAP slows down (green). How the speed ν , backtrack density and backtrack duration of a polymerase vary, depending on the concentrations of TFIIS, Spt4/5 and nucleosome remodelers, can be estimated as

$$\nu = \frac{\delta_{\text{nucl}}}{(\delta_{\text{nucl}} - 1)\tau_{\text{tot}}^{\text{cat}} + \tau_{\text{tot}}^{\text{nucl}}}. \quad (5.1)$$

Here $\tau_{\text{tot}}^{\text{nucl}}$ is the average stepping time of a polymerase at a site where the polymerase is facing the dyad of a nucleosome and $\tau_{\text{tot}}^{\text{cat}}$ is the average stepping time for a polymerase facing an empty lattice site. Both stepping times depend on concentration of transcription factors and can be estimated using the theory of continuum time random walks (see the Supplementary Material). The velocity is averaged over a nucleosomal region, with one site with a dyad and a long stepping time $\tau_{\text{tot}}^{\text{nucl}}$ and $\delta_{\text{nucl}} - 1$ fast sites without dyad and a shorter stepping time $\tau_{\text{tot}}^{\text{cat}}$. The calculation of the velocity in Equation 5.1 assumes that nucleosomes are all equally spaced, but in reality the spacing can vary due to eviction and rebinding of nucleosomes. Still, the fast reassembly of nucleosomes results in tightly and regularly packed nucleosomes [37], making Equation 5.1 a good estimate for the average velocity.

A polymerase has a higher probability to backtrack in front of a dyad, due to the increased barrier for translocation, $k_{\text{nucl}} < k_{\text{cat}}$. The probabilities to enter a backtrack at least once before taking a step with or without dyad on the next site are respectively given by $P_{\text{p}}^{\text{nucl}} = k_{\text{bt}}^*/(k_{\text{bt}}^* + k_{\text{nucl}})$ and $P_{\text{p}}^{\text{cat}} = k_{\text{bt}}^*/(k_{\text{bt}}^* + k_{\text{cat}})$. The average probability that a polymerase enters a backtrack at least once is again an average over sites with and without a dyad

$$\rho_{\text{p}} = 1 - [(1 - P_{\text{p}}^{\text{nucl}})(1 - P_{\text{p}}^{\text{cat}})^{\delta_{\text{rb}} - 1}]^{1/\delta_{\text{rb}}} \quad (5.2)$$

Where the term between the square brackets is the probability that RNAP does not backtrack in a nucleosomal region. The total number of backtracks in a nucleosomal region is $P_{\text{p}}^{\text{nucl}} + (\delta_{\text{rb}} - 1)P_{\text{p}}^{\text{cat}}$. Where we count an RNAP that re-enters a backtrack multiple times at the same site as one backtrack, such that a backtrack only ends once the RNAP has

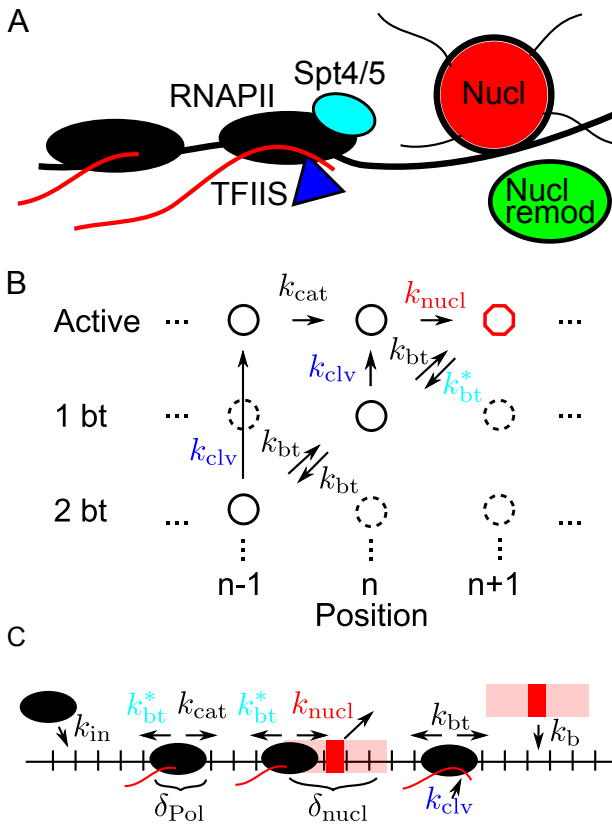


Figure 5.1: A model for transcription initiation and elongation with nucleosomes and various elongation factors. A) An active and a backtracked RNAPII and transcription factors. Part of the RNA of the backtracked RNAPII is threaded into the secondary channel and the transcription factor TFIIIS cleaves this RNA, returning the RNAPII to the active state. B) Schematic of the kinetic diagram for a single RNAP. C) Full model for transcription with all the rates, Table 5.1. RNAP (black) initiates at the start of a gene with a rate k_{in} . The catalyzation rate is k_{cat} when the next site is empty, k_{nucl} when the next site is a dyad and zero when the next site is occupied by a polymerase. A motor can enter a backtrack with rate k_{bt}^* when the previous site is empty and the transcription factor Spt4/5 can modify this rate. A backtracked RNAP (the most right one) hops forward and backward with rate k_{bt} and can return either by diffusing back to the active state, or directly by cleavage by TFIIIS with rate k_{clv} . Nucleosomes, pink, occupy δ_{nucl} lattice sites, but only the dyad, red, slows down an RNAP.

elongated the RNA by one nucleotide. When averaging over a nucleosomal region, the average backtrack duration is given by

$$T_p = \frac{(\delta_{nucl} - 1)P_p^{cat}\tau_p^{cat} + P_p^{nucl}\tau_p^{nucl}}{(\delta_{nucl} - 1)P_p^{cat} + P_p^{nucl}}, \quad (5.3)$$

where τ_p^{nucl} and τ_p^{cat} are the duration of a nucleotide addition step once RNAP is in the backtrack with and without a dyad on the next site respectively, (see the Supplementary Material).

Figures 5.2 B-D show the effect of Spt4/5, TFIIS and nucleosome remodeling on the average transcription velocity, pause density and pause duration for the physiological rates in Table 5.1. Nucleosome remodelers have a strong effect on pause duration, but barely change the pause density in physiological settings (Figure 5.2 B). Spt4/5 on the other hand can change the backtrack density per basepair by a factor two, while the effect on the pause duration is negligible (Figure 5.2 C). TFIIS can change the pause duration by a factor two (Figure 5.2 D). We only made a minor assumption when calculating the velocity in Equation 5.1, namely that an RNAP in a backtrack cannot encounter a dyad (because the same RNAP just evicted the histone) and the good agreement between theory and simulations shows that we have a good understanding for the dynamics of a single polymerase (Figure 5.2 B).

The solutions for a single RNAP do not describe the full dynamics of transcription. On heavily transcribed genes, multiple interacting RNAPs transcribe a gene at the same time [31]. To capture the dynamics on such genes, we need to include these interactions.

5.2.3. MACROSCOPIC EFFECTS OF RNAP COOPERATION

A backtracked RNAP can be rescued out of the backtrack by a trailing RNAP [38] and RNAPs cooperate to evict histones [10, 39]. Such cooperation could result in an increased average elongation rate with transcription initiation [22, 40], but the detailed mechanism of cooperation is not known. It might be that an active RNAP only biases the motion of a backtracked RNAP, preventing further backward motion, or that RNAP can actively push a backtracked RNAP forward [41]. Let N therefore be the number of backtracked RNAP that an active RNAP can push forward. We performed Monte Carlo simulations for a range of $N = 0 \dots 10$ to study the macroscopic effects of cooperation.

As expected, the velocity of multiple RNAPs is higher than for a single RNAP (Figure 5.3 A). When RNAP cannot push ($N = 0$) the velocity barely increases with transcription intensity and transcription is limited by the bulk dynamics for low cleavage rates (Figure 5.3 B). When RNAP can push ($N > 0$), the average RNAP velocity increases significantly with transcription intensity and transcription is initiation-limited, even for the highest measured transcription rates (Figure 5.3 C). The ability to push 1, 2 or more RNAPs does not significantly change the velocity, because it is rare for multiple backtracked RNAP to line up and being pushed forward (Figure 5.3 A). Since the exact choice of N does not change the dynamics, we choose $N = 2$ in the further analysis. Further, we only show data for which RNAPs do not hinder each other during initiation such that the rate of mRNA production is set by the initiation rate (see Figure 5.3 C and the Supplementary Material).

5.2.4. THE EFFECT OF ELONGATION FACTORS ON TRANSCRIPTION ON NUCLEOSOME COVERAGE AND OUTPUT DYNAMICS

Cooperation between RNAPs requires that they 'stick' together while evicting histones, because only when two RNAP are closely spaced, can the upstream RNAP effectively block backtracking of the downstream RNAP. The observation that polymerases cooperate thus implicitly suggests that when multiple polymerases transcribe a gene at the same time, they cluster along a gene, as we predicted in Chapter 3. However, the theory in this chapter did not include the effects of backtracking. Backtracking could potentially

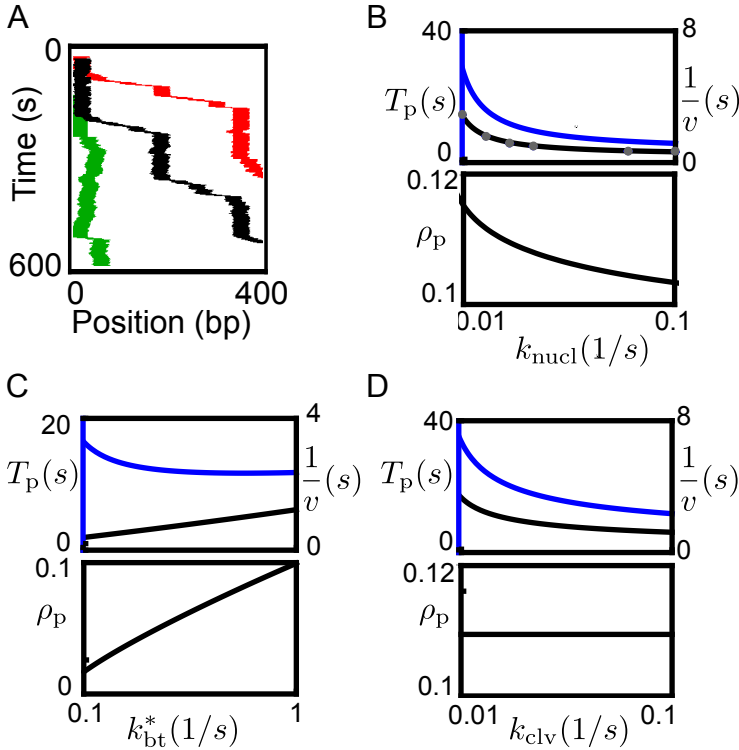


Figure 5.2: How the velocity and pausing dynamics of a single polymerase depends on elongation factors and nucleosome remodelers. A) A kymograph illustrating the effect of the different transcription factors. For the black trace, the backtrack entry rate equals the rate in the backtrack, $k_{bt}^* = k_{bt} = 1/s$ and the cleavage rate is $k_{clv} = 0.1/s$. When the concentration of Spt4/5 increases $k_{bt}^* < k_{bt}$, the polymerase speeds up (red trace), and when the concentration of TFIIS is reduced, $k_{clv} = 0.01/s$, the duration of backtracks increases and the polymerase slows down. B-D) On the left axis of the top panel the pause duration (blue, Equation 5.3) and on the right axis the average stepping time (black line, the inverse of Equation 5.1). In the lower panels the pause density is shown (Equation 5.2). B) When varying the nucleosome eviction rate, k_{nucl} , the change in velocity is mostly caused by a change in pause duration while the pause density barely changes for physiological parameters. The simulations (gray) dots perfectly follow the theory indicating that we have a good understanding of the single RNAP dynamics (the error bars are smaller than the symbol size). C) The average stepping time, pause duration and pause density when varying the backtrack entry rate. In physiological settings, Spt4/5 mostly changes the pause density, not the pause duration. D) The average stepping time, pause duration and pause density for varying concentrations of TFIIS. A change in the average stepping time is mostly caused by a change in the duration of backtracks.

destabilize a peloton, because an RNAP at the back of a peloton is not pushed forward by a trailing RNAP and could therefore 'peel off' from the peloton. We say that an RNAP has been peeled off from a peloton when a nucleosome has assembled between the RNAP and the peloton. As pelotons are defined as actively transcribing RNAPs without nucleosomes between them, (Chapter 3), a peloton has been split once an RNAP has peeled off.

Here we show that polymerases also form pelotons when backtracking is consid-

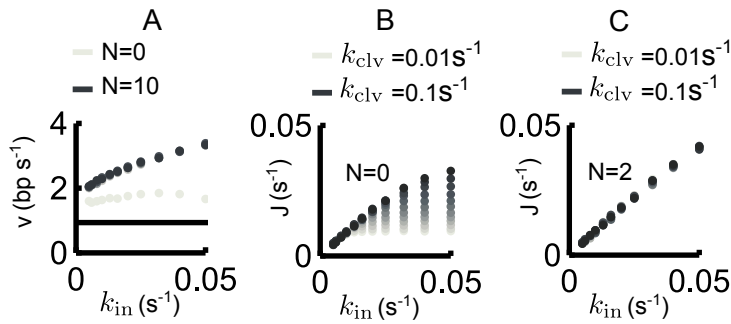


Figure 5.3: The effect of RNAP pushing. A) The RNAP velocity versus the transcription initiation rate for different values of N , the number of backtracked RNAPs that an active RNAP can push forward. Dots are results of Monte Carlo simulations and the solid line is the velocity of a single RNAP for the given parameters. For $N = 0$, the velocity does not increase with the initiation rate, but is still higher than for a single motor, because a backtracked RNAP can be prevented from further backtracking by a trailing RNAP. For higher $N = 2, 3, 4, 10$, the velocity does increase with initiation. The data points for $N > 0$ lie on top of each other (the dots for $N = 2, 3, 4$ lie below the dark gray dots for $N=10$), suggesting that the probability that an active RNAP pushes two more backtracked RNAPs is negligible for determining the average velocity. B) The flux versus the initiation rate for cleavage rates increasing from 0.01 s^{-1} to 0.1 s^{-1} for an RNAP that cannot push. For decreasing cleavage rates, the flux becomes independent of the initiation rate, which means that transcription is not initiation-limited anymore. C) The flux versus the initiation rate for the same range of cleavage rates as in B), but for an active RNAP that can push at most two backtracked RNAPs, $N = 2$. The flux equals the initiation rate for all simulations, indicating that transcription is low enough for RNAPs not to hinder each other during initiation. Note that there are multiple colors, for a range of k_{clv} , but that the data points overlap.

ered (Figure 5.4 A). RNAP rarely peels off from a peloton because nucleosomes slow down polymerases to a speed below the single polymerases speed (without cooperation). Therefore the polymerase at the front of a peloton limits the peloton speed and a polymerase at the back of a peloton can keep up with the peloton, even without cooperation. Though the probability of peeling off is small, it does happen occasionally (blue arrow in Figure 5.4 B) and another difference as compared to peloton formation without backtracking is that the spacing between polymerases within a peloton has increased.

The velocity of a peloton is higher than for a single polymerase (the blue line in Figure 5.4 A) and elongation factors can modulate the speed of a peloton. Figure 5.4 B-D have an increase in nucleosome remodeler, Spt4/5 and TFIIS concentration respectively as compared to Figure 5.4 A and all show an increase in the transcription speed. An increased speed leads to a decrease in polymerase density on the gene which would naively be expected to result in a higher nucleosome density. This basic picture is confirmed by the observation that nucleosome and RNAP coverage are always anti-correlated. Striking though, when plotting the nucleosome density versus the polymerase density, the data do not always collapse, Figure 5.4 E-G, indicating that the density of polymerases is not the only factor that determines the nucleosome density, but that polymerase organization also plays a role. We can understand this if we consider the reorganization of polymerases into pelotons. Every polymerase blocks δ_m nucleotides from nucleosome binding, but two polymerases that are closer than a distance $\Delta = \delta_{\text{tb}} + v/k_{\text{kb}}$ together, also block the space between them, because after the first polymerase has evicted the

nucleosome, the second arrives, before nucleosomes have had time to bind. We refer to the distance Δ as the roadblock shadow. Within a peloton, polymerases are closer to each other than the roadblock shadow (Chapter 3 and [21]) and if polymerases are getting closer, the space blocked per polymerase is smaller and the nucleosome density increases.

We can calculate how the nucleosome coverage depends on RNAP coverage for two extreme cases: large pelotons and no peloton formation. When RNAP form large and dense pelotons, there is always some space for nucleosomes to bind between pelotons, therefore nucleosomes are only completely depleted for an RNAP coverage that is the same as within a peloton (the red dashed line in Figures 5.4 E-G). When RNAP do not form pelotons, nucleosomes are depleted when the average space between RNAP is smaller than a nucleosome (blue dashed line in Figures 5.4 E-G). Both nucleosome remodelers and Spt4/5 can change the tendency of RNAP to form pelotons as they move the system between the red and the blue dashed line, while TFIIS only has a minor influence on the tendency to form pelotons.

For a varying nucleosome eviction rate, the effect of polymerase organization on nucleosome density is strongest (compare Figure 5.4 E to Figure 5.4 F and G) and for both low and high eviction rates, the nucleosome coverage can vary significantly while the RNAP coverage remains the same. For very high RNAP coverage (around 0.5), the nucleosome coverage flattens as a function of RNAP coverage. A likely explanation is that backtracks are suppressed for increasing RNAP coverage, such that RNAPs within a peloton are more closely spaced and addition of an RNAP has a smaller effect on nucleosome coverage. For Spt4/5 the effect of reorganization is negligible for smaller initiation rates, but increases with RNAP coverage (Figure 5.4 F), while for TFIIS the effect on reorganization is negligible for any RNAP coverage (Figure 5.4 G).

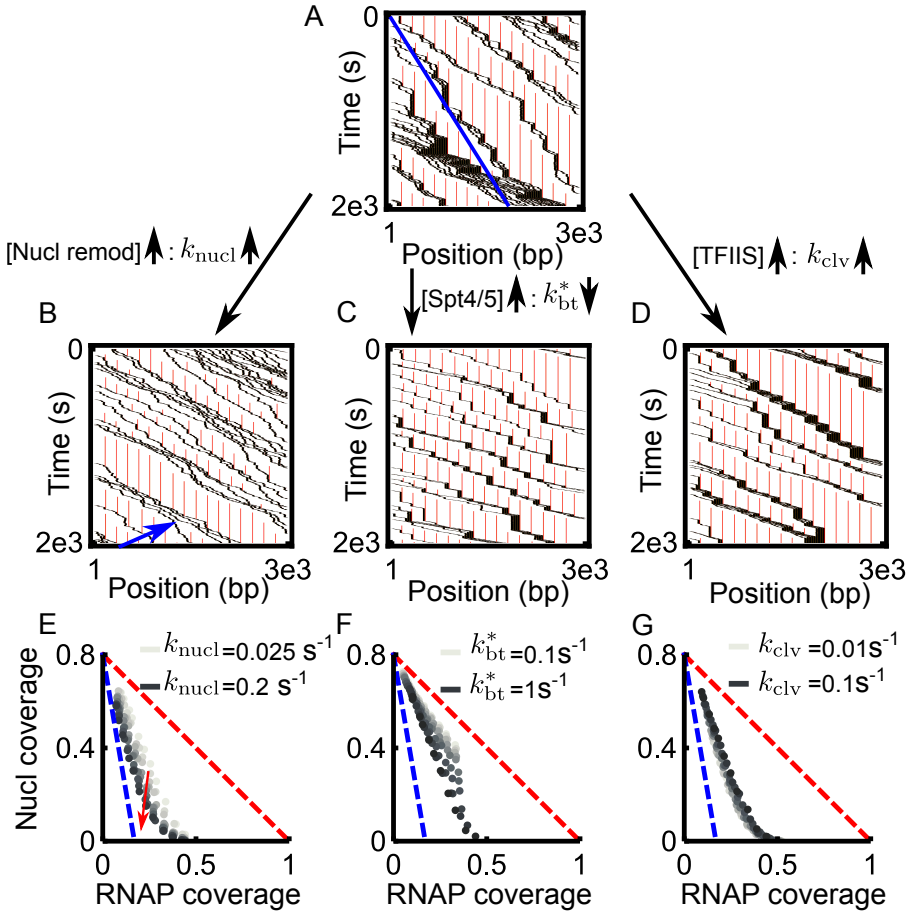


Figure 5.4: Elongation factors and nucleosome remodelers modify nucleosome coverage and polymerase organization on highly transcribed genes. A) Kymograph of a highly transcribed gene for $k_{\text{in}} = 0.01/s$, $k_{\text{nucl}} = 0.03/s$, $k_{\text{bt}}^* = 1/1$ and $k_{\text{clv}} = 0.01/s$. Pelotons are clearly visible and stable for these physiological rates. Due to peloton formation polymerases are faster than a single polymerase (blue line, Equation 5.1). B) Kymograph for an increased nucleosome eviction rate as compared to A), $k_{\text{nucl}} = 0.2/s$. RNAP moves faster and the pelotons are more spread out as compared to A). The blue arrow shows a polymerase peeling off from a peloton. C) Kymograph with a decreased backtrack entry rate as compared to A), $k_{\text{bt}}^* = 0.1$. As a result, RNAP moves faster and pelotons are more dense. D) Kymograph with an increased cleavage rate as compared to A). RNAP moves faster and pelotons are more dense. E) The nucleosome density for a range of nucleosome eviction rates, $k_{\text{in}} = 0.01/s$, $k_{\text{bt}}^* = 1/s$ and $k_{\text{clv}} = 0.01/s$. The initiation rate is constant along the red arrow. For large pelotons of RNAPs that are touching, the nucleosome coverage would approximately follow the red line, while for RNAPs that do not form pelotons, the nucleosome coverage would approximately follow the blue line. The data do not collapse on a line, showing that not only RNAP density, but also RNAP reorganization of pelotons is an important factor in determining the nucleosome coverage. The nucleosome coverage does not reach 1 for zero RNAP coverage, because after eviction and rebinding, nucleosomes are not optimally packed on the DNA. F) The nucleosome density for a range of backtrack entry rates or Spt4/5 concentration for $k_{\text{nucl}} = 0.03/s$ and $k_{\text{clv}} = 0.01/s$. For Spt4/5 the effect of reorganization becomes more important for large polymerase densities. From light to dark gray the nucleosome eviction rate increases logarithmically. G) The nucleosome density for a range of cleavage rates for $k_{\text{in}} = 0.01/s$, $k_{\text{nucl}} = 0.03/s$ and $k_{\text{bt}}^* = 1/s$. The collapse of the data illustrates that for varying initiation and cleavage rate, the polymerase density is the only variable that determines the nucleosome coverage.

5.3. DISCUSSION

To understand how elongation factors and nucleosome remodelers reorganize RNAP and nucleosomes along a gene, we studied a model for transcription including nucleosomes, multiple RNAPs and the effect of elongation factors and nucleosome remodelers on RNAP dynamics. Starting with a single polymerase, we calculated the average velocity, backtrack density and backtrack duration. We then performed Monte Carlo simulations to study cooperation between RNAPs, transcription dynamics and nucleosome density on a heavily transcribed genes with multiple RNAPs.

Nucleosome remodelers and TFIIS have a strong effect on pause duration and the elongation speed, while Spt4/5 has the strongest effect on pause density (Figure 5.2). Interestingly, the physiological range for the nucleosome eviction rate k_{nucl} is in the most tunable range of the velocity, Figure 5.2 B, suggesting the *in vivo* nucleosomal barriers have been chosen for optimal modulation of velocity.

The dynamic range for TFIIS is larger without cooperation (Figure 5.3 B), suggesting that the effect of elongation factors on the processivity of RNAP is especially important for low transcription levels, where the effect of cooperation is negligible. Cooperation between RNAPs on the other hand is a robust mechanism to keep transcription initiation-limited even for high initiation rates (Figure 5.3 C).

Surprisingly, the nucleosome density can be manipulated by transcription elongation factors and nucleosome remodelers without changing initiation rates or overall transcription levels. For physiological values, the rate of nucleosome eviction has a larger influence on nucleosome coverage than elongation factors, because it does not only influence the density, but also organization of polymerases along a gene. These new insights have consequences for the possible biological functions of elongation factors and nucleosome remodelers that we discuss below.

5.3.1. ELONGATION FACTORS MODIFY TRANSCRIPTIONAL BURSTS

The formation of pelotons along genes results in bursty transcription. When a peloton reaches the end of a gene, there is a sudden burst of mRNA production, after which there is no production until the next peloton arrives. When elongation factors reshape pelotons along the gene, they also reshape the transcription output dynamics. The changed RNAP distribution along the gene from Figure 5.4 A to Figures 5.4 B-D results in a different transcription output dynamics. Even when the times between initiation events are exponentially distributed, as in Figures 5.4 A-D, elongation factors can regulate the burst characteristics of a gene. The fluctuations in output dynamics are largest for dense pelotons. Increasing concentrations in Spt4/5 and TFIIS therefore result in larger amplitude fluctuations dynamics, while increased nucleosome eviction rates result in smaller amplitude fluctuations.

5.3.2. PIONEERING POLYMERASES

Multiple polymerases can be paused near a promoter proximal pause site, to allow for rapid activation of a gene [42]. Such pioneering polymerases have been suggested to leave an epigenetic mark and make genes more accessible for further transcription [10]. Since nucleosomes are evicted in two steps, efficient nucleosome eviction can only be achieved when the polymerases are close together. Our results suggest that a low nu-

cleosome eviction rate and a high cleavage rate are beneficial for effective clearing of a gene. Since nucleosomes are often remodeled or replaced during transcription, the next round of polymerases will be faster and further apart than the pioneering polymerases.

5.3.3. HEAVILY TRANSCRIBED GENES

During heavy transcription, multiple RNAPs transcribe the same gene at a time and the density of both histone dimers and tetramers decreases [43]. For such genes it could be beneficial to maximize the nucleosome density, because by maximizing nucleosome density, the DNA is optimally protected, for example against UV radiation. Given a rate of transcription initiation, the highest nucleosome coverage is reached for fast cleavage (more TFIIS), a low rate into the backtrack (more Spt4/5) and a low nucleosome eviction rate, Figure 5.4 E-G. Fast cleavage and a low backtrack entry rate also increase the transcription speed. An increase of the speed of RNAP does not directly increase the rate of mRNA production, as transcription is initiation limited, but it would allow for faster reinitiation in gene loops [44].

5.3.4. CONCLUSION AND OUTLOOK

Using Monte Carlo simulations we showed that elongation factors and nucleosome remodelers could have a completely overlooked role in manipulating the nucleosome density and regulation transcription output dynamics. Both effects are strongest on heavily transcribed genes, where multiple RNAPs transcribe a gene at the same time and the density of both histone dimers and tetramers decreases [43]. Experimental studies could further verify these hypotheses *in vivo* by measuring the nucleosome density using CHIP-seq [43], the transcription output dynamics [45] and by varying the RNAP elongation rate. A more direct method would be to measure the distribution of RNAPs along the gene, but existing *in vivo* techniques only measure average distributions [46], that cannot be used to detect pelotons. However, pelotons could be visualized *in vitro* using DNA curtains and fluorescence [47]. Studying the effect of nucleosome remodeling, for example by modifying nucleosome tails [13], is an interesting start as it has a stronger effect on polymerase reorganization than modifications on pausing. Furthermore, it is intriguing that nucleosome remodelers can control the nucleosome density, by speeding up and reorganizing RNAPs.

ACKNOWLEDGMENTS

MD and AAvdB acknowledge financial support from a TU Delft startup grant to MD. AAvdB further acknowledges financial support by the Netherlands Organisation for Scientific Research (NWO/OCW), as part of the Frontiers of Nanoscience program.

5.4. SUPPLEMENTARY MATERIAL

5.4.1. AVERAGE VELOCITY OF SINGLE RNAP WITH TRANSCRIPTION FACTORS AND NUCLEOSOME REMODELERS

The dynamics of an individual polymerase is modeled as a Markov chain with multiple states reflecting transcriptionally active and backtracked polymerases Figure 5.1 B. The average speed for transcription can be calculated using the theory for continuum time random walks (CTRW) [23]. In this theory steps are discrete, but waiting times are distributed exponentially: $P(t) = \sum k e^{-t \sum k}$, where t is the waiting time and $\sum k$ the sum of the rates leaving a node. In Laplace space, convolutions of the different waiting time distributions for one path become products. Let Ψ indicate the Laplace transforms of a waiting time distributions, $\Psi = \int_{t=0}^{\infty} P(t) dt$. The average waiting time is given by $\tau = -1/\Psi \partial \Psi / \partial s|_{s=0}$ and the probability for a transition to happen at any time is $\int_{t=0}^{\infty} P(t) dt = \Psi(s=0)$. To determine the first passage time τ from one active site to the next we sum over all the possible paths a polymerase could take to reach the next site, for example entering a backtrack, RNA cleavage, returning to the original site and making a step forward. The polymerase can backtrack and return to the original site an arbitrary number of times before making a step, therefore

$$\Psi_{\text{tot}}^{\text{cat/nucl}} = \sum_{n=0}^{\infty} (\Psi_{\text{bt}}^{\text{cat/nucl}} \Psi_{\text{back}})^n \Psi_{\text{cat/nucl}}, \quad (5.4)$$

where $\Psi_{\text{bt}}^{\text{cat/nucl}}$ is the Laplace transform of the waiting time distribution for going from the active state into the backtrack without or with a dyad on the next lattice site respectively. For illustration we give the complete derivation for $\Psi_{\text{bt}}^{\text{cat/nucl}}$: the probability to enter a backtrack, $k_{\text{bt}}/(k_{\text{bt}}^* + k_{\text{cat/nucl}})$ is multiplied by the Laplace transform of the total waiting time distribution for leaving the active site and the forward rate is k_{nucl} in front of a nucleosome and k_{cat} otherwise,

$$\begin{aligned} \Psi_{\text{bt}}^{\text{cat/nucl}} &= \frac{k_{\text{bt}}^*}{k_{\text{bt}}^* + k_{\text{cat/nucl}}} \int_0^{\infty} (k_{\text{bt}}^* + k_{\text{cat/nucl}}) e^{-(k_{\text{bt}}^* + k_{\text{cat/nucl}})t} e^{-st} dt \\ &= \frac{k_{\text{bt}}^*}{s + k_{\text{bt}}^* + k_{\text{cat/nucl}}}. \end{aligned} \quad (5.5)$$

$\Psi_{\text{cat/nucl}}$ is the probability distribution for making a step forward to an active site without or with nucleosome and can be derived analogously to $\Psi_{\text{bt}}^{\text{cat/nucl}}$,

$$\Psi_{\text{cat/nucl}} = \frac{k_{\text{cat/nucl}}}{s + k_{\text{bt}}^* + k_{\text{cat/nucl}}}. \quad (5.6)$$

The function Ψ_{back} in Equation 5.4 includes all the possible paths of returning back to the starting position

$$\Psi_{\text{back}} = \Psi_{\text{srv}} + \sum_{n=1}^{\infty} \Psi_{\text{clv,n}} (\Psi_{\text{tot}}^{\text{cat}})^n. \quad (5.7)$$

Ψ_{back} only depends on Ψ^{cat} and not on Ψ^{nucl} , because the polymerase just evicted the nucleosomes (dyads) behind it, and in physiological settings the polymerase does not backtrack far enough to encounter a newly bound nucleosome. Ψ_{srv} is the Laplace transform of the waiting time distribution to start from the first position of a backtrack and then leave the backtrack *without* cleaving, while $\Psi_{\text{clv},n}$ is the Laplace Transform of the waiting time distribution for starting at the first backtrack, labeled 1, and cleavage at position $n \geq 1$ in the backtrack, returning to an active site that is n positions behind the starting position. After cleavage, the polymerase has to pass n active sites to return to the original position and at each site the polymerase can backtrack, which is included in the term $(\Psi_{\text{tot}}^{\text{cat}})^n$

$$\Psi_{\text{srv}} = \frac{2k_{\text{bt}} + k_{\text{clv}} + s - \sqrt{k_{\text{clv}} + s} \sqrt{4k_{\text{bt}} + k_{\text{clv}} + s}}{2k_{\text{bt}}} \quad (5.8)$$

$$\Psi_{\text{clv}} = \frac{k_{\text{clv}}}{k_{\text{bt}}} \Psi_{\text{srv}}^n \quad (5.9)$$

Full derivations for Equation 5.8 and 5.9 are given in [23]. Equation 5.4 and 5.7 together give a self-consistent relation for $\Psi_{\text{tot}}^{\text{cat}}$ from which the first passage time to a site without a nucleosome can be obtained, Figure 5.5 B

$$\Psi_{\text{tot}}^{\text{cat}} = \frac{\Psi_{\text{cat}}}{1 - \Psi_{\text{bt}}^{\text{cat}} \Psi_{\text{srv}} (1 + \frac{k_{\text{clv}}}{k_{\text{bt}}} \frac{\Psi_{\text{tot}}^{\text{cat}}}{1 - \Psi_{\text{srv}} \Psi_{\text{tot}}^{\text{cat}}})} \quad (5.10)$$

$$\tau_{\text{tot}}^{\text{cat}} = -\frac{1}{\Psi_{\text{tot}}^{\text{cat}}} \frac{\partial \Psi_{\text{tot}}^{\text{cat}}}{\partial s} \Big|_{s=0} = \frac{2 + \frac{k_{\text{bt}}^*}{k_{\text{bt}}} \left(\sqrt{1 + 4 \frac{k_{\text{bt}}}{k_{\text{clv}}} - 1} \right)}{2(k_{\text{cat}} - k_{\text{bt}}^*)} \quad (5.11)$$

Equation 5.11 reduces to the first passage time without transcription factors and nucleosomes when $k_{\text{bt}}^* = k_{\text{bt}}$ [23]. The velocity in the absence of nucleosome is only positive (downstream) if the catalyzation rate is higher than the rate for entering a backtrack, $k_{\text{cat}} > k_{\text{bt}}^*$.

Using the expression for $\Psi_{\text{tot}}^{\text{cat}}$ we can now determine $\Psi_{\text{tot}}^{\text{nucl}}$ by combining Equation 5.4-5.9. The first passage time to a lattice site with a dyad is given by

$$\tau_{\text{tot}}^{\text{nucl}} = -\frac{1}{\Psi_{\text{tot}}^{\text{nucl}}} \frac{\partial \Psi_{\text{tot}}^{\text{nucl}}}{\partial s} \Big|_{s=0} = \tau_{\text{nucl}} + \frac{k_{\text{bt}}^*}{k_{\text{nucl}}} (\tau_{\text{bt}}^{\text{nucl}} + \tau_{\text{back}}), \quad (5.12)$$

where τ_{nucl} is the first passage time to a site with a dyad, without entering a backtrack, $\tau_{\text{bt}}^{\text{nucl}}$ is the first-passage time for entering a backtrack when the next site is a dyad and τ_{back} is the total first passage time for returning to the starting position after entering a backtrack, see Figure 5.5 C:

$$\tau_{\text{back}} = -\frac{1}{\Psi_{\text{back}}} \frac{\partial \Psi_{\text{back}}}{\partial s} \Big|_{s=0} = (1 + P_{\text{srv}}) \tau_{\text{srv}} + \tau_{\text{tot}}^{\text{cat}} \quad (5.13)$$

$$P_{\text{srv}} = \Psi_{\text{srv}}(s=0) = 1 + \frac{k_{\text{clv}}}{2k_{\text{bt}}} (1 - \sqrt{1 + 4k_{\text{bt}}/k_{\text{clv}}}) \quad (5.14)$$

$$\tau_{\text{srv}} = -\frac{1}{\Psi_{\text{srv}}} \frac{\partial \Psi_{\text{srv}}}{\partial s} \Big|_{s=0} = \frac{1}{k_{\text{clv}} \sqrt{1 + 4k_{\text{bt}}/k_{\text{clv}}}} \quad (5.15)$$

$$-\frac{1}{\Psi_{\text{bt}}^{\text{nucl}}} \frac{\partial \Psi_{\text{bt}}^{\text{nucl}}}{\partial s} \Big|_{s=0} = -\frac{1}{\Psi_{\text{nucl}}} \frac{\partial \Psi_{\text{nucl}}}{\partial s} \Big|_{s=0} \Rightarrow \tau_{\text{nucl}} = \tau_{\text{bt}}^{\text{nucl}} = \frac{1}{k_{\text{nucl}} + k_{\text{bt}}^*} \quad (5.16)$$

The average elongation speed can be obtained by averaging Equation 5.11 and 5.12 over a nucleosome, giving Equation 5.1 in the main text.

5.4.2. PAUSE DURATION

The duration of a pause, τ_{p} , is defined as the first passage time to the next site, given that the RNAP enters a backtrack at least once.

$$\Psi_{\text{p}}^{\text{cat/nucl}} = \Psi_{\text{back}} \sum_{n=0}^{\infty} (\Psi_{\text{bt}}^{\text{cat/nucl}} \Psi_{\text{back}})^n \Psi_{\text{cat/nucl}} \quad (5.17)$$

$$\tau_{\text{p}}^{\text{cat/nucl}} = -\frac{1}{\Psi_{\text{p}}^{\text{cat/nucl}}} \frac{\partial \Psi_{\text{p}}^{\text{cat/nucl}}}{\partial s} \Big|_{s=0} = \tau_{\text{tot}}^{\text{cat/nucl}} + \tau_{\text{back}} \quad (5.18)$$

The actual pause density is not simply given by $k_{\text{bt}}^*/(k_{\text{bt}}^* + k_{\text{cat}})$, because after cleavage, a polymerase can pass the same site a second time, doubling the probability to enter a pause at that site. However, in single molecule experiments, it is difficult to see whether a motor is in the active state or not. Therefore, we choose this definition of pause density and we define the pause duration as the time to get back to the starting site after entering a backtrack, including the possibility to cleave and backtrack again.

5.4.3. MONTE CARLO SIMULATIONS

We used fixed timestep Monte Carlo for simulating multiple polymerases and nucleosomes along a gene. The largest rate in our system is $k_{\text{cat}} + k_{\text{bt}}^*$ and we choose the timesteps such that $(k_{\text{cat}} + k_{\text{bt}}^*)\Delta t \approx 0.1$. The lattice is updated random sequentially: during every iteration step of duration Δt all lattice sites are updated once, in random order. A potentially initiating polymerase is also counted as a lattice site. To update the DNA lattice with RNAPs and nucleosomes several events can take place: an RNAP can move forward, evict a nucleosome, enter a backtracking state, move backward and forward in the backtracking state and get cleaved. To capture the effect of pushing, an active RNAP is able to push forward a maximum of two backtracking RNAPs.

For an active RNAP the following steps can happen:

- If the next site is free: step forward with rate k_{cat}
- If the next site is a nucleosome: evict nucleosome and step forward with rate k_{nucl}

- If the previous site is free: enter backtracking state with rate k_{bt}
- If the RNAP is located at the end of the lattice: terminated from the lattice with rate k_{exit} .
- If the next one (or two) sites are occupied by backtracking RNAP(s) and the second (or third) site is free: step forward with rate k_{cat} and also push the backtracking RNAP(s) forward
- If the next one (or two) sites are occupied by backtracking RNAP(s) and the second (or third) site is occupied by a nucleosome: evict the nucleosome with rate k_{nucl} and also step forward and push forward the backtracking RNAP(s)

For a backtracking RNAP there are some similar moving rules:

- If the next site is free: step forward with rate k_{bt}
- If the previous site is free: step backward with rate k_{bt}
- Can be cleaved and becomes an active RNAP with rate k_{clv} and remains at the same base pair position

All the sites which are not occupied by an RNAP or nucleosome are also updated. If there is a free site a nucleosome can bind with rate k_b . If the first site of the lattice is free a new RNAP can be initiated with rate k_{in} . The probability for a certain event to occur is $k\Delta t$, assuming all the necessary conditions are met. Imagine an RNAP is in between two free sites, thus it can move forward (probability $k_{cat}\Delta t$), enter a backtracking state (probability $k_{bt}\Delta t$) or nothing happens (probability $1 - k_{cat}\Delta t - k_{bt}\Delta t$). What is realized in the simulation is determined by a random number picked from a uniform distribution $[0, 1]$.

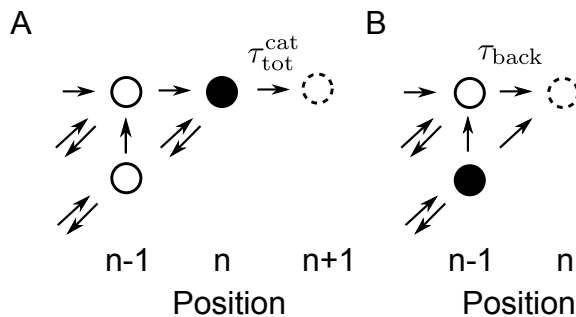


Figure 5.5: A) Schematic of all the possible paths included in τ_{tot}^{cat} , the first passage time without nucleosomes, from the filled to the dashed node, Equation 5.11. B) Schematic of all the possible paths included in τ_{back} , the first-passage time from one step into the backtrack, the filled node, to the starting position, the dashed node, Equation 5.13.

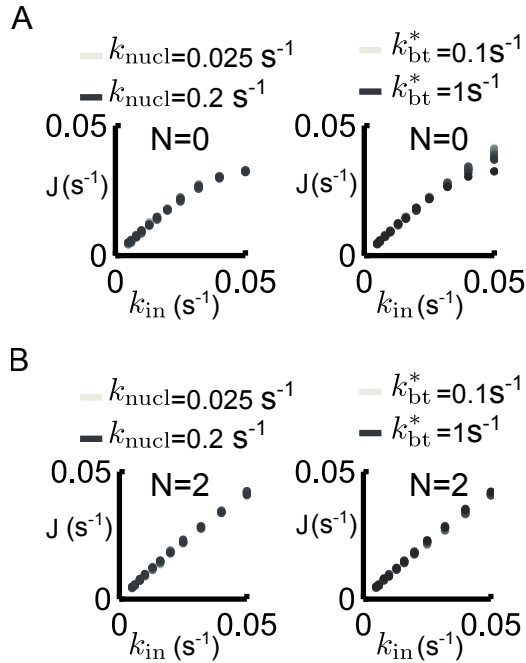


Figure 5.6: The flux versus the transcription initiation rate for the parameter values that are also used in the main text. The sweeps for the nucleosome eviction rate, k_{nucl} , are given in the left column and the sweeps for the backtracked entry rate, k_{bt}^* , in the right column for A) an RNAP that cannot push and B) and RNAP that can push two backtracked RNAPs. All these plots show that transcription is initiation-limited for the given parameter sweeps.

REFERENCES

- [1] I. Jonkers, H. Kwak, and J. T. Lis, *Promoter-proximal pausing of RNA polymerase II: emerging roles in metazoans*, *eLife* **3**, 1 (2014).
- [2] I. Jonkers and J. T. Lis, *Getting up to speed with transcription elongation by RNA polymerase II*. *Nature reviews. Molecular cell biology* **16**, 167 (2015).
- [3] C. Jeon and K. Agarwal, *Fidelity of RNA polymerase II transcription controlled by elongation factor TFIIIS*, *Proc. Natl. Acad. Sci. U. S. A.* **93**, 13677 (1996).
- [4] E. A. Galburt, S. W. Grill, A. Wiedmann, L. Lubkowska, J. Choy, E. Nogales, M. Kashlev, and C. Bustamante, *Backtracking determines the force sensitivity of RNAP II in a factor-dependent manner*. *Nature* **446**, 820 (2007).
- [5] D. S. Luse, L. C. Spangler, and A. Újvári, *Efficient and rapid nucleosome traversal by RNA polymerase II depends on a combination of transcript elongation factors*, *Journal of Biological Chemistry* **286**, 6040 (2011).
- [6] B. Li, M. Carey, and J. L. Workman, *The role of chromatin during transcription*. *Cell* **128**, 707 (2007).
- [7] R. J. Sims, R. Belotserkovskaya, and D. Reinberg, *Elongation by RNA polymerase II: the short and long of it*. *Genes & development* **18**, 2437 (2004).
- [8] B. Alberts, D. Bray, K. Hopkin, A. Johnson, J. Lewis, M. Raff, K. Roberts, and P. Walter, *Essential cell biology*, 2nd ed. (Garland Science, Taylor and Francis Group, 2004).
- [9] C. Hodges, L. Bintu, L. Lubkowska, M. Kashlev, and C. Bustamante, *Nucleosomal fluctuations govern the transcription dynamics of RNA polymerase II*. *Science (New York, N.Y.)* **325**, 626 (2009).
- [10] O. I. Kulaeva, F.-K. Hsieh, and V. M. Studitsky, *RNA polymerase complexes cooperate to relieve the nucleosomal barrier and evict histones*. *Proceedings of the National Academy of Sciences* **107**, 11325 (2010).
- [11] R. U. Protacio, G. Li, P. T. Lowary, and J. Widom, *Effects of histone tail domains on the rate of transcriptional elongation through a nucleosome*, *Mol Cell Biol* **20**, 8866 (2000).
- [12] A. Újvári, F. K. Hsieh, S. W. Luse, V. M. Studitsky, and D. S. Luse, *Histone N-terminal tails interfere with nucleosome traversal by RNA polymerase II*, *Journal of Biological Chemistry* **283**, 32236 (2008).
- [13] L. Bintu, T. Ishibashi, M. Dangkulwanich, Y.-Y. Wu, L. Lubkowska, M. Kashlev, and C. Bustamante, *Nucleosomal elements that control the topography of the barrier to transcription*. *Cell* **151**, 738 (2012).
- [14] K. M. Herbert, W. J. Greenleaf, and S. M. Block, *Single-molecule studies of RNA polymerase: motoring along*. *Annual review of biochemistry* **77**, 149 (2008).

- [15] M. H. Larson, R. Landick, and S. M. Block, *Single-molecule studies of RNA polymerase: one singular sensation, every little step it takes*. *Molecular cell* **41**, 249 (2011).
- [16] N. Komissarova and M. Kashlev, *Transcriptional arrest: Escherichia coli rna polymerase translocates backward, leaving the 3' end of the rna intact and extruded*, *Proceedings of the National Academy of Sciences* **94**, 1755 (1997), <http://www.pnas.org/content/94/5/1755.full.pdf>.
- [17] N. Komissarova and M. Kashlev, *Rna polymerase switches between inactivated and activated states by translocating back and forth along the dna and the rna*, *Journal of Biological Chemistry* **272**, 15329 (1997), <http://www.jbc.org/content/272/24/15329.full.pdf+html>.
- [18] J. W. Shaevitz, E. a. Abbondanzieri, R. Landick, and S. M. Block, *Backtracking by single RNA polymerase molecules observed at near-base-pair resolution*. *Nature* **426**, 684 (2003).
- [19] F. W. Martinez-Rucobo, S. Sainsbury, A. C. M. Cheung, and P. Cramer, *Architecture of the RNA polymerase-Spt4/5 complex and basis of universal transcription processivity*. *The EMBO journal* **30**, 1302 (2011).
- [20] M. J. Thomas, A. A. Platas, and D. K. Hawley, *Transcriptional fidelity and proofreading by rna polymerase ii*, *Cell*, *Cell* **93**, 627 (1998).
- [21] A. A. van den Berg and M. Depken, *Crowding-induced transcriptional bursts dictate nucleosome and polymerase density profiles along genes*, *Nucleic Acids Res* **45**, 7623–7632 (2017).
- [22] S. Klumpp, *Pausing and Backtracking in Transcription Under Dense Traffic Conditions*, *Journal of Statistical Physics* **142**, 1252 (2011).
- [23] M. Depken, J. M. R. Parrondo, and S. W. Grill, *Intermittent transcription dynamics for the rapid production of long transcripts of high fidelity*. *Cell reports* **5**, 521 (2013).
- [24] C. T. MacDonald, J. H. Gibbs, and A. C. Pipkin, *Kinetics of Biopolymerization on Nucleic Acid Templates*, *Biopolymers* **6**, 1 (1968).
- [25] I. Touloukhonov, J. Zhang, M. Palangat, and R. Landick, *A central role of the rna polymerase trigger loop in active-site rearrangement during transcriptional pausing*, *Molecular Cell*, *Molecular Cell* **27**, 406.
- [26] S. J. Greive and P. H. V. Hippel, *Thinking quantitatively about transcriptional regulation*. *Nature reviews. Molecular cell biology* **6**, 221 (2005).
- [27] K. Luger, a. W. Mäder, R. K. Richmond, D. F. Sargent, and T. J. Richmond, *Crystal structure of the nucleosome core particle at 2.8 Å resolution*. *Nature* **389**, 251 (1997).
- [28] K. Brogaard, L. Xi, J.-P. Wang, and J. Widom, *A map of nucleosome positions in yeast at base-pair resolution*. *Nature* **486**, 496 (2012).

- [29] K. C. Neuman, E. A. Abbondanzieri, R. Landick, J. Gelles, and S. M. Block, *Ubiquitous Transcriptional Pausing Is Independent of RNA Polymerase Backtracking*, *Cell* **115**, 437 (2003).
- [30] X. Darzacq, Y. Shav-Tal, V. de Turris, Y. Brody, S. M. Shenoy, R. D. Phair, and R. H. Singer, *In vivo dynamics of RNA polymerase II transcription*. *Nature structural & molecular biology* **14**, 796 (2007).
- [31] V. Pelechano, S. Chavez, and J. E. Perez-Ortin, *A Complete Set of Nascent Transcription Rates for Yeast Genes*, *PLoS one* **5**, e15442 (2010), [arXiv:1203.2655](https://arxiv.org/abs/1203.2655).
- [32] J. Mazurkiewicz, J. F. Kepert, and K. Rippe, *On the mechanism of nucleosome assembly by histone chaperone NAPI*. *The Journal of biological chemistry* **281**, 16462 (2006).
- [33] A. Worcel, S. Han, and M. L. Wong, *Assembly of newly replicated chromatin*. *Cell* **15**, 969 (1978).
- [34] M. Depken, E. A. Galburt, and S. W. Grill, *The origin of short transcriptional pauses*. *Biophysical journal* **96**, 2189 (2009).
- [35] A. Lisica, C. Engel, M. Jahnel, E. Roldán, E. A. Galburt, P. Cramer, and S. W. Grill, *Mechanisms of backtrack recovery by RNA polymerases I and II*, *Proc. Natl. Acad. Sci. U. S. A.* **113**, 2946 (2015).
- [36] K. M. Herbert, J. Zhou, R. a. Mooney, A. L. Porta, R. Landick, and S. M. Block, *E. coli NusG inhibits backtracking and accelerates pause-free transcription by promoting forward translocation of RNA polymerase*. *Journal of molecular biology* **399**, 17 (2010).
- [37] B. Osberg, J. Nuebler, P. Korber, and U. Gerland, *Replication-guided nucleosome packing and nucleosome breathing expedite the formation of dense arrays*, *Nucleic Acids Research* **42**, 13633 (2014).
- [38] V. Epshtein and E. Nudler, *Cooperation Between RNA Polymerase Molecules in Transcription Elongation*, *Science* **300**, 801 (2003).
- [39] J. Jin, L. Bai, D. S. Johnson, R. M. Fulbright, M. L. Kireeva, M. Kashlev, and M. D. Wang, *Synergistic action of RNA polymerases in overcoming the nucleosomal barrier*. *Nature structural & molecular biology* **17**, 745 (2010).
- [40] C. G. Danko, N. Hah, X. Luo, A. L. Martins, L. Core, J. T. Lis, A. Siepel, and W. L. Kraus, *Signaling Pathways Differentially Affect RNA Polymerase II Initiation, Pausing, and Elongation Rate in Cells*, *Molecular Cell* **50**, 212 (2013).
- [41] E. A. Galburt, J. M. R. Parrondo, and S. W. Grill, *RNA polymerase pushing*. *Biophysical chemistry* **157**, 43 (2011).
- [42] L. J. Core and J. T. Lis, *Transcription regulation through promoter-proximal pausing of RNA polymerase II*. *Science (New York, N.Y.)* **319**, 1791 (2008).

- [43] H. A. Cole, J. Ocampo, J. R. Iben, R. V. Chereji, and D. J. Clark, *Heavy transcription of yeast genes correlates with differential loss of histone H2B relative to H4 and queued RNA polymerases*. [Nucleic acids research](#) **42**, 12512 (2014).
- [44] S. M. Tan-Wong, J. B. Zaugg, J. Camblong, Z. Xu, D. W. Zhang, H. E. Mischo, A. Z. Ansari, N. M. Luscombe, L. M. Steinmetz, and N. J. Proudfoot, *Gene loops enhance transcriptional directionality*. [Science \(New York, N.Y.\)](#) **338**, 671 (2012).
- [45] K. Tantale, F. Mueller, A. Kozulic-Pirher, A. Lesne, J.-M. Victor, M.-C. Robert, S. Capozzi, R. Chouaib, V. Bäcker, J. Mateos-Langerak, X. Darzacq, C. Zimmer, E. Basyuk, and E. Bertrand, *A single-molecule view of transcription reveals convoys of RNA polymerases and multi-scale bursting*, [Nature Communications](#) **7**, 12248 (2016).
- [46] L. S. Churchman and J. S. Weissman, *Nascent transcript sequencing visualizes transcription at nucleotide resolution*. [Nature](#) **469**, 368 (2011).
- [47] A. Granéli, C. C. Yeykal, and E. C. Greene, *Organized Arrays of Individual DNA Molecules Tethered to Supported Lipid Bilayers*, [Langmuir](#) **22**, 292 (2006).

6

DISCUSSION AND PERSPECTIVES

Though the research in this thesis is motivated by fundamental questions about transcription, the studied model can be used to describe a broad class of systems where motors interact with roadblocks. The motor protein kinesin for example transports cargo along microtubules and trails can be describe by the bus route model [1] with pheromones as an effective attractive force between the ants. In this concluding chapter we discuss the broader implications of the results within and outside biology and possible future directions.

6.1. TRANSCRIPTIONAL BURSTS

Gene expression can be described as a stochastic process, and this stochasticity is often described as 'noise'. Noise in gene expression has both advantages and disadvantages. On the one hand, noise complicates precise regulation of genes, while on the other hand, noise can lead to cell-to-cell phenotypic variation, while genotypes are the same, which is essential for organisms to survive in fluctuating environments [2]. The distribution of times between protein production events varies from gene to gene. Often this distribution is bursty, with periods of low or no protein production and periods of high production [3].

As the first step in gene expression, transcription has the potential of inducing fluctuations in protein production. Indeed, transcription has been measured to be bursty and many mechanisms have been suggested as a source of bursts in transcription. Most mechanisms rely on bursty initiation, for example due to a promoter that is turned on and off by transcription factors, gene loops [4] or a small numbers of transcription factors [5]. Peloton formation is one of the few mechanisms that suggest the elongation phase as a source of bursting due to pausing of RNAP [6].

Whether or not transcriptional bursts propagate to result in bursts of protein production depends on the relative time scales of transcription and translation. The elongation induced bursts we present occur while a gene is constantly on, while on larger time scales, a gene also can switch to an off state resulting in initiation-induced bursts. The

times between elongation induced bursts and between events within a burst are therefore small compared to the time scales for initiation induced bursts. If both elongation- and initiation induced bursts are present, transcription would be bursty on multiple time scales, which has been observed experimentally [7]. For typical transcription and mRNA degradation rates, protein fluctuations are determined by time-averaged properties of fluctuations in mRNA and intrinsic noise is controlled at the translational level [8, 9]. When ignoring peloton formation, only low transcription levels result in bursts of protein production since every mRNA is translated multiple times, while for high transcription levels (where peloton formation is expected) noise is controlled at the translational level [8]. However, peloton formation has not been considered in previous studies on bursts in protein production. Future studies on gene regulatory networks can clarify the role of peloton formation in gene regulation.

6.2. SPECIFIC GENE TARGETING USING TRANSCRIPTION

The discovery of the CRISPR-Cas system, a prokaryotic immune system, allows for more effective gene editing than researchers could have dreamed of [10]. CRISPR-Cas allows for fast and specific cutting of genes and recent experiments show promising results for targeting genetic diseases in humans [11]. However, the Cas protein is hindered by nucleosomes [12, 13]. As discussed in Chapter 4, nucleosome eviction by RNAP is sometimes used to create access to a gene, for example for the RSS protein [14] and Rad26p [15]. The same trick could be used to create access for CRISPR-Cas. Using transcription has the advantage that genes can be targeted specifically as transcription factors ensure that only the targeted gene is exposed.

A histone should be evicted completely to expose DNA for gene editing. Using our model, the minimal transcription intensity can be determined, for which the DNA is exposed just enough to allow access for CRISPR-Cas. Due to peloton formation, RNAPs are efficient at evicting nucleosomes, but inducing too many spurious transcription events might have unforeseen side effects. For a quantitative prediction of gene targeting by CRISPR-Cas more details on protein dynamics should be included in the model.

6.3. THE ROBUSTNESS OF SPONTANEOUS PROCESSES

Due to the complexity of living systems, missing details are sometimes seen as a weakness of a model. However, rather than being a weakness, simplifications can sometimes help to see the big picture and allow us to see what processes happen spontaneously, without the need for additional regulation. Many biological details on transcription elongation [16] and histone exchange are missing in the theoretical model presented in this thesis. Yet, the agreement with *in vivo* experiments gives us confidence that these details do not influence the main conclusions of this thesis, that RNAP forms pelotons and that histone exchange can be dynamic. Our results show that peloton formation and histone replacement are robust processes that do not depend on specific regulatory factors, but rely on spontaneous, physical processes. This also makes sense from an evolutionary perspective: transcription factors do not make completely new reactions happen, they catalyze existing processes.

6.4. OPTIMIZING BUS TRAFFIC

The model for crowded transcription studied in this thesis is an extension of the Bus Route Model (BRM). The BRM has been studied extensively to model bus traffic [17, 18]. Many of these studies consider periodic boundary conditions and open boundaries are only studied using simulations [19]. To our knowledge, we are the first to give (heuristic) solutions for open boundary conditions and multiple roadblock species. These results can be used to optimize bus traffic. Unlike for transcription, the formation of pelotons of buses is unfavorable, because the presence of bus pelotons implies that the buses are not on schedule. The peloton size is minimized and the distance over which pelotons are formed is maximized when buses stop as short as possible for passengers (weak roadblock interactions). Pelotons disappear when there are so many passengers, that there is always somebody waiting at a bus stop or (obviously) when there are no passengers at all. Hence, scaling the number of buses with number of passengers would prevent peloton formation. The analytical solutions presented in this thesis could be used to further quantify these statements.

6.5. PHASE TRANSITIONS

We only discussed the sticky motor regime in the initiation-limited phase (Chapter 3), but the TASEP has three phases (Chapter 2) and in the bus route model, there is an additional phase transition, a jamming transition when the roadblock binding rate approaches zero, $k_b \rightarrow 0$ [20]. In open systems, this jamming transition is never reached as there is a constant flux of motors entering the system and they cannot all condense into one peloton over a finite distance. Therefore, we expect that the TASEP with roadblocks has the same three phases as the TASEP. Since transcription is initiation-limited, we never solved for the full phase diagram for the TASEP with roadblocks. For those who want to study the full phase diagram in more detail, we here use hand waving arguments to discuss the three phases in the limits of small and large pelotons. For convenience, we here consider motors and roadblocks of size 1. The shape of the phase diagram depends on the strength of motor-roadblock interactions and we here discuss the limit of weak and strong interactions.

For weak interactions between motors and roadblocks, pelotons are small and there is a roadblock between every pair of motors. Therefore the system behaves as a TASEP without roadblocks, but with a motor hopping rate that is set by the roadblock eviction rate, k_{tp} . Following the same reasoning as for the TASEP (Chapter 2), the phase transition between the initiation- and the bulk-limited phase will be at $k_{in}/k_{tp} = 1/2$, the phase transition between the termination- and the bulk-limited phase will at $k_{ter}/k_{tp} = 1/2$ and the phase transition between the initiation- and termination limited phase at $k_{in} = k_{ter}$ [1] (compare to Figure 2.2 C).

In the sticky motor regime, the distribution of motors changes from the start to the lattice to the point where steady state is reached, making the dynamics different than for the TASEP. We estimate the position of the phase boundary between the initiation- and maximal current limited phase by calculating when the flux in the initiation-limited phase matches the flux in the maximal current phase. The maximal current is the velocity of a peloton, $v = k_{tp}$ times the density at which the whole system has the same density

as a peloton, $\rho_1 = 1 - k_{tp}/k_{ip}$ [20] (Figure 3.6), giving

$$J_{\max} = k_{tp}(1 - k_{tp}/k_{ip}). \quad (6.1)$$

Next we have to determine what the initiation rate would be, when the input flux matches J_{\max} in the initiation-limited phase. At the start of the lattice, pelotons have not formed yet. For large pelotons, most motors at the start of the lattice have no roadblock in front of them, such that the average velocity is approximately k_{ip} and if we assume that the density of motors evolves slowly, such that the density is approximately the same at the first two lattice sites, conservation of motors dictates that $k_{in}(1 - \rho_{in}) = k_{ip}(1 - \rho_{in})$. Here ρ_{in} is the density at the first two lattice sites. Solving for ρ_{in} gives an input flux $J_{in} = k_{in}(1 - k_{in}/k_{ip})$. In the limit of large peloton formation, $k_{tp} < 0.5$, such that the lowest initiation rate at which the maximal current is reached is

$$k_{in} = k_{tp}, \quad (6.2)$$

which gives the position of the phase boundary between the initiation-limited and maximal current phase. Similarly, we can estimate the position of the phase boundary between the bulk-limited and the termination-limited phase. At the end of the gene, motors have formed pelotons, such that the flux out of the system is $J_{ter} = (1 - k_{tp}/k_{ip})k_{ter}$. If we ignore possible boundary effects near the termination site, the termination rate at the phase boundary between the termination-limited and maximal current phase is given by

$$k_{ter} = k_{tp}, \quad (6.3)$$

which gives the second phase boundary. Note that, the phase boundaries from the initiation- and termination-limited phase to the maximal current phase in the sticky motor regime have shifted to larger values as compared to the small peloton limit.

The position of the third phase boundary is more difficult to estimate, as it depends on shocks that are moving through the systems and the properties of these shocks could be influenced by the formation of pelotons. If peloton formation does not have a strong influence on the formation of shocks, the third phase boundary would be at the same position as without peloton formation, $k_{in} = k_{ter}$. We expect that for intermediate regimes of peloton formation the phase transitions will be in between those of the small peloton and sticky motor regime. Future research can reveal the phase diagram in more detail.

REFERENCES

- [1] A. Kunwar, A. John, K. Nishinari, A. Schadschneider, and D. Chowdhury, *Collective traffic-like movement of ants on a trail: dynamical phases and phase transitions*, *Journal of the Physical Society of Japan* **73**, 2979 (2004).
- [2] A. Raj and A. van Oudenaarden, *Nature, nurture, or chance: stochastic gene expression and its consequences*. *Cell* **135**, 216 (2008).
- [3] B. Munsky, G. Neuert, and A. van Oudenaarden, *Using gene expression noise to understand gene regulation*. *Science* **336**, 183 (2012).
- [4] D. Hebenstreit, *Are gene loops the cause of transcriptional noise?* *Trends in genetics* **29**, 333 (2013).
- [5] B. Meyer, O. Bénichou, Y. Kafri, and R. Voituriez, *Geometry-induced bursting dynamics in gene expression*. *Biophysical journal* **102**, 2186 (2012).
- [6] M. Dobrzynski and F. J. Bruggeman, *Elongation dynamics shape bursty transcription*, *PNAS* **106**, 2583 (2009).
- [7] K. Tantale, F. Mueller, A. Kozulic-Pirher, A. Lesne, J.-M. Victor, M.-C. Robert, S. Capozzi, R. Chouaib, V. Bäcker, J. Mateos-Langerak, X. Darzacq, C. Zimmer, E. Basyuk, and E. Bertrand, *A single-molecule view of transcription reveals convoys of RNA polymerases and multi-scale bursting*, *Nature Communications* **7**, 12248 (2016).
- [8] M. Thattai and A. V. Oudenaarden, *Intrinsic noise in gene regulatory networks*, *Proceedings of the National Academy of Sciences* **98**, 8614 (2001).
- [9] E. M. Ozbudak, M. Thattai, I. Kurtser, A. D. Grossman, and A. van Oudenaarden, *Regulation of noise in the expression of a single gene*. *Nature genetics* **31**, 69 (2002).
- [10] M. Jinek, K. Chylinski, I. Fonfara, M. Hauer, J. A. Doudna, and E. Charpentier, *A Programmable Dual-RNA – Guided DNA Endonuclease in Adaptive Bacterial Immunity*, *Science* **337**, 816 (2012).
- [11] L. Cong, F. A. Ran, D. Cox, S. Lin, R. Barretto, N. Habib, P. D. Hsu, X. Wu, W. Jiang, L. A. Marraffini, and F. Zhang, *Multiplex genome engineering using crispr/cas systems*, *Science* **339**, 819 (2013), <http://science.sciencemag.org/content/339/6121/819.full.pdf>.
- [12] M. A. Horlbeck, L. B. Witkowsky, B. Guglielmi, J. M. Replogle, L. A. Gilbert, J. E. Villalta, S. E. Torigoe, R. Tijan, and J. S. Weissman, *Nucleosomes impede Cas9 access to DNA in vivo and in vitro*, *eLife* **5**, e12677 (2016).
- [13] R. S. Isaac, F. Jiang, J. A. Doudna, W. A. Lim, G. J. Narlikar, and R. A. Almeida, *Nucleosome breathing and remodeling constrain CRISPR-Cas9 function*, *eLife* **5**, e13450 (2016).

- [14] S. Bevington and J. Boyes, *Transcription-coupled eviction of histones H2A/H2B governs V(D)J recombination*. [The EMBO journal](#) **32**, 1381 (2013).
- [15] S. Malik and S. R. Bhaumik, *Rad26p, a transcription-coupled repair factor, promotes the eviction and prevents the reassociation of histone H2A-H2B dimer during transcriptional elongation in vivo*, [Biochemistry](#) **51**, 5873 (2012).
- [16] I. Jonkers and J. T. Lis, *Getting up to speed with transcription elongation by RNA polymerase II*. [Nature reviews. Molecular cell biology](#) **16**, 167 (2015).
- [17] T. Nagatani, *The physics of traffic jams*, [Rep. Prog. Phys](#) **65**, 1331 (2002).
- [18] D. Chowdhury, L. Santen, and A. Schadschneider, *STATISTICAL PHYSICS OF VEHICULAR TRAFFIC AND SOME RELATED SYSTEMS*, [Physics Reports](#) **329**, 199 (2000).
- [19] Y.-J. Luo, B. Jia, X.-G. Li, C. Wang, and Z.-Y. Gao, *A realistic cellular automata model of bus route system based on open boundary*, [Transportation Research Part C: Emerging Technologies](#) **25**, 202 (2012).
- [20] O. J. O. Loan, M. R. Evans, and M. E. Cates, *Jamming transition in a homogeneous one-dimensional system : The bus route model*, [Physical Review E](#) **58**, 1404 (1998).

ACKNOWLEDGEMENTS

I hope you enjoyed reading (or skipping through) my thesis. This last chapter is where I can express my gratitude for all the people that have been directly or indirectly involved in the realization of this thesis.

Martin, I had the luxury of being your only student for some time, and you patiently introduced me into the field of theoretical biophysics. In the beginning it was sometimes difficult to keep up with your creativity and speed of thinking and I still remember the euphoric moment when for the first time you said 'I didn't think of that'. Now I am proud of the work that we did over the past years. Other than that, I enjoyed sharing thoughts on science, Buddhism, meditation, et cetera.

Ruben, you were sitting behind the desk next to me most of the time, and you are an excellent source of c++ programming tips, and, more importantly, I always enjoyed talking to you and wish you a lot of success with the next step in your career. Afshin and Tania, my other office mates, thanks for the company and the (non)scientific discussions. Renske and Rianda, you were two bright students that I had the honor to work with during my PhD. I am grateful for your commitment and enthusiasm and am certain that you both have a great career ahead of you. Misha, you are now the senior PhD student in the Depken group. You quickly found your way to the department and are becoming an expert on the Crispr-Cas system. I am sure you will have a wonderful thesis. Behrouz and Marc, the postdocs in the Depken group, it was always interesting to talk to you and Iasonas, we did not have much time to get to know each other, but I wish you good luck. Charl, I enjoyed working on and discussing your data on beta clamp binding during replication. Timon, Liedewij, Marie-Eve and Nynke, thank you for your support and for the scientific discussions.

BN was an inspirational environment with interesting people and I enjoyed being part of it. In the fear of forgetting somebody I hereby thank the entire BN department for the interesting, funny and weird conversations, for the amazing and memorable parties, for the inspiring, high quality talks, for the workshop sheep herding, for the creative environment, for the barbecue on the beach and, of course, for the awesome science.

Last, but not least, I have to thank the friends and family that have no clue what I have been doing over the past years. During my hours outside the university (or sometimes even during tea or lunch breaks) I always enjoyed your company : Erik(foto), Roland, my 'Physics Friends', Rosanne, Awital, Tjerk, Erwin, my 'friends from Rotterdam', Roos, Nadia, Els and Teun and of course Antoinette, Jaap (my parents), Peter en Jelle (my 'little' brothers). More and more I realize how important the people are that are close to you. Eric, we got to know each other during my PhD and now we are married and on an adventure in the US. Every day I am happy that we found each other, I luuuv you :)

CURRICULUM VITÆ

Aafke Andr e VAN DEN BERG

30-10-1987 Born in Rotterdam, The Netherlands.

EDUCATION AND RESEARCH PROJECTS

2000-2006 Libanon Lyceum Rotterdam

2006–2009 Bachelor Technische Natuurkunde (Honors)
Delft University of Technology, The Netherlands
Thesis: Modeling and Design
of Terahertz Quantum Cascade Laser Structures
Supervisors: dr J R Gao and Prof. T M Klapwijk

2009-2012 Master Applied Physics
Delft University of Technology, The Netherlands
Thesis: Time-bin Entanglement with a Single Quantum Dot
Supervisors: dr M Reimer and dr V Zwiller

2010-2011 Research Assistant
Delft University of Technology, The Netherlands
Project title: Probing DNA-YOYO-1 Interactions with
Single-Molecule Magnetic Tweezers
Supervisors: dr J Lipfert and Prof. N H Dekker

2012-2016 Ph.D. Theoretical Biophysics
Delft University of Technology, The Netherlands
Thesis: The Interplay between Polymerase Organization
and Nucleosome Density along DNA
Supervisor: dr M Depken
Promotor: Prof. N H Dekker

AWARDS

2009 Casimir prize

2015 1 st Poster prize Annual Meeting of the Dutch Biophysical Society

LIST OF PUBLICATIONS

5. **A.A. van den Berg**, R. Molenaar and M. Depken, *Transcription elongation factors modify nucleosome density and transcriptional bursts.*, in preparation [Chapter 5]
4. **A.A. van den Berg**, and M. Depken, *Principles of histone replacement in the wake of a transcribing RNAP*, in preparation [Chapter 4]
3. **A.A. van den Berg** and M. Depken, *Crowding-induced transcriptional bursts dictate nucleosome and polymerase density profiles along genes*, [Nucleic Acids Res \(2017\) 45 \(13\): 7623-7632](#) [Chapter 3]
2. M.C. Moolman, S.T. Krishnan, J.W.J. Kerssemakers, **A.A. van den Berg**, P. Tulinski, M. Depken, R. Reyes-Lamothe, D.J. Sherratt and N.H. Dekker *Slow unloading leads to DNA-bound β_2 -sliding clamp accumulation in live Escherichia coli cells*, [Nature Communications \(2014\) 5, 5820](#) .
1. M.A.M. Versteegh, M.E. Reimer, **A.A. van den Berg**, G. Juska, V. Dimastrodonato, A. Gocalinska, E. Pelucchi and V. Zwiller, *Single pairs of time-bin entangled photons*, [Phys. Rev. A \(2015\) 92, 033802](#) .