



HAL
open science

Elucidating the mechanism of Mfd-mediated mutagenesis from a single-molecule perspective

James Portman

► **To cite this version:**

James Portman. Elucidating the mechanism of Mfd-mediated mutagenesis from a single-molecule perspective. Molecular biology. Université Paris sciences et lettres, 2021. English. NNT : 2021UP-SLE039 . tel-03941562

HAL Id: tel-03941562

<https://theses.hal.science/tel-03941562v1>

Submitted on 16 Jan 2023

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



THÈSE DE DOCTORAT
DE L'UNIVERSITÉ PSL

Préparée à l'École Normale Supérieure de Paris

**Élucider le mécanisme de la mutagenèse médiée par Mfd
dans une perspective molécule-unique**

Soutenue par

James PORTMAN

Le 28 Juin 2021

Ecole doctorale n° 474

**Frontiers in Innovation,
Research and Education**

Spécialité

**Biologie Moléculaire et
Structurale et Biochimie,
Biophysique Moléculaire**

Composition du jury:

Giuseppe, BALDACCI *Président*
Professeur émérite, Institut Jacques Monod

Benoit, PALANCADE *Rapporteur*
Directeur de Recherche au CNRS, Institut Jacques Monod

Bianca, SCLAVI *Rapporteur*
Directrice de Recherche au CNRS, Sorbonne Université

Aziz, SANCAR *Examineur*
Professeur, University of North Carolina

Terence, STRICK *Directeur de thèse*
Professeur, École Normale Supérieure de Paris



Elucidating the mechanism of Mfd-mediated mutagenesis from a
single-molecule perspective

Abstract

The Mfd protein mediates the repair of DNA lesions on the template strand of actively-transcribed genes. In contrast, Mfd has recently emerged as an important driver of evolution through stress-induced mutagenesis. Evidence indicates Mfd contributes to stress-induced mutagenesis through the generation of R-loops, nucleic acid hybrids formed when an RNA invades duplex DNA to pair with its template sequence. We report real-time observations of co-transcriptional R-loop formation at single-molecule resolution. We show Mfd interacts simultaneously with elongating RNA polymerase and upstream DNA, tethering the two together and partitioning the DNA into distinct supercoiled domains. A highly negatively supercoiled domain forms between Mfd and RNA polymerase, with compensatory positive supercoiling appearing in front of RNA polymerase and behind Mfd. The nascent RNA invades this negatively supercoiled domain to form a stable R-loop that will drive mutagenesis. This mechanism theoretically enables any protein that simultaneously binds transcribing RNA polymerase and upstream DNA to stimulate R-loop formation and mutagenesis.

Acknowledgements

I have many people to thank, who played not only a role in my research, but who also helped me to begin my journey as a scientist. I must first thank Professor Terence Strick, my supervisor and mentor over the last three years. It has been a pleasure to learn from and be inspired by such a rigorous, passionate, and endlessly curious scientist, who always encouraged me to investigate interesting tangents, one of which became the focus of this thesis.

I would like to thank Professor Nigel Savery, who fostered my interest in DNA repair and mutagenesis during my Bachelor's degree, and Dr Rachid Rahmouni and Dr Bianca Sclavi, who formed my thesis advisory committee, for their advice and support during my PhD. I would like also to thank Professor Giuseppe Baldacci, Professor Aziz Sancar, Dr Benoit Palancade, and Dr Bianca Sclavi for their participation in my thesis jury. I would like to acknowledge the important collaborators Professor Nigel Savery, Professor Seth Darst, Professor Richard Ebright, Professor Mark Dillingham, Professor David Jeruzalmi, Professor Katsuhiko Murakami, Dr Jack Bollins, and Dr Gwendolyn Brouwer, for engaging in fascinating discussions and providing key data.

My employment through the Marie Skłodowska-Curie "DNA Repairman" network and my enrolment in the "Frontieres du Vivant" doctoral school immeasurably enriched my PhD experience. Frequent "DNA Repairman" meetings and conferences created a positive, open, and collaborative framework within which to conduct research. The importance of multi-component international collaborations like these cannot be underestimated.

I would like to thank the formidable Dorota Kostrz, Marc Nadal, Charlie Gosse, Francois Stransky, Camille Duboc, and all members of the Molecular Motors laboratory, past and present, for teaching me, indulging my experiment propositions, and their stimulating and uplifting discussions.

To my dearest friends, Claire Noble, Huw Edwardes-Evans, Imogen Senter, Joshua Buxton, Anthony Lourdiane, Francesca Fillipini, Lukas Hafner, Jonathan Bastard, Jakub Voznica, Bernard Bollignan, Asher Silverman, Hattie Collins, Maria Troullides, and Antony Leonida, this journey would not have been possible without them.

Similarly, I would like to thank Professor Avan Sayer for her advice and mentorship over the past three years.

And to Makayla Williams, whose relentless positivity, kind heart, and fantastic humour made the last two years complete.

Finally, I must thank my parents, Christopher and Deborah, my siblings, Katherine, David, William, and Sarah, and my grandparents, Michael, Dalzie, Kenneth, and Jacqueline. I would not be where I am without their encouragement and support.

Overview

I began my PhD hoping to understand how the bacterial DNA repair protein, Mfd, could stimulate mutagenesis. Mfd was structurally and biochemically well understood, and its role in transcription-coupled DNA repair was mostly characterised. However, at the time, it had become clear that Mfd could also stimulate mutagenesis. It was thought that transcriptional R-loops played a role in this pathway, and whilst much is known about the effects of R-loops (e.g., mutations), the mechanistic basis for their formation was elusive. Furthermore, Mfd-dependent mutagenesis accounts for roughly half of all stress-induced mutagenesis. Whilst the experiments of Luria and Delbrück in the 1950s revealed bacterial evolution to be driven by random mutation, soon after it was shown that directed, non-random, stress-induced mutation also occurred. It is still not clear mechanistically how stress-induced mutations occur. Therefore, understanding how Mfd stimulates mutagenesis will contribute to our understanding of how stress-induced mutagenesis occurs generally, which shapes how we comprehend evolution. The striking observation that Mfd mediates mutagenesis opened up many new questions: How does Mfd stimulate mutagenesis? How are R-loops formed? How does stress make a DNA repair protein stimulate mutagenesis? Are stress-induced mutagenesis mechanisms universal?

At that point, mostly ensemble methods had been used to study Mfd functions. Generally, ensemble molecular biology can reveal binding partners, co-factor dependencies, minimal substrate requirements and reaction kinetics. Indeed, much of what is known about DNA repair pathways has been elucidated through the well-conceived and clever use of ensemble methods. Despite this, ensemble methods do have limitations. They cannot easily be synchronised and so the results represent an average of all reactions – heterogeneity is averaged out and rare events are missed. Furthermore, ensemble methods are limited to the measurement of accessible and relatively long-lived products. As a result, single-molecule methods, and in particular the magnetic trap, are well-poised to supplement ensemble measurements because they can access

molecular heterogeneity, fast events, rare events, transient intermediates, and less tangible products such as transient topological changes or motor protein motion. Recently the magnetic trap has been extended to the study of transcription-coupled repair, demonstrating that this technique can go beyond single-molecules and explore complex, multi-component systems. Overall, the magnetic trap, in combination with genetics, were well-positioned tools to understand how Mfd mediates mutagenesis.

Understanding Mfd-mediated mutagenesis first merits the exploration of the fields with which it is inextricably linked. For instance, transcribing RNAP mediates not only gene expression but also acts as a damage-sensor (in transcription-coupled DNA repair) and can create R-loops when bound by Mfd (in Mfd-mediated mutagenesis). Transcription and R-loop formation both alter DNA topology, which in turn, through a multitude of feedback loops, modulates transcription. As a result, after beginning with the single-molecule approach and DNA mechanics, I will introduce DNA topology, R-loops, transcription, DNA repair pathways (focusing on transcription-coupled DNA repair), and mutagenesis.

Contents

Section I: Introduction	19
Chapter I: The Single-Molecule Approach	21
Single-Molecule Theory	21
Types of Single-Molecule Approaches	21
Chapter II: DNA Structure.....	29
Primary DNA Structure.....	29
DNA as a Single-Molecule	30
DNA Higher-Order Structure.....	36
Topoisomerases.....	40
RNA-DNA Hybrids (R-loops)	41
Chapter III: Transcription.....	45
Transcription Mechanism.....	45
RNAP Accessory Factors	49
Transcription and DNA Topology	50
Transcription Under Stress	53
Chapter IV: DNA Damage and Repair	55
Types of DNA Damage.....	55
Types of DNA Repair	57
Global Genome Repair and Transcription-Coupled Repair	60
Genetics of Transcription-Coupled Repair	60
Transcription-Coupled Repair Mechanism I.....	64
Mfd Structure.....	65
Transcription-Coupled Repair Mechanism II.....	67
Eukaryotic Nucleotide Excision Repair	74
Chapter V: Mutagenesis	77
Early Mutagenesis Experiments.....	77
Stress-Induced Mutagenesis.....	78
Spontaneous Mutagenesis	81
Section II: Results	85
Chapter I: Co-transcriptional R-loop formation by Mfd involves topological partitioning of DNA.....	87

Introduction.....	89
Results	91
Mfd interacts with elongating RNAP to form a tripartite supercoiled domain	91
DNA compaction without topological coupling is observed on nicked DNA molecules	93
Translocase-deficient MfdRA953 maintains the ability to form tripartite supercoiled domains.....	93
Mfd LR499, deficient in binding to RNAP, lacks the ability to form tripartite supercoiled domains.....	94
Backtracked RNAP is not a major source of tripartite supercoiled domain formation.	95
Formation of tripartite supercoiled domains leads to R-loops	95
Both wild-type and translocation-deficient Mfd support R-loop dependent mutagenesis in vivo	96
Discussion	98
Materials and Methods.....	102
Supplementary Information	115
Appendix I: Topological partitioning observed with a transcription factor.....	130
Chapter II: On the stability of stalled RNA polymerase and its removal by RapA.....	133
Introduction.....	135
Results	137
High torque removes stalled RNAP from positively supercoiled DNA	137
RapA removes stalled RNAP in a torque-sensitive manner.....	138
RapA removes RNAP stalled on positively, but not negatively, supercoiled DNA	139
Kinetics of RNAP removal by RapA.....	140
RapA is functionally inhibited by GreB.....	140
RapA takes longer to remove RNAPs stalled further away from the promoter	141
Discussion	142
Materials and Methods.....	144
Supplementary Information	155
Appendix II: <i>In vivo</i> RapA experiments	164
Section III: Conclusions and Perspectives	171
Section IV: Supplementary Protocols	179
Chapter I: Glass Surface Preparation and Functionalisation Protocol.....	181
Chapter II: Lambda-Red Recombination Protocol.....	183
Section V: Publications.....	189
Section VI: Bibliography	193

Section I: Introduction

Chapter I: The Single-Molecule Approach

Single-Molecule Theory

Biochemical experiments performed in 'ensemble' (also referred to as 'bulk' reactions; which both in this instance may refer to any *in vitro* reaction performed on a large scale – i.e., micromolar concentrations of protein interacting with nanomolar concentrations of DNA) allow us to access many aspects of the phenomena in question. Despite this, ensemble methods do have limitations. Firstly, they cannot easily be synchronised and so the results represent an average of all reactions – heterogeneity is averaged out, and rare events and transient intermediates are difficult to observe. As a result, single-molecule methods are well-poised to supplement ensemble measurements because they can access molecular heterogeneity, capture rare events and even detect fast events or transient intermediates. A notable difference between ensemble and single-molecule experimentation is the approach – ensemble experiments should be well-designed in order to piece apart a given process. Whilst single-molecule experiments must of course also be designed thoughtfully, there is already a lot of understanding to be gained from simply watching molecules in action. Ideally one combines ensemble and single-molecule approaches in a complementary fashion.

Types of Single-Molecule Approaches

The first single-molecule technique to be developed was the patch-clamp. Used to study ion transporters in neuronal cells, this approach uses a glass pipette to isolate a minute area of a neuronal cell membrane, containing on average one transmembrane transport protein [1]. One

can then monitor transport protein opening and closing in real-time by recording the current which flows through the channel when a potential difference is applied across the membrane.

Similarly and more recently, nanopore-based approaches rely on the embedding of a nanopore in a membrane and recording the current across the membrane [2]. This technology has been applied to DNA sequencing – each base has a different current disruption signal associated with it; and to protein-DNA interactions – both monitoring dissociation kinetics of DNA-binding proteins and the register of motors, such as RNAP, along the DNA [3]. In these experiments, a potential difference is applied across the membrane, driving DNA through the nanopore.

To probe the kinetics of protein binding to DNA, the nanopore is made smaller than the protein, meaning the protein will block the nanopore, until the DNA has been threaded through and the protein dissociates (the longer the protein arrests ion flow, the stronger it binds to DNA). Similarly, to probe transcriptional regulation, the RNAP can be positioned at the entrance of the nanopore and its position on DNA can be tracked with tens of picometres of precision [3]. However the force experienced by the DNA as it is pulled through the constriction is 25-50 piconewtons ($1 \text{ pN} = 10^{-12} \text{ N}$), on the order of the forces needed to stall RNAP, which limits the conclusions that can be drawn from these experiments [4].

Another historical single-molecule instrument is the atomic force microscope (AFM). AFM uses a sharp probe at the end of a cantilever that detects sample topography and/or interactions through monitoring sub-nanometre deflections of the probe from its expected motion using a laser reflecting off of the cantilever [5]. The AFM can be used in a variety of modes, the most common are: constant force contact mode – the tip moves across the sample to create an image; non-contact mode (or nano-contact mode) – the tip oscillates just above the sample within the attractive force regime; and tapping mode – less destructive than contact mode. Both the non-contact mode and tapping mode are capable of detecting forces on the 10-pN scale. Furthermore, the tip can be functionalised with protein to test interactions with the sample. Whilst the AFM can give very high (sub-nanometre) resolution, it is semi-destructive and relatively low throughput.

In contrast to the AFM, optical tweezers, acoustic force spectroscopy, and the magnetic trap are able to probe molecules on the sub-piconewton force scale. Whilst bond formation and breakage is accessed on the nanonewton scale (i.e., the force required to displace electrons over nanometres), these more sensitive techniques can access the piconewton-scale forces required to probe molecular motors (which use the energy derived from ATP hydrolysis, $\sim 25k_B T$, to move nanometres) and femtonewton-scale Brownian fluctuations.

Optical tweezers utilise a near-infrared laser to trap and control the position of two dielectric particles (e.g., polystyrene beads) with a DNA tether between them. The polystyrene bead experiences an attractive force due to the transfer of momentum from incident photon refraction [6]. The bead is held in the centre of the highly-focused laser beam due to the spring-like behaviour of the trap – the greater the distance the bead is from the centre of the trap, the greater the pulling force on the bead. One can then use a feedback loop to maintain the position of the bead and/or to observe the displacement of the bead to infer changes imposed on the DNA (e.g., bead-tethered RNAP transcribing towards surface-tethered DNA acts to pull the bead out of its trap, see Figure 1.1 [7]).

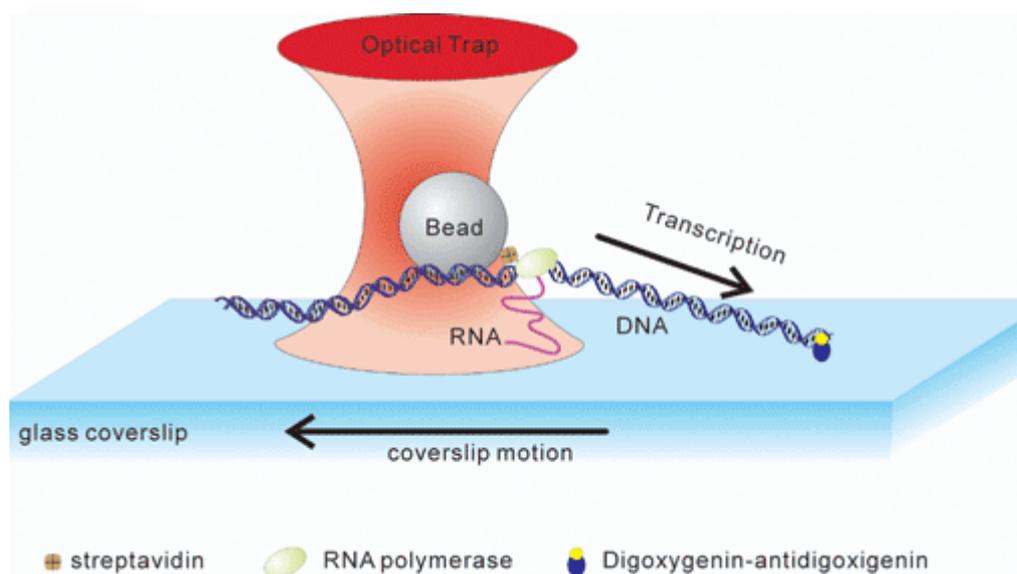


Figure 1.1: Bead-tethered RNAP transcribes along surface-tethered DNA [7].

The optical trap has been used extensively to study DNA-protein interactions as it has a wide range of applicable forces (hundreds of femtonewtons to hundreds of piconewtons), gives real-time output, and can be adapted for rapid solution changes (using multi-channel microfluidics one can move the DNA between channels in order to change the buffer conditions) [8]. However, the optical trap is a low-throughput technique that enables monitoring of only one molecule at a time.

Acoustic force spectroscopy (AFS) uses piezoelectric actuators to generate sound waves inside a microfluidic channel, creating energy potential wells that, similar to the optical trap, attract small particles such as polystyrene beads [9]. If the bead is surface-attached via a DNA molecule for example, the sound waves essentially exert a pulling force similar to the magnetic trap. The range of the AFS is from sub-fN to hundreds of pN and it can be used to probe many molecules in parallel. However, sound waves, comparatively to light waves, have a lower frequency and so there is poorer spatial control [10].

The magnetic trap is, like the AFS, a force-field approach used to monitor typically 50 single DNA molecules in unison. DNA molecules are multiply attached at one end to a glass coverslip and at the other end to a magnetic bead, ensuring a tethered DNA-bead system that is stable over days. A pair of magnets are held directly above the sample, and the sample is illuminated using an LED (Figure 1.2) [11]. The magnets generate a magnetic field, the vertical gradient of which exerts a vertical pulling force on the DNA-tethered beads that can range from the sub-fN to the tens of pN. Moving the magnets closer to the sample increases the extending force, while moving the magnets farther from the sample decreases the extending force. A camera focused through an objective lens in contact with the glass slide (from below, as the magnetic trap is an inverted microscope) monitors the diffraction pattern of the magnetic beads, which indicates their altitude (z position) above the surface.

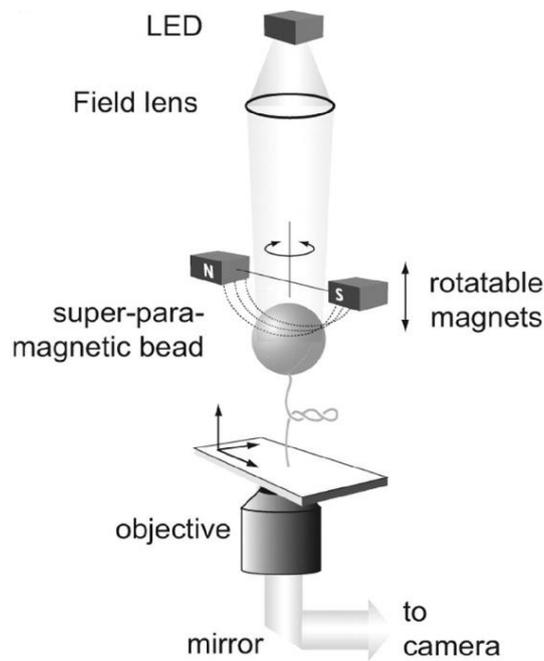


Figure 1.2: The magnetic trap set-up [11].

Independently of the applied force, the magnetic field orientation (horizontal at the sample) locks the magnetic bead orientation. Therefore, rotation of the magnets allows one to supercoil the DNA while the force remains constant. Beyond allowing us to study protein-DNA interactions under physiological levels of supercoiling, this accomplishes a number of things. Firstly, it causes the DNA extension to decrease as interwound plectonemic supercoils, or loops, appear along the DNA. These loops have a significant contour length given DNA stiffness, on the order of 50 nanometres [12]. Thus, secondly, at such a super-helicity (e.g., σ of $\sim\pm 0.03$), changes in DNA extension can be converted into changes in DNA topology, and so the DNA acts as its own signal detector and amplifier.

Consider the following example. At a super-helicity of zero, if a protein unwinds 10 base-pairs (bp) of DNA it creates an essentially undetectable change in DNA extension of a few nanometres. On the other hand, if the DNA has been supercoiled as mentioned then 10 bp of unwinding will titrate a supercoil, causing a 50 nm change in DNA extension as in this regime, the change in twist results in a compensatory change of plectonemic writhe. This makes the magnetic trap an extremely sensitive tool. More recently, the magnetic trap has been used to study protein-protein

and protein-drug interactions through the use of a junctured-DNA construct. Junctured-DNA facilitates repetitive interaction testing that can be probed at various buffer conditions, temperatures, and forces [13]. Taken together, this shows us that the magnetic trap is a high-throughput, robust, sensitive, and versatile tool to observe structural transitions in DNA, DNA-protein interactions, protein-protein interactions and protein-drug interactions.

A distinct group of methods use fluorescence to investigate the system in question at single-molecule resolution. Fluorescence can be used in its most simple form – focusing a laser of the correct wavelength in the sample to excite fluorescent molecules and observe their motions and interactions (the laser light is absorbed and light of a longer wavelength is emitted). Both DNA and proteins can be modified to contain fluorescence moieties that do not interfere with structure nor function. In fact, multiple separate lasers (typically up to three) can be used in order to image multiple different components at once. For example, in the ‘DNA curtains’ methodology DNA molecules are horizontally extended and attached at their ends to a coverslip surface. The molecules can be aligned such that they are all parallel to each other and exactly side-by-side, allowing for observation of a great many molecules simultaneously. Introduction of fluorescently labelled proteins into the experiment allows one to monitor their localization on the DNA [14]. Whilst this method does allow for the determination of target search mechanisms and classes of binding, it lacks functional insight and the DNA is linear so cannot supercoil. Similarly, colocalization of single-molecules spectroscopy (CoSMoS) uses this principle to simultaneously image DNA and multiple proteins in real-time [15]. Similar to the magnetic trap, CoSMoS tethers DNA to the surface of a flow cell which is imaged from below. Because one has such control over which proteins are tagged, whole cell lysates can be added to the flow cell and only the protein of interest is observed, meaning this method can access conformational heterogeneity normally lost through the protein purification process. CoSMoS can also utilise Förster Resonance Energy Transfer (FRET) to visualise two-component interactions (these two components can be protein and DNA, or two segments of the same DNA molecule or two domains in the same protein). FRET occurs when the emission spectra of one fluorophore overlaps with the absorbance spectra of a

second fluorophore, within a distance of 10 nm. The short distances over which FRET efficiently occurs allows FRET to be used as a molecular ruler, informing dynamically on the positions of the two fluorophores (Figure 1.3) [16].

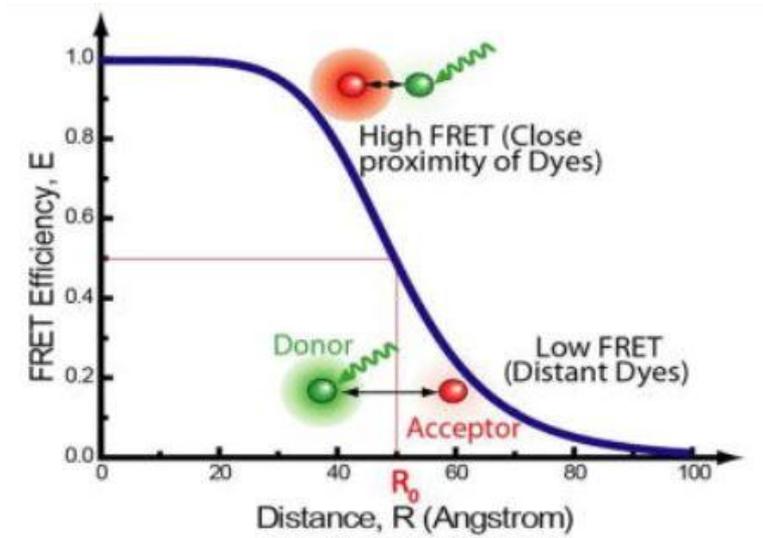


Figure 1.3: FRET efficiency as a function of distance between the two fluorophores [16].

The efficiency of energy transfer, E , can be defined as:

$$E = [1 + (\frac{R}{R_0})^6]^{-1}$$

Where R is the distance between the fluorophores and R_0 the Förster radius; when $R = R_0$ then $E = 0.5$. When the two fluorophores are close enough, as the laser excites the first fluorophore, this energy is passed on to the second fluorophore (via non-radiative energy transfer) which emits light of a separate distinctive wavelength. By observing the change of emitted light, one can infer interactions between the two fluorophores (Figure 1.4) [17].

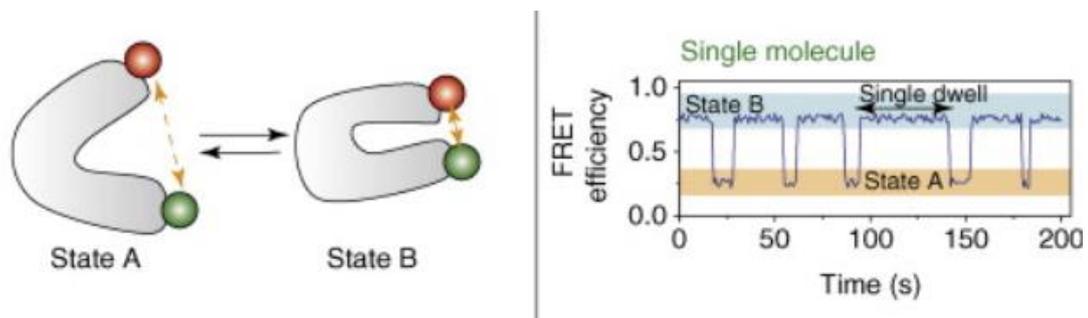


Figure 1.4: FRET efficiently between two fluorophores occurs when they are in close proximity (left panel; state B) as observed on the time-trace (right panel) [17].

To detect the weak signals emitted by single fluorescent molecules it is necessary to reduce background fluorescence. The single-molecule fluorescence approaches I have discussed tend to use total internal reflection fluorescence (TIRF) microscopy (TIRFM) to accomplish this. TIRFM selectively illuminates only a thin 'layer' of the sample closest to the objective lens. When the incident light is totally-internally reflected at the glass-oil or glass-water interface, an evanescent field is generated in the liquid which decays exponentially. Therefore, only fluorophores within ~100 nm of the glass surface are illuminated, increasing the signal-to-noise ratio. DNA curtains, CoSMoS and Nano-CoSM (nanomanipulation and colocalization of single-molecules) use TIRFM.

Nano-CoSM, a hybrid approach, is the combination of magnetic tweezers and TIRFM that enables the simultaneous monitoring of DNA extension and protein localisation and protein-protein/protein-DNA interaction [18]. Here, one can simultaneously monitor fluorescence and changes to DNA extension in order to probe complex composition, stoichiometry, and reaction assembly and disassembly.

Photo-activated localisation microscopy (PALM) uses photoactivatable-fluorescent proteins and can be combined with TIRFM. Here, a laser rapidly excites and photo-bleaches a few fluorophores at a time (within the evanescent field) which is easier to map due to the great signal-to-noise ratio. An image is generated over successive rounds and has been used notably for the determination of focal adhesion organisation in live cells [19].

Chapter II: DNA Structure

Primary DNA Structure

DNA is a right-handed anti-parallel double helix [20]. One DNA strand comprises a sugar-phosphate backbone with a nitrogenous base bound to a deoxyribose sugar. The nitrogenous base can be adenine, thymine, cytosine, or guanine. Two anti-parallel DNA strands come together via hydrogen bonding between the bases to form the double-helix (Figure 1.5) [20]. The ordering of A, T, C and G is the basis for information storage in cells. The 'central dogma' outlines the cellular information flow – DNA stores the information, and is used as a template to make an RNA copy, which is itself used as a template to make protein. DNA is transcribed into RNA by RNA polymerase and RNA has three fundamental differences in comparison to DNA. It is single-stranded, it uses uracil instead of thymine, and its constituent sugar is ribose, not deoxyribose. RNA is then translated into protein by the ribosome. This is termed translation because we shift from the language of nucleic acids to the language of amino acids. Amino acids are polymerised by the ribosome to make proteins.

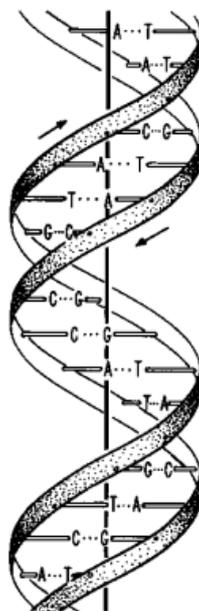


Figure 1.5: The double helix [20].

The double helix can exist in a variety of forms. It is typically in the right-handed B-form, but can also adopt a more compacted right-handed A-form, and even left-handed Z-forms. The A-form bears a C-3'-endo conformation, in which the carbon-3 of the sugar lies out of the plane, leading to slight tilting of each base-pair. As a result, the DNA in this conformation binds fewer water molecules. Z-form DNA is left-handed and the phosphate backbone appears to zig-zag. This conformation only occurs in specific scenarios (such as short sequences of alternating pyrimidines and purines; and high salt). *In vivo* most DNA is B-form [21].

DNA as a Single-Molecule

Before exploring higher-order DNA organisation, it is important to consider DNA as a physical polymer and interrogate its mechanical properties. The magnetic trap allows us to do this. The magnetic trap involves torsionally constraining DNA between a glass coverslip and magnetic bead. The DNA is extended using a pair of magnets above the sample and imaged using a camera, with the diffraction pattern of the magnetic bead informing us of its altitude (z position) [11]. Monitoring the altitude of the bead informs us on the conformational state of the DNA, for example DNA bending by a protein will reduce the end-to-end extension of the molecule and lower the magnetic bead z position. Firstly, we can observe that at low force, when the magnets are far away from the sample, the bead will fluctuate in the x and y directions (i.e., in the imaging plane) a lot more than at high force, when the magnets are closer to the sample. Using the following equation, we can estimate the extending force (F).

$$F = \frac{k_B T l}{\langle \delta x^2 \rangle}$$

where l = DNA extension, k_B = Boltzmann constant, T = temperature (in Kelvin), and $\langle \delta x^2 \rangle$ = mean-squared fluctuation in the x direction (alternatively and equivalently one can follow the fluctuations in the y direction) [12]. Thus by characterizing the position of the bead above the

surface (to determine l) and the amplitude of the bead's lateral fluctuations (to determine $\langle \delta x^2 \rangle$) one can obtain the extending force and thence plot the DNA extension as a function of extending force (Figure 1.6) [22].

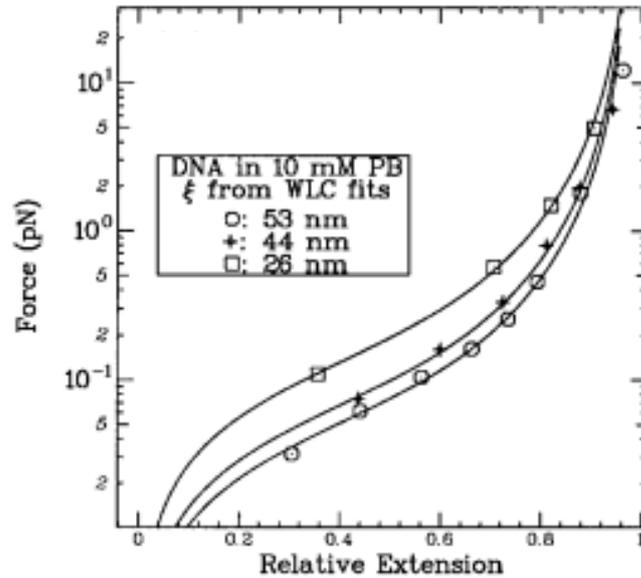


Figure 1.6: The force-extension relationship of a single non-torsionally-constrained DNA molecule (open circles) [22].

The circular points show the force-extension relationship for a single (torsionally-relaxed) DNA molecule which aligns very well to the estimation from the worm-like chain model with a persistence length of ~ 53 nm [23]. This corresponds to ~ 156 bp, meaning DNA fragments shorter than ~ 156 bp act as rigid rods. We can therefore consider our DNA molecules as a continuous chain of segments which can bend on the 50 nm length scale, each of which is undergoing Brownian motion and exploring all possible orientations due to thermal fluctuations. As the force is increased, the range of possible orientations decreases – as a result, there is a force due to entropy that acts against the extending force.

DNA topology is described by linking number (Lk). The linking number is the number of times the two DNA strands are intertwined. Lk is an integer and is constant for a closed system in which the DNA is not broken and re-sealed [24, 25]. The overall topology of a DNA molecule can be defined as:

$$Lk = Tw + Wr$$

Twist (T_w) is the number of times the DNA strands wind around each other and writhe (W_r) is the number of times the double-helix crosses over itself (a measure of 'higher-dimensional' twist). Lk_0 is the linking number of a fully relaxed DNA molecule and is defined as:

$$Lk_0 = \frac{bp}{10.5}$$

Therefore, in a torsionally-relaxed system (i.e., $W_r = 0$), $Lk = Lk_0 = T_w$. Rotating the magnets above the sample in the magnetic trap is akin to sequentially breaking, rotating, and re-sealing one end of the DNA, allowing the user to alter the linking number of the DNA. This is possible to do because the DNA is held at multiple points, on both strands, to the magnetic bead and to the surface via multiple biotin-streptavidin and digoxigenin-anti-digoxigenin linkages. If the DNA was held to the surface via a single biotin-streptavidin interaction, rotating the magnets would result in the DNA simply rotating around that single anchor point. Depending on the force and ionic conditions, changes in linking number will be mainly manifested as changes to twist or writhe (e.g., rotating the magnets at low forces typically means the change in linking number is satisfied by a change in writhe, whereas at higher forces the change in linking number is satisfied by a change in twist). We can quantify the difference in Lk between a supercoiled torsionally-constrained DNA molecule and the torsionally-relaxed molecule as:

$$\Delta Lk = Lk - Lk_0$$

The ratio between ΔLk and Lk_0 (which defines super-helical density, σ) allows for comparison between DNA molecules of different sizes:

$$\sigma = \Delta Lk / Lk_0$$

Therefore, if a DNA molecule in the magnetic trap has not been subject to rotation by the magnets, $Lk = Lk_0$ meaning $\Delta Lk = 0$ and super-helical density is 0. Writhe is referred to as plectonemic supercoiling (or simply, supercoiling) or toroidal wrapping. If DNA is wrapped around a protein structure, this is toroidal wrapping which can be left or right-handed. Toroidal wrapping can convert into plectonemes, which occur absent protein. Left-handed toroids convert to right-

handed plectonemes and right-handed toroids convert to left-handed plectonemes (see Figure 1.7) [26].

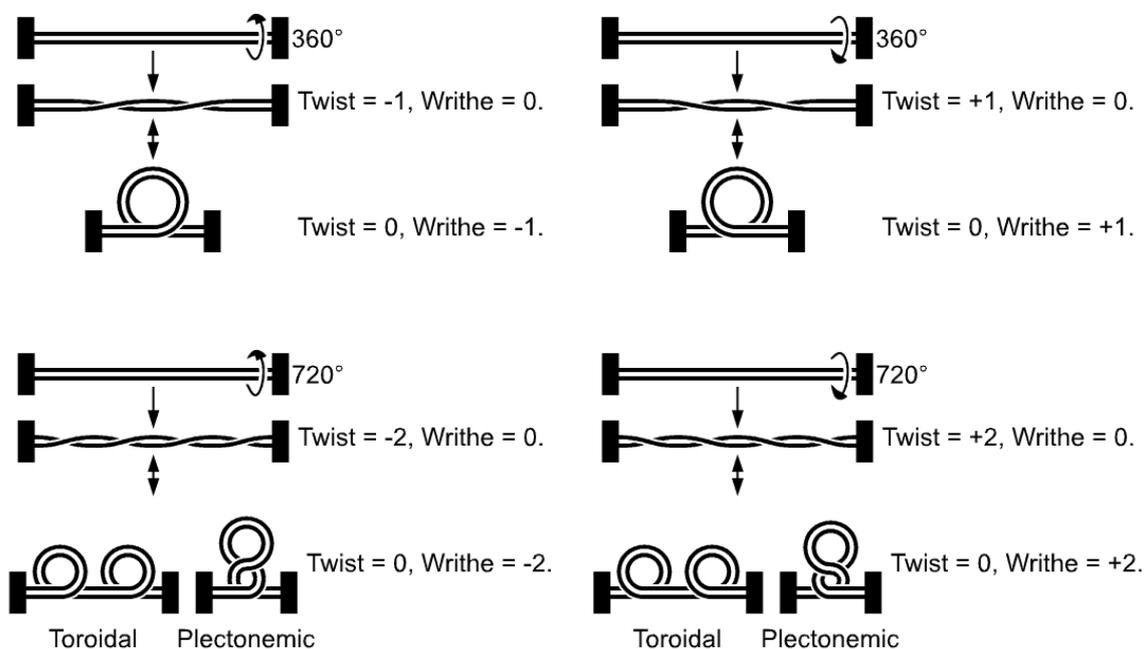


Figure 1.7: How twist and writhe inter-convert, and how toroidal and plectonemic supercoiling inter-convert [26].

The force-extension relationship for supercoiled DNA molecules was also determined (Figure 1.8) [22]. The upper and lower panels show the force-extension relationships for DNA of various levels of negative and positive supercoiling, respectively [22]. These show us that at low force (up to $\sim 0.5\text{pN}$) both positively and negatively supercoiled DNA contain plectonemic supercoils and have similar extensions (Figure 1.8, sketches a and c, respectively). At $\sim 0.5\text{pN}$, a transition occurs wherein negatively supercoiled DNA (of all values of σ) begins to act like a non torsionally-constrained molecule (see upper panel; all curves converge to the same line at $F > 0.5\text{pN}$). Presumably here all plectonemic supercoils are lost and the twist is altered to balance linking number (Figure 1.8, sketch b). A similar transition occurs for positively supercoiled DNA molecules at $\sim 3\text{pN}$ (Figure 1.8, sketch d). In the presence of 150mM NaCl , the force required to reach these transitions (i.e., 0.5pN for negatively supercoiled DNA and 3pN for positively

supercoiled DNA; both measured in the presence of 10mM phosphate buffer) rises to $\sim 3\text{pN}$ and $\sim 7\text{pN}$, respectively.

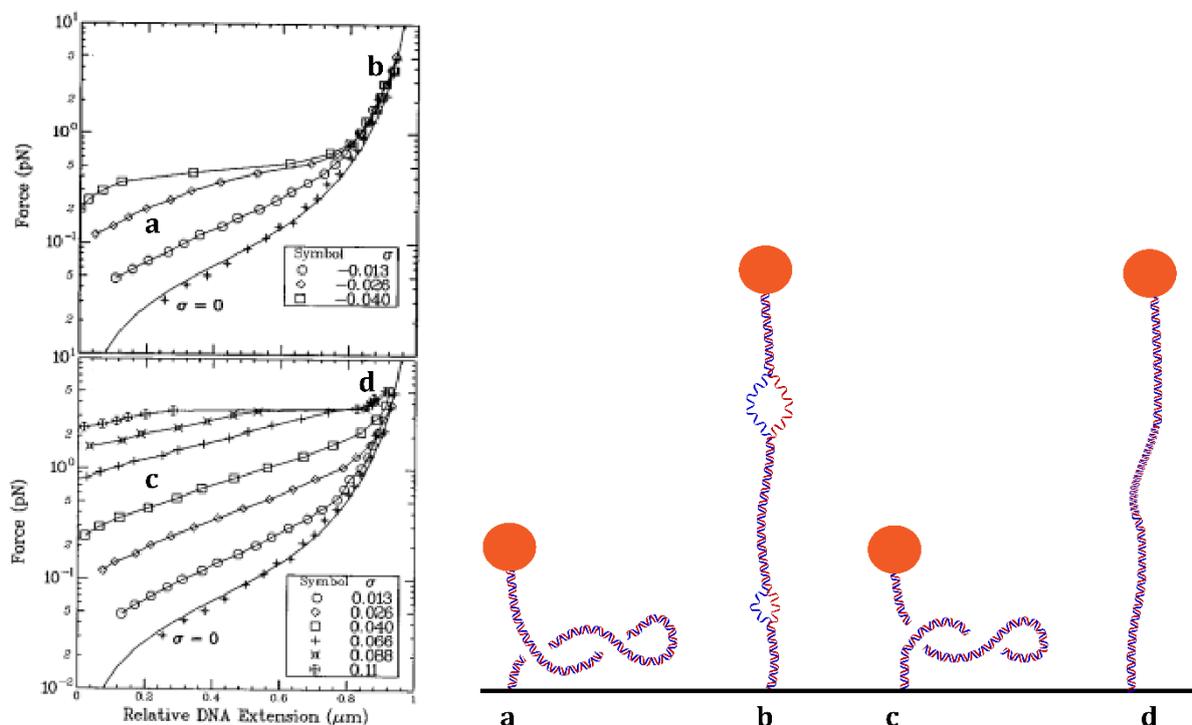


Figure 1.8: The force-extension relationships of many DNA molecules torsionally-constrained to differing extents [22].

In a contrasting experiment, the force was kept constant (either 0.2, 1.0 or 8.0pN, see Figure 1.9) and σ was changed incrementally, from positive supercoiling to negative supercoiling [22]. We observe that under low force ($\sim 0.2\text{pN}$) the curve is symmetric – for each complete rotation of the magnets, writhe (in the form of plectonemic supercoils) accumulates constantly which acts to reduce the end-to-end extension of the DNA (Figure 1.9, sketches b and c). At the slightly higher force of $\sim 1\text{pN}$, the curve is asymmetric – positive writhe continues to accumulate as it did at low force, whereas negative writhe doesn't form and the DNA extension is unaffected by magnet rotation. As before, presumably here the negative super-helical density results in unwinding – eventually the DNA may locally denature (fully unwind) regions in order to preserve as much DNA in B-form as possible. At higher forces ($\sim 8\text{pN}$) the DNA is essentially unresponsive to changes in super-helical density (Figure 1.9, sketches d and e) [22].

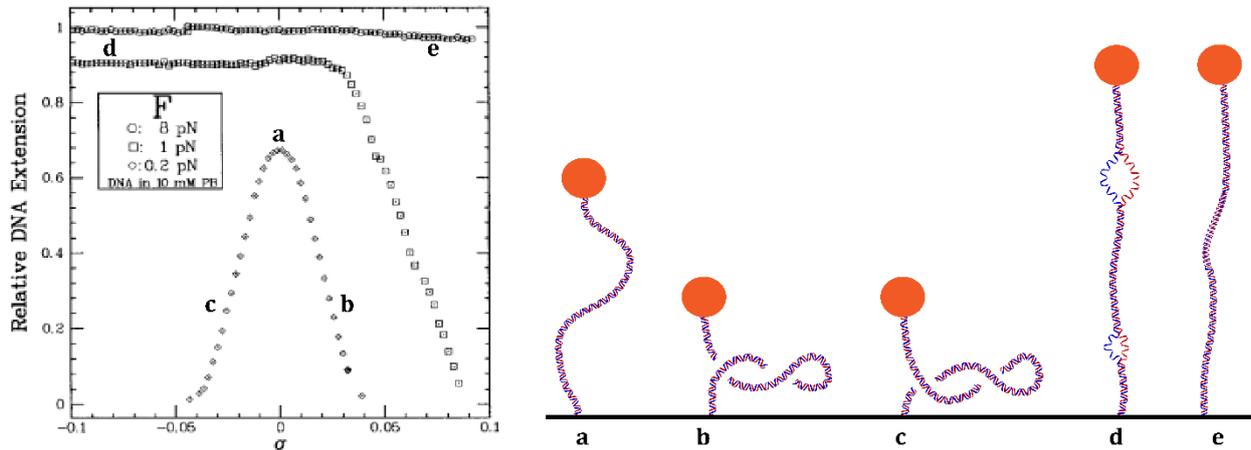


Figure 1.9: The rotation-extension curves of DNA molecules under differing constant forces [22].

In similar experiments across three low force points, the torque on the DNA molecule was monitored [27]. As Figure 1.10 shows, torque steadily increases as the DNA is rotated before a 'buckling transition' is reached. This is the point at which the DNA buckles and forms a first plectonemic supercoil. After that point, plectonemic supercoils grow at a constant rate and torque remains roughly stable [27]. We can see that the buckling transition is force-dependent. At 0.2pN the buckling transition occurs at ~ 5 turns, whereas at 0.9pN the buckling transition occurs at ~ 9 turns. In other words, the torque required to reach the buckling transition increases and the force increases. Note that these experiments were performed at 100mM NaCl, explaining the symmetric curve at 0.9pN. In this study the authors also probed the effect of salt concentration on the shape of the rotation-extension curve. At low salt concentrations of 20mM NaCl, DNA was less stable and would melt when underwound, leading to an asymmetric rotation-extension curve. Indeed, salt concentration greatly affects DNA properties due to its high negative charge. The number, density, valence and size of screening positive cations that are present act to modulate the melting propensity, as well as the bending stiffness, which affects plectoneme size [27-29].

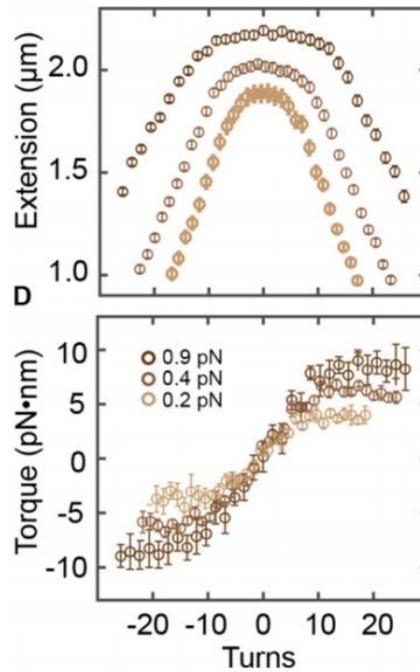


Figure 1.10: How torque changes during the rotation-extension curve at differing forces [27].

The magnetic trap has helped us to investigate the rigidity, force-extension relationship, force-supercoiling relationship, and structural transitions of DNA. This forms the platform from which we can now probe and understand DNA higher-order structure and DNA-protein interactions.

DNA Higher-Order Structure

On the genome scale, DNA is structured into domains that are topologically isolated. These topologically associated domains (TADs) appear to be evolutionarily conserved genome features [30]. There are many approaches to study genome structure (i.e., TAD organisation), but very limited ability to measure the topological state of TADs. Topology is challenging to measure genome-wide because it is very dynamic. Historically, topology was measured by studying the incorporation of psoralen, which intercalates preferentially between base-pairs of negatively supercoiled DNA [31]. However, psoralen intercalation is modulated by DNA sequence features and DNA-binding proteins, and is just now being used to probe the topology of TADs [32]. This

approach revealed the bacterial genome to maintain a σ of ~ -0.05 , yet this value is an average across the whole genome, comprising TADs that can have vastly different individual topologies [31].

Genome-wide structure and organisation can generally be studied from three perspectives. Imaging approaches, such as electron microscopy and fluorescence *in situ* hybridisation, allow us to broadly visualise genome organisation and the positions of specific loci. Biochemical approaches, such as chromatin immunoprecipitation (ChIP), chromatin isolation by RNA purification (ChIRP), and chromosome conformation capture (3C, and its derivatives), allow us to quantify genome regions that interact with protein, RNA, and each other, respectively. Genome editing approaches involve perturbing specific genome features and studying the outcome using the above techniques. 3C and its derivative techniques are perhaps the most commonly used method to probe genome structure and organisation on the Mbp scale, which is typically the size of TADs. Briefly, 3C involves crosslinking cells followed by restriction enzyme digestion, generating a series of chromatin fragments. DNA ligation follows, which preferentially ligates physically close DNA fragments held together by protein, and then these ligated DNA fragments are sequenced [33]. Here, 3C identifies DNA sequences that are separated sequentially but close physically due to protein interactions. For example, 3C has identified promoter-enhancer partners that are in close physical contact due to simultaneous protein binding and DNA looping [34].

The radius of gyration (R_G) of a given DNA molecule represents the average distance between the ends of the DNA and so provides an estimation of the physical space occupied by the DNA absent external forces. R_G can be estimated as $R_G \sim \sqrt{N} \cdot \xi$ where N is the number of segments (total DNA length/persistence length) and ξ is the persistence length. The R_G for human DNA is $\sim 340\mu\text{m}$ and for bacterial DNA is $\sim 7.5\mu\text{m}$. Therefore, given the size of a typical cell (a human neutrophil is $\sim 300\mu\text{m}^3$ [35], with the nucleus occupying about 10% of this volume, and a typical bacterial cell is $\sim 0.6\mu\text{m}^3$ [36], it is not surprising that significant DNA compaction must take place in order to

fit all the DNA in the space provided. This packaging is dynamic, to allow access to different parts of the genome dependent on the needs of the cell. We can describe DNA structure therefore on various scales of magnitude. In eukaryotes, DNA is wrapped around histones to form nucleosomes – roughly 150 bp (two negative supercoils) are spooled onto one histone octamer [37]. These nucleosomes can condense further into fibres and ultimately into a chromosome (Figure 1.11) [38].

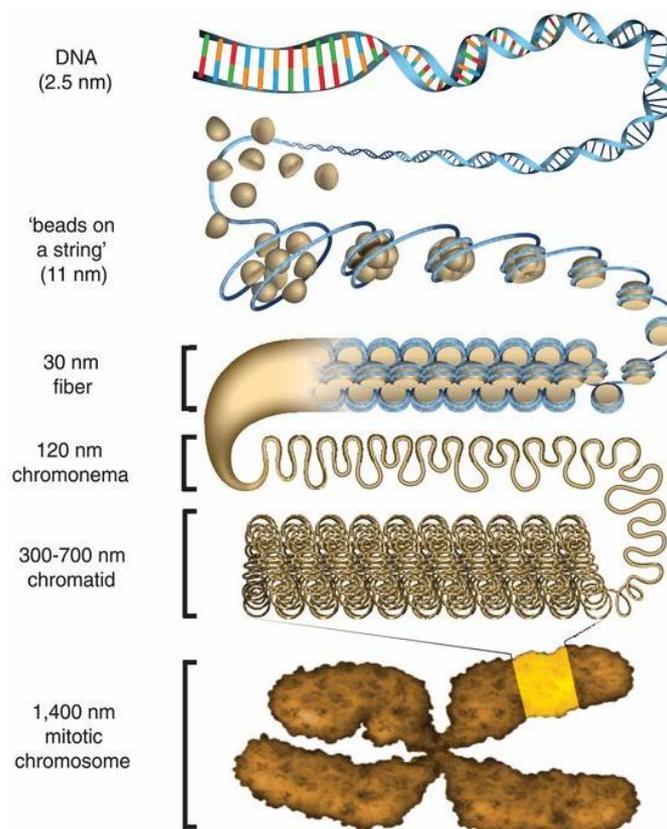


Figure 1.11: DNA structures across increasing scales of magnitude [38].

Eukaryotic chromatin structure may be more 'loose' in some areas (euchromatin), in comparison to more tightly condensed regions (heterochromatin) [39]. Euchromatin retains the capacity for transcriptional activity, and these two types of chromatin can interconvert depending on chemical modifications to histones [40]. It has been shown that TADs can act to section off heterochromatin from euchromatin [30]. In fact, this study showed TADs cover ~90% of the genome of mouse embryonic stem cells and that their organisation is mostly conserved

evolutionarily and across cell types [30]. TADs form in part due to loop extrusion by cohesin (a member of the structural maintenance of chromosomes (SMC) complex family; SMC1₂SMC3₂) and binding of architectural proteins [41, 42]. Once defined, TADs are separated and maintained by relatively short linkers referred to as topological boundary regions. Topological boundary regions show enrichment for binding of architectural proteins such as CCCTC-binding factor (CTCF). As CTCF binds DNA in a sequence-specific manner (although this is modulated by CpG methylation), this shows us that, at least in part, the genome controls its own organisation. In addition to CTCF, transcription start sites, transcription factors (TFIIIC is a well-studied example from *Saccharomyces cerevisiae*), global run-on sequencing data (a measure of the density and orientation of transcribing RNAP [43]), house-keeping genes, and tRNA genes were all enriched close to topological boundary regions [30, 44]. Conversely, H3K9me3, a heterochromatin marker, was depleted at topological boundary domains – two strong lines of evidence that transcription contributes to the maintenance of TAD boundaries. Lastly, the activation of short interspersed nuclear elements (SINEs; a class of retrotransposon) has been shown to expand and disperse CTCF binding sites which reprograms TAD organisation and influences evolution [45].

Prokaryotes don't have histones to organise around but do have an array of architectural proteins and a hierarchical structural organisation. Similar to the eukaryotic TAD, bacteria bear Mbp scale macrodomains [46]. *E. coli*, for example, has four macrodomains – ori, ter, left, right – which probably contribute to efficient DNA replication. Within these macrodomains are chromosomal interaction domains (CIDs) on the scale of hundreds of kbp [47]. Again, the organisation of CIDs seems to be set-up and maintained by highly-transcribed genes at boundaries. An estimated 400 'supercoiled domains', with an average length of ~10kb, are found within CIDs [48]. At this scale domain organisation is spatiotemporally dynamic and depends on stress levels, however supercoiled domains associated with the most highly-transcribed genes are essentially unaffected [47]. One could argue that these supercoiled domains essentially each represent a housekeeping gene (~300 genes were labelled essential in the Keio collection) whose domain supercoiling status and boundaries are maintained by high levels of transcription (which sets up

supercoiling boundaries) and RNAP phase separation [49]. Architectural proteins such as H-NS (forms DNA bridges), HU, IHF and Fis (generally compact DNA by introducing sharp bends) help prokaryotes to compact and organise DNA [50-52].

Topoisomerases

There exists a dynamic interplay between genome topology, gene expression and vulnerability to damage. For example, the more negatively supercoiled the genome is, the more accessible certain promoters become and so gene expression increases, but also increases the chances of R-loop formation and the incidence of single-stranded DNA regions, both of which can lead to DNA damage [53, 54]. All DNA processes have some form of dependence on DNA topology – replication requires negative supercoiling, and transcription involves unwinding and thus forms positive supercoiling. As a result, there are proteins whose role is to alter DNA topology, to convert DNA from one topological form to another. Topoisomerases react to maintain the supercoiling status of the genome within certain bounds – gyrase increases negative supercoiling whilst topol relaxes negative supercoiling (the promoters of the gyrase and topol genes are sensitive to negative supercoiling and this sets up a negative feedback loop – for example high negative supercoiling activates the topol promoter and topol expression results in the relaxation of negative supercoiling, which in turn activates the gyrase promoter, and so on) [55, 56]. There are many types of topoisomerases that have different roles – generally they are split into type I and type II topoisomerases (the former introduces transient single-strand breaks whilst the latter introduces transient double-strand breaks) [57]. The types of *E. coli* topoisomerases and their functions are summarised in Table 1 (adapted from [58]). Topoisomerases act not only to combat supercoiling changes, but also deal with knots, catenanes (intertwined rings of DNA, normally formed during replication of a circular genome) and Holliday junctions (four-armed DNA structures normally formed during DSB repair).

Enzyme name	Type	Gene(s)	Function(s)
Topoisomerase I	I	<i>topA</i>	Removes negative supercoils and decatenation
DNA gyrase	II	<i>gyrA</i> <i>gyrB</i>	Introduces negative supercoils Removes positive and negative supercoils and decatenation
Topoisomerase III	I	<i>topB</i>	Removes negative supercoils and decatenation
Topoisomerase IV	II	<i>parC</i> <i>parE</i>	Decatenation Removes positive and negative supercoils

Table 1: *E. coli* topoisomerases, their genes, and their functions.

RNA-DNA Hybrids (R-loops)

R-loops are implicated in DNA damage, mutagenesis and transcription and replication dysfunction, yet the mechanism of their formation remains elusive. An R-loop forms when RNA invades the double helix to base-pair to its complementary DNA sequence, leaving the other DNA strand displaced (see Figure 1.12) [59]. R-loops are typically associated with transcription, forming *in cis*, but have been known to form *in trans* [60].

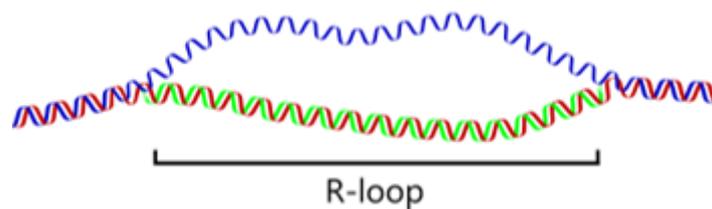


Figure 1.12: The R-loop (RNA in green, DNA strands in red and blue).

Various factors can increase the probability of R-loop formation. These all essentially favour local melting of the DNA which favours RNA invasion and hybridisation. Table 2 attempts to summarise these factors. As we can see, R-loop numbers and longevity can be finely-tuned, for example knock-down of an RNA/DNA helicase will favour R-loops, whilst over-expressing RNase H will dramatically decrease R-loops. Such approaches are available to establish a genetic background *in vivo* with the desired level of R-loops.

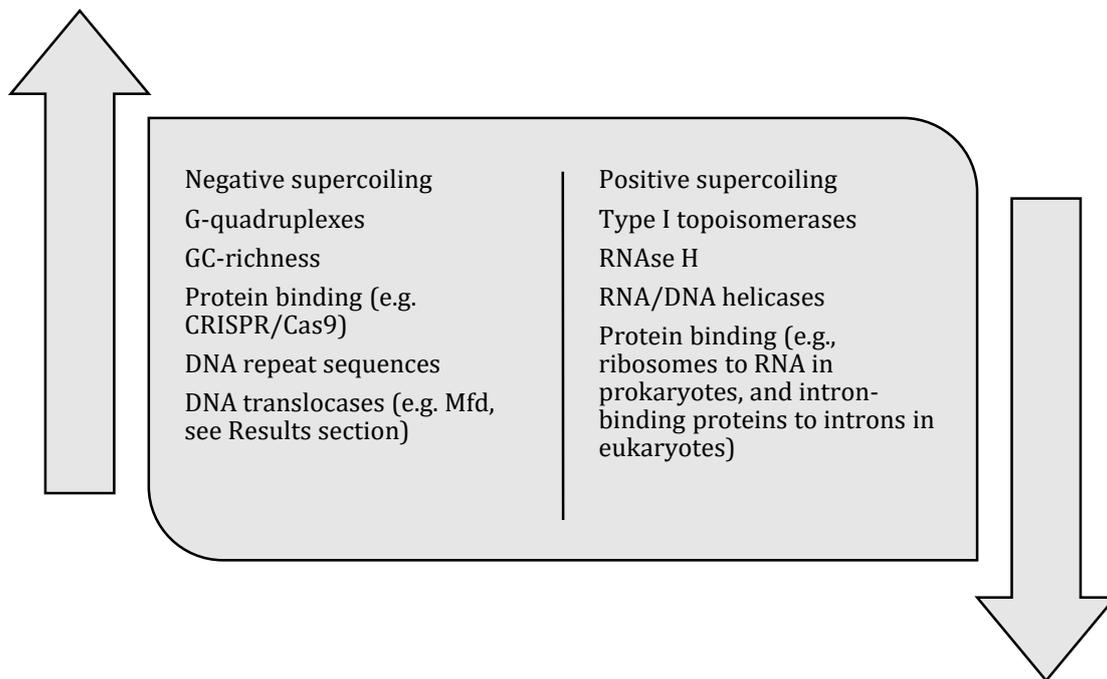


Table 2: Factors that act to increase (left panel) or decrease (right panel) either the number or longevity of an R-loop.

R-loops play a role in replication initiation (an R-loop acts as replication origin in some plasmids), transcription initiation (R-loops can either promote transcription by preventing epigenetic modifications to promoters, or inhibit transcription by preventing transcription factor binding), transcription termination (R-loops can stall RNAP and act as nucleation sites for transcription termination proteins), stress-induced mutagenesis (see Results section), topology management (see below), bacterial immunity (crRNA forms an R-loop to direct CRISPR/Cas9 machinery), and human immunity (class-switch recombination in immune cells) [61-65]. Indeed, R-loop enrichment at the 5' and 3' ends of genes has been observed, but their origin (for some genes lncRNA form an R-loop *in trans* to regulate transcription, whilst for others transcription itself forms the R-loop) and how they are regulated (again it has been observed that intrinsic DNA features, like GC-content, CpG islands, and repeats, as well as extrinsic cues, like nutritional stress, influence R-loop formation) remains poorly understood [66, 67].

The mutagenic aspects of R-loops will be discussed in detail later in the Introduction. Briefly, R-loops lead to genetic instability via mutations, recombination, and rearrangement. The displaced

ssDNA is highly prone to damage (deamination is 140-times more efficient on ssDNA compared to dsDNA) [68]. The R-loop can initiate origin-independent replication which is likely to be mutagenic or, if the replication machinery encounters a ssDNA break, lead to DSBs [63, 69]. Furthermore, R-loops can interfere with transcription and replication machineries, leading to transcription-replication conflicts which can lead indirectly to DSBs [70]. Interestingly, it was recently shown that the R-loop itself is a minor barrier to the replisome, it is the proteins that typically associate with R-loops that present a major obstacle [71].

Lastly, we can consider a role of R-loops in topological management. In prokaryotes and eukaryotes, R-loops essentially mitigate negative supercoiling. Once formed, the displaced ssDNA in an R-loop can become under-twisted (and potentially also form a left-handed helical wrap around the RNA-DNA hybrid). Thus an R-loop of size ~ 10.5 bp 'absorbs' one negative supercoil [72]. In fact, once nucleated the R-loop will extend as far as possible until all negative supercoiling has been absorbed. This phenomenon has been studied both *in vivo* and from a single-molecule perspective. Firstly, plasmid DNA isolated from *E. coli* topA mutants was found to be extremely negatively supercoiled and contain R-loops [73]. Without topoisomerase I to relax negative supercoils, transcription led to very negative super-helicity and many R-loops formed as a result. Clearly topoisomerase I is the main method by which cells deal with negative supercoils, as the topA mutant cells showed growth defects (presumably they had high levels of R-loop mediated mutagenesis). It was later shown that over-expression of RNase H partially complemented this defect, balancing R-loop levels and rescuing a relatively normal phenotype [74].

Under the magnetic trap, oligonucleotide hybridisation and negative supercoil relaxation can be observed through a change in molecular extension and a change in the profile of a rotation-extension curve (ssDNA was used here, however this experiment has been repeated recently with ssRNA and showed the same result (see Results section)) [75]. In order to allow hybridisation, DNA molecules were negatively supercoiled at low force and then a 2pN force was applied (above the transition force) which causes the molecule to be under-twisted. In fact, this is manifested as

a denaturation bubble in an AT-rich portion of the DNA, allowing the rest of the molecule to remain B-form. Then, a ssDNA oligonucleotide, homologous to the denatured region, was added. Subsequent force-extension and rotation-extension (shown in Figure 1.13) revealed the hybrid had formed and absorbed negative supercoils [75].

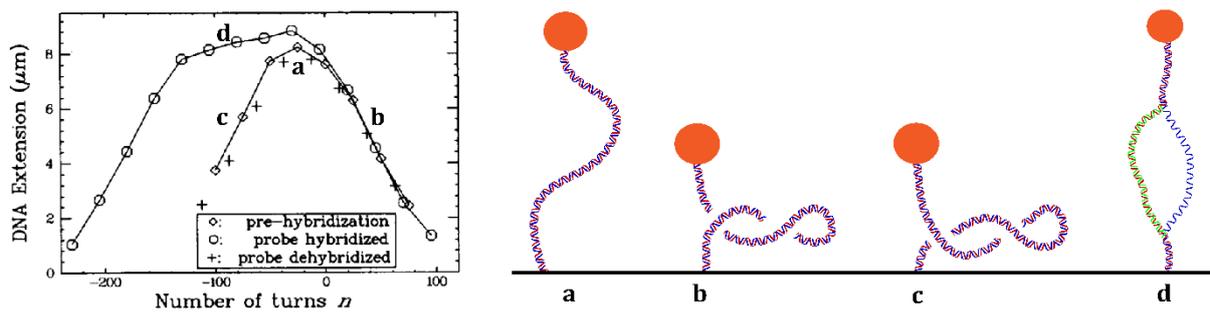


Figure 1.13: The rotation-extension curves before and after ssDNA probe hybridisation [75].

Before hybridisation the curve is symmetric, however once the oligonucleotide had hybridised, we see the left-hand portion of the curve has broadened by roughly 100 turns, indicating a roughly 1kb long hybrid had formed. During the acquisition at positive turn states, the hybrid is displaced as the DNA becomes re-wound and positively supercoiled, hence the right-hand portion of the curve is unaffected [75].

Chapter III: Transcription

Transcription Mechanism

So far, we have considered RNAP as a damage-sensor but in this section I will consider RNAP in terms of transcription and its relationship with DNA topology and regulatory factors. Generally, transcription begins with a promoter search followed by a complex sequence of promoter opening, promoter unwinding, scrunching and productive initiation (see Figure 1.14) [76]. The kinetics of these events will vary based on the given promoter – different promoters have different requirements in terms of sigma factors and initiation factors [77].

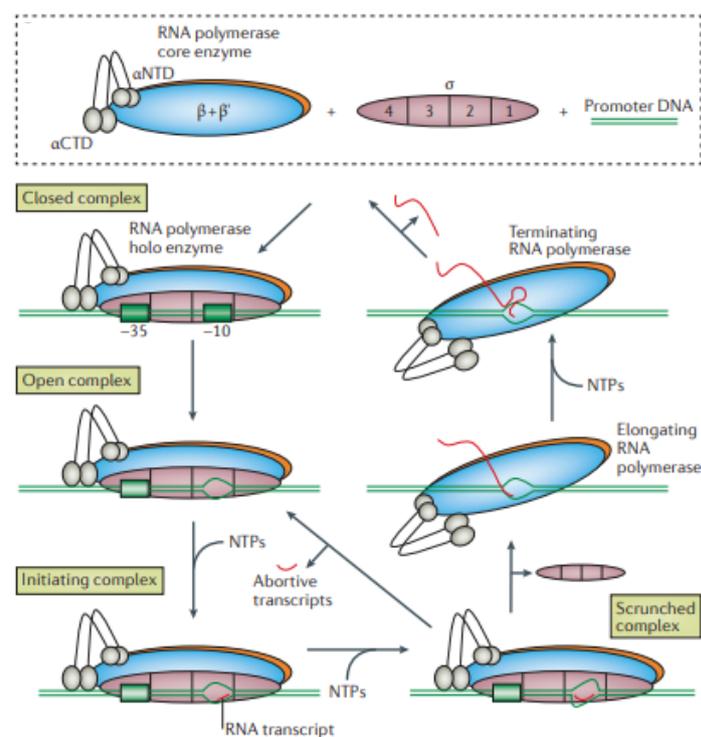


Figure 1.14: Initiation steps of transcription [76].

RNAP core is an $\alpha_2\beta\beta'\omega$ complex whereas the RNAP holoenzyme is an $\alpha_2\beta\beta'\omega\sigma$ complex. The sigma subunit is important for promoter recognition – different classes of promoters are recognised by different sigma variants (e.g., σ^{70} variants recognise promoters that typically drive

transcription of ‘housekeeping’ genes whilst σ^{32} variants recognise promoters that drive transcription of heat-shock genes) [78, 79]. Once RNAP has successfully initiated transcription we can refer to it as a transcription elongation complex (TEC). This acronym describes the RNAP-RNA-DNA hybrid in addition to any accessory factors. The TEC comprises a 9-nucleotide RNA-DNA hybrid within RNAP (a view of RNAP on DNA is shown in Figure 1.15) [80].

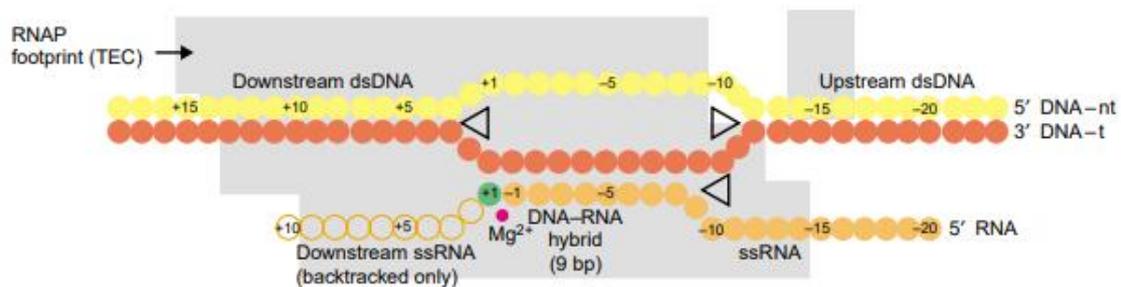


Figure 1.15: Position of RNAP on DNA during transcription (RNAP footprint shown in grey) [80].

Structurally speaking, RNAP forms a crab-claw shape with an internal channel for nucleic acids containing a conserved Asp triad that coordinates catalytic Mg^{2+} ; a secondary channel for activated rNTP entry/pyrophosphate exit; and an RNA exit channel [81]. The base at the downstream bubble junction (between the RNA-DNA hybrid and downstream duplex DNA; position +1) is located in the active site and is available to be bound by an incoming rNTP. Nucleotide addition begins with selection, the criteria for which are base-pair complementarity, steric fit, ribose sugar type, and presence of triphosphate. The reaction proceeds by an S_N2 mechanism that involves the Asp-coordinated Mg^{2+} ions. The conserved bridge helix and trigger loop, both of the β' subunit, are located near the active site. In the NTP-bound state, the trigger loop tips form trigger helices that, with the bridge helix, form a three-helix bundle. This bundle, along with the magnesium ions, correctly positions the NTP for catalysis [82]. An incorrect NTP prevents trigger loop folding and catalysis. As RNAP incorporates new rNTPs into the growing RNA chain it translocates along the DNA template strand via a Brownian ratchet mechanism, as shown in Figure 1.16) [83, 84]. Here, the active site of RNAP oscillates between pre-translocation

and post-translocation states, with NTP binding only possible in the latter state. NTP binding and incorporation into the RNA chain ratchets RNAP forwards by one base-pair, now again in the pre-translocation state.

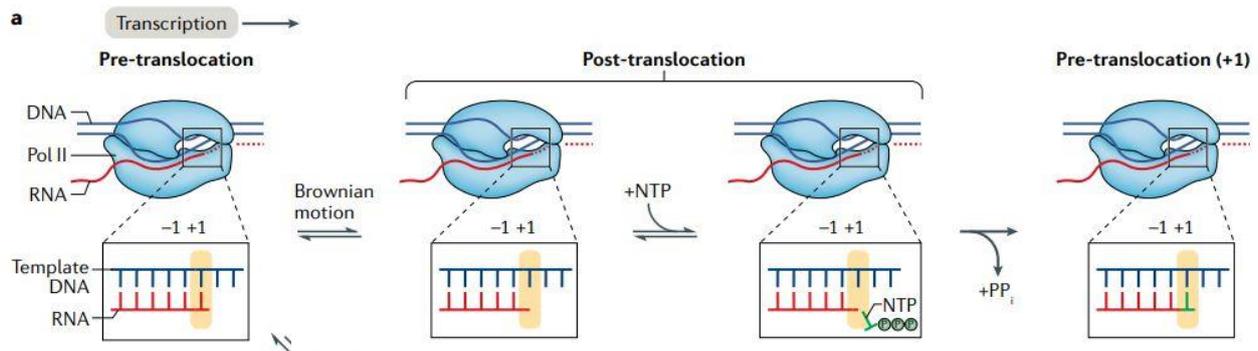


Figure 1.16: The RNAP active site register during NTP addition [84].

The bridge helix and trigger loop conformations also play a role in translocation. The RNA-DNA hybrid size remains constant because as a new DNA base-pair is unwound downstream, the most upstream RNA-DNA base-pair is broken and the DNA-DNA base-pair re-formed [85]. RNAP has an error rate of $\sim 10^{-5}$, in comparison with the $\sim 10^{-8}$ for DNA polymerases [86]. This may represent a key part of transcription-translation coupling as misincorporation-induced pausing can aid RNA folding and translation. At the same time, this relatively high error rate leads to phenotypic diversity that is transient but potentially important for survival [87]. RNAP can deal with misincorporation through RNA cleavage. RNAP possesses endonuclease activity but can also carry out exonuclease activity by backtracking, to position the RNA phosphodiester backbone in the active site, and binding a Gre factor, which coordinates the Mg^{2+} to enable cleavage [88].

Transcription of a single molecule of DNA by a single RNAP can be monitored using the magnetic trap. As mentioned, the use of supercoiled DNA ($\sigma \sim 0.05$) means that, due to the conservation of linking number, a relatively small change in twist is converted to a relatively large change in writhe. For example, the formation of open promoter complex (RP_o) involves the unwinding of ~ 10 bp (1 turn) of promoter DNA, to form the transcription bubble, which is balanced by a compensatory increase in writhe (positive plectonemic supercoiling). This gain of positive

supercoiling decreases the DNA extension by ~ 60 nm (Figure 1.17 (left) and Figure 1.17 (right; step 1)) [89]. Once in RP_o , RNAP begins to synthesise RNA and uses a scrunching mechanism to break free of its promoter contacts and begin productive initiation. During scrunching, RNAP pulls up to ~ 10 more downstream base-pairs into itself which results in a further gain of positive supercoiling, decreasing DNA extension by ~ 100 nm from baseline (Figure 1.17 (right; step 2)). Once the initially transcribing complex (RP_{itc}) has escaped from the promoter, the transcription bubble returns to a size of ~ 10 bp and so the DNA extension returns to a level ~ 55 nm from the baseline (Figure 1.17 (right; step 3)). The time RNAP spends in elongation phase (RD_e) depends on the length of the transcription cassette and the RNAP speed. Upon transcription termination, the transcription bubble closes, the positive supercoiling is lost, and the DNA returns to its baseline extension (Figure 1.17 (right; step 4)).

Since the DNA extension informs on the state of the RNAP, the magnetic trap provides a simple and convenient tool to study transcriptional regulation. For instance, RNAP can be stalled due to DNA damage or nucleotide starvation within the first ~ 25 nt of the transcription cassette, which sterically prohibits a second RNAP from binding the promoter. Here, the DNA extension remains ~ 55 nm from baseline unless the RNAP is removed by an accessory factor, identified by a return to baseline extension. At that point, the promoter is free to be bound by another RNAP, meaning one can observe iterative event cycles and obtain reliable statistics.

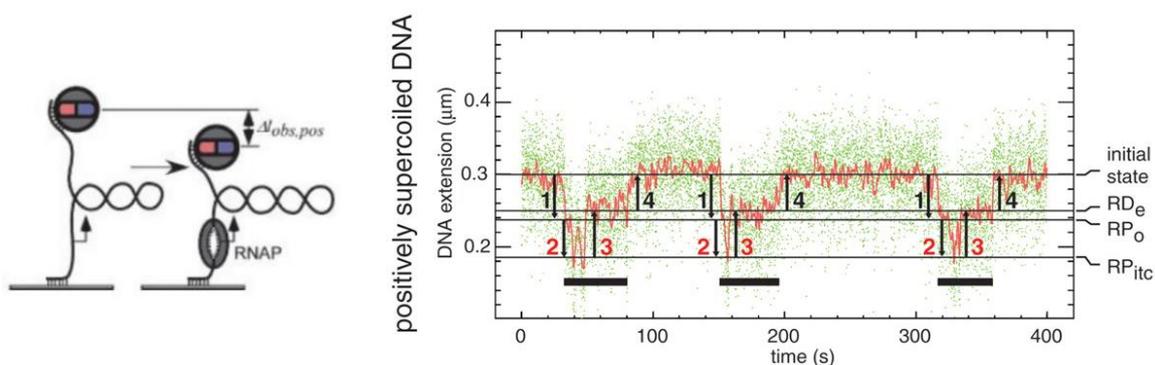


Figure 1.17: Left: Schematic of RNAP binding the promoter of a positively supercoiled DNA molecule. Right: Time-trace of three transcription cycles on a positively supercoiled DNA molecule [89].

RNAP Accessory Factors

The consensus elemental pause sequence for *E. coli* RNAP is $G_{-10}Y_1G_{+1}$ [90]. Backtracking and RNA hairpin formation stabilise an elemental pause and increase the residence time on DNA of the now catalytically-inactive RNAP conformer. RNAP accessory factors, such as the Gre factors, are necessary to deal with RNAPs that have paused. GreB, for instance, has an N-terminal coiled-coil domain that stretches into the RNAP active site via the secondary channel and a C-terminal globular domain that binds to RNAP β' subunit [91]. *In vivo* studies support biochemical data showing GreB increases transcription fidelity [92]. Other well-studied accessory factors are Mfd, NusA, NusG and UvrD. NusA has been shown to play a role in hairpin-stabilisation and so favours pausing in certain DNA regions [93]. NusG is found in all domains of life. When bound to RNAP, NusG can stabilise upstream DNA or may stimulate RNAP clamp closure in order to inhibit backtracking [94]. NusG also binds to either Rho, the termination factor, or NusE/S10, the ribosomal protein, via its C-terminal domain which allows cross-talk between transcription-translation coupling and transcription termination [95].

The same set of interactions that govern the early steps of TCR also allow Mfd to regulate transcription. As mentioned, once Mfd binds to stalled or paused RNAP and upstream DNA it begins to translocate forwards which builds up torque on the RNAP. If the RNAP is transiently paused, rather than damage-stalled, it will be re-activated and restart transcription before Mfd can commit the complex to TCR [96]. Recent evidence shows Mfd tends to regulate transcription in a context-dependent manner, depending on cell stress and the propensity of certain genes to form structured RNAs [97, 98]. Mfd's role in transcriptional regulation has examples in virulent contexts. For instance: in bacteriophage HK022 Mfd helps to remove RNAP stalled by Nun proteins; and in *Clostridium difficile*, Mfd controls the expression of toxins by removing RNAP stalled by toxin repressor proteins [99, 100]. Similarly, Mfd controls carbon catabolite repression by removing RNAP stalled at cre sites of hut and gnt operons by repressor proteins [101].

UvrD, in addition to its role in the later stages of NER and MMR, is capable of binding RNAP using a Tudor-like domain similar to that of Mfd [102]. Genetics studies show us that UvrD is one of a few helicases that act to resolve transcription- replication conflicts (see later), as the cell tolerate a single helicase knockout (e.g., UvrD or Rep), but is not viable upon double helicase knockout (e.g., UvrD and Rep) [103]. This is in agreement with observations that UvrD backtracks RNAP, which would help the replisome to progress through an RNAP roadblock in a head-on conflict [103].

Lastly, the RapA protein binds to RNAP near the RNA-exit channel but its role in the transcription cycle is poorly understood [104]. RapA was first identified because it consistently co-purifies with RNAP preparations from *E. coli* [105, 106] and further interrogation showed RapA had RNAP-dependent ATPase activity as well as nucleic acid-binding capabilities [107-109]. Studies have shown that RapA aids transcription reactions *in vitro*, but RapA has no effect on promoter binding, promoter escape, elongation or termination, with speculation that RapA helps by recycling RNAP in off-pathway conformations [110].

Transcription and DNA Topology

Transcription results in DNA topology changes which itself results in the modulation of transcription. Within the transcription-topology relationship, I will first consider how DNA topology affects transcription, then consider how transcription affects DNA topology.

Firstly, the supercoiling status of a given gene will favour or disfavour transcription initiation at certain promoters by altering the alignment of the -10 and -35 promoter elements or by changing the melting propensity (e.g., negative supercoiling can aid unwinding of certain promoters). Similarly high torque or negative supercoiling upstream of RNAP will favour backtracking during elongation phase [111]. Secondly, the presence of topological domains, whilst more important in

eukaryotes, is an issue for transcription. Not only do topological domains (e.g., TADs) essentially define and program which genes are accessible or not, these domains can essentially amplify supercoiling changes generated by transcription that would otherwise be dissipated in the chromosome. This in turn means that the use of one gene can affect its neighbouring genes unless they are topologically isolated from each other, as mentioned above [48]. H-NS forms bridges between two DNA segments that can topologically inhibit transcription elongation as they prevent rotation of DNA [112]. Further to DNA domains, phase separation of DNA and of RNAP has been proposed. RNAP clusters may phase separate to regions of DNA that are highly expressed and chromosomal sub-structures may phase separate in order to colocalise active genes [113]. This would reduce RNAP search times and increase transcription efficiency. Lastly, 'roadblocks' pose a constant and potentially problematic threat to transcription. A roadblock is anything that resides on DNA and acts as an obstacle to transcription, such as LacI or the replisome. When RNAP encounters a roadblock, it can be termed a conflict and the role of polymerase conflicts in mutagenesis will be discussed later. Many factors affect the outcome of a conflict, i.e., whether RNAP can overcome the obstacle and continues transcription or not. These include: the number of other RNAPs – multiple RNAPs cooperate to overcome obstacles [114, 115]; the presence of ribosomes or accessory factors – these can similarly help RNAP transcribe through the obstacle; the kinetics of binding of the roadblock – a long-lived roadblock poses more of a threat [116]; the propensity for RNAP backtracking – if the sequence-context favours pausing and backtracking then RNAP will have to be reactivated before proceeding past the roadblock; and Mfd – this accessory factor can help RNAP past the roadblock (via reactivating backtracked complexes) or catalyse RNAP removal [96, 101, 117, 118]. Conflicts commonly occur between RNAP and the replisome. RNAP can collide with the replisome in either a head-on or co-directional fashion, with the former being the worst as it leads to replisome stalling and incomplete transcription [119, 120]. As a result, essentially all highly-transcribed regions are found in an orientation co-directional with replication [121]. This, coupled with accessory proteins (e.g., UvrD which aids RNAP backtracking and Rep, the accessory replicative helicase,

which aids replisome bypass), reduces the frequency of head-on conflicts. The outcome of RNAP-replisome conflicts will be discussed in a later section.

During transcription there must be relative rotation of the TEC around the DNA axis in order for downstream template DNA to feed into the RNAP active site. For a variety of reasons, the TEC may not be able to rotate around the DNA. For example, due to viscous drag generated by the many ribosomes attached to the RNA, or due to co-transcriptional co-translational insertion of a transmembrane protein into the lipid bilayer [122]. Both DNA domains (downstream of RNAP and upstream of RNAP) have fixed linking numbers and during transcription RNAP transfers DNA, but not links, from the downstream domain into the upstream DNA domain. Therefore, the downstream DNA domain accumulates positive writhe (supercoiling) whilst the upstream DNA domain accumulates equal negative writhe (supercoiling). This is known as the twin supercoiled domain model, shown in Figure 1.18 [122, 123]. Considering a circular bacterial chromosome, it is theoretically possible that these two supercoiled domains diffuse, meet each other and annihilate; however, in reality topological domains act to limit large-scale diffusion, whilst topoI and gyrase act on and neutralise these supercoils.

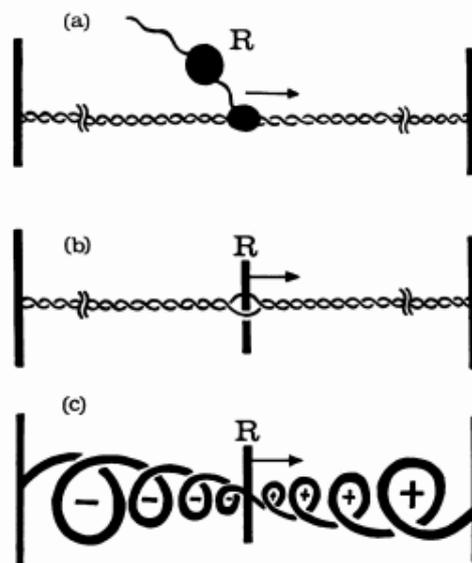


Figure 1.18: The twin-domain supercoiling model. RNAP with RNA and associated ribosomes form the transcription complex (R) which moves to the right on a torsionally-constrained DNA molecule. Waves of positive supercoils form in-front of R whilst waves of negative supercoils form behind R [122].

A consequence of the twin supercoiled domain model is that upstream genes in the vicinity may be upregulated as the negative supercoiling drives promoter unwinding, whereas the positive supercoiling inhibits transcription of downstream genes (as shown in Figure 1.19) [124].

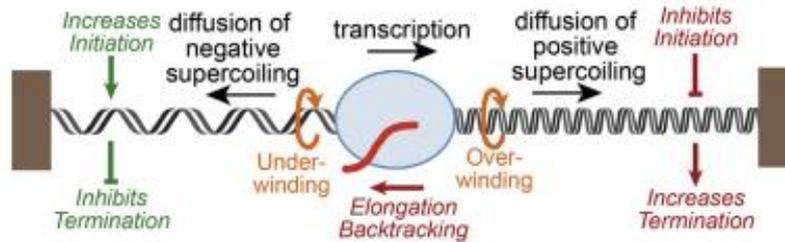


Figure 1.19: How topological ramifications of transcription can affect other promoters [124].

Transcription Under Stress

The stringent response is initiated in bacteria under nutrient stress and relies on the small molecule ppGpp (for simplicity this refers to both pppGpp and ppGpp, guanosine pentaphosphate and guanosine tetraphosphate, respectively). RelA and SpoT control ppGpp level [125]. The synthesis of ppGpp by RelA is activated by RelA binding to uncharged tRNA molecules in the ribosome active site, which is an indicator of nutrient stress [126]. ppGpp affects transcription along with the accessory factor DksA, typically acting at the initiation step, either in a positive or negative manner [127]. For example, ppGpp and DksA downregulate transcription from rRNA promoters [127]. ppGpp binds two sites on RNAP: binding site 1 is located at the interface between the β' and ω subunits; and binding site 2 is located at the interface between the β' subunit and DksA [128, 129]. The action of ppGpp at sites 1 and 2 is needed for full inhibition of transcription from certain promoters, whilst action at site 2 only is needed for upregulation of transcription for other promoters. Whilst there is much still to understand regarding how ppGpp and DksA affect initiation, it is thought that promoters are distinguished due to their initiation kinetics, which is a function of promoter sequence. Different promoters proceed through initiation steps with different rates and certain steps are sensitive to ppGpp and DksA therefore

promoters which proceed through these steps very quickly are largely unaffected by the stringent response. Alternatively, promoters that reside for longer in ppGpp/DksA-sensitive states are prone to modulation. In addition, ppGpp and DksA can affect transcription elongation – reducing the speed of elongation may help transcription-translation coupling in cells under stringent response [130].

Chapter IV: DNA Damage and Repair

Types of DNA Damage

The DNA backbone is more stable than the RNA backbone because DNA lacks the sugar 2' OH group (i.e. DNA has deoxy- rather than oxy-ribose) which is more prone to hydrolysis [131]. Thus, DNA has a longer-lived backbone but shorter-lived ribose-base bonds. The presence of a complementary base on the opposite strand of DNA means a labile ribose-base bond (the hydrolytic loss of the base is termed depurination or depyrimidination) is an acceptable problem, as DNA repair processes use the information on the opposing strand to fill-in the missing information [132]. DNA damage can be either repaired or, if not repaired, lead to mutation or cell death. Here I will discuss endogenous then exogenous DNA damage sources.

Figure 1.20 outlines the different ways DNA can be damaged via endogenous sources. Deamination can affect most bases (converting cytosine to uracil, 5-methylcytosine to thymine, guanine to xanthine, and adenine to hypoxanthine). Oxidation leads to numerous lesions such as the common 8-hydroxyguanine lesion, as well as single-strand breaks [133-135]. In some cases, the lesion will hydrogen bond preferentially to a different base, leading to post-replicative mutations. Alternatively, the lesion blocks the replication machinery leading to cell death. S-adenosylmethionine can methylate DNA in a non-enzymatic fashion (rather than enzymatic, such is the case for the generation of 5-methylcytosine), which can similarly block replication [136]. Double-strand breaks are rare but the most toxic form of lesion. They can be caused by endogenous factors such as reactive oxygen species (and also exogenous factors such as UV light, ionising radiation and chemotherapeutics) [137]. Finally, DNA damage can result from normal cellular processes – replication occasionally involves misincorporation (of uracil for example); and topoisomerases can form 'suicidal substrates' if they happen to act at unusual DNA structures [138, 139].

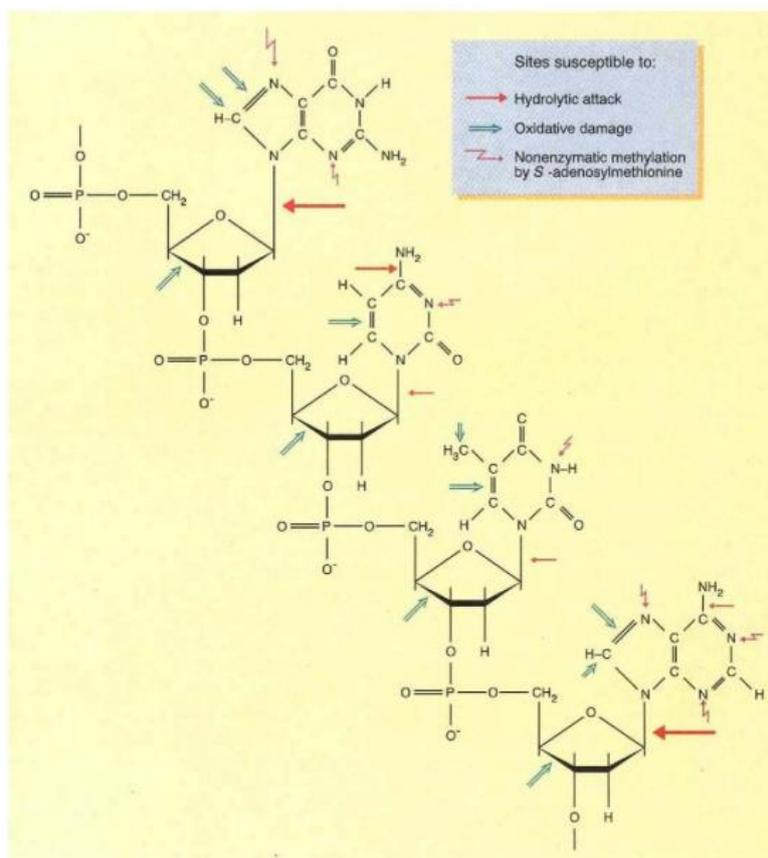


Figure 1.20: Areas of DNA that are susceptible to damage from endogenous sources [131].

Exogenous DNA damaging agents can be classed as chemical or physical (I will not discuss chemotherapeutics here). Chemical exogenous agents include: alkylating agents – generally diet-derived or environmental, alkylate bases with some being bifunctional leading to DNA cross-linking [85]; aromatic amines – from coal and cigarette smoke, attack guanine bases leading to persistent lesions which cause substitutions and frame-shifts [140]; polycyclic aromatic hydrocarbons – produced via incomplete combustion of organic matter, tend to intercalate and bind guanine residues, and are highly carcinogenic [141]. Physical exogenous agents include: ionising radiation – from a variety of sources, can directly lead to single or double-strand breaks [142], or indirectly lead to lesions via the generation of reactive oxygen species [143]; and UV radiation - comprising UVA, UVB and UVC (classed based on wavelength, with most UVC filtered by the ozone layer), mainly leads to cyclo-pyrimidine dimers (CPDs; two adjacent pyrimidines

bound covalently) and 6,4-photoproducts but can also lead to oxidative lesions and protein-DNA crosslinks [144-147]. CPD formation is shown in Figure 1.21 [148].

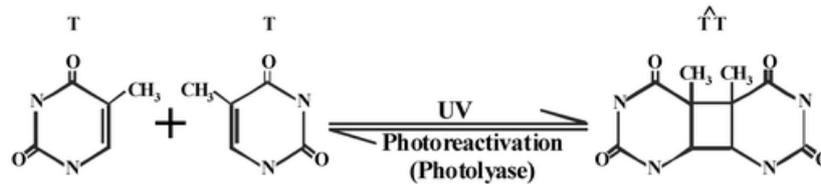


Figure 1.21: UV catalyses the formation of a CPD from two adjacent thymines [148].

Types of DNA Repair

Cells have a variety of ways to manage DNA damage which are generally universally conserved. As we have seen, DNA can be damaged in a variety of ways and cells take two main approaches to dealing with them: using pathways that are specific to a certain kind of DNA lesion; and using pathways that deal generally with abnormalities. The former pathway relies on recognition of the lesion itself, like in base excision repair, whilst the latter relies on the pathways ability to sense non-DNA-like chemistry, as seen in nucleotide excision repair. Cells also sometimes opt to ignore the lesion for a brief amount of time, regarding other cellular processes, such as replication, as more important and tending to the lesion after replication is complete. An example of this is trans-lesion synthesis. During replication, certain lesions can block the replisome. Rather than attempting to repair the lesion (bearing in mind this DNA is now single-stranded and so repair would lead to a troublesome single-strand break and potential replisome collapse), cells use trans-lesion polymerases that have more accommodative active sites and aren't stalled by lesions. Whilst this does increase the chances of misincorporation opposite the lesion, it allows the replisome to proceed past the lesion and for repair to take place post-replicatively [149].

Known pathways of DNA repair include: direct reversal, single-strand break repair (SSBR), double-strand break repair (DSBR), inter-strand crosslink repair (ICL), methyl-directed mis-

match repair (MMR), base excision repair (BER), and nucleotide excision repair (NER). These pathways are not mutually exclusive and many can repair the same type of lesion. These pathways and proteins are conserved but I will tend to focus on the bacterial systems.

Direct reversal is limited to a few specific lesions. UV-induced lesions can be directly repaired by photolyases in a reaction catalysed by UV light [150, 151]. Humans do not have competent photolyase activity but the NER pathway also deals with these types of lesions. O⁶-methylguanine is a common and potentially carcinogenic form of alkylation damage that can be repaired using direct reversal. The Ada protein recognises this adduct specifically and transfers the alkyl group onto itself (leaving the guanine base behind) in what is termed a suicide reaction because Ada is permanently modified [152, 153]. Modified Ada acts as a transcription factor for itself and other genes that are involved in the repair of O⁶-methylguanine [154].

Single-strand DNA breaks are generally dealt with by a group of enzymes which process the DNA ends and generate the necessary chemistry for the action of gap-filling DNA polymerases and/or DNA ligase. Double-strand DNA breaks can be repaired via non-homologous end-joining (NHEJ) or homologous repair (HR). The latter can only occur if there is a complementary DNA copy available. NHEJ begins rapidly with Ku70/80 heterodimer recruitment to the DNA ends [155]. This serves to protect the DNA ends from degradation and to act as a platform for further protein recruitment. Many proteins including DNA-PKcs, XRCC4, XLF and ligase IV ultimately form a large stable bridging complex. DNA ends are processed and ligated [156]. The joining of ends without a template makes NHEJ more error-prone in comparison to HR. HR is the dominant pathway in constantly dividing cells. Generally, HR first involves end resection by the functionally conserved RecBCD complex. These proteins unwind and degrade the DNA leaving 3' overhangs on either end [157]. Following resection, each 3' overhang is coated in the RecA protein to form a nucleoprotein filament that invades the homologous DNA template to form a D-loop [158]. DNA polymerase then fills in the gaps using the template DNA and DNA ligase completes the repair. The remaining Holliday junction is resolved by RuvABC.

The Fanconi Anaemia (FA) group of proteins deal with the repair of inter-strand crosslinks. Much of the pathway is not well understood and its cascade depends on the cell cycle [159]. If cells are non-replicative, NER in concert with trans-lesion DNA synthesis (TLS) can repair inter-strand crosslinks in a step-wise manner [160]. In replicative cells, the inter-strand crosslink stalls the replisome and the FA proteins work with TLS polymerases, HR machinery and NER machinery to repair the lesion [161].

MMR occurs in the seconds and minutes post-replication to fix any errors made during replication. MMR similarly deals with insertions and deletions of one to several base-pairs. The mis-pairing of bases presents a unique challenge because this error is not a lesion and cells cannot tell which base is the 'correct' one. Thankfully, in *E. coli* MMR relies on the fact that DNA is methylated at GATC sites and the methylation of the daughter strand (i.e., newly-synthesised strand) takes time. As a result, the DNA strand which is unmethylated is the DNA strand bearing the mismatched base [162]. Many mechanistic questions remain in this pathway, but we do know that MMR begins with MutS recognising the mismatch. MutL is then recruited and activated, and interacts with MutH that is located at a GATC site. This interaction induces DNA cleavage on the unmethylated DNA strand by MutH. Subsequent DNA unwinding by UvrD, DNA synthesis by DNA polymerase and re-sealing by DNA ligase complete the repair [163].

BER relies on an array of proteins that each recognise specific types of base lesion. These proteins are known as glycosylases and can be monofunctional (cleave out the damaged base) or bifunctional (cleave out the damage base and cleave the DNA backbone). Following removal of the lesion, this pathway proceeds either by short-patch or long-patch repair. If a monofunctional glycosylase acts, a second enzyme, APE1, cleaves the backbone and DNA polymerase acts to fill in the gap and DNA ligase seals the nick. This is short-patch repair. If a bifunctional glycosylase acts, when DNA polymerase begins gap-filling, strand-displacement occurs meaning additional processing by flap endonuclease is required before DNA ligase can seal the nick [164]. This is long-patch repair.

Global Genome Repair and Transcription-Coupled Repair

As we have seen, most DNA repair pathways focus on direct recognition of specific DNA lesions. The NER pathway however is capable of recognising an array of potential lesions due to their capacity to distort normal B-form DNA. The UV-induced CPD lesion is the canonical form of damage used to study NER. The Uvr genes were identified as responsible for NER as strains deficient in the Uvr loci were impaired in their ability to excise CPDs [165]. NER can be split into global genome repair (GGR) and transcription-coupled repair (TCR). In GGR, the DNA lesion is recognised directly by the UvrA₂B₂ complex, whereas in TCR, UvrAB is brought to the lesion site by the Mfd protein (also known as the transcription-repair coupling factor). During transcription, a lesion can stall RNAP indefinitely and Mfd couples RNAP removal to UvrAB recruitment [166]. TCR only occurs on the transcribed strand in DNA regions that are transcribed but is faster than GGR due to the rapid recruitment of UvrAB by Mfd. In both GGR and TCR, DNA damage verification by a UvrAB complex is followed by DNA incision by the dual endonuclease UvrC and removal of the lesion-containing oligonucleotide by UvrD helicase, and filling in by DNA polymerase and sealing by DNA ligase.

Genetics of Transcription-Coupled Repair

TCR was first studied from a genetic standpoint. When exposed to UV, the *E. coli* strain B would elongate and die. This is because UV induces expression of Sfi, the cell division inhibitor, which is usually degraded by the Lon protease after the cell has repaired the UV-induced lesions. Strain B lacks the Lon protease so Sfi persists and inhibits cell division, meaning cells elongate until they

eventually die [167]. *E. coli* strain B/r lacks Sfi and so doesn't elongate nor die in response to UV [167]. This UV-resistance made strain B/r a useful model organism to study TCR [168].

From B/r, a tryptophan auxotroph was isolated, named WP2 [169]. WP2 had an ochre nonsense mutation in the TrpE gene. UV light was used on WP2 to cause lesions that, if not repaired, led to mutations that could result in reversion to tryptophan prototrophy. This procedure was most effective (i.e. led to the most mutant prototrophs) when cells were moved to nutrient-rich conditions, allowing replication immediately post-UV [170]. Alternatively, if cells were made to wait (i.e., repress transcription and replication) before being moved to the nutrient-rich conditions they displayed far fewer prototrophs. This drop in prototrophs was termed mutation frequency decline (MFD). Further to this the strain WP2-S was isolated. WP2-S lacks the MFD phenotype [170]. Similarly, the tyrosine auxotroph WU36-10 was studied, and WU36-10-45 was isolated. WU36-10-45 lacks the MFD phenotype. It should be noted that WP2-S and WU36-10-45 still showed a gradual decline in prototrophs when they were made to wait before plating onto rich nutrient a sign that GGR was ongoing (pre-mutagenic lesions were being repaired during the waiting period).

The gene responsible for MFD (i.e., which is lacking or non-functional in WP2-S and WU36-10-45) was named Mfd. We now refer to the protein product of that gene as Mfd too. Mfd is functionally conserved in yeast as Rad26, in archaea as Eta, and in humans as Cockayne syndrome complementation group B (CSB; with protein partner CSA) [171-173].

Interestingly it was noted that “among auxotrophic sub-strains of *E. coli* B/r, isolated at random, only about 20 to 30% give relatively high yields of the MFD phenotype” [170]. In 1977, Bockrath and Palmer commented “in general, conditions favouring MFD have little effect on overall survival” [174]. Indeed in their comparison of WU36-10 (MFD+) and WU36-10-45 (MFD-) revertants, it is apparent that only a fraction of revertants – and in particular de novo nonsense suppressors – are affected by the MFD phenotype [174]. Other revertants – in particular backmutations— are repaired by mechanisms present at all time-points in the reaction (i.e., GGR).

A detailed analysis of auxotrophs helps to understand the genetic basis for these effects (Figure 1.22) [175]. These studies used strains made prototrophic by the appearance of a nonsense mutation in a key gene of the amino acid biosynthetic pathway [176, 177]. A nonsense mutation corresponds to the appearance of a stop codon in the mRNA – UAA (“ochre”), UGA (“opal”) or UAG (“amber”) – encoded by DNA “top strand” sequences TAA, TGA, or TAG, respectively. In the case of the WP2 strain, this corresponded to an ochre stop codon at the beginning of the TrpE gene required for tryptophan synthesis (denoted TrpE65(Oc), Figure 1.22A) [176]. For the WU36-10 strain, an ochre stop codon was found in the TyrA gene necessary for tyrosine synthesis (TyrA14(Oc)) [177]. These auxotrophs can revert to prototrophy if the nonsense mutation is itself undone by another mutation, and UV was the standard mutagen for such reversion assays – provided the base-pair change needed for reversion involves UV-susceptible adjacent pyrimidines (TT, TC, CT, or CC).

Thus, in the proper sequence context, the nonsense mutation can be undone by an unrepaired, UV-induced lesion and the error-prone replication it causes. Such UV-induced pre-mutational lesions typically result in either a direct backmutation to prototrophy (Figure 1.22B; in this case one speaks of a “true revertant”) or nonsense suppressor mutations in the anticodon loop of certain tRNAs (Figure 1.22C). However, pre-mutational lesions are repaired by the pathways under discussion and this repair leads to a reduction of the number of bacteria which revert to prototrophy (i.e., MFD). Pre-mutational lesions located on the coding strand can only be repaired by GGR, whereas pre-mutational lesions located on the template strand may be repaired in a manner which displays MFD on top of GGR.

Because the sequence context does not involve adjacent pyrimidines, UV-induced backmutation (or “true reversion”) from the TrpE65(Oc) TAA back to CAA should not take place (Figure 1.22B). As mentioned above, nonsense mutations are also often found to be undone by a compensatory (or ‘suppressor’) mutation in the anticodon loop of a tRNA; these are so-called de novo nonsense suppressor mutations (Figure 1.22C) [176]. For each stop codon there are nine tRNA anticodon

loops that are a single base-pair change away from becoming a de novo nonsense suppressor. For instance, the GlnU gene can become an ochre nonsense suppressor via UV exposure if a CPD is formed at the transcribed strand 'AAC TA' sequence, which then transitions, post-replication, to a 'AAT TA' (Figure 1.22C, red). Because CPDs and mutations located at the coding (non-transcribed) strand 'TT' cannot result in an ochre nonsense suppressor, a nonsense suppressor screen will only pick up the UV-induced lesions and mutations located on the transcribed strand. Because CPDs formed on the transcribed strand are prone to GGR as well as TCR, both repair pathways can, if given enough time, prevent those lesions from becoming a mutation and cause fewer prototrophs to appear. The exact balance between GGR and TCR at this locus will depend in particular on the transcriptional load at this locus. It is likely that the necessary transcription of these housekeeping genes ensures a sufficient level of RNAP on the gene to make it a good locus in which to observe MFD (TCR) as opposed to just GGR.

It is interesting to note that, in the case of the tyrosine auxotroph WU36-10 resulting from the TyrA14(Oc) mutation, it is the same GlnU gene which mediates de novo nonsense suppression as for the W2 tryptophan auxotroph – and these de novo nonsense suppressors also clearly display the fast MFD phenotype [170, 174]. Unlike the TrpE65(Oc) mutation however, the TyrA14(Oc) mutation can undergo true reversion via UV-induced lesions in a pyrimidine dimer on the coding strand, and also standard reversion via lesions in a pyrimidine dimer on the transcribed strand of the TyrA14(Oc) locus [174]. Indeed the coding strand sequence at this locus is GGC *TAA* TTA, where the nonsense codon resulting from appearance of a T is in italics and the coding strand shows adjacent pyrimidines in bold [177]. This explains why Bockrath et al. observe a slow GGR phenotype for true revertants (coding strand TAA>AAA) in the MFD+ strain [174]. Furthermore, it also likely explains why Witkin observed a slow GGR phenotype in WU36-10-45 (MFD-), but essentially no GGR phenotype in WP2-S (MFD-) [170]. Although the transcriptional load and pyrimidine contents on the coding or transcribed strands are already two parameters which must be instrumental in balancing the ratio between GGR and TCR, a third confounding factor has also recently emerged: pervasive transcription [178]. Pervasive transcription refers to the fact that

transcripts from both strands of entire genomes can be detected, albeit at highly differing levels. As a result, it is quite likely that some TCR of a gene's coding strand may also take place, depending on whether that locus sustains any level of antisense transcription from downstream genomic regions [179]. These recent genome-wide studies of TCR have nevertheless been able to show that Mfd is indeed the key component coupling transcription to repair.

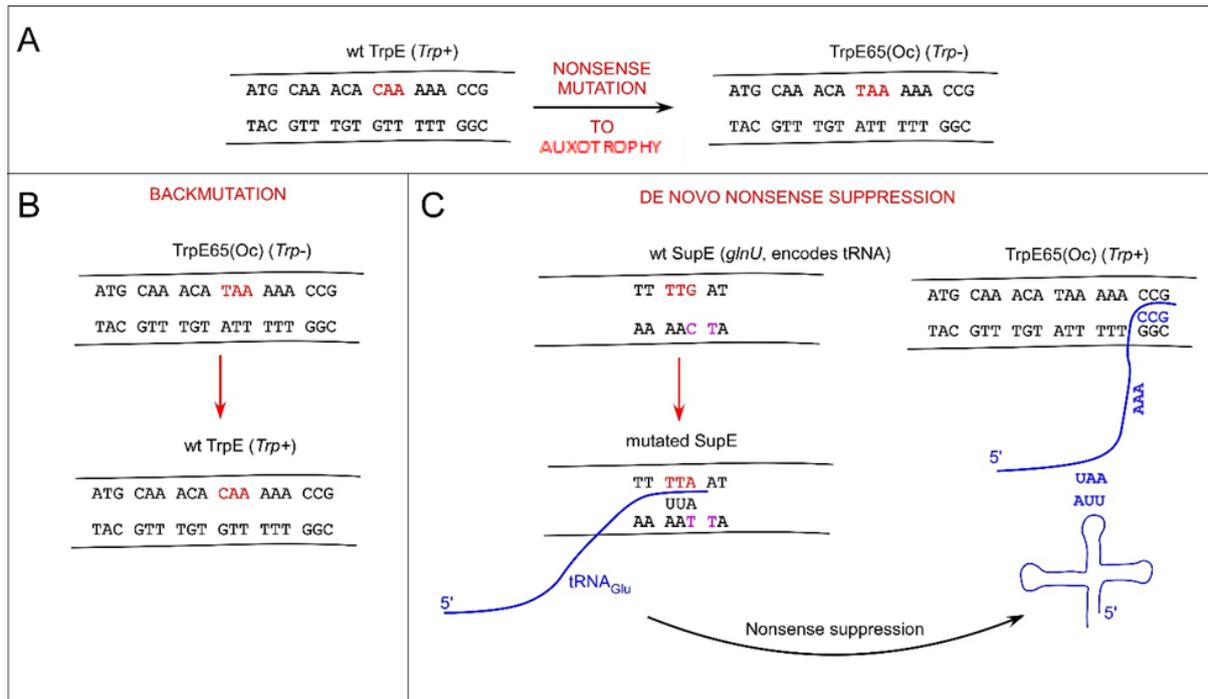


Figure 1.22: A) Genetic context of the nonsense mutation. B) How backmutation (true reversion) occurs. C) Mechanism of *de novo* nonsense suppression [175].

Transcription-Coupled Repair Mechanism I

TCR begins with the transcribing RNAP. Due to read-through and pervasive transcription RNAP can actually survey a large portion of the genome [178]. When RNAP encounters damage (such as a CPD) it stalls indefinitely [180]. This poses a problem to transcription (other RNAP on the same gene are now waiting behind), replication (the replisome can collide with RNAP) and indeed repair (with the damage in the active site of RNAP, it is essentially hidden from detection by

UvrAB). So, by physically removing RNAP from the lesion, Mfd not only partakes in TCR but also clearly has roles in transcription management and transcription-replication conflicts, as I will describe later.

Mfd Structure

Studying the structure of Mfd helps to understand its role in TCR (Figure 1.23) [181]. Mfd consists of 7 domains connected to each other by flexible linkers. It ultimately forms a cup-like structure with a 'latch' between domain 2 and domain 7 that renders the protein 'closed' – with the D2-D7 latch interaction all other domains are hindered in their function and so Mfd is auto-repressed [181, 182]. Domains 1 and 2 form the UvrB Homology Module (BHM) which, as its name implies, resembles the UvrA-interaction domain of UvrB and is the basis for the interaction between Mfd and UvrA [183]. Domain 3 is of unassigned function. Domain 4 forms a Tudor fold, known as the RNAP-interacting domain (RID), which binds to the β subunit of RNAP. Domains 5 and 6 form RecG-like DNA-binding and ATPase folds, assumed to enable 3' to 5' DNA translocation. Whilst these domains are in theory auto-repressed, some evidence indicates that Mfd does exist in an equilibrium between repressed and de-repressed states, although biased towards the former. Mfd can pull-down UvrAB from solution lacking RNAP; and Mfd displays poor triplex-forming oligonucleotide (TFO) displacement absent RNAP, indicating use of its translocation domains (this activity was greatly enhanced by disrupting the D2-D7 interface) [184, 185]. Various cryo-electron microscopy (cryo-EM) structures of DNA-bound Mfd revealed some structural changes that must occur within Mfd to bind DNA and the molecular contacts between Mfd and DNA [186]. Namely, two amino acid clusters at the 'tips' of D5 and D6 form two separate DNA contacts – a feature common to translocases as they cycle between contact sites to facilitate directional movement [183, 186]. Clearly Mfd exists in an equilibrium between open and closed states, allowing Mfd to bind dsDNA absent RNAP in this structural study. However, the biological

relevance of such structural information should be questioned, as we cannot validate the competence of such DNA-bound Mfd molecules for repair or transcriptional regulation.

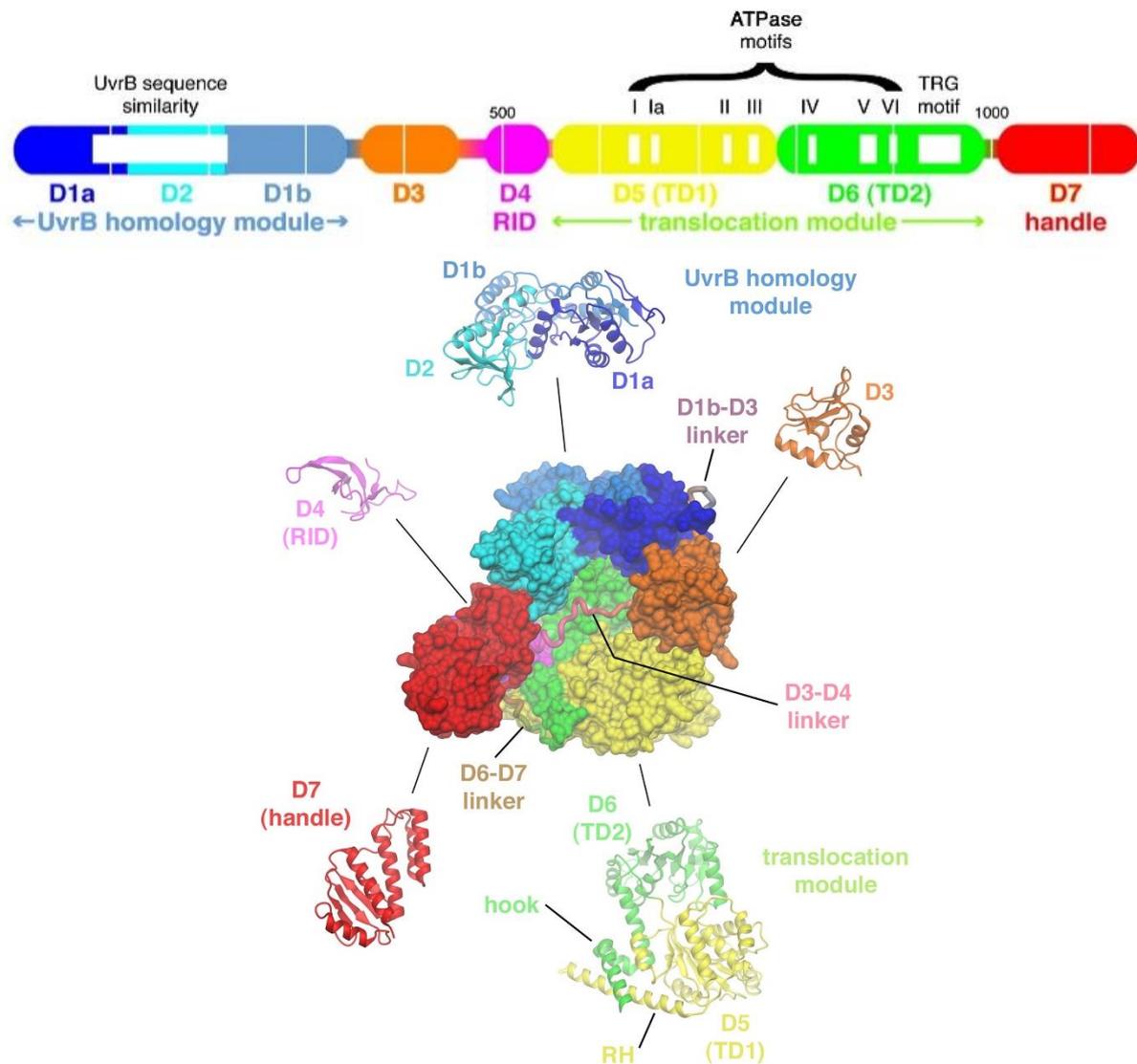


Figure 1.23: The domain organisation and three-dimensional structure of apo-Mfd [181].

Recently, seven Mfd-RNAP cryo-EM structures showed us the domain arrangements of Mfd as it interacts with RNAP and DNA during the initial stages of TCR [187]. Across these structures, the downstream DNA and RNAP show similar conformations, whilst upstream DNA and Mfd show major rearrangements. Once bound to RNAP and DNA, Mfd (all domains except the RID) undergo vast translations and rotations to topologically encircle the DNA (explaining earlier observations

of the vast processivity of Mfd engaged in *bona fide* TCR [18]). The physical route these domains take to wrap around the DNA is presumably not fast nor simple. Mfd translocation, along the minor groove, is expected to induce overwinding in the short stretch of DNA contained between Mfd and stalled RNAP, contributing to transcription bubble closure, mRNA release and RNAP dissociation [96]. Additionally, Mfd induces conformational changes in RNAP, most importantly clamp opening [187]. These structures further reflect the importance of ordering in multi-step pathways – only after binding RNAP and encircling the DNA does the BHM of Mfd become exposed. The consecutive structures indicate an ‘inchworm’ model of translocation, with the two DNA contacts cycling through states of ATP hydrolysis, γ -phosphate release, forward movement, formation of tight contacts, and ADP/ATP exchange [187].

Transcription-Coupled Repair Mechanism II

The steps subsequent to RNAP stalling but prior to UvrAB damage detection have been interrogated from biochemical, structural, and single-molecule standpoints, converging on the following mechanism. Mfd binds to the stalled RNAP and to DNA [185]. Mfd binds DNA upstream of RNAP, with a small stretch of DNA between the attachment points [96]. Mfd can now remove RNAP from DNA using translocation [96]. As understood from structural and single-molecule studies, Mfd controls the commitment of the pathway to TCR by distinguishing transiently paused RNAP from indefinitely stalled RNAP. The distance Mfd must travel (roughly 9 bp, which in part helps Mfd to form the topological clamp around DNA) and the step-wise generation of torque gives time for transiently paused RNAP to reactivate, move downstream and not be committed to TCR. Whereas only a damage-stalled RNAP will remain on the DNA long enough for Mfd to firstly form the full topological clamp (owing to the high processivity and long lifetime of the resulting Mfd-RNAP-DNA complex) and secondly to expose its BHM [187, 188]. Interestingly, RNAP is not lost to solution after being displaced from DNA. The RNA is lost from the complex, but RNAP

remains associated to the DNA via the β subunit-RID interaction. This Mfd-RNAP complex translocates in the same direction as transcription (i.e., in the 3' to 5' direction on the transcribed strand) with an extremely long lifetime absent UvrA [18, 189]. The translocation of the Mfd-RNAP explains observations that paused RNAP could promote repair enhancement of downstream lesions [190]. Translocation of the complex is relatively slow ($V_{\max} \sim 5\text{bp/s}$), but it is highly stable and serves to reveal the lesion to the cell and to act as a long-lived marker. The stability of the Mfd-RNAP complex underlies the importance of this process in the repair of lesions that impair gene expression. The BHM of Mfd acts as a relatively large surface for recruitment of UvrA. The high affinity UvrA has for the BHM of Mfd, in comparison to its affinity for a DNA lesion (roughly 20-200 times lesser) partly explains the enhanced rate of repair observed in TCR compared to GGR [189, 191]. Binding of UvrA, as part of a UvrAB complex, to the BHM causes the Mfd-RNAP complex to arrest translocation and ultimately dissociate. These experiments predict a transient RNAP-Mfd-UvrA-UvrA-UvrB complex during the short ($\sim 15\text{s}$) arrest phase [189]. Both Mfd and RNAP are lost to solution when the complex dissociates, leaving UvrAB on DNA. Regarding this Mfd-RNAP dissociation step, some questions are still to be answered. Depending on the cellular concentration of UvrA and UvrB, which in turn depends on environmental factors, the Mfd-RNAP complex will be interrupted at varying distances from the damage. This means that UvrAB is not actually conveniently deposited precisely on the lesion, rather UvrAB are recruited to the vicinity of the damage and must further verify its location.

The main hallmark of TCR is that it leads to faster repair of the transcribed strand of genes in comparison to the non-transcribed genes [192]. This is because RNAP is the starting point and so only identifies damage on the strand on which it moves. TCR is faster than GGR because of the molecular coupling between Mfd and UvrAB which results in fast, and presumably asymmetric, recruitment to the damage site vicinity. GGR shows equal rates of repair for both DNA strands because UvrAB searches in an un-biased, symmetric manner. It is generally thought that an asymmetric UvrAB complex must be recruited by Mfd to the DNA, as this would allow Mfd to 'inform' UvrAB on which DNA strand the damage is on, further contributing to an enhanced rate

of repair. Indeed, in theory Mfd can recruit just UvrA, UvrA₂ or UvrA₂B₁. However, in reality the first two options would ultimately lead to canonical GGR once recruited to the DNA, as each UvrA would probably be bound by a UvrB, with the third option leading to an alternative form of GGR in which the first steps have been skipped. Whilst estimations show most UvrA molecules *in vivo* likely exist associated with UvrB and so this third option is probable, the other options are still possible, highlighting the fact that repair processes can reach the final stages via a branched decision network rather than a linear pathway [193]. We can assume that the TCR branch of NER joins the GGR pathway at some point (probably after damage verification by UvrB) and then both pathways proceed in the same manner.

It's important to briefly consider pathways that aid NER by removing obstacles to UvrAB recognition but do not directly couple these processes – they can generally be referred to as “alleviation of transcription-based inhibition of repair”. The rho protein, known for rho-dependent transcription termination [194], can remove damage-stalled RNAP at certain genomic loci, thereby facilitating GGR [195]. The *rho mfd* double knockout shows severe growth defects and high UV-sensitivity in comparison to the single knockouts, indicating these proteins can augment each-others action to some degree [195]. UvrD, with its well-established role in the latter stages of NER, has been proposed to bind to damage-stalled RNAP and catalyse backwards translocation [196]. This study showed evidence for a UvrD-RNAP interaction, with UvrD mediated RNAP backtracking to reveal the lesion. The UvrD-RNAP interaction may be promoted by ppGpp and so perhaps is only relevant during the SOS response [197]. Whilst a UvrD-UvrB interaction has been reported [198] there is no clear coupling mechanism (as with rho, this is not TCR) and subsequent reports dispute these findings [179, 199, 200].

GGR begins with damage recognition and can recognise a wide array of lesions due to the structural distortion of the DNA helix (e.g., as seen in the thymine dimer shown in Figure 1.24) [201]. UvrA dimerises in an ATP-dependent manner and each UvrA in the UvrA₂ can bind a UvrB in an ATP-dependent manner, forming the UvrA₂B₂ damage-search complex [201, 202].



Figure 1.24: How a thymine-thymine dimer disrupts the DNA double helix (lesion-containing DNA overlaid in green) [201].

This complex scans the genome, typically displaying sliding motion on DNA that can be random, directed or paused. The UvrA₂B₂ complex has higher affinity for damaged DNA in comparison to non-damaged DNA, in part as a result of negative cooperativity between ATPase sites due to the presence of damage [203-205]. Although, both affinities are relatively high, sometimes leading to the repair of un-damaged DNA [206]. UvrA contains an insertion domain, with two conserved arginine residues that contact the DNA backbone, and a zinc-finger motif which is responsible for coupling the damage signal to the ATPase domains [207]. Figure 1.25 shows the UvrA₂B₂ complex bound to DNA [201].

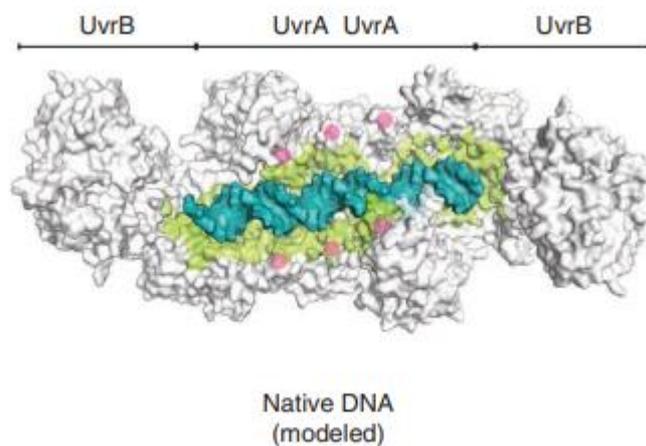


Figure 1.25: Structure of the UvrA₂B₂ hetero-tetramer on un-damaged DNA [201].

Once UvrAB is bound to the damaged DNA, UvrA departs leaving a pre-incision UvrB-DNA complex [208]. DNA is wrapped around UvrB, UvrB forms a small bubble and inserts a hairpin between the DNA strands [209, 210]. UvrB now translocates along DNA in an ATP-dependent manner and uses a base-flipping mechanism to interrogate each base [211, 212]. As this inserted hairpin moves in between the DNA strands, one base is flipped at a time into a small hydrophobic pocket within UvrB in which the base stacks on to a conserved phenylalanine residue [213]. This discriminatory pocket fits and allows through undamaged DNA but bulky lesions or CPDs cannot pass, leading to stalling of UvrB and recruitment of UvrC [211]. It is presumably at this point where TCR and GGR now proceed alike. UvrC carries out dual DNA incision, cleaving the DNA backbone 7 nucleotides 5' of the lesion and 3 or 4 nucleotides 3' of the lesion, leaving a damage-containing oligonucleotide still annealed to the double helix [214]. Following incision, UvrD is recruited through an interaction with UvrB [198]. UvrD is a superfamily I helicase that unwinds DNA 3' to 5' [215]. UvrD then unwinds the damage-containing oligonucleotide from duplex DNA, as well as ploughing off UvrC [198, 216]. DNA polymerase can then fill-in the short patch – this perhaps also results in the dissociation of UvrB [216]. DNA ligase completes the repair. A summary of the latter stages of GGR is shown in Figure 1.26 [217].

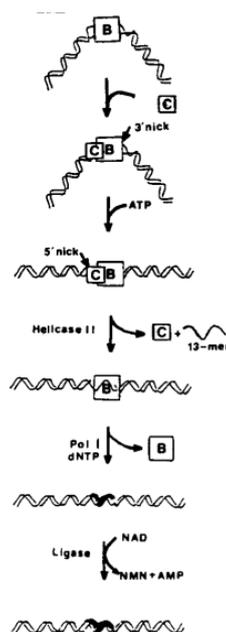


Figure 1.26: Putative model for the downstream steps of NER [217].

Recently NER has been studied *in vivo* using various fluorescently-tagged versions of UvrA and Mfd [218, 219]. Whilst these observations generally support conclusions drawn from *in vitro* data, there is additional complexity derived from imaging *in vivo* because multiple processes are spatio-temporally concurrent and pathways cannot be studied in isolation. Furthermore, we cannot identify the state of the substrate (e.g., elongating RNAP, paused RNAP, damage-stalled RNAP). As a result, data on Mfd for example reflects the sum-total of Mfd interactions in TCR as well in transcription regulation, transcription termination and tripartite-domain phenomena (these additional Mfd roles will be discussed later). So, the application of *in vivo* data to a mechanistic model for TCR is inherently risky and this must be borne in mind when comparing to isolated, re-constituted *in vitro* systems. Of course, this is not a predication that *in vivo* pathways proceed linearly and as described by our models, but a sobering reminder that *in vitro* data will tend to represent the path that is taken when all components are available at given concentrations and with no competing processes, whilst *in vivo* data will tend to represent the paths that are taken when competing processes and limited resources force the system to achieve the ultimate goal. Using the example of TCR, a lifetime of 18s was ascribed to fluorescent Mfd binding to RNAP in growing cells [220], but in cells lacking UvrA this lifetime increased to 29s [219]. This slight increase is in contrast to the roughly 15-fold lifetime increase observed *in vitro* (in fact this difference is surely greater as measurements of the true Mfd-RNAP translocase lifetime absent UvrA are limited by the length of DNA used) [189]. As we have seen, it's likely that *in vivo* the Mfd-RNAP complex dissociated due to another process and so this lifetime does not reflect the true fate of the Mfd-RNAP complex in TCR. Furthermore, if the state of the RNAP is not known, we must assume that fluorescent Mfd lifetime measurements *in vivo* represent Mfd engaged in TCR, transcription regulation (i.e., when RNAP is paused, not damage-stalled), and tripartite supercoiled domain formation (i.e., when RNAP is elongating, see Results section). Therefore, it is difficult to know what fraction of the fluorescent Mfd lifetime measurements reflect Mfd engaged in TCR. Interestingly in cells lacking UvrB the lifetime increased to 139s \pm 20 and cells expressing a hairpin-mutant UvrB showed similar lifetimes of 188s \pm 46, indicating that

the hand-off complex (the transient RNAP-Mfd-UvrA-UvrA-UvrB complex, observed to have a 15s lifetime *in vitro*) may rely on UvrB damage verification to trigger UvrA-UvrA-Mfd-RNAP dissociation, in addition to ATPase activity by UvrA. Distal ATP hydrolysis by UvrA is required for binding to Mfd but proximal ATP hydrolysis is needed for dissociation. Whilst UvrA is capable of dissociating the Mfd-RNAP complex alone, the process is faster with both UvrA and UvrB present. Regarding GGR, UvrA was observed to show two lifetimes absent Mfd and UvrB, a short 2s which reflects UvrA in damage-search mode, and a longer 24s which likely represents damage-bound UvrA because it changes in the presence of UvrB [218]. The UvrAB lifetime on DNA is ~9s and represents the time taken for UvrB to load correctly [218].

As mentioned previously, the ability of Mfd to act on RNAP allows Mfd to partake in non-TCR roles. Examples of such roles include rescue of backtracked RNAP [96, 117], carbon catabolite repression [101], aiding DSB repair [221], transcription-replication conflict resolution [222], transcription termination [99], recombination [223], gene expression control [100] and mutagenesis [63, 224, 225]. These examples, many of which are seen in many species, stem from the core functions of Mfd – RNAP-binding, DNA-binding, and translocation. Indeed, this represents a more general phenomena of pleiotropy – most proteins tend to have multiple functions. The gene responsible for such a protein is therefore less likely to be deleted or mutated because that would lead to myriad problems and loss of fitness [226, 227]. The role of Mfd in mutagenesis will be discussed in more detail later on.

Considering the great understanding we now have of TCR, how may we put into mechanistic context the observations of Witkin? In fact, we are presented with a paradox: post-UV exposure, *more* DNA is repaired (seen as a reduction of mutants) when bacteria are forced to *repress transcription* before plating on selective media. We may potentially reconcile this paradox with, at least, the following two notions. Firstly, it may be that the amount of transcription on tRNA is normally so high that TCR is inhibited, and only upon reducing the transcriptional load is TCR effective, leading to enhanced rates of repair and fewer mutants. Secondly, given the emerging

role of Mfd in mutagenesis one could envision that Mfd is responsible for some mutations that lead to prototrophy (either by true reversion or *de novo* nonsense suppression) and that repressing transcription reduces this phenomenon, leading to fewer mutants.

Eukaryotic Nucleotide Excision Repair

The human genetic disorder xeroderma pigmentosum (XP), characterised by sun sensitivity and freckling, is due to defects in UV-damage repair (i.e., defects in the NER pathway). As a result, many of the human NER genes are named according to the complementation group of XP they are responsible for (e.g., mutations to the XPA gene gives rise to XP group A).

NER, encompassing GGR and TCR, is well conserved across all domains of life. Whilst the number of proteins and level of fine-tuning may change, the basic rules of DNA repair are consistent – the lesion is identified, verified, cut-out and filled-in. As before, the GGR and TCR pathways begin differently but unify at the downstream steps.

In GGR, most lesions are recognised by XPC whilst other, less helix-distorting lesions (such as CPDs), are recognised by XPE and DDB1 which further kink the helix in order to recruit XPC [228]. XPC introduces a 42-degree bend and partially melts the DNA in order to insert a β -hairpin through the helix, allowing it to flip two un-damaged bases (opposite to the lesion) into a hydrophobic pocket [229]. XPC then recruits TFIIH [230]. This is the point at which GGR and TCR converge.

TCR, identified through the observation of strand bias in the repair of UV-induced lesions, has only been robustly observed in genes patrolled by RNAPII (i.e., genes encoding mRNA, although CSB has been shown to increase transcription efficiency at rRNA genes (transcribed by RNAPI)) [231]. CSB, a homodimer, is recruited to damage-stalled RNAPII and binds the Rbp2 subunit and upstream DNA, wrapping the DNA around itself [232]. CSB recruits CSA and many other proteins

(such as p300 and UVSSA) resulting in the ubiquitination of CSB and RNAPII [233, 234]. CSB is ubiquitinated at position K991 and interestingly disruption of this site does not impede TCR but does increase genetic instability (see later) [235]. UVSSA, in addition to CSA, stabilises the CSB-RNAPII complex. The additional complexity seen so far (the numerous protein partners recruited) can be partially explained by the chromatin context within which this pathway is occurring. Whilst transcription is typically associated with euchromatin, chromatin modification and rearrangement are still needed for efficient DNA repair (this has been referred to as the 'access-repair-restore' approach) [236]. As such, p300, the histone acetyl-transferase (these generally act to loosen chromatin), and HMGNI, the non-histone binding protein, together facilitate the nucleosome sliding that aids CSB [234].

The TFIIH complex comprises the XPB and XPD helicases which verify the presence of a lesion (XPD relies on its helicase activity to find the lesion) and extend the bubble to 20-30 nucleotides [228]. XPA, RPA and XPG are recruited to this 'pre-incision complex' – XPA binds the 5' edge of the bubble, RPA binds the ssDNA opposite to the lesion, and XPG binds TFIIH. Next, it has been proposed that XPA recruits XPF and XPF and XPG perform DNA incisions [228]. DNA polymerase simultaneously fills-in the gap and displaces the lesion-bearing oligonucleotide and TFIIH.

Human TCR is not very well understood and many questions remain, especially regarding if and/or how CSB displaces RNAPII from DNA. Similar to Mfd, CSB has myriad roles, for example in transcriptional regulation, BER, chromatin remodelling, intra-strand crosslink (ICL) repair and DSB repair. Comparing transcription rates with and without CSB reveals CSB to stimulate transcription elongation rate by three-fold [237]. CSB achieves this in part through transcriptional re-activation – like Mfd, CSB discriminates RNAP kinetically, only damage-stalled RNAP will enter the TCR pathway. Transiently paused RNAP will be reactivated by CSB. Similarly, CSB aids RNAP progression through nucleosomes and can initiate RNAP degradation (through ubiquitination, removing obstacles for other RNAP on the same gene) [238]. As mentioned, disruption of the K991 site on CSB leads to genetic instability. This occurs through poorly or un-

repaired oxidative lesions such as 8-oxoG. Cells of patients with CS accumulate 8-oxoG and 8-oxoA lesions over time and it appears CSB helps to repair 8-oxoG lesions in a number of ways. Firstly CSB appears to stimulate APE1, and secondly it was proposed that CSB is needed for efficient transcription of the OGG1 gene (whose protein product excises 8-oxoG) [239]. Thirdly, through a mainly unknown mechanism, CSB promotes BER in cells under oxidative stress in a PARP-mediated fashion (CSB is a lesser-known target of PARP) [240].

Finally, CSB has recently been implicated in mutagenesis, Not only does CSB appear to have roles in mutagenesis and anti-apoptosis, it seems cancerous cells take advantage of this and over-express CSB [241]. Interestingly, knocking down CSB resulted in apoptosis, reduced proliferation, and hyper-sensitivity to chemotherapeutics only in the cancerous cells, leaving non-cancerous cells unaffected [241]. CSB over-expression could be advantageous to a cancerous cell in a multitude of ways – for example increased DNA repair capacity or increased transcriptional regulation, but also through stress-induced mutagenesis. A recent study probed how the NER machinery links to R-loops that lead to genetic instability [242]. Sollier et al. began by establishing a cell line with high levels of R-loops by knocking down AQR, a helicase known to process R-loops. In this genetic background, it was discovered that the generation of (potentially mutagenic) DSBs was dependent on XPF and XPG. Going further, it was revealed that XPF/XPG-dependent processing of R-loops into DSBs was reliant on CSB, not XPC (and so was potentially a transcription-coupled process) [242]. One explanation is that CSB has another role in the processing of R-loops, and so knocking CSB out means R-loops aren't processed into DSBs but are resolved by other means. However, a unifying and easily testable alternative is that CSB interacts with RNAP to form R-loops, and therefore by knocking down CSB there are less R-loops to process into DSBs. As shown in the Results section, the bacterial homolog of CSB, Mfd, can interact with elongating RNAP to form R-loops that lead to mutagenesis. These observations support the hypothesis the CSB forms R-loops directly rather than processing them.

Chapter V: Mutagenesis

Early Mutagenesis Experiments

Whilst I have devoted much space to the discussion of DNA repair, it is important to remember that it is in fact the balance of repair and mutagenesis that drives survival and evolution. Mutations lead to diversity in a population which is a fundamental aspect of evolution, and cells and populations utilise a variety of methods to control their genomic stability.

Experiments by Luria and Delbrück in the 1940s gave the first insights into how bacterial mutations arise [168]. If one plates a culture of *E. coli* on to LB-agar containing bacteriophage, most bacteria will be killed but a very few will survive. These bacteria will remain resistant to the same bacteriophage for many generations – clearly, they contain mutations that confer bacteriophage resistance. At the time two hypotheses sought to explain this phenomenon: the mutation hypothesis and the acquired immunity hypothesis. The random mutation hypothesis states there is a small probability for any bacterium to mutate from bacteriophage-sensitive to bacteriophage-resistant, therefore their descendants will inherit this phenotype and at any given time during the culture there will be resistant clones. The acquired immunity hypothesis states that all bacteria are bacteriophage-sensitive but there is a small probability for survival, based on some pre-determined characteristics, and this immunity is passed onto descendants. Therefore, analysing many repeats of the above experiment using small culture volumes, one would expect from the mutation hypothesis to observe most plates to show no colonies, with few plates showing many colonies (characterised by a very high variance in comparison to the mean) [168]. Whereas based on the acquired immunity hypothesis, one would expect all plates to have roughly an equally small number of colonies (characterised by a variance equal to the mean). This difference is outlined in Figure 1.27 [243]. Indeed, when Luria and Delbrück performed these ‘fluctuation tests’ they observed an extremely high variance – some plates gave no colonies as

these cultures didn't happen to obtain the advantageous mutation, whilst other plates gave many colonies, indicating that mutations to bacteriophage-resistance had occurred before plating, leading to clonal resistant populations [168]. The conclusion was that evolution is driven by the mutation hypothesis, independent of stress. Although we now know additional layers of regulation (e.g., stress-induced mutagenesis, see below) act to fine-tune mutation rates and evolution.

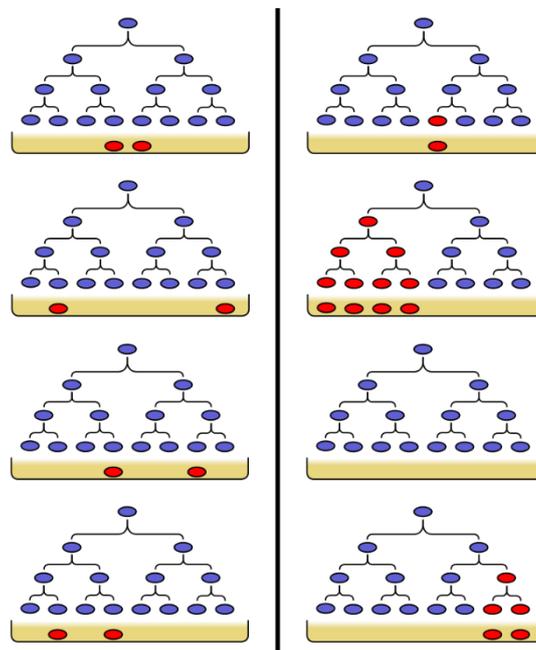


Figure 1.27: The acquired immunity hypothesis (left) expects a Poissonian colony representation (variance equal to the mean) whilst the spontaneous mutation hypothesis (right) expects a colony representation whose variance is much higher than the mean [243].

Stress-Induced Mutagenesis

Soon after this work it was proposed that bacteria might be able to control their mutation rates – increasing them in times of stress [244]. Cairns, Overbaugh and Miller (and later Cairns and Foster) plated Lac⁻ *E. coli* strains (specifically, *E. coli* Δ lac F' lacZ_{amber} which cannot metabolise lactose due to the mutated lacZ gene) on to LB-agar containing lactose and monitored colony appearance over many days [244, 245]. Two phases of colony appearance were observed. Across

the first two days, colonies appeared in agreement with the observations of Luria and Delbrück (displaying a very high variance; in this case the stressor is nutrient-based rather than bacteriophage presence). Beyond day 2, colonies appeared at a constant linear rate which was completely dependent on the presence of lactose. One can even verify these late mutations are indeed stress-induced by waiting a given number of days before adding lactose to the minimal media and observing that the colony appearance distribution is delayed by the same number of days [244]. This indicates two classes of mutations – non-adaptive mutation occurring during growth at random; and adaptive (stress-induced) mutation occurring during growth on selective media. Interestingly, using *E. coli* $\Delta lac F' lacZ_{amber}$ cells that also lacked UvrB showed the same mutation rate in the first two days, but a roughly five-fold increase in number of stress-induced mutants [244]. The first class of mutagenesis comprises deletions, duplications and frame-shifts whilst the second class of mutagenesis is mostly -1 bp frame-shifts (due to point mutation) [246]. The other mechanism contributing to stress-induced mutagenesis is amplification – cells accumulate 20 or more copies of the 'leaky' lac gene [247]. These stress-induced frame-shift mutations are highly dependent on DNA recombination and the stress-induced sigma factor RpoS, and mostly dependent on error-prone DNA polymerases [248]. It was later shown that Mfd is needed for this stress-induced mutagenesis because Mfd⁻ cells displayed roughly half the level of stress-induced mutagenesis in comparison to wild-type cells [63]. The current model for how stress-induced mutagenesis occurs relies on the Mfd-dependent generation of co-transcriptional R-loops, the mechanism for which is the focus of this thesis. This has been confirmed as RNase HI knock-out greatly increases the number of colonies observed, and knocking Mfd out in this background greatly reduced the number of colonies observed – verifying that R-loops are a key intermediate in the stress-induced mutagenesis pathway and Mfd is one of the main generators of R-loops [63]. In support of this, another study showed Mfd mutation suppresses RNase HI essentiality [249]. The R-loop then serves as a focus for origin-independent replication (cells at this point, on selective media, are non-replicating). When the replication fork encounters a ssDNA nick, the resultant fork collapse leads to a DSB which can result in mutagenesis via point mutation

(arising through HR-mediated replication restart using error-prone DNA polymerase) or amplification (micro-homology mediated ectopic strand annealing followed by replication restart) [63]. These experiments strictly probed mutagenesis on the plasmid-based Lac gene, and it's important to remember that this mechanism may be regulated differently when occurring on the chromosome.

Mfd has also been proposed to mediate mutagenesis, in many organisms, potentially through alternative mechanisms. In the soil bacteria *Pseudomonas putida* Mfd plays a role in UV-induced and stationary-phase mutagenesis, but does not affect mutation frequency in exponential phase [250]. In *Campylobacter jejuni* antibiotic treatment induces Mfd expression, and it has been shown that in this context Mfd over-expression increases spontaneous mutations and leads to antibiotic resistance, whereas Mfd⁻ strains are 100-fold less likely to become antibiotic resistant [225]. In nutritionally-stressed stationary-phase *Bacillus subtilis*, Mfd can be mutagenic even absent DNA lesions. Here, the Merrikh group showed that Mfd can work with polY1, increasing mutation frequency on genes on the lagging strand (whose RNAP is inevitably involved in head-on conflicts) [206]. Models to explain this include: head-on RNAP-replisome conflicts creating exposed ssDNA that is prone to damage and the repair, involving Mfd and PolY1, is error-prone; or post-replication, RNAPs may encounter more damage than usual (as discontinuous replication creates more lesions), also resulting in error-prone TCR. More recently, the Merrikh group has probed Mfd's mutagenic role in *Shigella typhirium*, *Pseudomonas aeruginosa*, *Bacillus subtilis* and *Mycobacterium tuberculosis* subjected to antibiotic stress [251]. Consistent with the earlier findings, they find that Mfd⁺ bacteria evolve resistance to antibiotics at rates two- to five-fold higher than Mfd⁻ strains, supporting a role for Mfd as an evolvability factor. Finally, it was shown that species cross-complementation maintained Mfd's mutagenic properties, implying that this mutagenic role of Mfd is common and evolutionarily conserved [251]. It should be noted that these observations stand in contrast to the original studies in the field, in which appearance of resistance to chloramphenicol was nearly systematically examined, but with no apparent difference between Mfd⁺ and Mfd⁻ strains [170]. This may be due to the fact that the classical

experiments exposed the bacteria to both UV and antibiotic, possibly titrating available Mfd onto stalled RNAP for DNA repair. It is certainly clear that picking apart the opposing context-dependent roles of Mfd is difficult. Reductionist *in vitro* systems allow us to understand how Mfd acts when solely performing TCR or transcriptional reactivation or mutagenesis, but only *in vivo* experiments can inform us on how the cell uses environmental cues to modulate the many roles of Mfd.

Whilst Mfd is a key mediator of stress-induced mutagenesis, it's important to remember cells deal with stress in a wholesale manner – utilising many approaches with the goal of survival. Nutritional stress, which along with drastic temperature change, low pH and increased osmotic pressure leads to the general stress response, which is mediated by RpoS. RpoS is a sigma factor that re-programs the transcriptional profile of the cell to alter the expression of roughly 200 genes [252]. Importantly, RpoS increases the expression of error-prone DNA polymerases and decreases expression of MMR proteins, favouring mutagenesis. The SOS response is triggered due to DNA damage which generates ssDNA (either directly or as an intermediate in lesion repair). RecA binds ssDNA which stimulates the auto-cleavage of LexA leading to the de-repression of the SOS response genes [252]. The strength by which LexA represses a gene controls the order of de-repression, meaning firstly NER genes are up-regulated and, if needed, later on genes involved in arresting cell division, promoting filamentation, and error-prone TLS are up-regulated [253]. Other stress responses include the heat-shock response and stringent response.

Spontaneous Mutagenesis

Spontaneous mutagenesis in bacteria gives rise to a basal evolution rate that can be modulated based on environmental stressors. A source of spontaneous mutagenesis is DNA replication. Not only does the most high-fidelity DNA polymerase still make errors, conflicts between the DNA replication machinery and the transcription machinery lead to mutations. As we have seen, cells

tend to align highly-transcribed genes in the same direction as replication in order to reduce the number of head-on conflicts that occur. Whilst it is not fully understood, it was observed that these conflicts lead to indels (insertions, normally gene duplications, and deletions) and single base substitutions in the location on the DNA where the two protein complexes initially contact [254]. Furthermore, the ssDNA exposed in the replication machinery, transcription machinery and in R-loops provides a substrate for spontaneous mutagenesis through its intrinsic propensity for cytosine deamination. Promoters are especially vulnerable to this as during transcription initiation both DNA strands are exposed and single-stranded – the T₋₇ position in the transcribed strand of the -10 element in a hot-spot for AT>GC transversions. The adenine opposite to the thymine can be deaminated to hypoxanthine, which base-pairs with cytosine during replication [255]. Finally, it was observed that roughly one percent of exponential-phase *E. coli* have induced their SOS response, causing mutations through error-prone TLS [256].

One can consider mutation rates not of individuals, but rather of populations. One may assume that all bacteria share roughly the same mutation rate, but instead we observe high variability in the mutation rates of individuals in a population. We know that if the mutation rate increases, fitness decreases (the rate of deleterious mutations is $2-8 \times 10^{-4}$ whilst the rate of beneficial mutation is 2×10^{-9}) [257, 258]. Therefore the rate of mutation of an individual must be balanced between the risk of deleterious mutation and the fitness costs associated with reducing the mutation rate [259]. One would therefore predict that ‘mutators’ (i.e., cells with a high mutation rate, normally due to mutations to MMR proteins) are rarely found, although interestingly mutators are readily found in nature. We observe that mutators represent about one percent of a given population, but this value can increase if that population overcomes a stressor because there is linkage between a favourable mutation and the mutator allele(s) [260]. However, mutators seem to rapidly specialise to a given niche so quickly that they are unable to cope with a further environmental change. Overall, weak mutators, which represent about 20% of the population, are slow to adapt but persist for much longer [260]. Some species (e.g., *Haemophilus*) potentially bear genome regions that display a high mutation rate (‘mutator region’) in

comparison to the rest of the genome. This is probably due to the DNA context (DNA repeats are prone to replication slippage and indels, and genome regions devoid of GATC sites ('GATC deserts') display reduced MMR capability) and is only beneficial in specific scenarios (e.g., if the protein product of a gene in the mutator region is involved in virulence or immune evasion) [261].

The last mechanism of spontaneous mutation I will discuss is horizontal gene transfer. This form of adaptation has gained interest in recent years because it is a mechanism by which antibiotic resistant bacteria can share their resistance [262]. Generally bacteria share mobile genetic elements and transposons through conjugation which involves a secretion-system to transport ssDNA through a direct cell-to-cell connection [263]. Once in the cytoplasm HR mediates the incorporation of this DNA into the genome. Similarly, competent bacteria can take up naked DNA from their surroundings and incorporate it into their genome. Bacteriophage can also transfer mobile genetic elements through a process known as transduction – when bacteriophage infect their host, they insert their DNA into the host genome. As new phage particles are made and DNA is packaged into them, flanking fragments of the host genome are included and therefore passed on to the subsequent host.

Section II: Results

Chapter I: Co-transcriptional R-loop formation by Mfd involves topological partitioning of DNA

This chapter focuses on my research into the mechanism by which Mfd mediates mutagenesis. As we have seen earlier, the fluctuation tests of Luria and Delbruck established spontaneous mutation theory, but soon after it was revealed that stress-induced mutagenesis also occurred under certain conditions. Simply put, if we plate a small culture of *E. coli* onto selective media, the colony representation in the first two days displays a high variance indicative of spontaneous mutation during growth in non-selective liquid media. Whereas following day two, colonies appear at a constant rate and this is due to stress-induced mutagenesis. We are now aware that Mfd is responsible for roughly half of the stress-induced mutations, and that Mfd achieves this through the generation of R-loops [63]. It appears that the role of Mfd in stress-induced mutagenesis is conserved, as this phenomenon has been observed in many bacteria and in human cells. The human homologue of Mfd, CSB, recently emerged as a driver of mutagenesis in human cancers. Taken together, the mechanism of mutagenesis by Mfd is potentially conserved from bacteria to humans and so revealing it has implications for both bacteria anti-microbial resistance and chemotherapeutics. Many studies have shown how R-loops perturb genetic stability – whilst not always necessarily mutagenic, increasing R-loop numbers favours disruptive and mutagenic pathways. Thus, my research has focused on how Mfd forms R-loops and how this relates to genetic instability and genome topology. It is presented in the form of a manuscript below.

Co-transcriptional R-loop formation by Mfd involves topological partitioning of DNA

James R. Portman^{1,2}, Gwendolyn M. Brouwer^{2,3}, Jack Bollins³, Nigel J. Savery^{2,3}, and Terence R. Strick^{*,1,2,4}

¹Institut de Biologie de l'École Normale Supérieure, PSL Université, INSERM, CNRS

²Horizons 2020 Innovative Training Network, DNAREPAIRMAN

³DNA-Protein Interactions Unit, University of Bristol

⁴Equipe Labellisée de la Ligue Nationale Contre le Cancer

*Lead Contact

Abstract

R-loops are nucleic acid hybrids which form when an RNA invades duplex DNA to pair with its template sequence. Although they are implicated in a growing number of gene regulatory processes, their mechanistic origins remain unclear. We here report real-time observations of co-transcriptional R-loop formation at single-molecule resolution and propose a novel mechanism for their formation. We show that the bacterial Mfd protein can simultaneously interact with both elongating RNA polymerase and upstream DNA, tethering the two together and partitioning the DNA into distinct supercoiled domains. A highly negatively supercoiled domain forms in between Mfd and RNA polymerase, and compensatory positive supercoiling appears in front of the RNA polymerase and behind Mfd. The nascent RNA invades the negatively supercoiled domain and forms a stable R-loop that can drive mutagenesis. This mechanism theoretically enables any protein that simultaneously binds an actively translocating RNA polymerase and upstream DNA to stimulate R-loop formation.

Introduction

R-loops are triple-stranded structures in which one strand of a DNA double-helix is annealed to a complementary RNA strand rather than its complementary strand of deoxyribonucleic acid. The superior stability of the resulting RNA-DNA hybrid over that of the DNA-DNA hybrid [59] explains why these R-loops are hot-spots for mutagenesis. However, the mechanism whereby an RNA could successfully invade and anneal to regular B-form DNA remains mysterious. Because the process is thermodynamically favoured by both the stability of the RNA-DNA hybrid as well as the negative DNA supercoiling of the genome, it is likely that limiting nucleation of such invasion is central to preventing large-scale genome-wide R-loop formation. An example of a pathway involving R-loops that leads to genetic instability is Mfd-dependent mutagenesis.

Mfd is historically well-characterised as the transcription-repair coupling factor. Transcription-coupled repair (TCR) deals with bulky lesions in the transcribed strand of actively-transcribed DNA [185, 192]. Mfd binds to RNA polymerase (RNAP) stalled irreversibly at a DNA lesion, and simultaneously binds to DNA immediately upstream RNAP [96]. Mfd uses ATP binding and hydrolysis to bind tightly to and translocate along DNA and against stalled RNAP thus removing RNAP from the DNA, however remarkably RNAP remains thereafter bound to Mfd as part of a slowly-translocating complex [96, 183, 188]. Upon recruitment of the UvrAB complex the Mfd-RNAP complex dissociates from the DNA [18, 189]. Repair proceeds as UvrB recruits UvrC which nicks the DNA either side of the lesion, UvrD unwinds the damage-containing oligonucleotide, and DNA polymerase and ligase complete repair [214, 264].

Interestingly, recent work has shown that Mfd can also induce mutagenesis, helping many bacterial organisms evolve to overcome various stresses [63, 206, 224, 225, 251]. A proposed

mechanism for this begins with Mfd-dependent R-loop formation [63, 224, 251]. Here we show that a mechanism by which R-loops are formed involves interactions between Mfd and elongating RNAP, and Mfd and upstream DNA. This results in the formation of a closed topological domain, between Mfd and RNAP, in which negative supercoiling builds up over time as RNAP continues to translocate. The nascent RNA readily invades this negatively supercoiled domain to form an R-loop. Tellingly, a derivative of Mfd that is incapable of translocation or of displacing RNAP from lesions is competent for R-loop formation confirming that the mechanistic requirements of R-loop formation differ from those of TCR. This represents a potentially universal model of R-loop formation.

Results

Mfd interacts with elongating RNAP to form a tripartite supercoiled domain

To observe the interaction of Mfd with elongating RNAP we used a 4.6 kbp DNA construct containing a 1 kbp transcription cassette. This DNA construct was manipulated using a magnetic trap, allowing us to observe transcription by individual RNAPs under saturating NTP conditions (Fig. 2.1a and Fig. 2.1b) as already described [89]. When Mfd was added we observe transcriptional events interrupted by dramatic bursts of positive DNA supercoiling in which the DNA extension rapidly decreases in a linear fashion and at a rate of ~ 100 nm/s (Fig. 2.1c and Supp. Fig. 2.1). The same behaviour was also observed using a 2 kbp DNA construct bearing a distinct 100 bp transcription cassette (Supp. Fig. 2.1e). No such behaviour was observed when Mfd was omitted (Supp. Fig. 2.2).

These events are consistent with topological partitioning of the DNA into three domains in a manner similar to that proposed by Liu and Wang [122, 123]. Mfd simultaneously binds to RNAP and to upstream DNA in a robust and long-lived manner [18, 188]. The former interaction is mediated by the RNAP-interaction domain of Mfd which binds to a conserved site on RNAP's β subunit [265]; the latter interaction is mediated by the RecG-like domains of Mfd which allow it to translocate along DNA. This dual interaction generates three distinct topological domains in the DNA: one "internal" domain accounting for the DNA in between Mfd and RNAP, and two "external" domains corresponding to the remaining portion of the DNA held between bead and surface in the magnetic trap ("in-front" of RNAP and "behind" Mfd). During transcription DNA is transferred from the external domains, which shrink in length, to the internal domain, which grows in length. Because Mfd acts as an anchor to prevent RNAP from rotating about the DNA helix as it transcribes, the DNA linking number in each of the two domains remains unchanged.

The internal domain therefore grows in length but has a fixed linking number. This means it accumulates a negative supercoil for every turn of DNA transcribed, while the external domains conversely accumulate a positive supercoil. Because the external domains are topologically accessible to the magnetic trap, this positive supercoil pulls the bead towards the microscope surface by ~ 55 nm, the length of a supercoil under the low extending force used here (see Materials and Methods). With transcription occurring at ~ 20 bases/s and transcription of 10.5 bases causing accumulation of one supercoil [89], this predicts a rate of descent of ~ 100 nm/s, consistent with our observations.

According to this model, the bead's rate of descent should depend on the net velocity difference between RNAP and Mfd. In other words, the difference in velocity between Mfd and RNAP governs the rate at which the internal domain grows and becomes negatively supercoiled. This negative supercoiling is balanced by positive supercoiling in the external domains which pulls the bead to the surface. As a result, the velocities of the proteins control the rate of bead descent. We therefore measured this rate for both high and low NTP concentrations (Fig. 2.1d). In both cases Mfd is expected to translocate at maximum velocity (~ 4 bp/s [18]) due to its strong affinity for ATP ($K_M^{ATP} \sim 10$ μ M [189]), and so only the velocity of RNAP will change. As we decrease NTP concentration the distribution of bead descent rates indeed shifts to lower values (Fig. 2.1d). This shows we can modulate the speed of RNAP to alter the rate of supercoil accumulation within the tripartite supercoiled domain.

We note that adding the anti-backtracking factor GreB whilst keeping NTP concentration constant shifted velocity distributions to higher values. This is presumably because GreB reduces the time RNAP spends in backtracked paused states thus increasing the average velocity (Supp. Fig. 2.3). We also note that control experiments carried out with only wild-type Mfd and ATP occasionally result in similar events, however in such cases the bead's average rate of descent is only ~ 4 nt/s

(Supp. Fig. 2.4). These Mfd-only events also occur in experiments with RNAP and NTPs, but are sufficiently rare and kinetically distinct that they did not interfere with tripartite supercoiled domain measurements. However, when these control experiments were performed on negatively supercoiled DNA the frequency of such events increased considerably, interfering with measurements of Mfd-RNAP interactions. For these reasons experiments on negatively supercoiled DNA were next pursued with the MfdRA953 translocation-deficient mutant (see below).

DNA compaction without topological coupling is observed on nicked DNA molecules

In nicked DNA molecules we could still observe a decrease in DNA extension due to the formation of the internal domain (Supp. Fig. 2.5). Indeed, as before, the differential velocity of the two motors reduces the length of DNA in the external domains and increases the length of DNA in the isolated internal domain, causing the DNA extension to decrease. However, in this case no topological coupling takes place to positively supercoil the external domain. As a result, the rate of bead descent is observed to be roughly twenty times lower than on supercoiled molecules, on the order of 5 nm/s. Given DNA structure and mechanics, this corresponds to a differential velocity of approximately 20 bp/s, consistent with expectations [12].

Translocase-deficient MfdRA953 maintains the ability to form tripartite supercoiled domains

To further test the validity of our model, we used a 'translocase-deficient' Mfd mutant, MfdRA953. The arginine to alanine substitution at position 953 perturbs the Rec-G like module's 'TRG' motif, rendering Mfd unable to translocate on DNA [266]. MfdRA953 and ATP alone on either positively or negatively supercoiled DNA shows no events unlike wild-type Mfd (Supp. Fig. 2.6). This mutant

protein can still bind to RNAP and DNA so it should give rise to tripartite supercoiled domain events [266]. Indeed, when transcribing RNAP is exposed to MfdRA953 we observe events similar to those observed with wild-type Mfd (Fig. 2.2a). Although our expectation was that we would observe higher differential velocities with this mutant, we in fact observe a lower velocity differential compared to wild-type (Supp. Fig. 2.7). This could possibly be explained by sliding of MfdRA953 on DNA in the same direction as RNAP.

We also tested MfdRA953-RNAP interactions on DNA that was negatively supercoiled prior to addition of components. In this case, the positive supercoils generated in the external topological domains first annihilate the negative supercoils already present in the DNA, causing an increase in the bead's altitude with respect to the surface. If the tripartite domain lasted long enough the bead altitude crossed the threshold of maximal DNA extension (i.e., the DNA's torsionally relaxed state) before decreasing as net positive supercoils formed in the external domains (Fig. 2.2b). In addition the chances of RNAP backtracking increase with the degree of negative supercoiling immediately upstream of RNAP [111]. Thus, using GreB to reactivate backtracked complexes increases the fraction of events that pass through the maximal extension state (see Table 3) and go on to form net positive supercoils [267] in the external topological domains.

Mfd LR499, deficient in binding to RNAP, lacks the ability to form tripartite supercoiled domains

To verify that Mfd must bind RNAP for this process to take place, we tested an Mfd mutant, MfdLR499, deficient in its RNAP-binding ability [181]. MfdLR499 in the presence of ATP alone could be observed to interact with positively supercoiled DNA in a manner similar to that seen with wild-type Mfd (Supp. Fig. 2.8; top panel). However, no topological partitioning was observed using MfdLR499 with RNAP and NTPs on positively supercoiled (Supp. Fig. 2.8; bottom panel).

We therefore confirm that RNAP-binding is absolutely required for the tripartite supercoiled domain phenomena.

Backtracked RNAP is not a major source of tripartite supercoiled domain formation.

Lastly, we enquired whether elongating or transiently paused RNAP is the initial substrate for Mfd, rather than RNAP durably arrested in a backtracked state. We therefore defined a metric, N_{complete} , corresponding to the number of complete transcription cycles which occur on a given DNA before the next transcription cycle is interrupted by formation of a tripartite supercoiled domain. If backtracked RNAP is the main substrate for Mfd, then addition of GreB should increase $\langle N_{\text{complete}} \rangle$ as GreB lowers the time RNAP spends in the backtracked state. Experiments carried out in the presence or absence of 50 nM GreB (Fig. 2.2c) show that $\langle N_{\text{complete}} \rangle = 4$ in both cases. We conclude that Mfd initially interacts with elongating or transiently paused RNAP rather than backtracked RNAP.

Formation of tripartite supercoiled domains leads to R-loops

We next consider whether the negatively supercoiled internal domain contained between Mfd and RNAP could be a substrate for R-loop formation [75]. In order to observe R-loops, experiments were carried out using MfdRA953 on negatively supercoiled DNA. This is because an R-loop formed in a positively supercoiled background will be ejected when the domain dissolves (e.g. upon transcription termination or dissociation of Mfd), yet will be maintained if formed in a negatively supercoiled background [75]. Indeed, events observed on negatively supercoiled DNA with MfdRA953 essentially always display a new intermediate state (labelled 5) which appears upon dissolution of the tripartite supercoiled domain and which is followed by a

return to the baseline state (Fig. 2.3a and Supp. Fig. 2.9). This intermediate state is extremely long-lived, with a mean lifetime on the order of $\sim 13,000$ seconds (Fig. 2.3b and Supp. Fig. 2.10a). It is also extremely sensitive to RNase HI, which specifically degrades the RNA in RNA-DNA hybrids [268], displaying a mean lifetime of only ~ 15 seconds in the presence of $0.025\text{U}/\mu\text{l}$ RNase HI (Fig. 2.3b and Supp. Fig. 2.10b).

Furthermore, we successfully interrogated the topology of the intermediate state. We awaited formation of the intermediate state and measured the supercoiling response of the DNA while in this state (Fig. 2.3c). This supercoiling response measurement is termed a 'rotation scan' and involves iteratively rotating the magnets and recording the extension of the molecule (Supp. Fig. 2.11). In Fig. 2.3c the DNA rotation scan taken whilst the DNA is in the intermediate state is shown in red and begins by exploring the response of DNA in the intermediate state to negative supercoiling. Indeed, once this scan proceeds into the positively supercoiled regime, the R-loop is ejected [75]. Overlaid in blue is a second scan, taken immediately after the first scan. The difference in the negative portion of the scan is indicative of the presence (and subsequent absence) of an R-loop, as shown in other single-molecule hybridisation experiments [75]. According to these single-molecule hybridisation experiments the size of the shift between the two states corresponds to the size of the R-loop formed. For the molecule in Fig. 2.3c, the shift of ~ 7 turns indicates the formation of a ~ 70 bp R-loop.

Both wild-type and translocation-deficient Mfd support R-loop dependent mutagenesis *in vivo*

In order to probe our model *in vivo* we used a Lac revertant assay [63, 245]. This assay quantifies starvation stress-induced mutagenesis by measuring the emergence of bacteria that have mutated a defective *lacZ* gene and are thus able to utilise lactose as the sole carbon source. The chromosomal *lac* operon of these bacteria is deleted and they contain an F' plasmid bearing a *lacI*-

lacZ fusion which produces non-functional β -galactosidase due to a +1-frameshift mutation in the *lacZ* gene. This strain cannot use lactose as a carbon source unless a point mutation or amplification occurs [247].

First, as seen by others [63], we confirmed that deletion of the gene encoding RNase H promoted mutation in the wild-type background. Knockout of Mfd caused a decrease in mutation relative to wild-type, and in the knockout background further deletion of the gene encoding RNase H had no further effect on mutation rates (Supp. Fig. 2.12). As described previously these results indicate that R-loops are generated by Mfd and are intermediates in the Mfd-dependent mutagenesis pathway [63].

We next asked whether the effect of *mfd* deletion on R-loop-dependent mutagenesis could be complemented by expressing wild-type or mutant Mfd proteins from a plasmid (*pET21a^{Lac}*) under control of the wild-type *mfd* promoter. Fig. 2.4 shows that Δmfd bacteria transformed with vector expressing wild-type Mfd showed a higher mutation frequency than those transformed with an empty vector. Expression of MfdRA953, which lacks DNA translocation activity and is unable to displace stalled RNAP from DNA, also complemented the *mfd* deletion and increased mutation frequency to a level similar to that seen with wild-type Mfd. In contrast, expression of MfdLR499, which is unable to bind stably to RNAP [181], had no effect on the observed mutation frequency in this assay. These results indicate that, in accordance with the single-molecule findings, the RNAP-binding function of Mfd is essential for the phenomenon whilst its translocase function is not required, and suggest that the mechanism for R-loop formation identified in our single-molecule experiments *in vitro* underpins mutagenic R-loop formation *in vivo*.

Discussion

Our work demonstrates robust co-transcriptional formation of an R-loop within the negatively supercoiled portion of a novel tripartite supercoiled domain. We reveal the key mechanistic features enabling Mfd, a factor first identified for its role in transcription-coupled DNA repair [166], to ultimately be involved in R-loop mediated mutagenesis and adaptation to stress. The mechanism is based on already well-defined aspects of Mfd biochemistry, namely binding to RNAP and binding to DNA. This ability of Mfd to simultaneously interact with both elongating RNAP and DNA results in the formation of a tripartite supercoiled domain including massive negative supercoiling in the relatively small domain bookended by RNAP and Mfd.

Given that Mfd binds to DNA immediately upstream of RNAP [96] the linking number of this domain when it is formed is small, on the order of one or two. As a result, transcription of e.g. only 10 bp by RNAP is expected to result in a massive degree of negative supercoiling in this domain (normalized supercoiling $\sigma \sim -0.5$), conditions under which DNA is not expected to remain double-stranded [12]. This negatively supercoiled DNA is a powerful substrate for the formation of R-loops which are expected to be extremely stable in the context of negative chromosomal supercoiling. At the same time the exact process of R-loop annealing remains poorly understood. The genetics work of Wimberly et al. shows that reduced accessibility of the RNA 5' end resulting from ribosome loading indeed reduces the frequency of this phenomenon but does not eliminate it [63]. R-loop formation is presumably a stochastic process which occurs in competition with occupancy of the RNA by ribosomes.

Mfd's role in R-loop formation is separate and distinct from its role in TCR. In TCR, Mfd translocates against irreversibly stalled RNAP to dislodge it from the lesion site. Prior magnetic

trap experiments did not observe R-loop formation because they were probing displacement of irreversibly stalled RNAP. For those assays RNAP was immobilised at +20 from the transcription start site, i.e., immediately after promoter escape, leaving no opportunity for formation of topological domains of any significant size. We also note that to observe this novel process in as many 25% of transcription events, we used an Mfd concentration compatible with the upper range of Mfd concentrations reported *in vivo* (from 100 to 500 nM [269, 270]). Thus, although infrequent this phenomenon is robust, as underscored by its importance *in vivo*.

Interestingly, recent optical trapping experiments have also probed the movement of Mfd along the DNA contour, both alone and in the presence of RNAP [117]. However, the optical trapping experiments implemented could not probe DNA topology or R-loop formation, nor ascertain whether there was a physical link between Mfd and RNAP that persisted during translocation of the two motors. Thus at least three specific classes of translocating Mfd complexes have been identified to date: Mfd alone (Supp. Fig. 2.4 and [117]), Mfd associated with a transcription elongation complex ([89, 219] and this work) and Mfd associated with RNA polymerase that has been dissociated from a stall site and no longer retains RNA [18, 188, 189].

Taken together, the collection of single-molecule data obtained to date converge on the fact that the velocity of Mfd, on the order of ~ 5 bp/s, is essentially the same in all of these contexts (Fig. 2.1d, Supp. Fig. 2.4 and [18, 117, 189]). We note, however, that the processivity of Mfd translocating alone on DNA tends to be short, in the range of a few hundred bp (Supp. Fig. 2.4 and [117]). The processivity of Mfd translocating on DNA in the context of TCR can reach into the thousands of bp if downstream GGR factors UvrA and UvrB are not present to remove stable Mfd-RNAP complexes from DNA before carrying out downstream steps of DNA repair [18, 189]. Mfd processivity could not be measured in the experiments presented here because of the limited size of the transcription unit. The activity of Mfd in each of these three classes likely differs as a

function of the competence for transcription of the targeted RNAP and the availability of the downstream repair components.

Ultimately the topological mechanism we identify stands in contrast to the classical twin-supercoiled domain phenomenon first discussed by Liu and Wang [122] in which RNAP, prevented from rotating about the helix due to ribosomal torsional drag for example, induces positive supercoiling downstream and negative supercoiling upstream of RNAP. Although transcription has since been broadly recognised as a potential driver of negative supercoiling behind of RNAP, the large size of chromosomal topological domains (or TADs) is likely to dilute this effect. Thus, most importantly, in the twin domain model upstream negative supercoiling is dissipated on the kbp scale, whereas in the context of the tripartite supercoiled domain negative supercoiling is highly focused onto a very small region of DNA. Secondly, our model also has important consequences for overall chromosome topology. Whereas positive supercoils form only downstream of RNAP in the twin-domain model, they form both upstream and downstream of RNAP in the triple-domain model, with the potential of greatly perturbing the overall topology of TADs [41].

Our work highlights the docking site of Mfd on RNAP β subunit as a potential target for inhibiting evolution of antimicrobial resistance. RNAP-associated DNA translocases are found across all kingdoms of life [171, 172, 271] and the differential-translocation mechanism identified in this work may drive R-loop formation in species beyond *E. coli*. In humans for instance, the CSB homolog of Mfd is found to be overexpressed in cancers and drives tumour survival in the face of chemotherapeutic agents [241].

Finally, RNAP-binding proteins have not yet been mechanistically linked to R-loop formation. Concomitant binding to RNAP and DNA will initially generate a small topological domain and then focus negative supercoiling in this small domain, resulting in its near-complete unwinding. The relatively small size of this domain likely makes it harder for topoisomerases to locate it via diffusion. These features are likely to enhance successful R-loop formation *in vivo*, possibly explaining how Mfd acquired this pro-mutagenic function. Yet as we have shown, DNA translocation by the secondary protein is not a requirement for this phenomenon. Therefore, we propose that numerous transcription factors, which simultaneously bind RNAP and upstream DNA, could be involved in promoter proximal R-loop formation via this mechanism, with ramifications for genetic regulation and stability.

Materials and Methods

DNA constructs

The DNA construct used in single-molecule experiments is 4.6 kbp long, bearing a “transcription cassette” (T5 N25 promoter, roughly 900 bp of transcript, followed by a tr2 terminator, see Supplementary Information for details and full sequence), and was cloned into home-made vector pET21aΔMCS via XbaI and SbfI sites. The resulting plasmid was extracted from bacterial cultures using a Nucleobond Xtra Maxiprep kit (Macherey-Nagel), digested with XbaI, SbfI-HF, and EcoRI (New England Biolabs), and the resulting 4.6 kbp band purified via agarose gel electrophoresis followed by extraction from the gel (NucleoSpin Gel and PCR Clean-up, Macherey-Nagel). This purified DNA was ligated at the XbaI end to a 1 kbp DNA handle bearing multiple biotin groups on both strands, and ligated at the SbfI end to a 1 kbp DNA handle bearing multiple digoxigenin groups on both strands (see [272] for details). These handles are home-made, using the *T. aquaticus rpoC* gene as template for a PCR reaction using the following primers:

5' GAGAGACCTGCAGGACATCAAGGACGAGGTGTGGG 3'

5' GAGAGATCTAGATCCTCAAAGTTCTTGAAGACCGCCTGG 3'

The PCR reaction is done twice using the Expand High Fidelity PCR system (Roche), once with dUTP-biotin (Roche) in the dNTP mix, and the other with alkali-stable dUTP-digoxigenin (Roche) in the dNTP mix. The biotin-labelled DNA was digested with XbaI whilst the digoxigenin-labelled DNA was digested with SbfI-HF. As in [272], this DNA is tethered first to streptavidin-coated 1- μ m diameter superparamagnetic beads (Dynabeads MyOne C1, Thermofisher), and then

deposited onto a borosilicate glass coverslip (Menzel Gläser) functionalised with anti-digoxigenin (Roche). The DNA-bead mixture incubates on the surface for 10 minutes, after which free magnetic beads and DNA are washed out with buffer.

Proteins

RNAP core, σ^{70} , GreB and wild-type Mfd were purified as previously described [18, 188]. Core RNAP was saturated with a fivefold excess of σ^{70} to ensure maintenance as holoenzyme. MfdRA953 and MfdLR499 were made by site-directed mutagenesis (QuikChange II, Agilent Technologies) on the pAD6 wild-type Mfd expression plasmid using the following primers:

MfdRA953:

5' CACGATCTGGAGATTGCGGGCGGGGTGAACTG 3'

5' CAGTTCACCCGCGCCCGCAATCTCCAGATCGTG 3'

MfdLR499:

5' GGAATGACCACGAGGGAAGCGGGTGGCATTACTGG 3'

5' CCAGTAATGCCACCCGCTTCCCTCGTGGTCATTCC 3'

Successful mutants were verified by sequencing, and expressed and purified exactly as wild-type Mfd. Protein concentrations were determined using the Folin-Lowry assay and spectrophotometry.

Flow cell preparation

Flow cells derivatized with anti-digoxigenin for single-molecule assays were prepared as previously described [272]. Different surfaces were used for experiments with wild-type Mfd and MfdRA953, to avoid cross-contamination.

Reaction conditions

Single-molecule assays were performed at 34°C in reaction buffer containing 40mM K-Hepes pH 8.0, 100mM KCl, 8mM MgCl₂, 0.5mg/ml BSA, 0.1% w/v Tween 20, and 1mM 1,4-Dithiothreitol [189]. Unless stated otherwise, concentrations of components in reactions were as follows: 300pM RNAP holoenzyme (including a total of 1.5 nM of σ^{70}), 500nM Mfd/MfdRA953/MfdLR499, 50nM GreB, 1mM ATP, 100 μ M CTP, 100 μ M UTP, 100 μ M GTP. Collecting the specified number of events, n , typically requires more than three experimental runs involving normally 20-50 DNA molecules simultaneously.

Tethered-DNA assays

Only single molecules of intact, unnicked double-stranded DNA are initially selected for experimentation, as verified by their ability to supercoil and the form of the rotation-extension curve 40. DNA was under a constant extending force of \sim 0.3pN and with +10 turns or -12 turns (for experiments done at positive or negative supercoiling, respectively). For addition of components, DNA was positively supercoiled and the force was increased to \sim 2pN to prevent DNA molecules becoming stuck to the surface under the flow of components. Once the flow equilibrated the force was reduced and the DNA supercoiled to the relevant value. Tripartite supercoiled domain events were chosen based on the following criteria: RNAP initiation must be

observed prior to the event, and the event amplitude must be greater than 200 nm (to avoid recording the theoretically possible rapid successive initiation of two RNAPs).

Lac revertant assay

The Δmfd *E. coli* strain (PJH813) [63] was transformed with a Mfd expressing plasmid (pETMfd2^{Lac-}, pETMfd2LR499^{Lac-}, pETMfd2RA953^{Lac-}) or pET21a^{Lac-}. The pETMfd2^{Lac-} and pET21a^{Lac-} plasmids are derivatives of pETMfd2 [181] (and variants expressing Mfd with the indicated substitutions) and pET21a (Novagen). They were created by excising the *lacI^q* gene from the parent plasmids using restriction enzymes FspAI and EcoNI and re-ligating the plasmid backbone following blunt-ending with Klenow DNA polymerase. The *mfd* gene in these plasmids is expressed from the native *mfd* promoter. All cultures and plates were supplemented with 100 µg/ml ampicillin throughout the experiment to maintain the plasmid. A single colony was grown until prolonged stationary phase (~36-48 hours) at 37 °C in M9 media supplemented with 0.1 % glycerol and 20 µg/ml thiamine. 100 µl of culture was plated onto M9 agar supplemented with 0.1 % lactose and 20 µg/ml thiamine and incubated at 37 °C for 7 days. New Lac⁺ CFUs were counted on each day. Previous research shows that colonies arising before day 2 mutated during liquid growth and thus were not counted [63]. The number of CFUs plated was determined by serial dilutions on LB agar. Control experiments using wild-type (SMR4562), Δmfd (PJH813), $\Delta rnhA$ (PJH683) and $\Delta rnhA\Delta mfd$ (PJH946) *E. coli* strains 13, were carried out as above, however as there was no plasmid incorporation ampicillin was omitted throughout.

Data acquisition and analysis

Data were collected on a homebuilt magnetic trap using the PicoJai software suite (PicoTwist SARL). Real-time bead tracking was carried out at 30 Hz; in time-traces raw bead position is shown in blue and ~ 1 s-average is shown in red. Histograms were fit to a single exponential or Gaussian fit as stated. Number of events (n), and the values of the fits with errors are stated in the figure legends. Data analysis was done using the Xvin software suite (PicoTwist SARL). Velocities were measured manually, by locating the linear portion of an event (linear section of the averaged data, with the same averaging values used in all cases), and calculating the gradient of this line. This value was converted from nm/s to nt/s as $55 \text{ nm} = 1 \text{ turn} = 10.5 \text{ bp}$ [22].

Acknowledgements

We thank Peter Hastings for supplying the various reporter strains used in the mutagenesis experiments, Strick lab members and members of the Bristol DNA-protein interactions Unit for stimulating discussions and Xi Yang for assistance with figure design software. J.R.P. and G.M.B. are supported by graduate scholarships of the Horizons 2020 Innovative Training Network “DNAREPAIRMAN”. J.B. was supported by a BBSRC SWBio DTP scholarship. J.R.P. acknowledges support from the Frontières du Vivant – Programme Bettencourt doctoral program. Research on this topic in the Strick lab was supported by grants from the French ANR (PrTxConf, ANR-17-CE11-0042), the Ligue Nationale Contre le Cancer program for Core Research Teams (“Equipes Labellisées”), the NanoRep grant funded by Paris Sciences et Lettres University as well as core funding from the CNRS, the University of Paris, the Ecole Normale Supérieure, and Paris Sciences and Lettres University.

Author Contributions

J.R.P. and T.R.S. designed single-molecule experiments and J.R.P. performed single-molecule experiments, and J.R.P. and T.R.S. performed single-molecule analysis. G.M.B., J.B. and N.J.S. designed and interpreted Lac revertant assays and G.M.B. and J.B. performed Lac revertant assays. J.R.P., N.J.S, and T.R.S. conceived the project and wrote the manuscript. N.J.S and T.R.S. supervised the project.

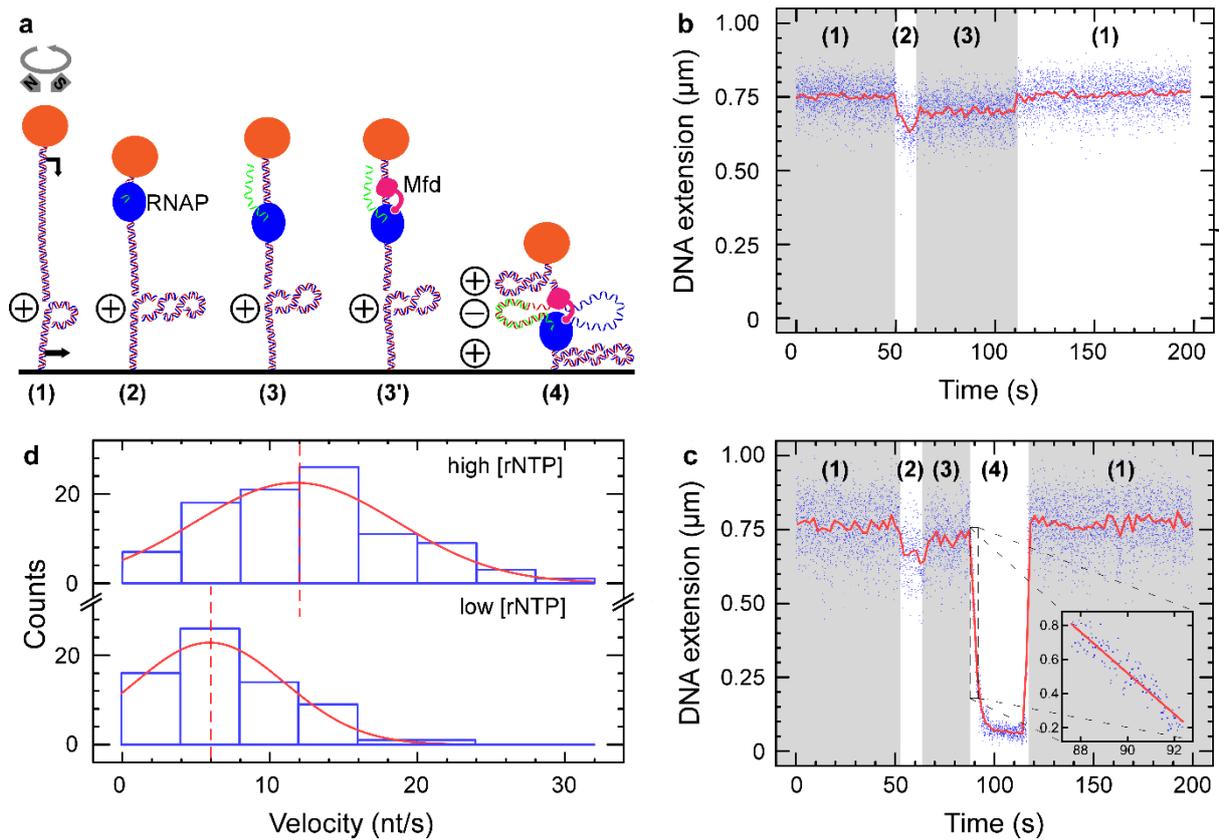


Figure 2.1 - Mfd interacts with elongating RNAP to form a tripartite supercoiled domain

a) Experimental model. The 4.6 kbp DNA bears a promoter (black arrow, top) and terminator (black arrow, bottom) separated by ~900 bp of DNA and is tethered between a magnetically-trapped bead (orange) and glass surface. DNA is positively supercoiled prior to addition of components (RNAP: cyan, Mfd: pink). During synthesis of RNA (green), Mfd can bind to elongating RNAP and DNA to form a tripartite supercoiled domain. **b)** RNAP-only time-trace showing (1) positively supercoiled DNA. (2) As RNAP initiates it scrunches 16 and unwinds ~2 turns of DNA, reducing DNA extension by ~100 nm. (3) Successful conversion to an elongation complex results in an unwound bubble of ~9 bp and so DNA extension is reduced by only ~50 nm compared to baseline. Upon transcription termination DNA returns to (1) baseline state. **c)** RNAP and Mfd time-trace showing (1) positively supercoiled DNA. (2) Transcription initiation and (3) conversion to the elongation state begins as in b), but then (4) Mfd binds to RNAP and DNA to form the tripartite supercoiled domain. Here ongoing elongation by RNAP causes a gain of

positive supercoiling in the external domains, pulling the bead down towards the surface. Finally, loss of Mfd-RNAP interaction or RNAP termination causes the tripartite domain to reunite and dissolve, returning the DNA to (1) baseline state. Blue points – raw data (30 Hz); red line – averaged data (1 s filtering). Inset focuses on the period of rapid bead descent at the beginning of phase (4) with linear fit (red line) used to measure bead velocity. **d)** Bead descent rates in nm/s (see inset in prior panel) are converted to differential velocities of motor proteins (in bp/s) using the DNA supercoiling response [12, 75] and represented as histograms (blue) fit to Gaussian distributions (red). (Top panel) At high NTP concentration ($[UTP] = [GTP] = [CTP] = [ATP] = 1\text{mM}$) we obtain a mean differential velocity of 12 ± 0.8 nt/s (SEM, $n = 96$ events). (Bottom panel) At lower NTP concentration ($[UTP] = [GTP] = [CTP] = [ATP] = 50 \mu\text{M}$) we obtain a mean differential velocity of 6 ± 1.0 nt/s (SEM, $n = 67$ events). Dashed lines indicate mean values.

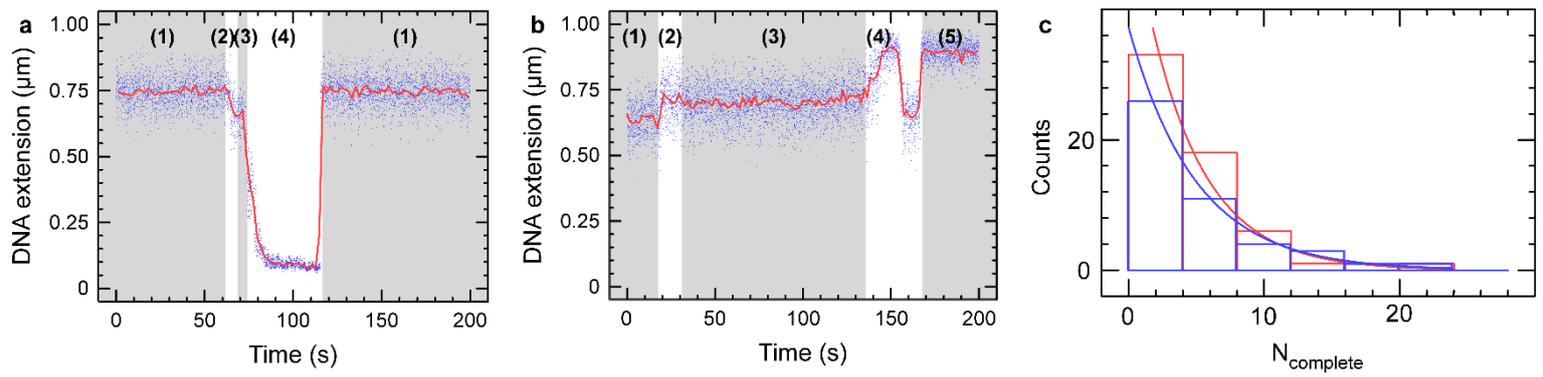


Figure 2.2: Translocase-deficient MfdRA953 maintains the ability to form tripartite supercoiled domain

a) Time-trace showing (1) positively supercoiled DNA, (2) RNAP initiation and (3) elongation. In (4) MfdRA953 binds to elongating RNAP and DNA to form the tripartite supercoiled domain. As RNAP terminates the domains dissolve and annihilate and the DNA returns to (1) baseline extension. **b)** Time-trace showing (1) negatively supercoiled DNA, (2) RNAP initiation and (3) elongation. These signals are in the opposing direction to a) because introduction of positive supercoiling into negatively supercoiled DNA leads to an increase in DNA extension. (4) MfdRA953 binds to elongating RNAP and DNA to form the tripartite supercoiled domain. The gain of positive supercoils first annihilates the negative supercoils (DNA extension increases) before reaching the maximal extension state, after which point there is a net gain of positive supercoils which causes the bead to descend towards the surface. When RNAP terminates or Mfd dissociates from either RNAP or DNA, (5) a new high-extension intermediate state is observed for negatively supercoiled DNA substrate only. **c)** The frequency of tripartite supercoiled domain events is fit to an exponential distribution. N_{complete} is defined for each DNA molecule as the number of uninterrupted RNAP transcription cycles completed on that given DNA molecule before a tripartite supercoiled domain event interrupts a transcription cycle; a low N_{complete} indicates an elevated frequency of transcription events displaying tripartite supercoiled domain formation. With 50 nM GreB (blue) the average of N_{complete} is 4.2 ± 0.6 (SEM, $n = 59$) and without GreB (red) the average of N_{complete} is 5.0 ± 0.9 (SEM, $n = 46$).

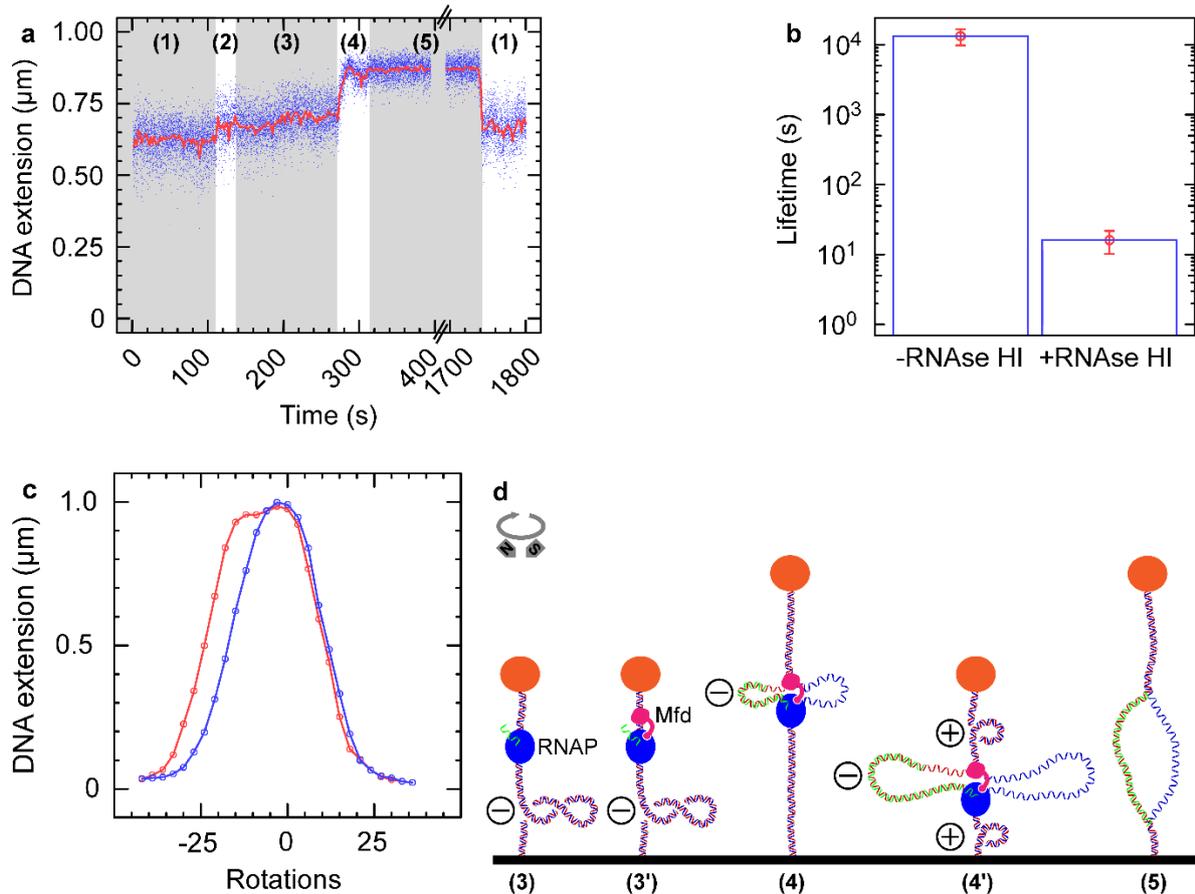


Figure 2.3: Formation of tripartite supercoiled domains leads to R-loops

a) Time-trace showing (1) negatively supercoiled DNA. (2) As RNAP initiates it scrunches and unwinds ~ 2 turns of DNA, increasing DNA extension by ~ 100 nm. After promoter escape, (3) elongation proceeds with an unwound bubble of ~ 9 bp and so DNA extension is increased by only ~ 50 nm compared to baseline. Then MfdRA953 binds to RNAP and DNA to form the tripartite supercoiled domain. Ongoing elongation by RNAP causes a gain of positive supercoiling in the external domains, (4) pulling the bead to the surface after passage through the maximal extension state. There follows (5) a new state in the DNA extension after the tripartite supercoiled domain event ends. This state (5) is the maximal-extension (i.e. torsionally relaxed) state of the DNA for the extending force used, and is consistent with the presence of an R-loop [75]. Whilst the DNA is in the maximal-extension state, it is possible for more RNAPs to initiate transcription, and so once the intermediate state ends the DNA does not always return exactly to a baseline extension, but

can instead return to a near-baseline extension (1') indicative of a new ongoing transcription cycle, as we see here. **b)** The lifetime of the maximal extension "R-loop" state without RNase HI ($n = 29$; average = 13000 ± 3000 s (SEM)) and with $0.025\text{U}/\mu\text{l}$ RNase HI ($n = 17$; average = 16 ± 6 s (SEM)), see Supp. Fig. 2.10 for distributions and fits. **c)** Rotation scan of a DNA molecule captured while in the intermediate state (5). A first scan (in red) is taken immediately following formation of the intermediate state, i.e., whilst the DNA is in the R-loop state (5), from -42 turns through to +33 turns. Once the DNA becomes positively supercoiled the R-loop is ejected, and a second scan (blue) is taken from +33 turns to -42 turns. The difference between the two curves in the negative branch of the curve is indicative of an R-loop [75]. A rotation scan taken before the experiment began is shown in Supp. Fig. 2.11. **d)** Experimental model. (3) RNAP transcribing on negatively supercoiled DNA is (3') bound by Mfd to form the tripartite supercoiled domain. (4) Positive supercoiling builds up in the external domain which annihilates the negative supercoiling, increasing end-to-end extension until a maximal extension state is reached. (4') As net positive supercoils are generated the end-to-end extension again decreases until (5) one of the motors dissociates revealing the R-loop maintained in the DNA.

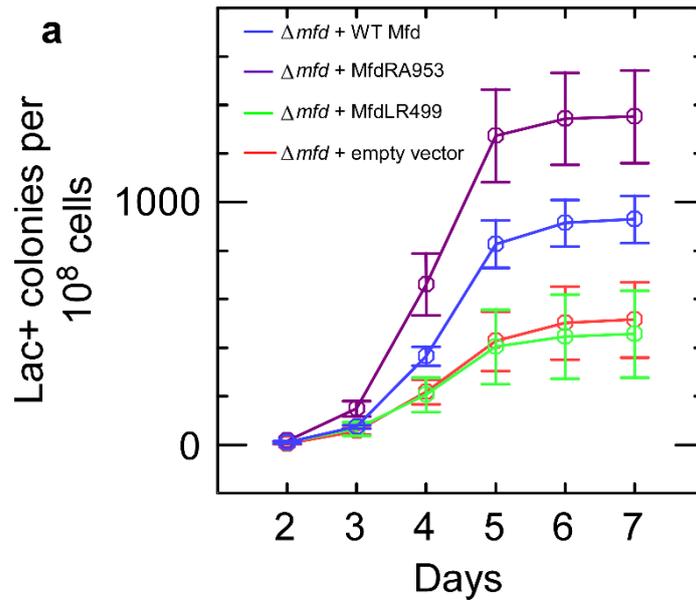


Figure 2.4: Wild-type Mfd and MfdRA953 induce mutations via R-loops in cells under starvation stress

Number of mutants (number of colonies counted on minimal lactose media) observed over time in strains lacking Mfd (PJH813 + $pET21a^{Lac}$; red) or expressing wild-type Mfd (PJH813 + $pET21a^{Lac}$ -Mfd; blue), MfdRA953 (PJH813 + $pET21a^{Lac}$ -MfdRA953; magenta) or MfdLR499 (PJH813 + $pET21a^{Lac}$ -MfdLR499; green). Colonies seen before day 2 arise due to replication-associated mutations whilst in liquid media and so are not counted [245]. Data is an average of three independent cultures and error bars represent SEM.

[GreB] (nM)	Total number of events (n)	Number of events passing through maximal extension state	Percentage of events passing through maximal extension state (%)
0	32	7	22 ± 8 (SEM)
50	73	24	33 ± 7 (SEM)

Table 3: Fraction of events with MfdRA953 initiated on negatively supercoiled DNA and that pass through the maximal extension state, resulting in positively supercoiled DNA in the external domain. GreB re-activates backtracked RNAP to enable more events to proceed further and pass the maximal extension state.

Supplementary Information

DNA Sequences

The transcribed portion of the DNA construct used in the single-molecule experiments consists of a 970 bp transcription cassette containing the T5 phage early N25 promoter, followed by a roughly 900 bp transcript taken from a genic region of *E. coli*, followed by the intrinsic tR2 terminator from phage lambda. The transcription cassette sequence is, from 5' to 3':

```
ggtaccTTCGAGGGAAATCATAAAAAATTTATTGCTTTCAGGAAAATTTTCTGTATAATAGATTCAT
AAATTTGAGCGGCGGTGGCAGCGGAATTCGAGGGCAGTTGCGGTCTGTGGAACCCACCGAGTGAAAGTG
TGGATGCAGCCCTGTTGCCAACTTTACCCGTGGCAATGCCCGCGCAGACGATCTGGTACGCAATAACG
GCTATGCCGCCAACGCCATCCAGCTGCATCAGGATCATATCGTCGGGTCTTTTTTCCGGCTCAGTCATC
GCCCAAGCTGGCGCTATCTGGGCATCGGGGAGGAAGAAGCCCGTGCCTTTTTCCCGCGAGGTTGAAGCG
GCATGGAAAGAGTTTGCCGAGGATGACTGCTGCTGCATTGACGTTGAGCGAAAACGCACGTTTACCAT
GATGATTCGGGAAGGTGTGGCCATGCACGCCTTTAACGGTGAAGTGTTCGTTTCAGGCCACCTGGGATA
CCAGTTCGTCGCGGCTTTTTCCGGACACAGTTCGGGATGGTCAGCCCGAAGCGCATCAGCAACCCGAACA
ATACCGGCGACAGCCGGAAGTCCCGTGCCGGTGTGCAGATTAATGACAGCGGTGCGGCGCTGGGATAT
TACGTCAGCGAGGACGGGTATCCTGGCTGGATGCCGCAGAAATGGACATGGATACCCCGTGAGTTACC
CGGCGGTGCGCGCTCGTTCATTCACGTTTTTTGAACCCGTGGAGGACGGGCAGACTCGCGGTGCAAATG
TGTTTTACAGCGTGATGGAGCAGATGAAGATGCTCGACACGCTGCAGAACACGCAGCTGCAGAGCGCC
ATTGTGAAGGCGATGTATGCCGCCACCATTGAGAGTGAGCTGGATACGCAGTCAGCGATGGATTTTAT
TCTGGGCGCGAACAGTCAGGAGCAGCGGGAAAGGCTGACCAAGCTTACAAAACGCTCTGGTTCGGCCTG
GTACTAGTGTAGCACTCTGTGGGATATCCTGAAAAGCCCCCGGAAGATGCATCTTCCGGGGGCTTTTT
TTTTGGTTCggtacc
```

where the -35 and -10 elements of the N25 promoter are highlighted in cyan, the transcription start site is highlighted in red, and the tR2 terminator is highlighted in magenta. In lowercase are shown the KpnI sites used to insert this transcription cassette into the unique KpnI site of a “backbone” 3.6 kbp segment taken from the *T. aquaticus rpoC* gene. The backbone DNA segment was obtained by PCR amplification of the *rpoC* gene using the following primers containing XbaI and SbfI sites (underlined):

5' GAGAGACCTGCAGGGAAGTCCGCAAGGTCCGCAT 3'

5' GAGAGATCTAGAACAGGTCGTAGCCGTAGCAC 3'

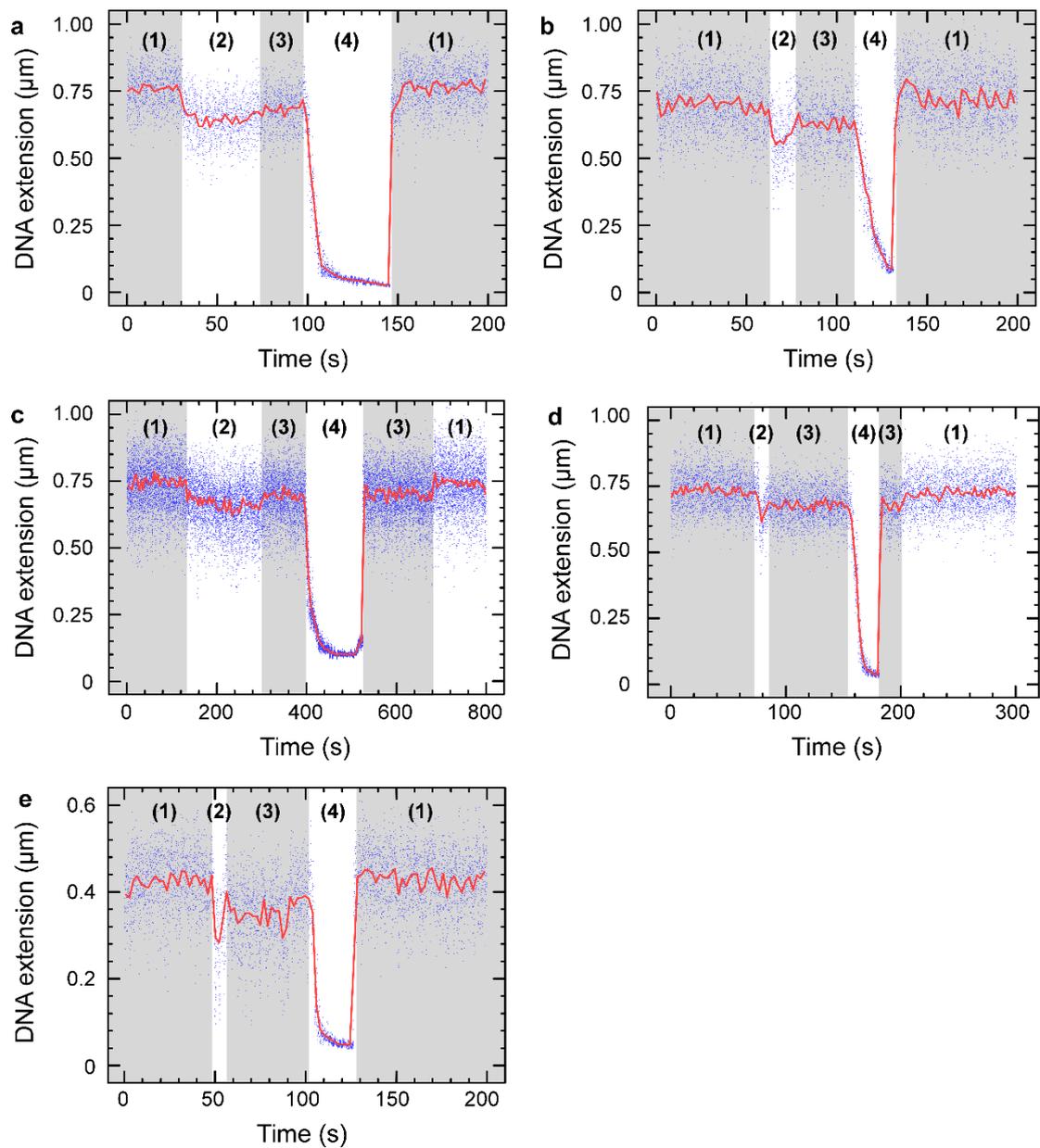
and cloned into the unique XbaI and SbfI sites of homemade vector pET21aΔMCS, itself obtained by PCR amplification of pET21a (Novagen) using the following primers:

5' GAGAGACCTGCAGGTATCGCCGACATCACCGATGGGGAAGATCG 3'

5' GAGAGATCTAGAGGCTGCTGCCACCGCTGAGCAATAACTAGC 3'

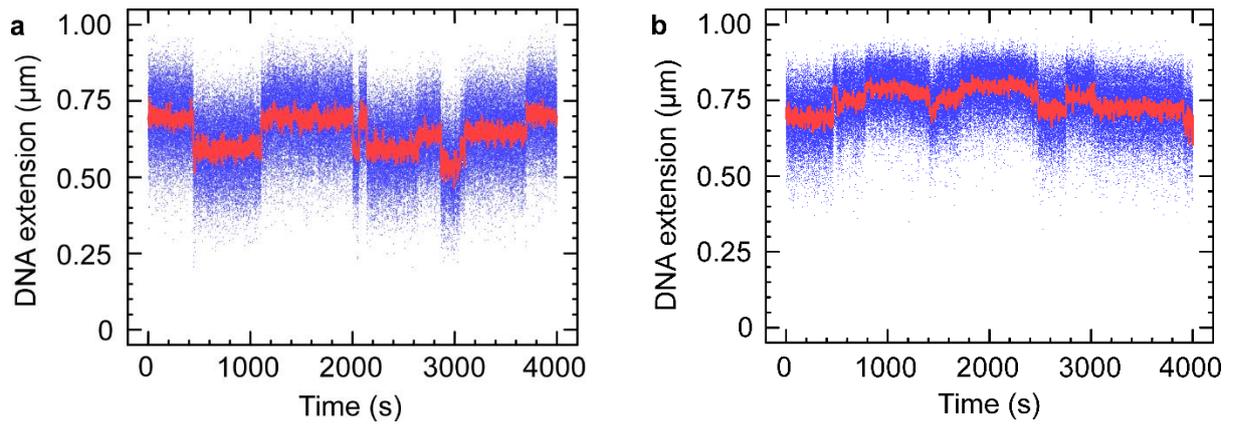
followed by digestion with XbaI and SbfI.

The DNA construct bearing an unrelated 90 bp transcript (Supp. Fig. 2.1e) is described in [189].

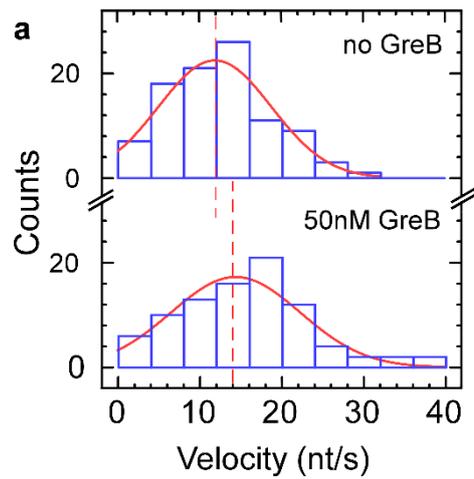


Supplementary Figure 2.1: a-d) Additional time-traces showing Mfd interacts with elongating RNAP to form a tripartite supercoiled domain. RNAP and Mfd time-traces begin with (1) positively supercoiled DNA. RNAP then (2) initiates transcription and upon promoter escape converts into (3) an elongation complex, whereupon (4) Mfd binds to RNAP and DNA to form the tripartite supercoiled domain. Ongoing elongation by RNAP causes a gain of positive supercoiling in the external domains, pulling the bead down towards the surface. Finally, (a, b) RNAP termination causes dissolution of topological domains, directly returning the DNA to (1) the baseline state. Alternatively, (c, d) loss of Mfd-RNAP interaction or Mfd dissociation from DNA

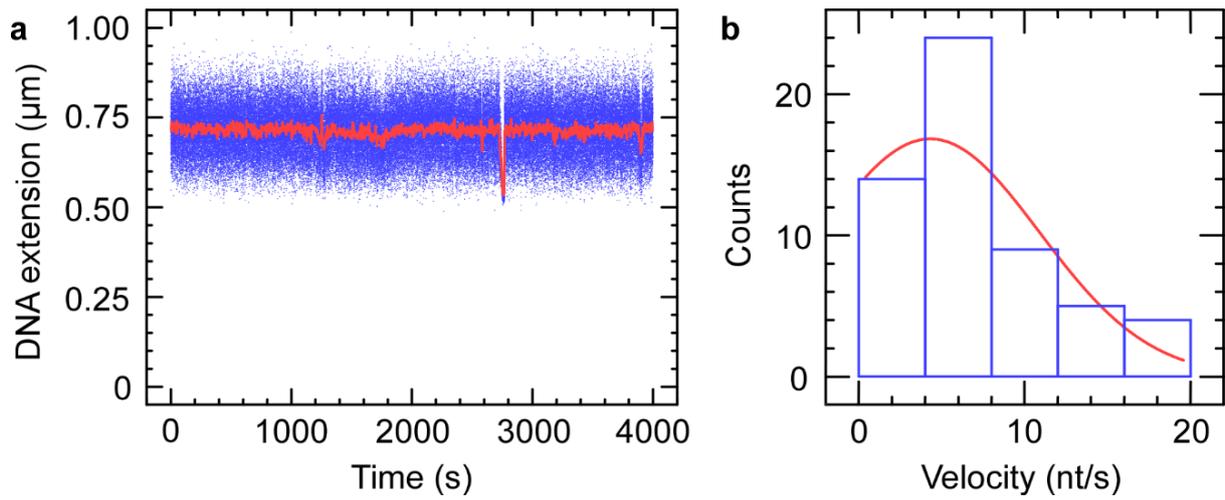
dissolves the topological domains and restores (3) the elongation state which, upon intrinsic transcription termination at tR2, returns the DNA to (1) the baseline state. Blue points – raw data (30 Hz); red line – averaged data (1 s filtering). **e)** Tripartite supercoiled domain formation observed on an unrelated 90 bp transcript sequence with positive supercoiling and the 1-2-3-4 pattern of panels (a, b).



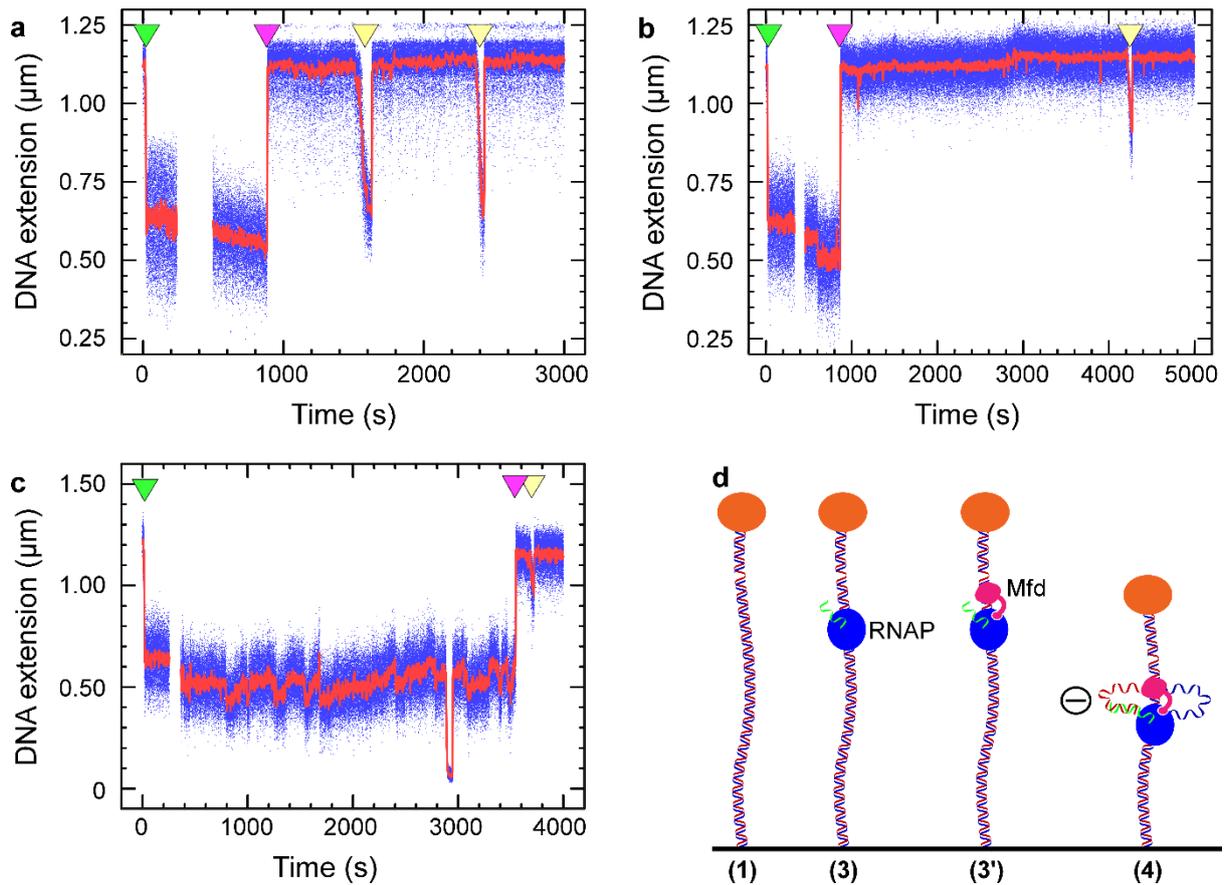
Supplementary Figure 2.2: No Mfd controls. **a)** Time-trace showing iterative cycles of transcription by RNAP on positively supercoiled DNA. Across 33 DNA molecules over more than 18000s for each molecule, no topological domain-like events were observed. **b)** Time-trace showing iterative cycles of transcription by RNAP on negatively supercoiled DNA. Across 36 DNA molecules over more than 20000s for each molecule, no topological domain-like events were observed. On negatively supercoiled DNA transcription termination is less efficient and RNAP molecules can build up on the DNA, so occasionally via many ~ 50 nm increases the DNA would reach the maximal extension state. This was readily distinguishable from events observed with Mfd present, and reason for our criteria of only measuring events with amplitudes greater than 200 nm (see Materials and Methods).



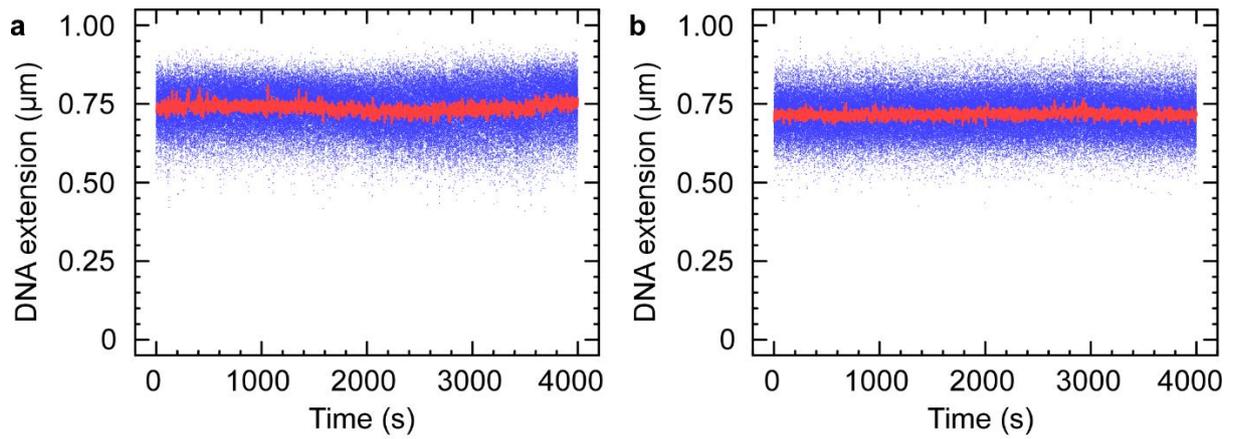
Supplementary Figure 2.3: Velocity distributions obtained with wild-type Mfd and RNAP in the absence or presence of GreB. Both experiments were performed at high NTP concentration ([UTP] = [GTP] = [CTP] = [ATP] = 1mM). Histograms of velocities measured (blue) and Gaussian fits (red): (top panel) in the absence of GreB we obtain a mean differential velocity of 12 ± 0.8 nt/s (SEM, $n = 96$ events); (bottom panel) in the presence of 50nM GreB we obtain a mean differential velocity of 14 ± 1 nt/s (SEM, $n = 88$ events).



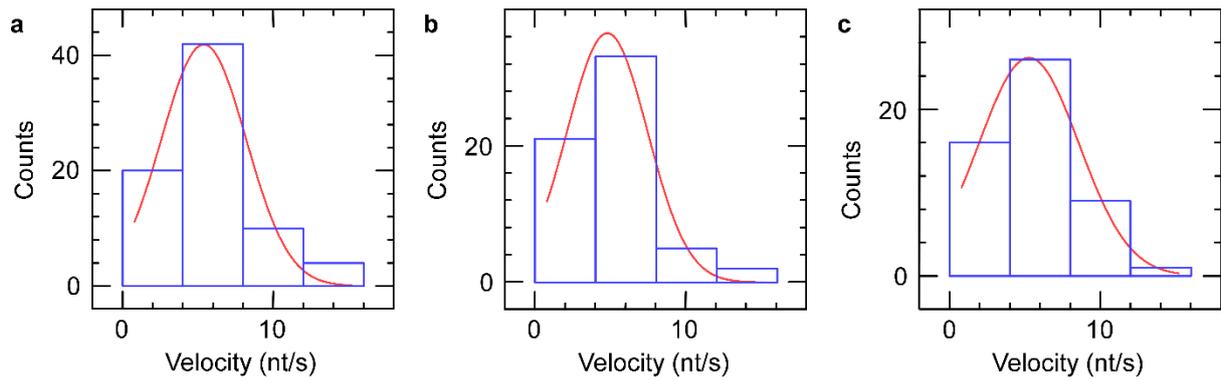
Supplementary Figure 2.4: Wild-type Mfd alone control. **a)** Time-trace of 500nM wild-type Mfd and 1mM ATP on positively supercoiled DNA. Apparent unwinding events that lead to an apparent transient gain of positive supercoils are observed (here seen at ~2800s). **b)** Histogram of velocities derived from these unwinding events and Gaussian fit (red, $n = 56$; average = 4.3 ± 2.6 nt/s (SEM)).



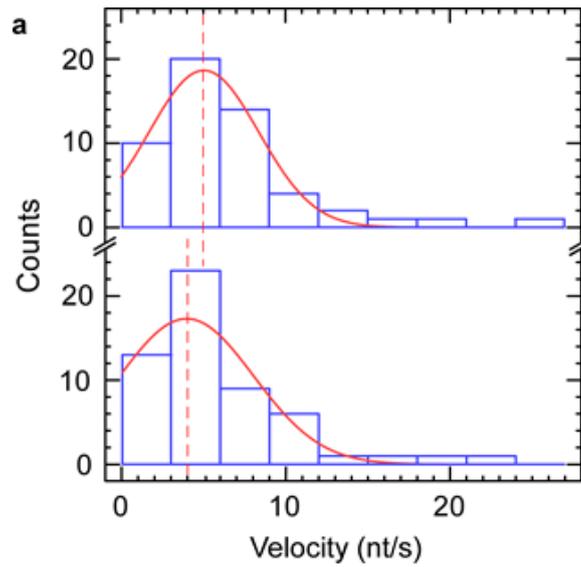
Supplementary Figure 2.5: (a-c) Time-traces showing DNA compaction in nicked DNA molecules. The green arrow shows the point in time at which the DNA molecule is initially supercoiled from a maximal extension state. The break in data points is when RNAP, Mfd and NTPs were injected into the flow cell. The magenta arrow indicates the point at which the molecule becomes nicked, identified by the rapid loss of supercoils and the return to the maximal extension state. The yellow arrows indicate formation of the tripartite domains wherefore the differential velocity of the two motors reduces the length of DNA in the external domains and increases the length of DNA in the internal domain, causing the DNA extension to decrease (as shown in the model presented in **(d)**). Rates of bead descent for the events shown are a) 9 nm/s and 14 nm/s; b) 6 nm/s; and c) 7 nm/s. At the 0.3 pN extending force used, the DNA end-to-end extension factor is about 0.7x contour length, allowing us to convert bead velocities from nm/s into bp/s, with values of 38 bp/s, 59 bp/s, 25 bp/s and 29 bp/s, respectively.



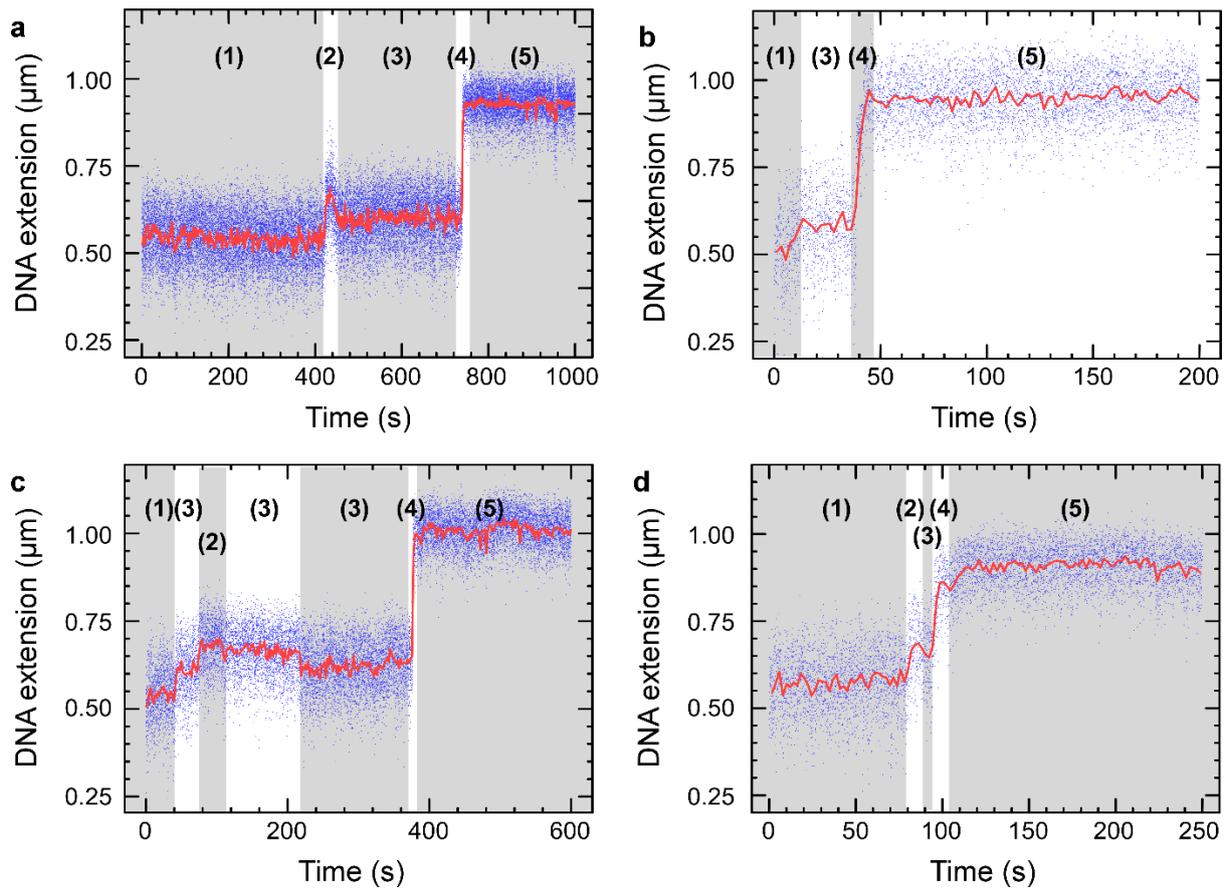
Supplementary Figure 2.6: MfdRA953 only controls. **a)** Time-trace of positively supercoiled DNA with 500nM MfdRA953 and 1mM ATP. Across 20 DNA molecules over more than 20000s, no topological domain-like events were observed. **b)** Time-trace of negatively supercoiled DNA with 500nM MfdRA953 and 1mM ATP. Across 20 DNA molecules over more than 28000s, no topological domain-like events were observed.



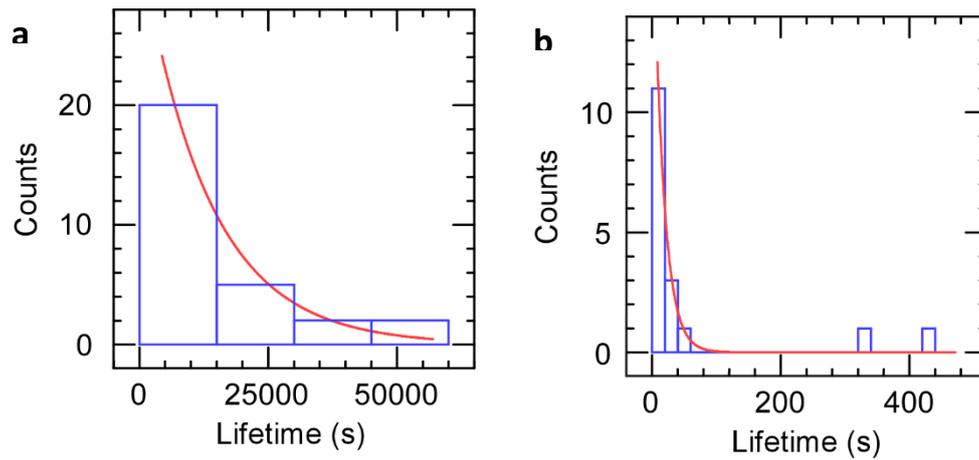
Supplementary Figure 2.7: Velocity distributions for bead descent obtained using MfdRA953 and RNAP. **a)** Histogram of velocities measured on positively supercoiled DNA (blue) and Gaussian fit (red, $n = 61$; average = 4.8 ± 0.4 nt/s (SEM)). **b)** Histogram of velocities measured on negatively supercoiled DNA (blue) and Gaussian fit (red, $n = 38$; average = 5.4 ± 0.5 nt/s (SEM)). **c)** Histogram of velocities measured on negatively supercoiled DNA in the presence of 50nM GreB (blue) and Gaussian fit (red, $n = 52$; average = 5.3 ± 0.6 nt/s (SEM)).



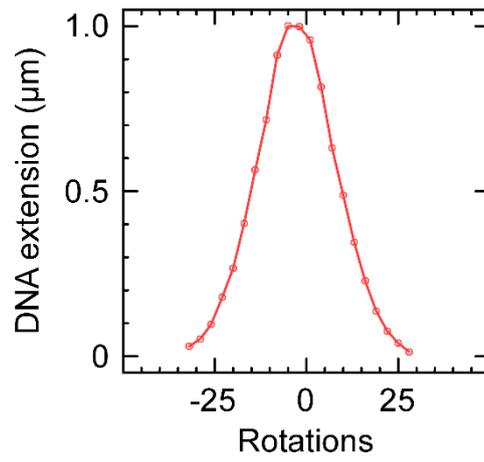
Supplementary Figure 2.8: Velocity distributions for bead descent (blue) obtained using MfdLR499 (top panel) without RNAP and (bottom panel) with RNAP. Top panel – MfdLR499 with 1mM ATP on positively supercoiled DNA ($n = 53$; average = 5.0 ± 0.6 nt/s SEM from Gaussian fit in red). Bottom panel – MfdLR499 with RNAP and 1mM each NTP on positively supercoiled DNA ($n = 55$; average = 4.0 ± 1.1 nt/s SEM from Gaussian fit in red). There is no significant difference between the two distribution averages, indicating no RNAP-dependent events.



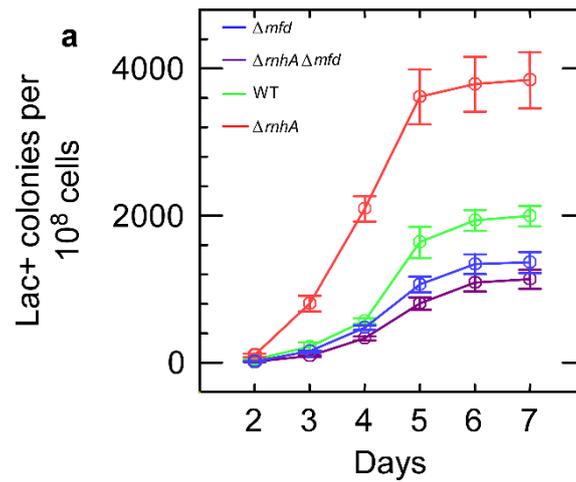
Supplementary Figure 2.9: Additional time-traces showing formation of tripartite supercoiled domains leading to R-loops. Time-traces begin showing (1) negatively supercoiled DNA. As RNAP initiates it (2) scrunches and unwinds ~ 2 turns of DNA, increasing DNA extension by ~ 100 nm compared to baseline 16. Successful promoter escape leads to formation of (3) the elongation complex characterized by a ~ 9 bp transcription bubble, and so DNA extension is increased by only ~ 50 nm compared to baseline. Then (4) MfdRA953 binds to RNAP and DNA to form the tripartite supercoiled domain. Ongoing elongation by RNAP causes a gain of positive supercoiling in the external domains, annihilating initial negative supercoiling and allowing the bead extension to increase to the maximal extension state. The long-lived state (5) is the maximal-extension (i.e., torsionally relaxed) state of the DNA for the extending force used, and is consistent with the presence of an R-loop which can be probed as in Fig. 2.3c.



Supplementary Figure 2.10: Histograms of R-loop lifetimes. **a)** Histogram of R-loop lifetimes measured in the absence of RNase HI (blue) and exponential fit (red, $n = 29$; average = $13000\text{s} \pm 3000$ (SEM)). 50nM GreB was present in all experiments. **b)** Histogram of R-loop lifetimes measured in the presence of $0.025\text{U}/\mu\text{l}$ RNase HI (blue) and exponential fit (red, $n = 17$; average = 16 ± 6 s (SEM)).



Supplementary Figure 2.11: Rotation vs extension curve for the DNA molecule shown in Fig. 2.3c prior to the addition of proteins.



Supplementary Figure 2.12: Lac revertant assay control. Cells lacking RNase HI (PJH683; red) show about double the mutagenesis compared to cells with RNase HI (SMR4562; green). Cells lacking both Mfd and RNase HI (PJH946; magenta) show similar levels of mutagenesis as cells lacking Mfd (PJH813 + *pET21a^{Lac}*; blue), indicating R-loops are involved in the Mfd-dependent mutagenesis pathway. Colonies seen before day 2 arise due to replication-associated mutations whilst in liquid media and so are not counted [245]. Data is an average of three independent cultures and error bars represent SEM.

Appendix I: Topological partitioning observed with a transcription factor

We next searched for evidence for our hypothesis that the tripartite supercoiled domain formation was a universal phenomenon. We identified the catabolite activator protein (CAP; also referred to as the cAMP receptor protein, CRP) as a good candidate to test. CAP is a well-studied homo-dimeric transcription factor that, in the presence of cAMP, stimulates formation of the RNAP-promoter closed complex and subsequent isomerisation into the open complex. The promoters that CAP acts at bear the consensus sequence isomerisation into the open complex. The promoters that CAP acts at bear the consensus sequence 5'AATGTGATCTAGATCACATTT3' which can be located upstream of the core promoter (class I CAP-dependent promoters) or overlapping the core promoter (class II CAP-dependent promoters, e.g., the CC promoter) [273, 274]. The mutant CAPKN52 has a roughly 50 times stronger interaction with RNAP, making it an attractive protein to first test.

The DNA construct used was derived from that used earlier (Section II Chapter I Supplementary Information). The ~5kb DNA bears a CC promoter and tR2 terminator separated by ~900 bp of transcription cassette. The sequence is as follows:

```
gagctcgggtacccgggatcaggtaaatgtgatgtacatcacatggatccccctcactcctgccataattctgatagaattcGAGGGCA
GTTGCGGTTCGTGGAACCCACCGAGTGAAAGTGTGGATGCAGCCCTGTTGCCCACTTTACCCGTGGCA
ATGCCC GCGCAGACGATCTGGTACGCAATAACGGCTATGCCGCCAACGCCATCCAGCTGCATCAGGATC
ATATCGTCCGGTCTTTTTTCCGGCTCAGTCATCGCCCAAGCTGGCGCTATCTGGGCATCGGGGAGGAA
GAAGCCCGTGCCTTTTCCCGCGAGGTTGAAGCGGCATGGAAAGAGTTTGCCGAGGATGACTGCTGCTG
CATTGACGTTGAGCGAAAACGCACGTTTACCATGATGATTCGGGAAGGTGTGGCCATGCACGCCTTTA
ACGGTGAAGTGTTCGTTCCAGCCACCTGGGATACCAGTTCGTCGCGGCTTTTCCGGACACAGTTCCGGA
TGGTCAGCCCGAAGCGCATCAGCAACCCGAACAATACCGGCGACAGCCGGAAGTGCCTGCGGGTGTG
CAGATTAATGACAGCGGTGCGGCGCTGGGATATTACGTCAGCGAGGACGGGTATCCTGGCTGGATGCC
GCAGAAATGGACATGGATACCCCGTGAGTTACCCGGCGGTCGCGCCTCGTTCATTCACGTTTTTGAAC
```

CCGTGGAGGACGGGCAGACTCGCGGTGCAAATGTGTTTTACAGCGTGATGGAGCAGATGAAGATGCTC
 GACACGCTGCAGAACACGCAGCTGCAGAGCGCCATTGTGAAGGCGATGTATGCCGCCACCATTGAGAG
 TGAGCTGGATACGCAGTCAGCGATGGATTTTATTCTGGGCGGAACAGTCAGGAGCAGCGGGAAAGGC
 TGACCAAGCTTACAAAACGCTCTGGTCGGCCTGGTACTAGTGTAGCACTCTGTGGGATATCCTGgaaag
 cccccggaagatgcatcttccgggggctttttttt

The CAPKN52 mutant was purified exactly as described in [275] and stored as single-use 5 μ l aliquots at -80°C.

As before, DNA molecules were verified and negatively supercoiled prior to the addition of components (300pM RNAP holoenzyme, 50nM CAPKN52, 1mM ATP, 1mM GTP, 1mM CTP, 1mM UTP, 400 μ M cAMP). As Figure 2.13 shows, we observed events consistent with the formation of tripartite supercoiled domain formation. In contrast to Mfd, these events always occur at the very start of transcription because CAP is already bound to RNAP once transcription begins. These events seem to be shorter-lived and we have not yet attempted to observe R-loops via the rotation-extension scan analysis, both due to the low numbers of events recorded to date. This is clear evidence that tripartite supercoiled domain phenomena are conserved between transcription factors. Following this, the resultant R-loop's presence and lifetime should be verified and measured.

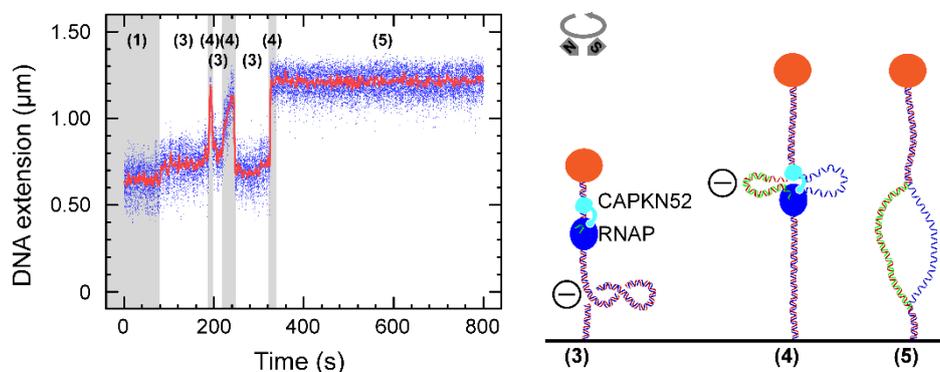


Figure 2.13: Time-trace (left) showing three tripartite supercoiled domain events on CC promoter-bearing DNA with CAPKN52 and experimental model (right).

Chapter II: On the stability of stalled RNA polymerase and its removal by RapA

RapA was discovered in the 1960s because it consistently co-purified with RNAP [276]. Ensemble methods have since extended our understanding of the structure and binding partners of RapA. Certainly, RapA is involved in the transcription cycle and probably relies on nascent RNA-binding [104, 277, 278]. Yet, a comprehension of what RapA does *in vivo* and how it does it remains to be found. Thus, we sought to observe RNAP-RapA interactions from a single-molecule perspective with the goal of understanding 1) at which stage of transcription RapA acts; and 2) the mechanism of this action. In doing so, we also probed a second key question: what is the inherent stability of a stalled TEC absent accessory factors? Whilst a lot is known about how various accessory factors or DNA sequences (e.g., hairpins) affect RNAP stability, little is known about the stability of a lone TEC on DNA and how it is affected by supercoiling and torque. My research on the stability of the TEC and how it is affected by RapA is presented in the form of a manuscript below.

On the stability of stalled RNA polymerase and its removal by RapA

James R. Portman^{1,2}, Mohammad Z. Qayyum³, Katsuhiko S. Murakami³ and Terence R. Strick^{*1,2,4}

¹Institut de Biologie de l'Ecole Normale Supérieure, PSL Université, INSERM, CNRS

²Horizons 2020 Innovative Training Network, DNAREPAIRMAN

³Department of Biochemistry and Molecular Biology, The Pennsylvania State University

⁴Equipe Labellisée de la Ligue Nationale Contre le Cancer

*Lead Contact

Abstract

Stalling of the transcription elongation complex formed by DNA, RNA polymerase (RNAP), and RNA presents a serious obstacle to concurrent processes due to the extremely high stability of the DNA-bound polymerase. RapA, known to remove RNAP from DNA in an ATP-dependent fashion, was identified over 50 years ago as an abundant binding partner of RNAP, however its mechanism of action remains unknown. Here we use single-molecule magnetic trapping assays to characterize RapA activity and begin to specify its mechanism of action. We first show that stalled RNAP resides on DNA for times on the order of 10^6 seconds, and that positive supercoiling of DNA reduces this lifetime. We further show that the RapA protein stimulates dissociation of stalled RNAP from positively supercoiled DNA but not negatively supercoiled DNA. We observe that RapA activity is torque-sensitive, is inhibited by GreB, and depends on RNA length. We propose that stalled RNAP is dislodged from DNA by RapA via backtracking in a torque- and supercoiling-dependent manner, and we therefore suggest that *in vivo* RapA is responsible for resolving conflicts between polymerase molecular motors.

Introduction

Because RNAs cannot be truncated, RNA polymerase is characterized by both very elevated processivity and a very stable association to DNA in the elongation phase [279]. This is made possible in part by the closure of the RNAP's "crab-claw" or "clamp" domains about the DNA. As a result, RNAP stalled during elongation does not easily dissociate from DNA but instead forms possibly the most stable, formidable obstacle to other proteins engaged with the DNA, such as DNA polymerases, repair proteins, and recombinases. As a result, no one has yet robustly and quantitatively characterised the dissociation of a stalled RNAP from DNA absent accessory factors. The presence of numerous factors whose main purpose is to remove or remodel stalled RNAP shows us firstly that these complexes are indeed very stable and secondly that there is a pressing need to remove them from DNA in a timely manner, to repair the stall cause and/or to enable other motors access to DNA [96, 189, 280].

RapA was the first bacterial homolog found to be of the SWI/SNF family of eukaryotic proteins that are generally associated with chromatin remodelling [106]. The RapA protein was identified because it consistently co-purifies with RNAP preparations from *E. coli* [105, 106]. RapA is a seven-domain 110-kDa protein with RNAP-binding, nucleic acid-binding and ATPase activities [104, 108-110, 277]. RapA binds to core RNAP near the RNA exit channel and this binding stimulates RapA ATPase activity [104, 110]. It has been shown that RapA activates transcription *in vitro*, but RapA does not affect promoter binding, promoter escape, elongation nor termination [110]. It has been speculated that RapA activates transcription by recycling RNAP that is in an off-pathway state [110]. Further studies showed the existence of a transient RapA-RNA interaction, leading to the model that RapA may remodel the RNA in the off-pathway post-termination complex which ultimately leads to complex dissociation [277, 278]. Little is understood on the

precise mechanism of RapA action, and furthermore the presence and importance of such atypical RNAP states *in vivo* is not known.

In this study we use a single-molecule approach to interrogate the mechanism by which stalled RNAP is removed from DNA. We show that RNAP stalled on positively supercoiled DNA is removed when very high torque is applied, and that RapA accelerates this process. It appears that the torque-dependence of the kinetics of RNAP removal are similar when RapA is absent or present, indicating that there is a common underlying mechanism. This removal occurs only on DNA which is positively supercoiled and never on DNA that is negatively supercoiled. We propose RapA removes RNAP by backtracking because firstly RapA is functionally inhibited by the anti-backtracking factor GreB, and secondly it takes RapA progressively longer to remove RNAP stalled progressively further downstream from the promoter. Our data leads us to the conclusion that the biological role of RapA has evaded detection primarily because it acts on RNAP, on meaningful timescales, in specific transient topological environments.

Results

High torque removes stalled RNAP from positively supercoiled DNA

To probe the stability of stalled RNAP on DNA, we used a single-molecule assay in which RNAP is stalled at position +20 from the transcription start site (+1) due to nucleotide starvation [188, 189]. Because the footprint of elongating RNAP is approximately 25 bases [281, 282], stalling RNAP at +20 sterically prohibits a second RNAP from binding to the promoter, thus we observe a single RNAP bound to each DNA molecule which, in the absence of accessory factors and at a relatively low force (~ 0.3 pN), remains bound indefinitely for both positively and negatively supercoiled DNA (Fig. 3.1a and Supp. Fig. 3.1) [188]. For positively supercoiled DNA, we found that by increasing the torque applied to the molecule, we could reliably observe RNAP dissociation on the scale of tens of minutes (Fig. 3.1b). We did not observe dissociation of RNAP from negatively supercoiled DNA over tens of hours, and because elevated negative torque denatures DNA, we could not probe the effect of torque in these conditions (Supp. Fig. 3.1).

To collect data, we therefore cycled positively supercoiled DNA between a low force state, in which RNAP can bind to and stall on DNA, and a high force state, in which RNAP is forced to dissociate (Fig. 3.1b). Because elevated positive torque inhibits transcription initiation [283], RNAP cannot re-initiate transcription until we return to the low force state again. This 'force-cycling' assay allowed us to repeatedly measure the time that the stalled RNAP-DNA elongation complex (RDe) resides on DNA before dissociation (denoted as the RDe lifetime). RDe lifetimes follow single-exponential distributions (Supp. Fig. 3.2) from which a mean lifetime $\langle t_{\text{RDe}} \rangle$ could be obtained by fitting. By repeating these measurements at three different torques we obtain a plot showing a linear relation between $\ln(k_{\text{RDe}})$ and the applied torque (Fig. 3.2c), where $k_{\text{RDe}} =$

$1/\langle t_{\text{RDe}} \rangle$. Extrapolating this line to zero torque indicates that stalled RNAP remains bound to torsionally-relaxed DNA for 10^6 seconds (~ 12 days).

The slope of this line reflects the torque-sensitivity of RNAP dissociation as described by a simple Boltzmann law accounting for the system's mechanical energy:

$$k_{\text{RDe}}(\Gamma) = k_{\text{RDe}}(0)\exp(\Gamma\theta/k_{\text{B}}T)$$

where $k(0)$ is the zero-torque dissociation rate obtained above, Γ is the applied torque (see materials and methods), θ is the extent of DNA rewinding in the transcription bubble that corresponds to the transition state intermediate to RNAP dissociation, and k_{B} is Boltzmann's constant and T the temperature in degrees Kelvin. From this analysis we obtain $\theta \sim 30^\circ$, corresponding to rewinding of approximately one base-pair. This suggests that the transcription bubble, although extremely stable against torque, is nevertheless "brittle" in the sense that a small amount of rewinding can upset the entire structure.

RapA removes stalled RNAP in a torque-sensitive manner

To see how RapA affects RDe lifetime, we added RapA to the 'force-cycling' assay described above. We observed RapA to drastically shorten RDe lifetimes (Fig. 3.2a), allowing us to observe dissociation of stalled elongation within a reasonable amount of time even at low levels of positive torque. In fact, it was possible to observe iterative RNAP binding and dissociation by RapA at a constant low force (0.3pN and 0.7pN, see Fig. 2b; above ~ 0.9 pN RNAP could not reliably initiate transcription). In this low force 'recycling assay' RNAP can initiate transcription and stall, then

only after dissociation by RapA is the promoter free to be bound by a subsequent RNAP. RDe lifetimes were measured across five torque values and all followed single-exponential distributions (Supp. Fig. 3.3). The average of each lifetime distribution was obtained by fitting and plotted, along with RNAP-alone data, as $\ln(k_{RDe})$ vs torque (Fig. 3.2c). Clearly RapA stimulates RNAP dissociation from positively supercoiled DNA and this process occurs faster under higher torque. Interestingly, we observe very similar gradients for the line fits of Fig. 3.2c, giving values of $\theta = 0.15$ (30 degrees) in the absence of RapA and $\Gamma\theta/k_B T = 0.13$ (27 degrees) in its presence, indicating the underlying process by which RNAP dissociates from DNA is the same. Overall, RapA lowers the energy barrier between the RNAP-bound and RNAP-free states by ~ 3.4 kcal/mol, accelerating the dissociation of RNAP from DNA.

RapA removes RNAP stalled on positively, but not negatively, supercoiled DNA

Following the observation that RapA can remove stalled RNAP from DNA at low torques, we next wanted to explore the relationship between RapA activity and DNA supercoiling. The initial ‘force-cycling’ assay had to be performed on positively supercoiled DNA as negatively supercoiled DNA forms denaturation bubbles at forces above ~ 0.3 pN that would interfere with our assays [75]. However, the low force ‘recycling assay’ can be carried out on either positively or negatively supercoiled DNA. As previously mentioned, in the absence of RapA RDe lifetime on either positively or negatively supercoiled DNA at low force (~ 0.3 pN) was consistently longer than the experimental time-frame (i.e., days). As Figure 3.2b and Figure 3.3a (top panel) show, RapA drastically shortens the RDe lifetime on positively supercoiled DNA. Strikingly, when the DNA was negatively supercoiled, RapA had no effect on RDe lifetime (Fig. 3.3a bottom panel). We were only able to observe RapA remove stalled RNAP from positively supercoiled DNA and never on negatively supercoiled DNA (Supp. Fig. 3.1).

Kinetics of RNAP removal by RapA

To interrogate the kinetics of RapA action as seen by analysing the RDe lifetime we used positively supercoiled DNA in the low force ‘recycling assay’ (Fig. 3.3a top panel). We measured RDe lifetime as a function of RapA (Fig. 3.3b) and ATP concentrations (Fig. 3.3c). At saturating RapA and ATP concentration, a rather slow V_{\max} of 0.066s^{-1} was obtained for dissociation of RNAP from the substrate previously mentioned. The K_m of RapA for RNAP is 7.1nM and the K_m of the RapA-RNAP complex for ATP is $165\mu\text{M}$, generally in agreement with previously published bulk biochemical data [106]. RapA and ATP alone (no RNAP) on positively supercoiled DNA gave rise to rare and very short-lived increases in DNA extension (Supp. Fig. 3.4) indicating potentially weak over-winding or right-handed wrapping of DNA.

RapA is functionally inhibited by GreB

In order to understand the mechanism by which RapA removes RNAP from DNA, we sought to observe how the addition of the anti-backtracking factor GreB may affect RapA activity. If RapA removes RNAP by backtracking, then GreB should counteract that because it can reactivate backtracked RNAP, enabling it to elongate down to the stall site again [88, 284]. We added 50nM GreB to the recycling assay and observed that in this condition more than 50% of RNAP remained bound indefinitely, as if RapA activity were lost. Based on previous experiments in this study, we expect to have almost 100% of RNAP removed by RapA (Table 4), however in the presence of GreB, more than half of the RNAP remain bound for the duration of the experiment. For the RNAP molecules that are removed by Rap A in the presence of GreB, the $\langle t_{\text{RDe}} \rangle$ is roughly doubled (Fig. 3.4a and Supp. Fig. 3.5). As structural data shows us these proteins do not share or overlap binding sites on RNAP [104, 285] this is strong evidence to indicate GreB functionally opposes the action of RapA.

RapA takes longer to remove RNAPs stalled further away from the promoter

To further test the backtracking model, we designed DNA constructs with stall sites at different distances from the promoter. If RapA removes RNAP by backtracking, then RDe lifetime should increase as we increase the distance RNAP must backtrack to be dissociated. To test this, we built two additional DNA constructs, one with a stall site at +36 from the TSS and one with a stall site at +83 from the TSS. The latter DNA construct uses nucleotide starvation to stall the RNAP at +15, following which we wash out all RNAP in solution and add in all four ribonucleotides (and RapA) which allows RNAP to rapidly transcribe up to a stall-inducing cyclopyrimidine dimer (CPD) at +83. It has been shown previously that a CPD also irreversibly stalls RNAP indefinitely [188] and we have shown that RapA action on RNAP is independent of stall cause (Supp. Fig. 3.6). We compared $\langle t_{RDe} \rangle$ with each DNA construct and observe clearly that RapA takes longer to remove RNAP stalled further away from the promoter (Fig. 3.3b). This data further supports a backtracking model.

Discussion

Many studies have shown how accessory factors may activate or remodel RNAP at various stages of transcription [166, 195, 286-288]. In particular, many factors can act on RNAP that is paused or stalled in elongation phase, however little is known about the intrinsic stability of the RNAP-RNA-DNA complex alone. Indeed, we have never observed stalled RNAP to dissociate from DNA at the low forces typically used for single-molecule transcription assays. However, at higher force, thus higher torque, we have observed the dissociation of stalled RNAP from DNA in the absence of accessory factors.

Furthermore, we have shown that RapA accelerates the dissociation of stalled RNAP from DNA. Two lines of evidence point towards a mechanism involving backtracking. Firstly, RapA is functionally inhibited by the anti-backtracking factor GreB; and secondly, increasing the stall site distance from the promoter, thus increasing the length of the RNA that must be threaded back through the RNAP, increases the time taken for RapA to remove RNAP. We propose that RapA binds to stalled RNAP and catalyses backtracking until the RNA is released via its 5' end from the RNAP-RNA-DNA complex, leaving on the DNA a core RNAP-RapA complex which is intrinsically unstable and dissociates rapidly.

We note that $\langle t_{\text{RD}e} \rangle$ increases faster than linearly with increasing stall site distance from the promoter. If RNAP were to be removed via processive backtracking by a single RapA, i.e. via a succession of irreversible steps, one would expect $\langle t_{\text{RD}e} \rangle$ to increase linearly with distance. The observed non-linearity could be explained by a model wherein RapA only backtracks stalled RNAP by one base-pair, allowing another slower process to complete the backtracking. For

example, backtracked RNAP is known to undergo a random walk, and this could give rise to non-linear time-dependencies.

Importantly, our observations of backtracking by RapA occurs only on DNA that is positively supercoiled. We have never observed RapA removing RNAP from negatively supercoiled DNA. Furthermore, within the positively supercoiled regime, increasing the torque resulted in faster removal of RNAP by RapA. In contrast, the bacterial nucleoid is generally negatively supercoiled and positive supercoiling is generated only in specific contexts such as in front of the replication fork or within highly-transcribed genes, and particularly when these contexts are coincident. For instance, RapA could conceivably act to remove RNAPs that come to be stalled in front of the replisome, or perhaps even because of the positive torque they experience if located just in front of the replisome. Indeed DNA is known to be positively supercoiled in front of the replication fork [289], and positive torque of ~ 18 pN.nm in front of RNAP is known to stall RNAP [267]. In these conditions that same positive torque could allow RapA to rapidly backtrack RNAP by one base-pair in a manner specific to these loci. This will prevent RNAP from attempting to carry out nucleotide addition, allowing the replisome to win the conflict and chase the RNAP off the 5' end of its RNA in just a few seconds.

Additionally, given what is known about how salt concentration affects the mechanical properties of DNA, it is possible that our measurements of supercoiling and torque-sensitivity represent more generally a dependence of RapA on certain DNA conformations that are also observed under high salt concentrations. Indeed, recent observations show that RNAP dissociates from DNA due to osmotic shock [290] and it could be this transient and specific topological cellular scenario in which RapA also acts.

Materials and Methods

DNA constructs

The 20_stall DNA construct is part of a home-made vector, pET21ΔMCS (referred to as pET21ΔMCS_20stall). A PCR is performed on this vector using PCR primers bearing one Xba and one Sbf restriction site, enabling subsequent ligation of the 2kb PCR product into a pUC18 vector. This vector was then digested with XbaI and SbfI, and the 2kb band gel purified. The sequence of the transcription cassette is as follows (promoter elements underlined and stalling site in bold):

5'TTGCTTTCAGGAAAATTTTCTGTATAATAAGCTTATAAATTTGAGAGAGGAGACCAAATATGGCT
GGTTCTCGCACTAGT3'

This DNA was ligated at the XbaI end to a 1kb DNA handle bearing many biotin groups on both strands, and ligated at the SbfI end to a 1kb DNA handle bearing many digoxigenin groups on both strands. These handles are home-made, using the TaqRPOC gene as template for a PCR reaction using the following primers:

5'GAGAGACCTGCAGGACATCAAGGACGAGGTGTGGG3'

5'GAGAGATCTAGATCCTCAAAGTTCTTGAAGACCGCCTGG3'

The PCR reaction is done twice, one with dUTP-biotin in the dNTP mix, and the other with dUTP-digoxigenin in the dNTP mix. The biotin-labelled DNA was cut with XbaI whilst the digoxigenin-labelled DNA was cut with SbfI-HF.

The 36_stall and 83_stall DNA constructs were generated by altering the parent 20_stall construct. In both cases, the pET21ΔMCS_20stall vector was digested with HindIII and SpeI, and the open vector was gel-purified. The following DNA fragments were ligated into the open vector to generate the relevant construct:

36_stall:

5'AGCTTATAAATTTGAGGAGAATAATTGTAGAGGAGAGAGTCCAAATATGGCTGGTTCTCGCA3'

83_stall:

5'AGCTAATAAATTTGAGGAGACCAAATATGGCTGGTTCTCGACGGTCTTCTCCATGCCAGGCGAAGC
TTAGAGAAA3'

The 83_stall construct was designed to destroy the original HindIII site and create a new one further downstream, enabling the insertion of the CPD-bearing oligo by restriction-digestion and ligation using the HindIII and SpeI sites. The sequence of the CPD-bearing oligo is:

5'CTAGAGGAGAACCAGCCATATTTXXTCTCCTCTCTCAAATTTATT3' (where XX is the CPD)

Ligated DNA is stored at -20°C, with a working stock of 50pM stored at 4°C. This DNA is ligated rapidly first to streptavidin-coated magnetic beads, and then added to the glass coverslip. The DNA-bead mixture incubates on the surface for 10 minutes, after which free magnetic beads and free DNA are washed out with buffer.

Proteins

E. coli core RNAP and σ^{70} were purified as previously described and core RNAP was saturated with a fivefold excess of σ^{70} [188]. *E. coli* RapA was purified from BL21(DE3) cells transformed with pQE80L expression vector (encoding N-terminally His₆-tagged full-length RapA) grown at 37°C to OD = 0.6 and induced using 1mM IPTG for 4 hours at 30°C [107]. Cells were lysed at 4°C (10mM Tris pH 8.0 (@ 4°C), 5% glycerol, 1M NaCl, 0.5mM beta-mercaptoethanol, and protease inhibitor) using an emulsiflex homogeniser and samples were kept at 4° for the duration of the purification. Cell lysate was applied to a Nickel column, washed with up to 15mM imidazole before elution with 200mM imidazole. The eluate was applied to a Q column, washed and eluted over a 0-500mM NaCl gradient with RapA eluting at ~300mM NaCl. The eluate was concentrated and ran on a gel-filtration column (GF200) equilibrated in 10mM Tris pH8.0 (@ 4°), 5% glycerol, 100mM NaCl, 0.1mM EDTA and 5mM 1,4-Dithiotheritol. Single-use 10µl aliquots were snap-frozen in liquid nitrogen and stored at -80°C.

Flow cell preparation

Flow cells derivatized with anti-digoxigenin were prepared as previously described [272].

Reaction conditions

Single-molecule assays were performed at 34°C in reaction buffer containing 40mM K-Hepes pH=8.0, 100mM KCl, 8mM MgCl₂, 0.5mg/ml BSA, 0.1% w/v Tween 20, and 1mM 1,4-Dithiotheritol. Unless stated otherwise, concentrations of components in reactions were as follows: 100pM RNAP holoenzyme, 100nM RapA, 1mM ATP, 200μM GTP and 200μM UTP. Collecting the specific number of events, *n*, typically requires more than three experimental runs involving normally 20-50 DNA molecules simultaneously.

Tethered-DNA assays

Only single molecules of intact double-stranded DNA are chosen for experimentation, which are verified by the ability to supercoil and the shape of the rotation-extension curve. For addition of components, DNA was positively supercoiled and the force was increased to ~2pN to prevent DNA molecules becoming stuck to the surface under the flow of components. Once the flow equilibrated the force was reduced. For the recycling assay, the DNA was maintained at a constant force of 0.3pN at positive supercoiling. For the force-cycling assay, the positively supercoiled DNA was kept at 0.3pN for 400s to allow RNAP to bind and stall, then the force increased to the relevant value for many thousands of seconds.

Data acquisition and analysis

Histograms were fit to a single exponential. Number of events (n) and the values of the fits with errors are stated in the figure legends. All data analysis was done using Xvin software (PicoTwist).

Torque was estimated as follows: $torque = \sqrt{k.T.\xi.F}$ where k = Boltzmann constant, T = temperature, ξ = persistence length and F = force.

Acknowledgements

J.R.P. acknowledges support from the Frontières du Vivant – Programme Bettencourt doctoral program. J.R.P. is supported by a graduate scholarship of the Horizons 2020 Innovative Training Network “DNAREPAIRMAN”. Research on this topic in the Strick lab is supported by grants from the French ANR (PrTxConf, ANR-17-CE11-0042), the Ligue Nationale Contre le Cancer program for Core Research Teams (“Equipes Labellisées”), the NanoRep grant funded by Paris Sciences et Lettres University as well as core funding from the CNRS, the University of Paris, the Ecole Normale Supérieure, and Paris Sciences and Lettres University.

Author Contributions

J.R.P. performed single-molecule experiments and analysed data. M.Z.Q. and K.S.M. provided reagents. J.R.P. and T.R.S designed experiments and wrote the paper.

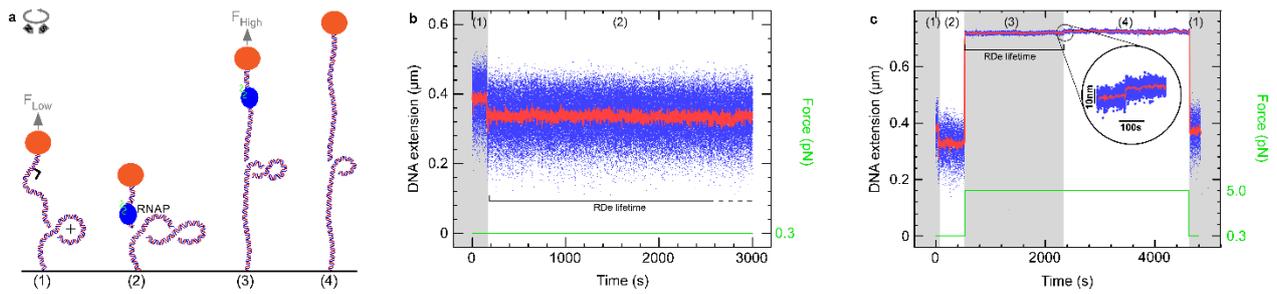


Figure 3.1: High torque removes stalled RNAP from DNA

a) Experimental setup. The 2kb DNA bears a promoter (black arrow) and is tethered between a magnetically-trapped bead (orange) and glass surface (black). DNA is positively supercoiled and at low force before components (RNAP and ribonucleotides except CTP) are added (1). RNAP binds to the promoter, initiates transcription and stalls at +20, causing the DNA extension to decrease. At low force it remains bound here indefinitely (2). The force is increased, causing the DNA molecule to extend and increase torque (3). RNAP dissociates from DNA, causing the DNA extension to increase (4). **b)** Time-trace with RNAP and limiting ribonucleotides on positively supercoiled DNA (1). As RNAP initiates transcription it scrunches and unwinds ~ 2 turns of DNA, reducing DNA extension by ~ 100 nm. Successful elongation proceeds with an unwound bubble of ~ 9 bp and so DNA extension is reduced by only ~ 50 nm compared to baseline (2). As RNAP reaches +20 it stalls indefinitely. Each state corresponds to the model in part a. **c)** Time-trace with RNAP and limiting ribonucleotides in the force-cycling assay. Positively supercoiled DNA is initially at low force (1). Components are added (~ 1000 s) and RNAP initiates transcription and stalls (2). The force is increased (3) and we observe RNAP dissociation (inset) as the DNA returns to the baseline state (4). RNAP dissociation is confirmed because when the system returns to the low force state, the DNA extension is the same as it was prior to RNAP binding (1).

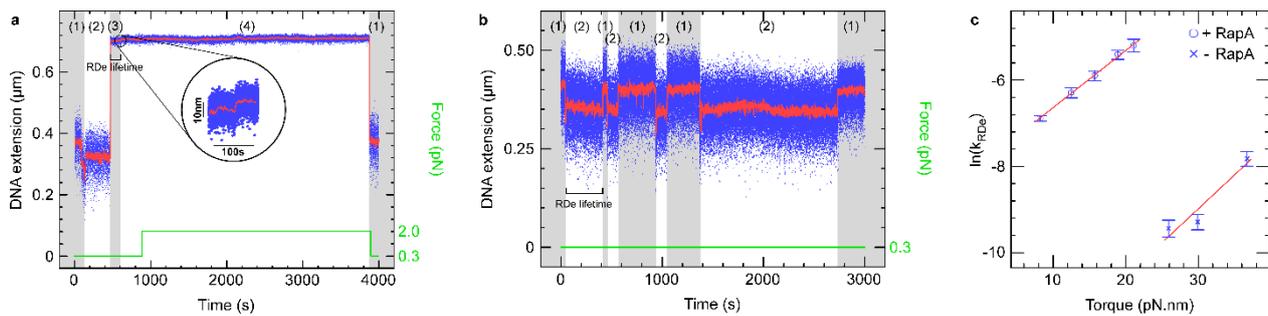


Figure 3.2: RapA removes stalled RNAP in a torque-sensitive manner

a) Time-trace with RNAP, limiting ribonucleotides and RapA in the force-cycling assay. As in Fig. 3.1b, RNAP initiates and stalls on DNA. The force is increased and we observe rapid RNAP removal, which can be confirmed as the molecule extension returns to baseline upon returning to low force. **b)** Time-trace with RNAP, limiting ribonucleotides and RapA on positively supercoiled DNA in the low force recycling assay. RNAP initiates, stalls and is removed by RapA in an iterative cycle, with RDe lifetimes measured. **c)** Plot of $\ln(k_{RDe})$ vs torque. RDe lifetimes were measured with and without RapA at relevant forces and fit according to a single-exponential (Supp. Fig. 3.3 and Supp. Fig. 3.2), with the mean values used here. 5 force points were taken with RapA and the gradient of the fit is 0.13, whereas 3 force points were taken without RapA and the gradient of the fit is 0.15.

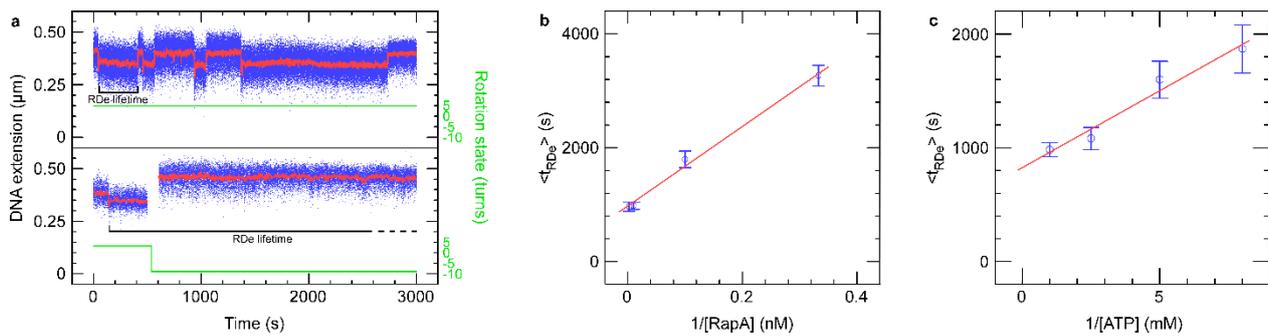


Figure 3.3: RapA removes RNA polymerase stalled on positively, but not negatively, supercoiled DNA

a) Top panel: As shown in Figure 3.2b, time-trace showing the iterative stalling and removal of RNAP from positively supercoiled DNA in the presence of RapA. Bottom panel: Time-trace showing first the stalling of RNAP on positively supercoiled DNA, followed by a washing step to remove free RNAP and the addition of RapA and ATP (at ~500s), then the DNA was negatively supercoiled. RNAP remains stalled at +20 indefinitely. **b)** Plot of $\langle t_{RDe} \rangle$ vs RapA concentration. RDe lifetimes were measured at various RapA concentrations (but saturating ATP concentrations) and fit according to a single-exponential (Supp. Fig. 3.7), with the mean values plotted here. $V_{max} = 980s$, $K_m = 7.2nM$ and $K_{cat} = 1.1 \times 10^{-3} s^{-1}$. **c)** Plot of $\langle t_{RDe} \rangle$ vs ATP concentration. RDe lifetimes were measured at various ATP concentrations (but saturating RapA concentrations) and fit according to a single-exponential (Supp. Fig. 3.8), with the mean values plotted here. The K_m of RapA-RNAP for ATP is $165\mu M$.

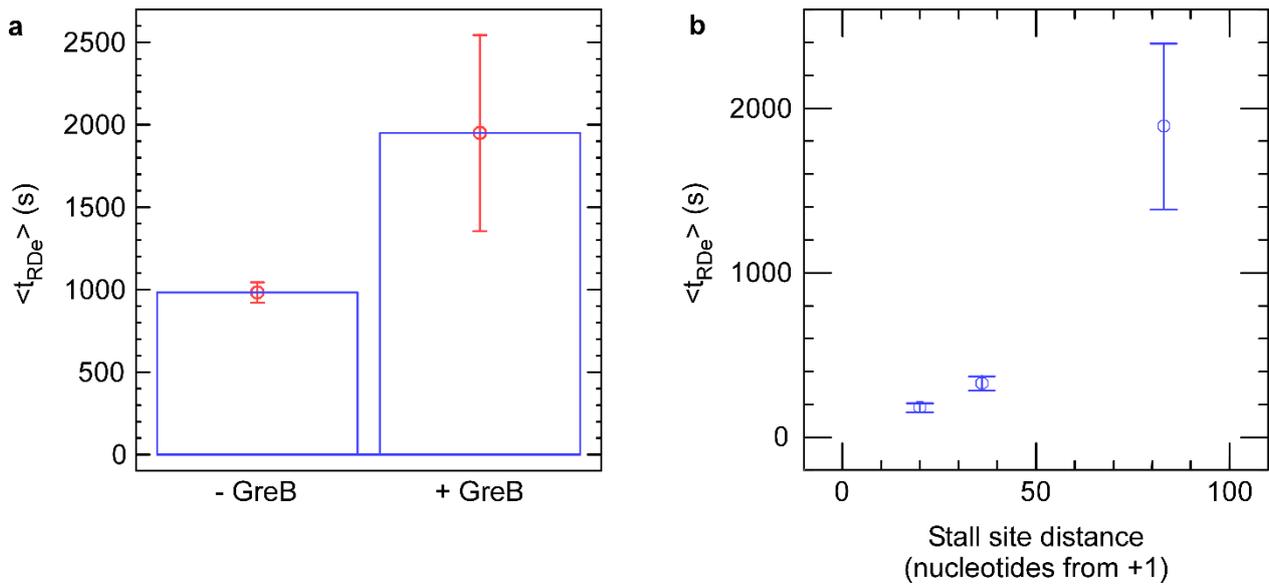


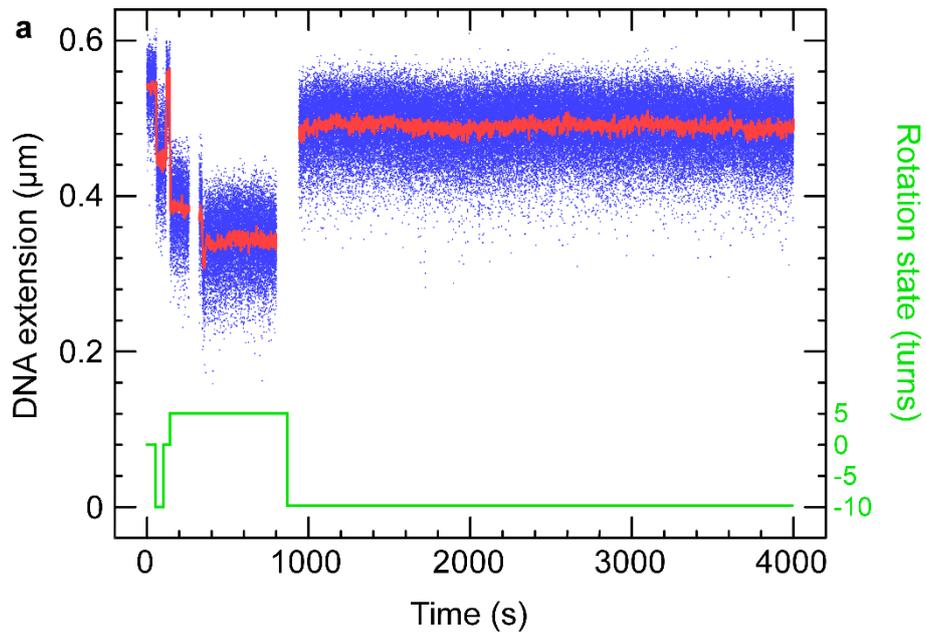
Figure 3.4: RapA removes RNAP by backtracking

a) GreB slows RapA removal of RNAP. Using the force-cycling assay RapA removes 97% of stalled RNAP with a $\langle t_{RDe} \rangle$ of 983 ± 62 s. In the same assay but in the presence of GreB RapA removes 44% of stalled RNAP with a $\langle t_{RDe} \rangle$ of 1951 ± 594 s. GreB roughly halves the number of RNAP affected by RapA and roughly doubles the $\langle t_{RDe} \rangle$ of the RNAP that are removed by RapA. **b)** Plot of $\langle t_{RDe} \rangle$ vs stall site distance from promoter. RDe lifetimes were measured with RNAP, limiting ribonucleotides, saturating RapA and saturating ATP in the force-cycling assay. These lifetimes were fit according to a single-exponential (Supp. Fig. 3.9), with the mean values shown here. As the distance between the stall site and the promoter increases, the longer it takes RapA to remove the RNAP.

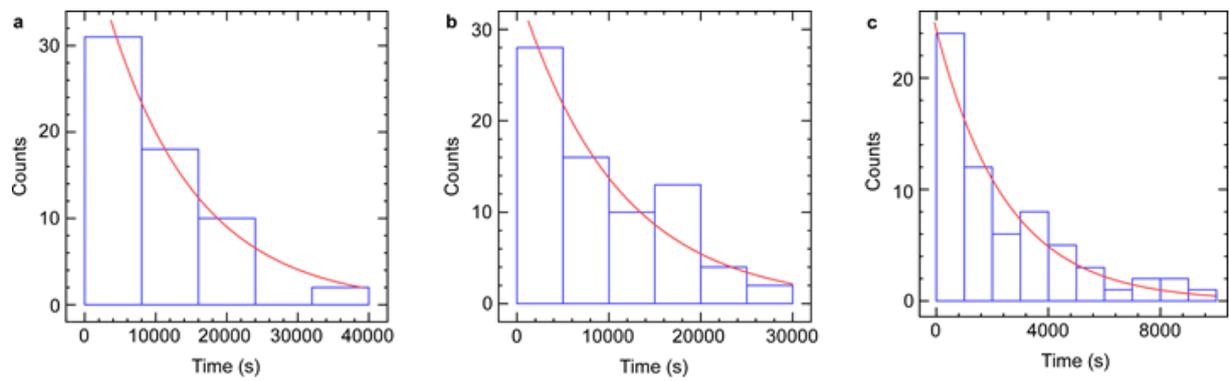
[GreB] (nM)	Number of molecules	Number of molecules with RNA polymerase initiation	Percentage of molecules with RNA polymerase initiation (%)	Number of molecules with RNA polymerase removed	Percentage of molecules with RNA polymerase removed (%)	$\langle t_{RD_e} \rangle$ (s)
0	84	74	88	72	97	983 ±62
50	53	50	94	22	44	1951 ±594

Table 4: GreB inhibits RNAP removal by RapA. The percentage of RNAP molecules that are removed by RapA goes from 97% in the absence of GreB to 44% in the presence of 50nM GreB.

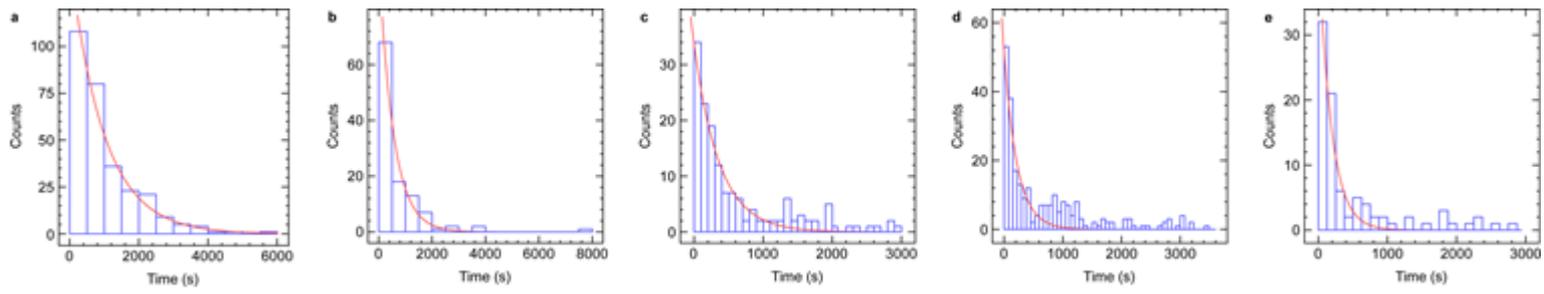
Supplementary Information



Supplementary Figure 3.1: Time-trace with RNAP and limiting ribonucleotides on negatively supercoiled DNA. Components are injected ($\sim 300\text{s}$) and RNAP stalls on positively supercoiled DNA. The DNA is then negatively supercoiled ($\sim 900\text{s}$). The difference in DNA extension at the -10 rotation state confirms RNAP remained stalled on DNA during the negative supercoiling. Across 40 molecules over 20 hours all RNAP molecules remained bound to DNA.



Supplementary Figure 3.2: RDe lifetime distributions at different forces without RapA. **a)** Histogram of RDe lifetimes measured at 3.0pN force ($n = 61$; average = $12590s \pm 2430$). **b)** Histogram of RDe lifetimes measured at 4.0pN force ($n = 73$; average = $10810s \pm 1920$). **c)** Histogram of RDe lifetimes measured at 6.0pN force ($n = 64$; average = $2500s \pm 426$).



Supplementary Figure 3.3: RDe lifetime distributions at different forces with RapA. **a)**

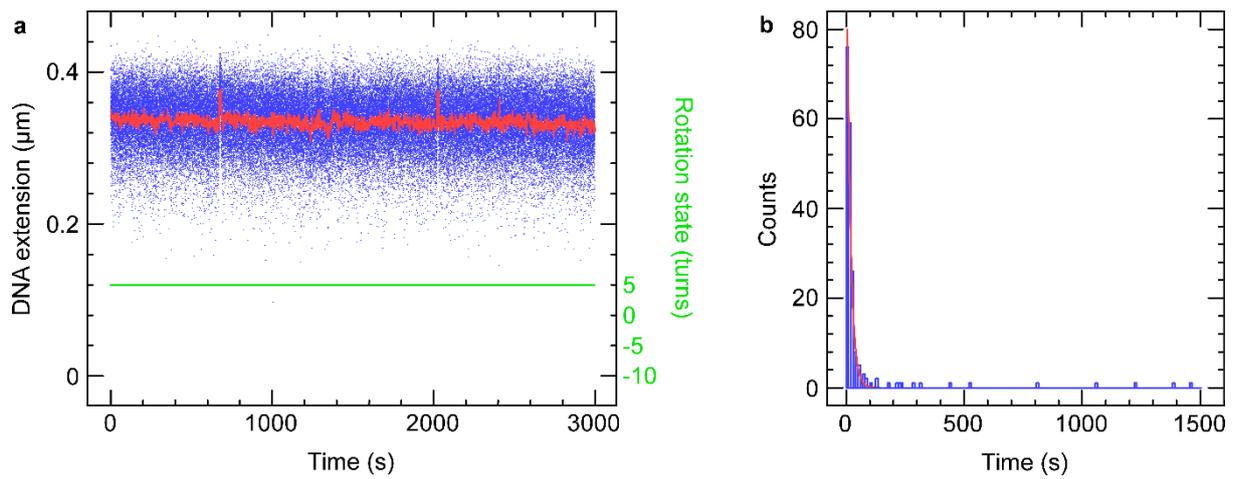
Histogram of RDe lifetimes measured at 0.3pN force ($n = 289$; average = $983s \pm 61.9$). **b)**

Histogram of RDe lifetimes measured at 0.7pN force ($n = 112$; average = $542s \pm 63.9$). **c)**

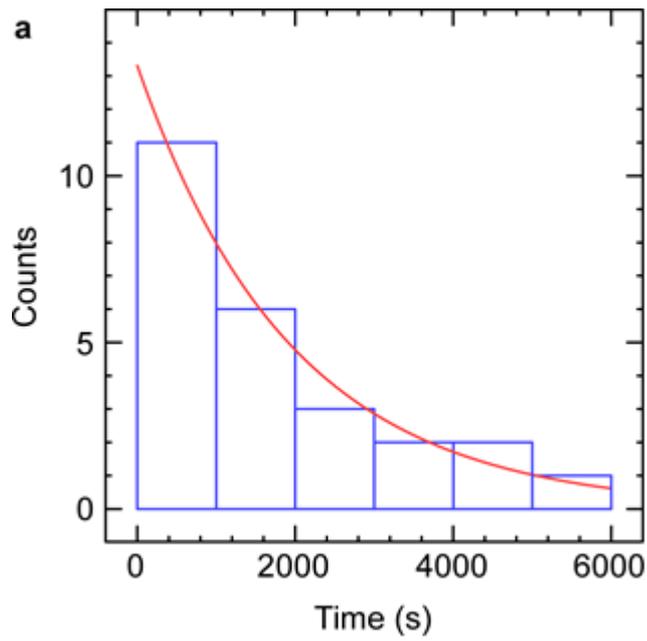
Histogram of RDe lifetimes measured at 1.1pN force ($n = 150$; average = $365s \pm 41.6$). **d)**

Histogram of RDe lifetimes measured at 1.6pN force ($n = 248$; average = $223s \pm 23.9$). **e)**

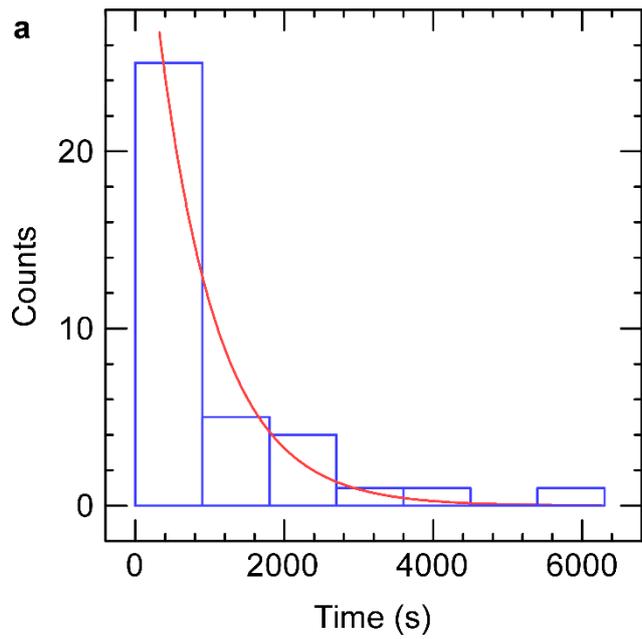
Histogram of RDe lifetimes measured at 2.0pN force ($n = 87$; average = $181s \pm 27.5$).



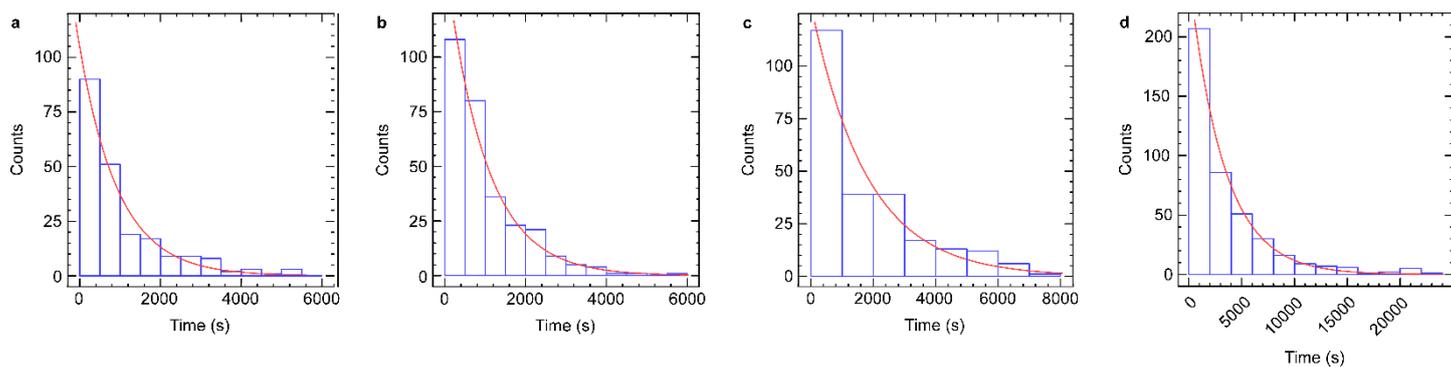
Supplementary Figure 3.4: RapA and ATP alone on DNA. **a)** Time-trace on positively supercoiled DNA with 100nM RapA and 1mM ATP. DNA extension increase events are seen at ~700s and ~2000s. **b)** Histogram of these event lifetimes ($n = 205$; average = $15.5\text{s} \pm 1.24$).



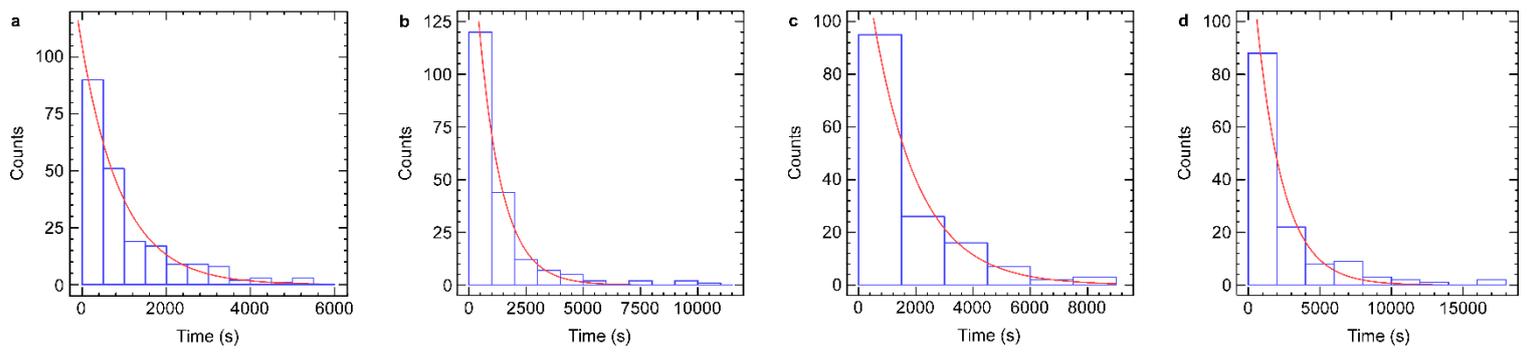
Supplementary Figure 3.5: RDe lifetime distribution with 50nM GreB. **a)** Histogram of RDe lifetimes measured in the force cycling assay (2pN; saturating ATP and RapA concentration) with 50nM GreB present ($n = 25$; average = $1950\text{s} \pm 594$).



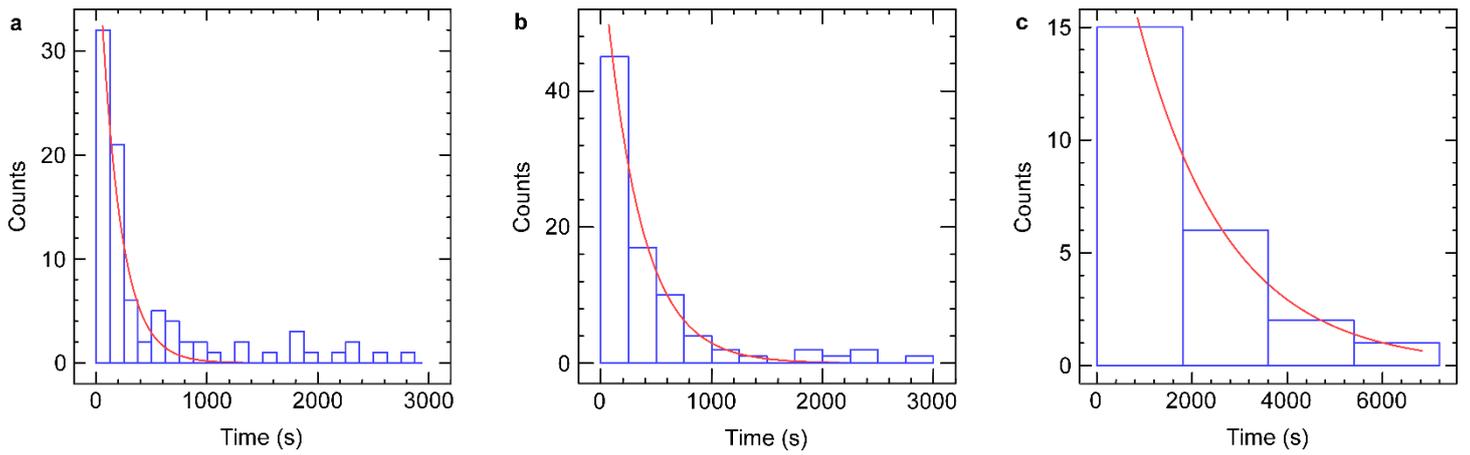
Supplementary Figure 3.6: RDe lifetime distribution with CPD stalling. **a)** Histogram of RDe lifetimes measured in the force cycling assay (2pN; saturating ATP and RapA concentration) using a CPD to stall RNAP at +20 ($n = 37$; average = $796\text{s} \pm 179$). This is not significantly different from the average RDe lifetime obtained under the same conditions but with ribonucleotide starvation as the stalling cause.



Supplementary Figure 3.7: RDe lifetime distributions at different RapA concentrations. **a)** Histogram of RDe lifetimes measured with 500nM RapA ($n = 212$; average = $965\text{s} \pm 82.5$). **b)** Histogram of RDe lifetimes measured with 100nM RapA ($n = 289$; average = $983\text{s} \pm 61.9$). **c)** Histogram of RDe lifetimes measured with 10nM RapA ($n = 244$; average = $1800\text{s} \pm 144$). **d)** Histogram of RDe lifetimes measured with 3nM RapA ($n = 421$; average = $3270\text{s} \pm 182$).



Supplementary Figure 3.8: RDe lifetime distributions at different ATP concentrations. **a)** Histogram of RDe lifetimes measured with 1mM ATP ($n = 289$; average = $983\text{s} \pm 61.9$). **b)** Histogram of RDe lifetimes measured with $400\mu\text{M}$ ATP ($n = 195$; average = $1080\text{s} \pm 97.1$). **c)** Histogram of RDe lifetimes measured with $200\mu\text{M}$ ATP ($n = 149$; average = $1600\text{s} \pm 163$). **d)** Histogram of RDe lifetimes measured with $125\mu\text{M}$ ATP ($n = 135$; average = $1870\text{s} \pm 212$).



Supplementary Figure 3.9: RDe lifetime distributions at different stalling distances. **a)** Histogram of RDe lifetimes measured at 2.0pN with RNAP stalling at +20 ($n = 87$; average = $181s \pm 27.5$). **b)** Histogram of RDe lifetimes measured at 2.0pN with RNAP stalling at +36 ($n = 85$; average = $328s \pm 43.4$). **c)** Histogram of RDe lifetimes measured at 2.0pN with RNAP stalling at +83 ($n = 24$; average = $1890s \pm 505$).

Appendix II: *In vivo* RapA experiments

Following these single-molecule observations, one of our hypotheses for the *in vivo* role of RapA was that it mediates conflict resolution. Several lines of evidence led us to this hypothesis. Firstly, RapA seemed only to remove stalled RNAP at very high torques that would only be experienced *in vivo* when RNAP encountered the replisome. Secondly, RapA only works on positively supercoiled DNA which is similarly encountered in-front of the replisome. As a result, we can refine our hypothesis to state that RapA mediates resolution of head-on RNAP-replisome conflicts. Thirdly, anecdotal evidence indicates that hyperthermophiles, which are thought to contain positively supercoiled genomes, show periodic expression of RapA homologues. In contrast to the constitutive expression of RapA that *E. coli* shows, hyperthermophiles only express RapA just before replication. One could interpret this by stating hyperthermophiles avoid the constitutive removal of RNAP by RapA through controlling RapA expression. Only prior to replication, where RapA can clear the genome of RNAP obstacles, is RapA expressed, because it would be problematic otherwise. UvrD and Rep are conflict resolution factors and it has been shown that cells lacking just UvrD or Rep are viable, whereas the UvrD-Rep double knock-out is not viable. In a similar vein, we sought to generate various single and double knock-out *E. coli* strains to see if they displayed growth defects that would help to understand RapA's role *in vivo*. Table 5 outlines the strains I used and generated.

Strain	Identifier (if applicable)	Reference
Wild-type	MG1655	Commercially available
TB28 (MG1655 Δ lacIZYA:: $\langle\rangle$)	TB28	[291]
TB28 Δ RapA (MG1655 Δ lacIZYA:: $\langle\rangle$ Δ rapa::kan)		This work
TB28 Δ Rep (MG1655 Δ lacIZYA:: $\langle\rangle$ Δ rep::cat)	N6577	[292]
TB28 Δ Mfd (MG1655 Δ lacIZYA:: $\langle\rangle$ Δ mfd:: $\langle\rangle$)	KM386	Gift from Nigel Savery

TB28Δ <i>UvrD</i> (MG1655Δ <i>lacIZYA</i> ::<>Δ <i>uvrD</i> :: <i>dhfr</i>)	N6632	[292]
TB28Δ <i>Rep</i> Δ <i>RapA</i> (MG1655Δ <i>lacIZYA</i> ::<>Δ <i>rep</i> :: <i>cat</i> Δ <i>rapA</i> :: <i>kan</i>)		This work
TB28Δ <i>Mfd</i> Δ <i>RapA</i> (MG1655Δ <i>lacIZYA</i> ::<>Δ <i>mfd</i> ::<>Δ <i>rapA</i> :: <i>kan</i>)		This work
TB28Δ <i>UvrD</i> Δ <i>RapA</i> (MG1655Δ <i>lacIZYA</i> ::<>Δ <i>uvrD</i> :: <i>dhfr</i> Δ <i>rapA</i> :: <i>kan</i>)		This work

Table 5: *E. coli* strains used in this study.

These strains were generated using lambda-red recombination. A detailed protocol for lambda-red recombination of *E. coli* can be found in the Supplementary Protocols. Briefly, lambda-red recombination relies on three plasmids (pKD46 – contains arabinose-inducible lambda-red recombinase; pcp20 – contains FLP recombinase; and pKD13 – contains Kanamycin cassette flanked by FRT sites). As shown in Figure 3.10, one first transforms the parent strain with pKD46 and then transforms with a PCR product (which bears the Kan^R gene, FRT sites and homology regions (H1 and H2) to the gene of interest). The recombinase swaps the gene of interest with the PCR product. Next, optionally one can transform with pcp20 to remove the Kan^R cassette.

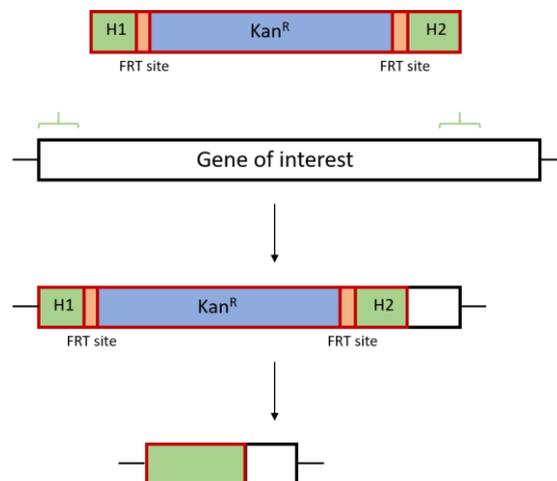


Figure 3.10: Overall work-flow of lambda-red recombination for the generation of *E. coli* knockouts

For the generation of RapA null mutants, we used the following primers in the PCR reaction that uses pKD13 as the template:

1) RapA_pKD13_Fw_full

5'ATAAGCACGGTAGCGGGTGGTTATTGCCTGCAATTAAAGATATAGAGCCGAACACATATGATTCCG
GGGATCCGTCGACC3'

2) RapA_pKD13_Rv_full

5'GCGGTGGATTGTAGTTTTCCATCCCCATTTTCGGCTCCGTTACTGATGCGTTACAACGATTGTAGGC
TGGAGCTGCTTCG3'

3) RapA_pKD13_Rv_CT_trunc

5'GCAGAGCTTCCAGACGAGACAGCTCGGCAGACAGTTTTTCGTCGGCTTCGTTACGCGCTGTGTAGGC
TGGAGCTGCTTCG3'

Initially I made two versions of RapA null mutants. The first (a) was a full gene removal whereas the second (b) left in the CTD that contained a putative promoter of a downstream gene. It was possible to leave the CTD in because it didn't contain any ATPase folds or interaction domains. During initial testing there was no difference between either strain. The agarose gel in Figure 3.11 shows the verification by PCR of the successful lambda-red recombination – the RapA gene was successfully replaced by the Kan^R cassette in every strain.

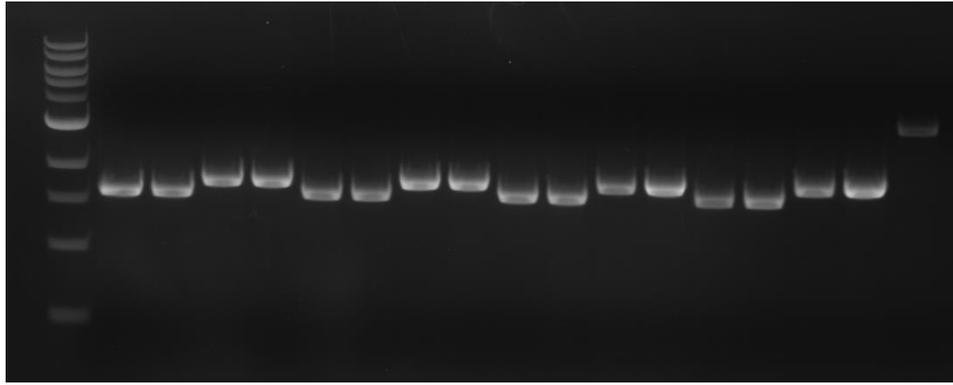


Figure 3.11: 1% agarose gel showing the products of colony PCR reactions performed using primers just upstream and just downstream of the RapA gene location.

Lane 1: NEB 1kb ladder

Lanes 2 and 3: TB28 Δ RapA(*a*)

Lane 4 and 5: TB28 Δ RapA(*b*)

Lanes 6 and 7: TB28 Δ Mfd Δ RapA(*a*)

Lanes 8 and 9: TB28 Δ Mfd Δ RapA(*b*)

Lanes 10 and 11: TB28 Δ Rep Δ RapA(*a*)

Lanes 12 and 13: TB28 Δ Rep Δ RapA(*b*)

Lanes 14 and 15: TB28 Δ UvrD Δ RapA(*a*)

Lanes 16 and 17: TB28 Δ UvrD Δ RapA(*b*)

Lane 18: TB28

Every strain was viable and was compared in growth in liquid media and solid media. In agreement with previous results, there was no appreciable difference in the growth of these strains in liquid rich media, as shown in Figure 3.11 [106]. Regarding growth on solid agar, again there was no appreciable difference in colony size or number. Interestingly, we did not observe markedly slow growth of the RapA null mutant on high salt plates, as was observed by others (see Table 6) [110]. Similarly, UV irradiation tests showed RapA null mutants were no more sensitive than their parent, in agreement with previous findings (not shown) [108].

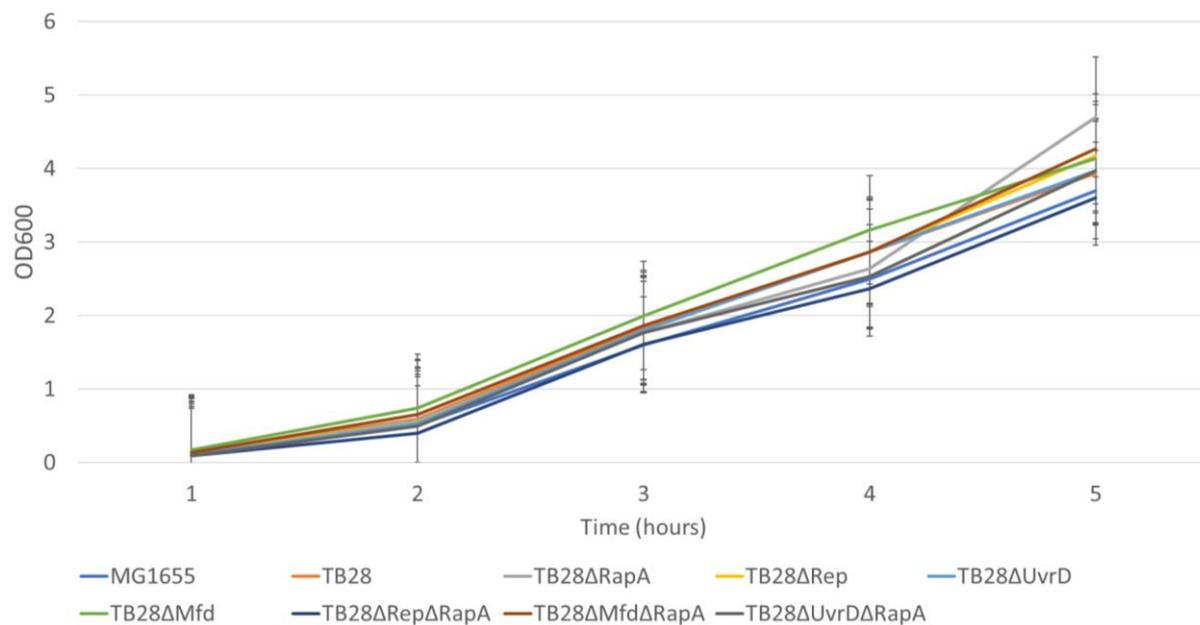


Figure 3.11: Simple growth curve performed in LB at 37 degrees. Numbers represent averages of three separate experiments.

	MG1655	TB28	TB28Δ <i>RapA</i>
LB			
LB + 0.75M NaCl			
LB + 1M NaCl			

Table 6: Images of agar plates containing varying concentrations of salt after plating and incubation at 37 degrees for 100 hours.

In summary, we generated a RapA single knock-out and three RapA double knock-outs through lambda-red recombination. These strains showed no clear growth defects in liquid or on solid media, and no increased UV sensitivity. Another interpretation of the single-molecule findings is that RapA plays a role in the osmotic shock response. It was recently shown that during osmotic shock, all transcription halts and RNAP are removed from DNA [290]. There is then a recovery period of roughly twenty minutes (see Figure 3.12), following which RNAP slowly accumulate again on DNA, first to osmotic shock response genes [290]. One testable theory therefore is that our observations of supercoiling and torque-sensitivity of RapA represent a dependence on a range of DNA properties that also occur under osmotic shock.

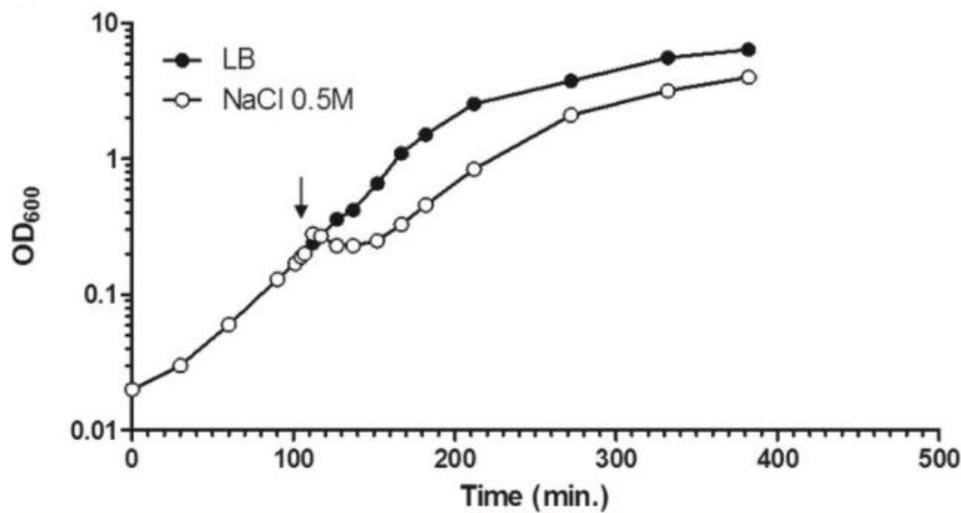


Figure 3.12: The osmotic shock response. Two cultures are grown in parallel and at ~100 minutes (downwards arrow) one culture is subjected to osmotic shock. After a recovery period of ~20 minutes, cells continue to grow [290].

The experiment would therefore be to monitor and compare optical density over time of two cultures subjected to osmotic shock; one wild-type and one RapA null culture. If RapA plays a role in removing RNAP from DNA due to osmotic shock, then one would expect the RapA null cells to have an extended recovery period in comparison to the wild-type cells.

Section III: Conclusions and Perspectives

Understanding the mechanism by which Mfd mediates mutagenesis leads us to review and re-examine many avenues of enquiry, and to propose new ones. Not only does this mechanism represent a new mode of Mfd function and motility, it reveals a potentially universal mechanism by which other proteins form R-loops which has implications for genomic stability, topology, and therapeutics.

We may now revisit the early genetic observations of MFD by Witkin and of stress-induced mutagenesis by Cairns, Overbaugh and Foster (and many others thereof). Firstly, revealing that mutagenesis by Mfd relies on active transcription allows us to propose a second resolution to the apparent paradox that the MFD observations present (see Introduction: post-UV exposure, *more* DNA is repaired (seen as a reduction of mutants) when bacteria are forced to *repress transcription* before plating on selective media). This appears paradoxical given Mfd's role in TCR, but it could be reconciled given Mfd's role in mutagenesis. Mfd may be responsible for a proportion of mutations that lead to prototrophy, but this action is inhibited when cells are grown in minimal media before plating on selective media – halting transcription drastically reduces tripartite supercoiled domain phenomenon and therefore reduces Mfd-mediated mutagenesis. Secondly, the observation of roughly five times more late (stress-induced) mutants in UvrB null cells may be reconciled considering Mfd's role in mutagenesis [244]. In UvrB null cells, the lifetime of UvrA on DNA increases (*in vivo* observations in the absence of exogenous DNA damage show UvrA resides on DNA for ~100 seconds absent UvrB, compared to either ~2 seconds (fast species) or ~12 seconds (slower species) with UvrB) [218]. If we assume UvrA interacting with Mfd that is engaged with transcribing RNAP (i.e., forming a tripartite supercoiled domain) is one way to release Mfd, then we can conclude that if UvrA is quenched on DNA, Mfd is freer to form R-loops. Therefore, in UvrB null cells, Mfd stimulates greater R-loop formation and more stress-induced mutagenesis is observed.

If we extend these observations of the bacterial system to the human system, we can offer an explanation for the observed decrease in DSBs upon CSB knockdown. As mentioned in the

Introduction, Sollier et al. proposed that the reduction of DSBs following CSB knockdown was because CSB played a role in the processing of R-loops [242]. A revised explanation would be that like Mfd, CSB is responsible for the formation of the R-loops and so naturally, reducing CSB levels reduces R-loop levels.

To see where this new Mfd mechanism sits within the known landscape of Mfd functions, I will here summarise and group the modes of Mfd motility. The first mode of Mfd motility, Mfd alone on DNA, has been observed or inferred from many perspectives. Structural studies have shown the conformation of translocating Mfd alone on DNA and single-molecule studies have probed the processivity and lifetime of Mfd alone on DNA (Supp. Fig. 2.4) [117, 186]. This Mfd species is rare and a common explanation is that it arises through conformational heterogeneity of apo-Mfd in solution. A small percentage of auto-inhibited Mfd become able to bind DNA without RNAP as a result of its domain flexibility. However, this species is typically not processive, is short-lived, and doesn't play a major role in repair. The second mode of Mfd motility, Mfd associated with a TEC, represents Mfd engaged in transcription reactivation, transcriptional regulation, or tripartite supercoiled domain phenomenon [96, 99, 220]. In contrast, here Mfd relies on interaction with a TEC in order to bind and translocate on DNA. The state of the RNAP (i.e., reactivated or terminated) prevents the complex from committing to TCR, which represents the final mode of Mfd motility. The third mode of Mfd motility is Mfd topologically encircling DNA in complex with RNAP. The RNA is not present and RNAP is not transcriptionally active. This Mfd-RNAP complex is committed to TCR and translocates on DNA waiting to be interrupted by a UvrA(B) complex. The topological clamp around DNA imparts unto the complex a high processivity which is clearly necessary for a protein complex acting as a signal for repair. In every case Mfd moves at ~5 bp/s. How Mfd mediates mutagenesis relies on robust and well-studied aspects of Mfd biochemistry, namely DNA binding and RNAP binding. One could argue that having a role in repair gives Mfd markedly robust binding abilities that best place it to stimulate tripartite supercoiled domain phenomena. As mentioned, the properties required to form tripartite supercoiled domains are

simultaneous RNAP binding and upstream DNA binding. These properties are shared across many transcription factors and, as is shown in Appendix I, means that R-loop formation through the generation of a tripartite supercoiled domain is potentially a conserved mechanism. Indeed, whilst we have observed the tripartite supercoiled domain phenomenon with another protein, it was necessary to use a mutant version that bound more strongly to DNA in order to observe events roughly as frequently as with Mfd. It appears that R-loops resulting from tripartite supercoiled domain phenomena may be conserved, but the frequency at which they occur (e.g., one in five transcription events for Mfd at 500nM concentration) may depend on the relative DNA or RNAP binding strength. As mentioned, playing a role in TCR grants Mfd strong DNA and RNAP binding abilities, that position Mfd well to greatly stimulate tripartite supercoiled domain formation.

Nonetheless, we have observed tripartite supercoiled domain formation in two separate systems. The now plausible idea that tripartite supercoiled domain phenomena is conserved in transcription factors brings to the fore questions regarding genetic instability and genome topology. We know that roughly half of stress-induced mutagenesis is due to Mfd, and whilst we can speculate that tripartite supercoiled domain formation accounts for the rest, we do not know the proteins responsible nor the mechanistic origins of all stress-induced mutations. Revealing to what degree all other candidate proteins mediate tripartite supercoiled domain formation is immediately important. Similarly, understanding how tripartite supercoiled domain phenomena are regulated is key. For Mfd, it is likely that UvrA binding is a regulatory mechanism.

The extension of the tripartite supercoiled domain mechanism to transcription factors has implications for genome topology. This model, in contrast to the twin supercoiled domain model of Liu and Wang, produces positive supercoiling both downstream and upstream rather than positive supercoiling downstream and negative supercoiling upstream. One may propose that upstream positive supercoiling is an important feature for genes located at or near topological boundary regions, preventing waves of negative supercoiling from diffusing outside of the

topological domain and activating promoters of other genes [30]. Therefore, a supplemental role of tripartite supercoiled domains arising from transcription factors may be to establish or maintain topological domains and to stem the spread of negative supercoiling that would otherwise occur.

Identifying the regulatory framework that governs mutagenesis by Mfd is important when we consider Mfd as a drug target. As mentioned, Mfd represents an attractive drug target for combinatorial therapies – in addition to the drug of choice, an Mfd inhibitor would act to reduce the frequency of resistance by inhibiting this key evolvability factor (relevant for both bacterial infections and human cancers) [241, 251]. Thus, designing a drug to perturb the conserved β subunit-RID interaction offers a route to inhibit Mfd's role in TCR and mutagenesis which reduces the evolutionary capacity of the targeted cell.

Regarding RNAP itself, we have probed for the first time the inherent stability of a stalled TEC on DNA and have shown the extremely long lifetime previously observed drastically shortens when high torque is applied, which has important ramifications for transcriptional regulation during topological stresses. The mechanism by which the TEC dissociates from DNA (after backtracking or not) is not clear, but it is clearly stimulated by RapA. The dependence of RapA on supercoiling and torque indicates RapA acts in a specific cellular context. The fact that RapA is expressed constantly and at high levels might imply that RapA needs to be immediately present to respond to a stressor. The cause of stress may be osmotic shock or heat shock, or indeed anything that leads to a change in the mechanical properties of DNA that would impede transcription.

Section IV: Supplementary Protocols

Chapter I: Glass Surface Preparation and Functionalisation Protocol

This protocol is adapted from [272]. It begins with cleaning of the glass slides with water and ethanol, followed by sonication with KOH. Slides are then washed thoroughly with water and dried with nitrogen gas and treated in the plasma oven. Then, a layer of adhesion promoter is applied, followed by a layer of polystyrene. The flow chamber is then assembled and functionalised with anti-digoxigenin. The steps of drying, plasma treatment, and spin-coating and extremely sensitive to humidity and so should be performed fast and whilst wearing a mask.

1. Rinse the slide holder and the bin with ethanol, and then fill with ethanol.
2. Rinse the slides, holding them by the side only, with ethanol, milli-Q (MQ) water, then ethanol, then place in slide holder – leave at least a two-space gap between them.
3. Sonicate for 10 minutes, during this time make 0.5L of 1M KOH.
4. Empty the bin, rinse twice with water (lift slide holder up/down 10x each time), refill with 1M KOH and sonicate for 15 minutes (check slide spacing is maintained after each sonication).
5. Empty KOH, rinse with water 2x and refill with KOH, sonicate for 15 minutes.
6. Take to MQ tap, rinse very well to remove all KOH traces, put back into slide holder full of water.
7. Prepare for the upcoming steps (8 through 12) – set heating plate to 110°C, make fresh aliquot of TI Prime Adhesion Promoter (argon the stock), prepare two glass pipettes, and prepare two petri dishes with kimwippes inside (the following steps – drying, plasma cleaning, primer coating and heating – are time-sensitive so ensure you have time and can proceed well).

8. Wear a mask to reduce local humidity and dry slides well with Nitrogen gas (dry holed slides first, then non-holed slides after), put dried slides into a petri dish containing a kimwipe.
9. Take to plasma cleaner and place all slides inside, set to high and begin timer (1.5 minutes) when the inside turns pink.
10. Set slide on spin coater, apply vacuum and coat slides with TI Prime, then spin.
11. Immediately put slides on 120°C heating plate face-up for 2 minutes.
12. Take back to spin coater and coat with polystyrene (0.5%, high MW).
13. Assemble slide chambers with two layers of pre-cut parafilm (treated sides face inwards), put assembled slides on heating plate and melt parafilm using glass rod.
14. Put slides into a petri dish, wet the kimwipe with water and add anti-digoxigenin to slide channel and make petri dish airtight using parafilm, leave at 37°C overnight.
15. The next morning, wash slides with blocking buffer (20mM HEPES pH 7.5, 100mM KCl, 5mM MgCl₂, 10% Tween, 10mg/ml BSA, 2mM DTT) and leave to block for 1.5 hours.
16. Wash with your chosen reaction buffer and store in petri dish (with kimwipe and water), covered with parafilm to seal, in the fridge.
17. Before use, add input/output wells using glue.

Chapter II: Lambda-Red Recombination Protocol

Notes:

- Always double check incubation temperatures. pKD46 and pcp20 have temperature-sensitive origins of replication. If cells are incubated at higher temperatures the plasmid will be lost.
- Make glycerol stocks of all the intermediate steps.

(Steps 1 and 2 can be performed in parallel)

Step 1: Transform host strain with pKD46

1. Pick a single colony of *E. coli* MG1655 and inoculate in 5ml of LB. Grow overnight at 37°C.
2. Start a new culture by performing a 1:100 dilution (50µl) of the overnight culture in 5ml of fresh LB. Grow at 37°C until an OD of 0.5-0.8 is reached (2-2.5 hours). Place one electroporation cuvette at -20°C.
3. Spin down 1ml of culture for 5 min at 6000 rpm at 4°C. Discard the supernatant and resuspend the cells in 1ml of cold MiliQ water. Keep the cells on ice all the time. Repeat this step 2 times. After the second time, resuspend the cells in 100µl of cold MQ water.
4. Add 1µl (50-70ng) of pKD46 to the cell suspension. Mix gently and transfer to the cold electroporation cuvette.
5. Dry the cuvette with a tissue and place in the electroporator (select EC 2 and ms). Electroporate cells (ms should be between 5.0-6.0) and immediately add 900µl of prewarmed LB (at 30°C). Incubate for 1 hour at 30°C.
6. After incubation, plate 100µl on a LB + Ampicilin plate. Spin down the rest of the cells for 5 min at 6000 rpm at room temperature and remove most of the supernatant (leave 50-

100µl). Resuspend the cells and plate on a LB + Ampicilin plate. Incubate the plates overnight at 30°C.

7. Pick 2 colonies and purify on a new LB + Ampicilin plate. Grow at 30°C. Store plates at 4°C.

Step 2: Kanamycin cassette amplification

1. Perform a PCR amplification of the kanamycin cassette using Phusion High fidelity polymerase. Use the following protocol:

Reagent	Concentration (Stock)	Volume
MQ		86.5µl
HF Phusion Buffer	5X	30µl
dNTPs	2mM	15µl
Forward Primer	10mM	6.0µl
Reverse Primer	10mM	6.0µl
Phusion Polymerase	2U/µl	1.5µl

Split the Mix in three PCR tubes (49µl). Close the last one and label as negative control. Add 1µl of pKD13 (1-5ng) to each of the other tubes and start the PCR reaction using the following program:

Initial denaturation 98°C 2 min

Cycles (30):

– Denaturation	98°C 10 s
– Annealing	60°C 30 s
– Extension	72°C 1 min
Final extension	72°C 10 min
Hold	10°C

- Analyse the PCR products in a 1% agarose gel. A product of approximately 1.4 kbp should be observed.
- Add 1µl of DpnI (20U/µl) to each of the tubes and incubate for 1 hour at 37°C.
- Mix the two tubes and purify using a PCR purification Kit. IMPORTANT: Perform the final elution in 30-35µl of MilliQ water (Do not use elution buffer).
- Store the purified PCR product at -20°C.

Step 3: Kanamycin cassette recombination

- Pick a single colony of *E. coli* MG1655 + pKD46 and inoculate in 5ml of LB. Grow overnight at 30°C. You can use this culture the next day to make glycerol stocks.
- Start a new culture by performing a 1:100 dilution (50µl) of the overnight culture in 5ml of fresh LB. Grow at 30°C until an OD of 0.3-0.5 is reached (2-2.5 hours). Induce production of recombination proteins by adding 5µl of 20% arabinose. Place back in the incubator for 20 minutes. Place one electroporation cuvette at -20°C.
- Spin down 4ml of culture for 5 minutes at 6000 rpm at 4°C. Discard the supernatant and resuspend cells 4ml of cold MQ water. Keep the cells on ice all the time. Repeat this step 2 times. After the second time resuspend the cells in 10 µl of cold MQ water.

8. Add 3-5 μ l of PCR product from step 2. Mix gently and transfer to the cold electroporation cuvette.
9. Dry the cuvette with a tissue and place in the electroporator (select EC 2 and ms). Electroporate cells (ms should be between 4.5-6.0) and immediately add 900 μ l of prewarmed LB (at 37°C). Incubate for 3 hours at 37°C.
10. After incubation, plate 100 μ l on a LB + Kanamycin plate. Spin down the rest of the cells for 5 minutes at 6000 rpm at room temperature and remove most of the supernatant (leave 50-100 μ l). Resuspend the cells and plate on a LB + Kanamycin plate. Incubate the plates overnight at 37°C.
4. Verify successful insertion of the kanamycin cassette by colony PCR using the control primers. Check 5-7 colonies. Include a negative control (no template) and a WT control (colony PCR of MG1655 WT). Use a LB + Kanamycin plate to grow the selected colonies. Incubate this plate at 37°C. If the colony PCR is performed in the early morning, the plate of selected colonies should have grown by the late afternoon. If not leave overnight.
5. Purify 2 colonies on a LB + Kanamycin plate. Grow at 37°C. pKD46 should be lost after the cells are grown at 37°C. Check successful loss of pKD46 by plating the selected colonies also on LB + ampicillin (no growth).

Step 3: Kanamycin cassette excision

1. Pick a single colony of *E. coli* MG1655 x::kn and inoculate in 5ml of LB. Grow overnight at 30°C. You can use this culture the next day to make glycerol stocks.
2. Start a new culture by performing a 1:100 dilution (50 μ l) of the overnight culture in 5ml of fresh LB. Grow at 37°C until an OD of 0.5-0.8 is reached (2-2.5 hours). Place one electroporation cuvette at -20°C.

3. Spin down 1ml of culture for 5 minutes at 6000 rpm at 4°C. Discard the supernatant and resuspend cells 1ml of cold MQ water. Keep the cells on ice all the time. Repeat this step 2 times. After the second time resuspend the cells in 100µl of cold MiliQ water.
4. Add 1µl (50-70ng) of pKD46 to the cell suspension. Mix gently and transfer to the cold electroporation cuvette.
5. Dry the cuvette with a tissue and place in the electroporator (select EC 2 and ms). Electroporate cells (ms should be between 5.0-6.0) and immediately add 900µl of prewarmed LB (at 30°C). Incubate for 1 hour at 30°C.
6. After incubation, plate 100µl on a LB + Ampicilin plate. Spin down the rest of the cells for 5 min at 6000 rpm at room temperature and remove most of the supernatant (leave 50-100µl). Resuspend the cells and plate on a LB + Ampicilin plate. Incubate the plates overnight at 30°C.
7. Verify successful excision of the kanamycin cassette by colony PCR using the control primers. Check 5-7 colonies. Include a negative control (no template), a WT control (colony PCR of MG1655 WT) and a x:Kn control (colony of MG1655 x:kn). Use a LB (no antibiotic) plate to grow the selected colonies. Incubate this plate at 37°C. If the colony PCR is performed in the early morning, the plate of selected colonies should have grown by the late afternoon. If not leave overnight.
8. Purify 2 colonies on a LB plate. Grow at 37°C. pKD46 should be lost after the cells are grown at 37°C. Check successful loss of pKD46 by plating the selected colonies also on LB + ampicillin (No growth).
9. Pick a single colony of *E. coli* MG1655 ΔX and inoculate in 5ml of LB. Grow overnight at 30°C. You can use this culture the next day to make glycerol stocks.

Section V: Publications

Portman, J.R. and Strick, T.R. *Transcription-Coupled Repair and Complex Biology*. J Mol Biol, 2018. **430**(22): p. 4496-4512.

Strick, T.R. and Portman, J.R. *Transcription-Coupled Repair: From Cells to Single Molecules and Back Again*. J Mol Biol, 2019. **431**(20): p. 4093-4102.

Portman, J.R., Brouwer, G.M., Bollins, J., Savery, N.J. and Strick, T.R. *Cotranscriptional R-loop formation by Mfd involves topological partitioning of DNA*. Proceedings of the National Academy of Sciences of the United States of America, 2021. **118**(15).

Portman, J.R., Qayyum, M.Z., Murakami, K.S. and Strick, T.R. *On the stability of stalled RNA polymerase and its removal by RapA*. Manuscript submitted for publication, 2021.

Section VI: Bibliography

1. Sakmann, B. and E. Neher, *Patch clamp techniques for studying ionic channels in excitable membranes*. Annual review of physiology, 1984. **46**.
2. Varongchayakul, N., et al., *Single-molecule protein sensing in a nanopore: a tutorial*. Chemical Society reviews, 2018. **47**(23).
3. Laszlo, A.H., I.M. Derrington, and J.H. Gundlach, *MspA nanopore as a single-molecule tool: From sequencing to SPRNT*. Methods (San Diego, Calif.), 2016. **105**.
4. Wang, M.D., et al., *Force and velocity measured for single molecules of RNA polymerase*. Science (New York, N.Y.), 1998. **282**(5390).
5. Trache, A. and G.A. Meininger, *Atomic force microscopy (AFM)*. Current protocols in microbiology, 2008. **Chapter 2**.
6. Ashkin, A. and J.M. Dziedzic, *Optical levitation of liquid drops by radiation pressure*. Science (New York, N.Y.), 1975. **187**(4181).
7. Kimura, Y. and P.R. Bianco, *Single molecule studies of DNA binding proteins using optical tweezers*. The Analyst, 2006. **131**(8).
8. Liu, B., R.J. Baskin, and S.C. Kowalczykowski, *DNA unwinding heterogeneity by RecBCD results from static molecules able to equilibrate*. Nature, 2013. **500**(7463).
9. Ozcelik, A., et al., *Acoustic tweezers for the life sciences*. Nature methods, 2018. **15**(12).
10. Neuman, K.C. and A. Nagy, *Single-molecule force spectroscopy: optical tweezers, magnetic tweezers and atomic force microscopy*. Nature methods, 2008. **5**(6).
11. Lionnet, T., et al., *Magnetic trap construction*. Cold Spring Harb Protoc, 2012. **2012**(1): p. 133-8.
12. Strick, T.R., et al., *The elasticity of a single supercoiled DNA molecule*. Science, 1996. **271**(5257): p. 1835-7.
13. Gosse, C., T.R. Strick, and D. Kostrz, *Molecular scaffolds: when DNA becomes the hardware for single-molecule investigations*. Current opinion in chemical biology, 2019. **53**.
14. Collins, B.E., et al., *DNA curtains: novel tools for imaging protein-nucleic acid interactions at the single-molecule level*. Methods in cell biology, 2014. **123**.
15. Friedman, L.J., J. Chung, and J. Gelles, *Viewing dynamic assembly of molecular complexes by multi-wavelength single-molecule fluorescence*. Biophysical journal, 2006. **91**(3).
16. Roy, R., S. Hohng, and T. Ha, *A practical guide to single-molecule FRET*. Nature methods, 2008. **5**(6).
17. Kapanidis, A.N. and T. Strick, *Biology, one molecule at a time*. Trends Biochem Sci, 2009. **34**(5): p. 234-43.
18. Graves, E.T., et al., *A dynamic DNA-repair complex observed by correlative single-molecule nanomanipulation and fluorescence*. Nat Struct Mol Biol, 2015. **22**(6): p. 452-7.
19. Deschout, H., et al., *Complementarity of PALM and SOFI for super-resolution live-cell imaging of focal adhesions*. Nature communications, 2016. **7**.
20. Watson, J.D. and F.H. Crick, *Molecular structure of nucleic acids; a structure for deoxyribose nucleic acid*. Nature, 1953. **171**(4356): p. 737-8.
21. Berg, J.M., J.L. Tymoczko, and L. Stryer, *Biochemistry*. 5 ed. 2002: W H Freeman.
22. Strick, T.R., et al., *Behavior of supercoiled DNA*. Biophys J, 1998. **74**(4): p. 2016-28.
23. Bustamante, C., et al., *Entropic elasticity of lambda-phage DNA*. Science (New York, N.Y.), 1994. **265**(5178).
24. Fuller, F.B., *The Writhing Number of a Space Curve*. 1971.
25. Crick, F.H., *Linking numbers and nucleosomes*. Proc Natl Acad Sci U S A, 1976. **73**(8): p. 2639-43.
26. *DNA supercoil*. [01/03/2021]; Available from: https://en.wikipedia.org/wiki/DNA_supercoil.
27. Kriegel, F., et al., *Probing the salt dependence of the torsional stiffness of DNA by multiplexed magnetic torque tweezers*. Nucleic acids research, 2017. **45**(10).
28. Odijk, T., *Polyelectrolytes near the rod limit*. Journal of Polymer Science, 1977. **15**: p. 477-483.
29. Allemand, J.F., et al., *Stretched and overwound DNA forms a Pauling-like structure with exposed bases*. Proceedings of the National Academy of Sciences of the United States of America, 1998. **95**(24).
30. Dixon, J.R., et al., *Topological domains in mammalian genomes identified by analysis of chromatin interactions*. Nature, 2012. **485**(7398).

31. Sinden, R.R., J.O. Carlson, and D.E. Pettijohn, *Torsional tension in the DNA double helix measured with trimethylpsoralen in living E. coli cells: analogous measurements in insect and human cells.* Cell, 1980. **21**(3).
32. Kouzine, F., L. Baranello, and D. Levens, *The Use of Psoralen Photobinding to Study Transcription-Induced Supercoiling.* Methods in molecular biology (Clifton, N.J.), 2018. **1703**.
33. Ramani, V., J. Shendure, and Z. Duan, *Understanding Spatial Genome Organization: Methods and Insights.* Genomics, proteomics & bioinformatics, 2016. **14**(1).
34. He, B., et al., *Global view of enhancer-promoter interactome in human cells.* Proceedings of the National Academy of Sciences of the United States of America, 2014. **111**(21).
35. Rosengren, S., P.M. Henson, and G.S. Worthen, *Migration-associated volume changes in neutrophils facilitate the migratory process in vitro.* Am J Physiol, 1994. **267**(6 Pt 1): p. C1623-32.
36. Kubitschek, H.E., *Cell volume increase in Escherichia coli after shifts to richer media.* J Bacteriol, 1990. **172**(1): p. 94-101.
37. Kornberg, R.D., *Chromatin structure: a repeating unit of histones and DNA.* Science (New York, N.Y.), 1974. **184**(4139).
38. Ou, H.D., et al., *ChromEMT: Visualizing 3D chromatin structure and compaction in interphase and mitotic cells.* 2017.
39. Heitz, E., *Das heterochromatin der moose.* 1928: Bornträger.
40. Scott, K.C., S.L. Merrett, and H.F. Willard, *A heterochromatin barrier partitions the fission yeast centromere into discrete chromatin domains.* Curr Biol, 2006. **16**(2): p. 119-29.
41. Nuebler, J., et al., *Chromatin organization by an interplay of loop extrusion and compartmental segregation.* Proc Natl Acad Sci U S A, 2018. **115**(29): p. 6697-706.
42. Arzate-Mejía, R.G., et al., *In situ dissection of domain boundaries affect genome topology and gene transcription in Drosophila.* Nature communications, 2020. **11**(1).
43. Min, I.M., et al., *Regulating RNA polymerase pausing and transcription elongation in embryonic stem cells.* Genes & development, 2011. **25**(7).
44. Ebersole, T., et al., *tRNA genes protect a reporter gene from epigenetic silencing in mouse cells.* Cell cycle (Georgetown, Tex.), 2011. **10**(16).
45. Schmidt, D., et al., *Waves of retrotransposon expansion remodel genome organization and CTCF binding in multiple mammalian lineages.* Cell, 2012. **148**(1-2).
46. Espeli, O., R. Mercier, and F. Boccard, *DNA dynamics vary according to macrodomain topography in the E. coli chromosome.* Molecular microbiology, 2008. **68**(6).
47. Badrinarayanan, A., T.B. Le, and M.T. Laub, *Bacterial chromosome organization and segregation.* Annual review of cell and developmental biology, 2015. **31**.
48. Postow, L., et al., *Topological domain structure of the Escherichia coli chromosome.* Genes & development, 2004. **18**(14).
49. Ladouceur, A.M., et al., *Clusters of bacterial RNA polymerase are biomolecular condensates that assemble through liquid-liquid phase separation.* Proceedings of the National Academy of Sciences of the United States of America, 2020. **117**(31).
50. Rice, P.A., et al., *Crystal structure of an IHF-DNA complex: a protein-induced DNA U-turn.* Cell, 1996. **87**(7): p. 1295-306.
51. Murphy, L.D. and S.B. Zimmerman, *Isolation and characterization of spermidine nucleoids from Escherichia coli.* J Struct Biol, 1997. **119**(3): p. 321-35.
52. Bradley, M.D., et al., *Effects of Fis on Escherichia coli gene expression during different growth stages.* Microbiology, 2007. **153**(Pt 9): p. 2922-2940.
53. Brahms, J.G., et al., *Activation and inhibition of transcription by supercoiling.* J Mol Biol, 1985. **181**(4): p. 455-65.
54. Gan, W., et al., *R-loop-mediated genomic instability is caused by impairment of replication fork progression.* Genes Dev, 2011. **25**(19): p. 2041-56.
55. Tse-Dinh, Y.C., *Regulation of the Escherichia coli DNA topoisomerase I gene by DNA supercoiling.* Nucleic Acids Res, 1985. **13**(13): p. 4751-63.
56. Menzel, R. and M. Gellert, *Modulation of transcription by DNA supercoiling: a deletion analysis of the Escherichia coli gyrA and gyrB promoters.* Proc Natl Acad Sci U S A, 1987. **84**(12): p. 4185-9.
57. Bush, N.G., K. Evans-Roberts, and A. Maxwell, *DNA Topoisomerases.* EcoSal Plus, 2015. **6**(2).

58. Hooper, D.C., *Bacterial topoisomerases, anti-topoisomerases, and anti-topoisomerase resistance*. Clinical infectious diseases : an official publication of the Infectious Diseases Society of America, 1998. **27 Suppl 1**.
59. Thomas, M., R.L. White, and R.W. Davis, *Hybridization of RNA to double-stranded DNA: formation of R-loops*. Proc Natl Acad Sci U S A, 1976. **73**(7): p. 2294-8.
60. Wahba, L., S.K. Gore, and D. Koshland, *The homologous recombination machinery modulates the formation of RNA-DNA hybrids and associated chromosome instability*. Elife, 2013. **2**: p. e00505.
61. Reaban, M.E. and J.A. Griffin, *Induction of RNA-stabilized DNA conformers by transcription of an immunoglobulin switch region*. Nature, 1990. **348**(6299): p. 342-4.
62. Baker, T.A. and A. Kornberg, *Transcriptional activation of initiation of replication from the E. coli chromosomal origin: an RNA-DNA hybrid near oriC*. Cell, 1988. **55**(1): p. 113-23.
63. Wimberly, H., et al., *R-loops and nicks initiate DNA breakage and genome instability in non-growing Escherichia coli*. Nat Commun, 2013. **4**: p. 2115.
64. Szczelkun, M.D., et al., *Direct observation of R-loop formation by single RNA-guided Cas9 and Cascade effector complexes*. Proc Natl Acad Sci U S A, 2014. **111**(27): p. 9798-803.
65. Crossley, M.P., M. Bocek, and K.A. Cimprich, *R-Loops as Cellular Regulators and Genomic Threats*. Mol Cell, 2019. **73**(3): p. 398-411.
66. Skourti-Stathaki, K., N.J. Proudfoot, and N. Gromak, *Human senataxin resolves RNA/DNA hybrids formed at transcriptional pause sites to promote Xrn2-dependent termination*. Molecular cell, 2011. **42**(6).
67. Ginno, P.A., et al., *R-loop formation is a distinctive characteristic of unmethylated human CpG island promoters*. Molecular cell, 2012. **45**(6).
68. Frederico, L.A., T.A. Kunkel, and B.R. Shaw, *A sensitive genetic assay for the detection of cytosine deamination: determination of rate constants and the activation energy*. Biochemistry, 1990. **29**(10).
69. Aguilera, A. and T. García-Muse, *R loops: from transcription byproducts to threats to genome stability*. Molecular cell, 2012. **46**(2).
70. Merrikh, H., et al., *Replication-transcription conflicts in bacteria*. Nature reviews. Microbiology, 2012. **10**(7).
71. Brüning, J.G. and K.J. Marians, *Replisome bypass of transcription complexes and R-loops*. Nucleic acids research, 2020. **48**(18).
72. Chedin, F. and C.J. Benham, *Emerging roles for R-loop structures in the management of topological stress*. The Journal of biological chemistry, 2020. **295**(14).
73. Drolet, M., X. Bi, and L.F. Liu, *Hypernegative supercoiling of the DNA template during transcription elongation in vitro*. The Journal of biological chemistry, 1994. **269**(3).
74. Drolet, M., et al., *Overexpression of RNase H partially complements the growth defect of an Escherichia coli delta topA mutant: R-loop formation is a major problem in the absence of DNA topoisomerase I*. Proceedings of the National Academy of Sciences of the United States of America, 1995. **92**(8).
75. Strick, T.R., V. Croquette, and D. Bensimon, *Homologous pairing in stretched supercoiled DNA*. Proc Natl Acad Sci U S A, 1998. **95**(18): p. 10579-83.
76. Browning, D.F. and S.J. Busby, *Local and global regulation of transcription initiation in bacteria*. Nat Rev Microbiol, 2016. **14**(10): p. 638-50.
77. Feklistov, A., et al., *Bacterial sigma factors: a historical, structural, and genomic perspective*. Annu Rev Microbiol, 2014. **68**: p. 357-76.
78. Travers, A.A. and R.R. Burgess, *Cyclic re-use of the RNA polymerase sigma factor*. Nature, 1969. **222**(5193): p. 537-40.
79. Cowing, D.W., et al., *Consensus sequence for Escherichia coli heat shock gene promoters*. Proc Natl Acad Sci U S A, 1985. **82**(9): p. 2679-83.
80. Darst, S.A., *Bacterial RNA polymerase*. Curr Opin Struct Biol, 2001. **11**(2): p. 155-62.
81. Batada, N.N., et al., *Diffusion of nucleoside triphosphates and role of the entry site to the RNA polymerase II active center*. 2004.
82. Mishanina, T.V., et al., *Trigger loop of RNA polymerase is a positional, not acid-base, catalyst for both transcription and proofreading*. Proc Natl Acad Sci U S A, 2017. **114**(26): p. E5103-e5112.

83. Abbondanzieri, E.A., et al., *Direct observation of base-pair stepping by RNA polymerase*. Nature, 2005. **438**(7067): p. 460-5.
84. Noe Gonzalez, M., D. Blears, and J.Q. Svejstrup, *Causes and consequences of RNA polymerase II stalling during transcript elongation*. Nature reviews. Molecular cell biology, 2021. **22**(1).
85. Lawley, P.D., *Effects of some chemical mutagens and carcinogens on nucleic acids*. Prog Nucleic Acid Res Mol Biol, 1966. **5**: p. 89-131.
86. Rosenberger, R.F. and G. Foskett, *An estimate of the frequency of in vivo transcriptional errors at a nonsense codon in Escherichia coli*. Mol Gen Genet, 1981. **183**(3): p. 561-3.
87. Workentine, M.L., et al., *Phenotypic heterogeneity of Pseudomonas aeruginosa populations in a cystic fibrosis patient*. PLoS One, 2013. **8**(4): p. e60225.
88. Borukhov, S., V. Sagitov, and A. Goldfarb, *Transcript Cleavage Factors From E. Coli*. Cell, 1993. **72**(3).
89. Revyakin, A., et al., *Abortive initiation and productive initiation by RNA polymerase involve DNA scrunching*. Science, 2006. **314**(5802): p. 1139-43.
90. Larson, M.H., et al., *A pause sequence enriched at translation start sites drives transcription dynamics in vivo*. Science, 2014. **344**(6187): p. 1042-7.
91. Laptenko, O., et al., *Transcript cleavage factors GreA and GreB act as transient catalytic components of RNA polymerase*. Embo j, 2003. **22**(23): p. 6322-34.
92. Bubunenko, M.G., et al., *A Cre Transcription Fidelity Reporter Identifies GreA as a Major RNA Proofreading Factor in Escherichia coli*. Genetics, 2017. **206**(1): p. 179-187.
93. Touloukhonov, I., I. Artsimovitch, and R. Landick, *Allosteric control of RNA polymerase by a site that contacts nascent RNA hairpins*. Science, 2001. **292**(5517): p. 730-3.
94. Turtola, M. and G.A. Belogurov, *NusG inhibits RNA polymerase backtracking by stabilizing the minimal transcription bubble*. Elife, 2016. **5**.
95. Burmann, B.M., et al., *A NusE:NusG complex links transcription and translation*. Science, 2010. **328**(5977): p. 501-4.
96. Park, J.S., M.T. Marr, and J.W. Roberts, *E. coli Transcription repair coupling factor (Mfd protein) rescues arrested complexes by promoting forward translocation*. Cell, 2002. **109**(6): p. 757-67.
97. Ragheb, M.N., et al., *Mfd regulates RNA polymerase association with hard-to-transcribe regions in vivo, especially those with structured RNAs*. Proceedings of the National Academy of Sciences of the United States of America, 2021. **118**(1).
98. Martin, H.A., et al., *Mfd Affects Global Transcription and the Physiology of Stressed Bacillus subtilis Cells*. Frontiers in microbiology, 2021. **12**.
99. Washburn, R.S., Y. Wang, and M.E. Gottesman, *Role of E.coli transcription-repair coupling factor Mfd in Nun-mediated transcription termination*. J Mol Biol, 2003. **329**(4): p. 655-62.
100. Willing, S.E., et al., *Increased toxin expression in a Clostridium difficile mfd mutant*. BMC Microbiol, 2015. **15**: p. 280.
101. Zalieckas, J.M., et al., *Transcription-repair coupling factor is involved in carbon catabolite repression of the Bacillus subtilis hut and gnt operons*. Mol Microbiol, 1998. **27**(5): p. 1031-8.
102. Sanders, K., et al., *The structure and function of an RNA polymerase interaction domain in the PcrA/UvrD helicase*. Nucleic acids research, 2017. **45**(7).
103. Guy, C.P., et al., *Rep provides a second motor at the replisome to promote duplication of protein-bound DNA*. Molecular cell, 2009. **36**(4).
104. Liu, B., Y. Zuo, and T.A. Steitz, *Structural basis for transcription reactivation by RapA*. Proc Natl Acad Sci U S A, 2015. **112**(7): p. 2006-10.
105. Muzzin, O., et al., *Disruption of Escherichia coli hepA, an RNA polymerase-associated protein, causes UV sensitivity*. J Biol Chem, 1998. **273**(24): p. 15157-61.
106. Sukhodolets, M.V. and D.J. Jin, *RapA, a novel RNA polymerase-associated protein, is a bacterial homolog of SWI2/SNF2*. J Biol Chem, 1998. **273**(12): p. 7018-23.
107. Shaw, G., et al., *Structure of RapA, a Swi2/Snf2 protein that recycles RNA polymerase during transcription*. Structure, 2008. **16**(9): p. 1417-27.
108. Sukhodolets, M.V. and D.J. Jin, *Interaction between RNA Polymerase and RapA, a Bacterial Homolog of the SWI/SNF Protein Family*. 2000.

109. Kakar, S., et al., *Allosteric Activation of Bacterial Swi2/Snf2 (Switch/Sucrose Non-fermentable) Protein RapA by RNA Polymerase: BIOCHEMICAL AND STRUCTURAL STUDIES*. J Biol Chem, 2015. **290**(39): p. 23656-69.
110. Sukhodolets, M.V., et al., *RapA, a bacterial homolog of SWI2/SNF2, stimulates RNA polymerase recycling in transcription*. 2001.
111. Ma, J., L. Bai, and M.D. Wang, *Transcription under torsion*. Science, 2013. **340**(6140): p. 1580-3.
112. Kotlajich, M.V., et al., *Bridged filaments of histone-like nucleoid structuring protein pause RNA polymerase and aid termination in bacteria*. Elife, 2015. **4**.
113. Janissen, R., et al., *Global DNA Compaction in Stationary-Phase Bacteria Does Not Affect Transcription*. Cell, 2018. **174**(5): p. 1188-1199.e14.
114. Epshtein, V., et al., *Transcription through the roadblocks: the role of RNA polymerase cooperation*. Embo j, 2003. **22**(18): p. 4719-27.
115. Epshtein, V. and E. Nudler, *Cooperation between RNA polymerase molecules in transcription elongation*. Science, 2003. **300**(5620): p. 801-5.
116. Sellitti, M.A., P.A. Pavco, and D.A. Steege, *lac repressor blocks in vivo transcription of lac control region DNA*. Proc Natl Acad Sci U S A, 1987. **84**(10): p. 3199-203.
117. Le, T.T., et al., *Mfd Dynamically Regulates Transcription via a Release and Catch-Up Mechanism*. Cell, 2017.
118. Belitsky, B.R. and A.L. Sonenshein, *Roadblock repression of transcription by Bacillus subtilis CodY*. J Mol Biol, 2011. **411**(4): p. 729-43.
119. Mirkin, E.V. and S.M. Mirkin, *Mechanisms of transcription-replication collisions in bacteria*. Mol Cell Biol, 2005. **25**(3): p. 888-95.
120. Rocha, E.P. and A. Danchin, *Essentiality, not expressiveness, drives gene-strand bias in bacteria*. Nat Genet, 2003. **34**(4): p. 377-8.
121. Srivatsan, A., et al., *Co-orientation of replication and transcription preserves genome integrity*. PLoS genetics, 2010. **6**(1).
122. Liu, L.F. and J.C. Wang, *Supercoiling of the DNA template during transcription*. Proc Natl Acad Sci U S A, 1987. **84**(20): p. 7024-7.
123. Wu, H.Y., et al., *Transcription generates positively and negatively supercoiled domains in the template*. Cell, 1988. **53**(3): p. 433-40.
124. Shen, B.A. and R. Landick, *Transcription of Bacterial Chromatin*. J Mol Biol, 2019. **431**(20): p. 4040-4066.
125. Fiil, N. and J.D. Friesen, *Isolation of "relaxed" mutants of Escherichia coli*. J Bacteriol, 1968. **95**(2): p. 729-31.
126. Haseltine, W.A. and R. Block, *Synthesis of guanosine tetra- and pentaphosphate requires the presence of a codon-specific, uncharged transfer ribonucleic acid in the acceptor site of ribosomes*. Proc Natl Acad Sci U S A, 1973. **70**(5): p. 1564-8.
127. Paul, B.J., et al., *DksA: a critical component of the transcription initiation machinery that potentiates the regulation of rRNA promoters by ppGpp and the initiating NTP*. Cell, 2004. **118**(3): p. 311-22.
128. Ross, W., et al., *ppGpp Binding to a Site at the RNAP-DksA Interface Accounts for Its Dramatic Effects on Transcription Initiation during the Stringent Response*. Mol Cell, 2016. **62**(6): p. 811-823.
129. Zuo, Y., Y. Wang, and T.A. Steitz, *The mechanism of E. coli RNA polymerase regulation by ppGpp is suggested by the structure of their complex*. Mol Cell, 2013. **50**(3): p. 430-6.
130. Zhang, Y., et al., *DksA guards elongating RNA polymerase against ribosome-stalling-induced arrest*. Mol Cell, 2014. **53**(5): p. 766-78.
131. Lindahl, T., *Instability and decay of the primary structure of DNA*. Nature, 1993. **362**(6422): p. 709-15.
132. Dianov, G., A. Price, and T. Lindahl, *Generation of single-nucleotide repair patches following excision of uracil residues from DNA*. Mol Cell Biol, 1992. **12**(4): p. 1605-12.
133. Wallace, S.S., *AP endonucleases and DNA glycosylases that recognize oxidative DNA damage*. Environ Mol Mutagen, 1988. **12**(4): p. 431-77.

134. Lindahl, T. and B. Nyberg, *Heat-induced deamination of cytosine residues in deoxyribonucleic acid*. *Biochemistry*, 1974. **13**(16): p. 3405-10.
135. Henner, W.D., S.M. Grunberg, and W.A. Haseltine, *Enzyme action at 3' termini of ionizing radiation-induced DNA strand breaks*. *J Biol Chem*, 1983. **258**(24): p. 15198-205.
136. Rydberg, B. and T. Lindahl, *Nonenzymatic methylation of DNA by the intracellular methyl group donor S-adenosyl-L-methionine is a potentially mutagenic reaction*. *Embo j*, 1982. **1**(2): p. 211-6.
137. Mehta, A. and J.E. Haber, *Sources of DNA Double-Strand Breaks and Models of Recombinational DNA Repair*, in *Cold Spring Harb Perspect Biol*. 2014.
138. Pourquier, P. and Y. Pommier, *Topoisomerase I-mediated DNA damage*. *Adv Cancer Res*, 2001. **80**: p. 189-216.
139. Andersen, S., et al., *Incorporation of dUMP into DNA is a major source of spontaneous DNA damage, while excision of uracil is not required for cytotoxicity of fluoropyrimidines in mouse embryonic fibroblasts*. *Carcinogenesis*, 2005. **26**(3): p. 547-55.
140. Mah, M.C., et al., *Mutations induced by aminofluorene-DNA adducts during replication in human cells*. *Carcinogenesis*, 1989. **10**(12): p. 2321-8.
141. Cosman, M., et al., *Solution conformation of the major adduct between the carcinogen (+)-anti-benzof[a]pyrene diol epoxide and DNA*. *Proc Natl Acad Sci U S A*, 1992. **89**(5): p. 1914-8.
142. Henner, W.D., S.M. Grunberg, and W.A. Haseltine, *Sites and structure of gamma radiation-induced DNA strand breaks*. *J Biol Chem*, 1982. **257**(19): p. 11750-4.
143. Vignard, J., G. Mirey, and B. Salles, *Ionizing-radiation induced DNA double-strand breaks: a direct and indirect lighting up*. *Radiother Oncol*, 2013. **108**(3): p. 362-9.
144. Peak, M.J. and J.G. Peak, *DNA-to-protein crosslinks and backbone breaks caused by far- and near-ultraviolet, and visible light radiations in mammalian cells*. *Basic Life Sci*, 1986. **38**: p. 193-202.
145. Epe, B., *Genotoxicity of singlet oxygen*. *Chem Biol Interact*, 1991. **80**(3): p. 239-60.
146. Beukers, R. and W. Berends, *The effects of U.V.-irradiation on nucleic acids and their components*. *Biochimica et biophysica acta*, 1961. **49**: p. 181-189.
147. Chatterjee, N. and G.C. Walker, *Mechanisms of DNA damage, repair, and mutagenesis*. *Environ Mol Mutagen*, 2017. **58**(5): p. 235-263.
148. Sinha, R.P. and D.P. Hader, *UV-induced DNA damage and repair: a review*. *Photochem Photobiol Sci*, 2002. **1**(4): p. 225-36.
149. Sale, J.E., *Translesion DNA synthesis and mutagenesis in eukaryotes*. *Cold Spring Harb Perspect Biol*, 2013. **5**(3): p. a012708.
150. Sancar, G.B., F.W. Smith, and A. Sancar, *Binding of Escherichia coli DNA photolyase to UV-irradiated DNA*. *Biochemistry*, 1985. **24**(8).
151. Sancar, A., F.W. Smith, and G.B. Sancar, *Purification of Escherichia coli DNA photolyase*. *The Journal of biological chemistry*, 1984. **259**(9).
152. Yi, C. and C. He, *DNA repair by reversal of DNA damage*. *Cold Spring Harb Perspect Biol*, 2013. **5**(1): p. a012575.
153. Demple, B., et al., *Active site and complete sequence of the suicidal methyltransferase that counters alkylation mutagenesis*. *Proc Natl Acad Sci U S A*, 1985. **82**(9): p. 2688-92.
154. Nakabeppu, Y. and M. Sekiguchi, *Regulatory mechanisms for induction of synthesis of repair enzymes in response to alkylating agents: ada protein acts as a transcriptional regulator*. *Proceedings of the National Academy of Sciences of the United States of America*, 1986. **83**(17).
155. Pang, D., et al., *Ku proteins join DNA fragments as shown by atomic force microscopy*. *Cancer Res*, 1997. **57**(8): p. 1412-5.
156. Ma, Y., et al., *Hairpin opening and overhang processing by an Artemis/DNA-dependent protein kinase complex in nonhomologous end joining and V(D)J recombination*. *Cell*, 2002. **108**(6): p. 781-94.
157. Dillingham, M.S., M. Spies, and S.C. Kowalczykowski, *RecBCD enzyme is a bipolar DNA helicase*. *Nature*, 2003. **423**(6942): p. 893-7.
158. Roman, L.J., D.A. Dixon, and S.C. Kowalczykowski, *RecBCD-dependent joint molecule formation promoted by the Escherichia coli RecA and SSB proteins*. *Proc Natl Acad Sci U S A*, 1991. **88**(8): p. 3367-71.

159. Mi, J. and G.M. Kupfer, *The Fanconi anemia core complex associates with chromatin during S phase*. Blood, 2005. **105**(2): p. 759-66.
160. Bessho, T., D. Mu, and A. Sancar, *Initiation of DNA interstrand cross-link repair in humans: the nucleotide excision repair system makes dual incisions 5' to the cross-linked base and removes a 22- to 28-nucleotide-long damage-free strand*. Mol Cell Biol, 1997. **17**(12): p. 6822-30.
161. Raschle, M., et al., *Mechanism of replication-coupled DNA interstrand crosslink repair*. Cell, 2008. **134**(6): p. 969-80.
162. Laengle-Rouault, F., G. Maenhaut-Michel, and M. Radman, *GATC sequence and mismatch repair in Escherichia coli*. Embo j, 1986. **5**(8): p. 2009-13.
163. Lenhart, J.S., et al., *Mismatch repair in Gram-positive bacteria*. Res Microbiol, 2016. **167**(1): p. 4-12.
164. Odell, I.D., S.S. Wallace, and D.S. Pederson, *Rules of engagement for base excision repair in chromatin*. J Cell Physiol, 2013. **228**(2): p. 258-66.
165. Boyce, R.P. and P. Howard-Flanders, *RELEASE OF ULTRAVIOLET LIGHT-INDUCED THYMINE DIMERS FROM DNA IN E. COLI K-12*. Proc Natl Acad Sci U S A, 1964. **51**: p. 293-300.
166. Selby, C.P., E.M. Witkin, and A. Sancar, *Escherichia coli mfd mutant deficient in "mutation frequency decline" lacks strand-specific repair: in vitro complementation with purified coupling factor*. Proc Natl Acad Sci U S A, 1991. **88**(24): p. 11574-8.
167. Witkin, E., *It was heaven: an interview with Evelyn Witkin*. Interview by Jane Gitschier. PLoS genetics, 2012. **8**(10).
168. Luria, S.E. and M. Delbruck, *Mutations of Bacteria from Virus Sensitivity to Virus Resistance*. Genetics, 1943. **28**(6): p. 491-511.
169. Witkin, E.M., *THE EFFECT OF ACRIFLAVINE ON PHOTOREVERSAL OF LETHAL AND MUTAGENIC DAMAGE PRODUCED IN BACTERIA BY ULTRAVIOLET LIGHT*. Proc Natl Acad Sci U S A, 1963. **50**: p. 425-30.
170. Witkin, E.M., *Radiation-induced mutations and their repair*. Science, 1966. **152**(3727): p. 1345-53.
171. Walker, J.E., O. Luyties, and T.J. Santangelo, *Factor-dependent archaeal transcription termination*. Proc Natl Acad Sci U S A, 2017. **114**(33): p. 6767-6773.
172. Tijsterman, M. and J. Brouwer, *Rad26, the yeast homolog of the cockayne syndrome B gene product, counteracts inhibition of DNA repair due to RNA polymerase II transcription*. J Biol Chem, 1999. **274**(3): p. 1199-202.
173. van Hoffen, A., et al., *Deficient repair of the transcribed strand of active genes in Cockayne's syndrome cells*. Nucleic Acids Res, 1993. **21**(25): p. 5890-5.
174. Bockrath, R.C. and J.E. Palmer, *Differential repair of premutational UV-lesions at tRNA genes in E. coli*. Molecular & general genetics : MGG, 1977. **156**(2).
175. Strick, T.R. and J.R. Portman, *Transcription-Coupled Repair: From Cells to Single Molecules and Back Again*. J Mol Biol, 2019.
176. Ohta, T., et al., *Characterization of Trp(+) reversions in Escherichia coli strain WP2uvrA*. Mutagenesis, 2002. **17**(4): p. 313-6.
177. Li, B.H., et al., *Diverse backmutations at an ochre defect in the tyrA gene sequence of E. coli B/r*. Mutation research, 1991. **246**(1).
178. Wade, J.T. and D.C. Grainger, *Pervasive transcription: illuminating the dark matter of bacterial transcriptomes*, in Nat Rev Microbiol. 2014: England. p. 647-53.
179. Adebali, O., et al., *Genome-wide transcription-coupled repair in Escherichia coli is mediated by the Mfd translocase*. Proc Natl Acad Sci U S A, 2017. **114**(11): p. E2116-e2125.
180. Howan, K., et al., *Stopped in its tracks: the RNA polymerase molecular motor as a robust sensor of DNA damage*. DNA Repair (Amst), 2014. **20**: p. 49-57.
181. Deaconescu, A.M., et al., *Structural basis for bacterial transcription-coupled DNA repair*. Cell, 2006. **124**(3): p. 507-20.
182. Manelyte, L., et al., *Regulation and rate enhancement during transcription-coupled DNA repair*. Mol Cell, 2010. **40**(5): p. 714-24.
183. Selby, C.P. and A. Sancar, *Structure and function of transcription-repair coupling factor. I. Structural domains and binding properties*. J Biol Chem, 1995. **270**(9): p. 4882-9.

184. Smith, A.J., M.D. Szczelkun, and N.J. Savery, *Controlling the motor activity of a transcription-repair coupling factor: autoinhibition and the role of RNA polymerase*. Nucleic Acids Res, 2007. **35**(6): p. 1802-11.
185. Selby, C.P. and A. Sancar, *Molecular mechanism of transcription-repair coupling*. Science, 1993. **260**(5104): p. 53-8.
186. Brugger, C., et al., *Molecular determinants for dsDNA translocation by the transcription-repair coupling and evolvability factor Mfd*. Nature communications, 2020. **11**(1).
187. Kang, J.Y., et al., *Structural basis for transcription complex disruption by the Mfd translocase*. eLife, 2021. **10**.
188. Howan, K., et al., *Initiation of transcription-coupled repair characterized at single-molecule resolution*. Nature, 2012. **490**(7420): p. 431-4.
189. Fan, J., et al., *Reconstruction of bacterial transcription-coupled repair at single-molecule resolution*. Nature, 2016. **536**(7615): p. 234-7.
190. Haines, N.M., et al., *Stalled transcription complexes promote DNA repair at a distance*. Proc Natl Acad Sci U S A, 2014. **111**(11): p. 4037-42.
191. Van Houten, B., et al., *DNase I footprint of ABC excinuclease*. J Biol Chem, 1987. **262**(27): p. 13180-7.
192. Mellon, I. and P.C. Hanawalt, *Induction of the Escherichia coli lactose operon selectively increases repair of its transcribed DNA strand*. Nature, 1989. **342**(6245): p. 95-8.
193. Stracy, M., et al., *Single-molecule imaging of UvrA and UvrB recruitment to DNA lesions in living Escherichia coli*. Nat Commun, 2016. **7**: p. 12568.
194. Roberts, J.W., *Termination factor for RNA synthesis*. Nature, 1969. **224**(5225): p. 1168-74.
195. Jain, S., R. Gupta, and R. Sen, *Rho-dependent transcription termination in bacteria recycles RNA polymerases stalled at DNA lesions*. Nat Commun, 2019. **10**.
196. Epshtein, V., et al., *UvrD facilitates DNA repair by pulling RNA polymerase backwards*. Nature, 2014. **505**(7483): p. 372-7.
197. Kamarthapu, V., et al., *ppGpp couples transcription to DNA repair in E. coli*. Science, 2016. **352**(6288): p. 993-6.
198. Ahn, B., *A physical interaction of UvrD with nucleotide excision repair protein UvrB*. Mol Cells, 2000. **10**(5): p. 592-7.
199. Adebali, O., A. Sancar, and C.P. Selby, *Mfd translocase is necessary and sufficient for transcription-coupled repair in Escherichia coli*. J Biol Chem, 2017. **292**(45): p. 18386-18391.
200. Strick, T.R. and N.J. Savery, *Understanding bias in DNA repair*. Proc Natl Acad Sci U S A, 2017. **114**(11): p. 2791-2793.
201. Pakotiprapha, D., et al., *Structure and mechanism of the UvrA-UvrB DNA damage sensor*. Nat Struct Mol Biol, 2012. **19**(3): p. 291-8.
202. Grossman, L., et al., *Repair of DNA-containing pyrimidine dimers*. Faseb j, 1988. **2**(11): p. 2696-701.
203. Mazur, S.J. and L. Grossman, *Dimerization of Escherichia coli UvrA and its binding to undamaged and ultraviolet light damaged DNA*. Biochemistry, 1991. **30**(18): p. 4432-43.
204. Kad, N.M., et al., *Collaborative dynamic DNA scanning by nucleotide excision repair proteins investigated by single- molecule imaging of quantum-dot-labeled proteins*. Mol Cell, 2010. **37**(5): p. 702-13.
205. Barnett, J.T. and N.M. Kad, *Understanding the coupling between DNA damage detection and UvrA's ATPase using bulk and single molecule kinetics*. Faseb j, 2019. **33**(1): p. 763-769.
206. Million-Weaver, S., et al., *An underlying mechanism for the increased mutagenesis of lagging-strand genes in Bacillus subtilis*. Proc Natl Acad Sci U S A, 2015. **112**(10): p. 1096-105.
207. Wagner, K., G.F. Moolenaar, and N. Goosen, *Role of the insertion domain and the zinc-finger motif of Escherichia coli UvrA in damage recognition and ATP hydrolysis*. DNA Repair (Amst), 2011. **10**(5): p. 483-96.
208. Orren, D.K. and A. Sancar, *The (A)BC excinuclease of Escherichia coli has only the UvrB and UvrC subunits in the incision complex*. Proc Natl Acad Sci U S A, 1989. **86**(14): p. 5237-41.
209. Theis, K., et al., *Crystal structure of UvrB, a DNA helicase adapted for nucleotide excision repair*. Embo j, 1999. **18**(24): p. 6899-907.

210. Verhoeven, E.E., et al., *Architecture of nucleotide excision repair complexes: DNA is wrapped by UvrB before and after damage recognition*. *Embo j*, 2001. **20**(3): p. 601-11.
211. Lee, S.J., R.J. Sung, and G.L. Verdine, *Mechanism of DNA Lesion Homing and Recognition by the Uvr Nucleotide Excision Repair System*. Research (Wash D C), 2019. **2019**: p. 5641746.
212. Malta, E., et al., *Functions of base flipping in E. coli nucleotide excision repair*. *DNA Repair (Amst)*, 2008. **7**(10): p. 1647-58.
213. Truglio, J.J., et al., *Structural basis for DNA recognition and processing by UvrB*. *Nat Struct Mol Biol*, 2006. **13**(4): p. 360-4.
214. Sancar, A. and W.D. Rupp, *A novel repair enzyme: UVRABC excision nuclease of Escherichia coli cuts a DNA strand on both sides of the damaged region*. *Cell*, 1983. **33**(1): p. 249-60.
215. Matson, S.W. and J.W. George, *DNA helicase II of Escherichia coli. Characterization of the single-stranded DNA-dependent NTPase and helicase activities*. *J Biol Chem*, 1987. **262**(5): p. 2066-76.
216. Orren, D.K., et al., *Post-incision steps of nucleotide excision repair in Escherichia coli. Disassembly of the UvrBC-DNA complex by helicase II and DNA polymerase I*. *J Biol Chem*, 1992. **267**(2): p. 780-8.
217. Sancar, A. and M.S. Tang, *Nucleotide excision repair*. *Photochem Photobiol*, 1993. **57**(5): p. 905-21.
218. Ghodke, H., H.N. Ho, and A.M. van Oijen, *Single-molecule live-cell imaging visualizes parallel pathways of prokaryotic nucleotide excision repair*. *Nat Commun*, 2020. **11**(1): p. 1477.
219. Ho, H.N., A.M. van Oijen, and H. Ghodke, *Single-molecule imaging reveals molecular coupling between transcription and DNA repair machinery in live cells*. *Nat Commun*, 2020. **11**(1): p. 1478.
220. Ho, H.N., A.M. van Oijen, and H. Ghodke, *The transcription-repair coupling factor Mfd associates with RNA polymerase in the absence of exogenous damage*. *Nat Commun*, 2018. **9**(1): p. 1570.
221. Lee, G.H., et al., *The Helicobacter pylori Mfd protein is important for antibiotic resistance and DNA repair*. *Diagn Microbiol Infect Dis*, 2009. **65**(4): p. 454-6.
222. Ragheb, M. and H. Merrikh, *The enigmatic role of Mfd in replication-transcription conflicts in bacteria*. *DNA Repair (Amst)*, 2019. **81**: p. 102659.
223. Ayora, S., et al., *The Mfd protein of Bacillus subtilis 168 is involved in both transcription-coupled DNA repair and DNA recombination*. *J Mol Biol*, 1996. **256**(2): p. 301-18.
224. Gomez-Marroquin, M., et al., *Stationary-Phase Mutagenesis in Stressed Bacillus subtilis Cells Operates by Mfd-Dependent Mutagenic Pathways*. *Genes (Basel)*, 2016. **7**(7).
225. Han, J., et al., *Key role of Mfd in the development of fluoroquinolone resistance in Campylobacter jejuni*. *PLoS Pathog*, 2008. **4**(6): p. e1000083.
226. Cooper, T.F., E.A. Ostrowski, and M. Travisano, *A negative relationship between mutation pleiotropy and fitness effect in yeast*. *Evolution; international journal of organic evolution*, 2007. **61**(6).
227. Otto, S.P., *Two steps forward, one step back: the pleiotropic effects of favoured alleles*. *Proceedings. Biological sciences*, 2004. **271**(1540).
228. Spivak, G., *Nucleotide excision repair in humans*. *DNA repair*, 2015. **36**.
229. Chen, X., et al., *Kinetic gating mechanism of DNA damage recognition by Rad4/XPC*. *Nature communications*, 2015. **6**.
230. Drapkin, R., et al., *Dual role of TFIIH in DNA excision repair and in transcription by RNA polymerase II*. *Nature*, 1994. **368**(6473).
231. Bradsher, J., et al., *CSB is a component of RNA pol I transcription*. *Molecular cell*, 2002. **10**(4).
232. Xu, J., et al., *Structural basis for the initiation of eukaryotic transcription-coupled DNA repair*. *Nature*, 2017. **551**(7682): p. 653-657.
233. Beerens, N., et al., *The CSB protein actively wraps DNA*. *The Journal of biological chemistry*, 2005. **280**(6).
234. Tiwari, V., et al., *Current and emerging roles of Cockayne syndrome group B (CSB) protein*. *Nucleic acids research*, 2021.
235. Ranes, M., et al., *A ubiquitylation site in Cockayne syndrome B required for repair of oxidative DNA damage, but not for transcription-coupled nucleotide excision repair*. *Nucleic acids research*, 2016. **44**(11).

236. Smerdon, M.J., *DNA repair and the role of chromatin structure*. Current opinion in cell biology, 1991. **3**(3).
237. Selby, C.P. and A. Sancar, *Cockayne syndrome group B protein enhances elongation by RNA polymerase II*. Proceedings of the National Academy of Sciences of the United States of America, 1997. **94**(21).
238. Xu, J., et al., *Cockayne syndrome B protein acts as an ATP-dependent processivity factor that helps RNA polymerase II overcome nucleosome barriers*. Proceedings of the National Academy of Sciences of the United States of America, 2020. **117**(41).
239. Dianov, G., et al., *Repair of 8-oxoguanine in DNA is deficient in Cockayne syndrome group B cells*. Nucleic acids research, 1999. **27**(5).
240. Alemasova, E.E. and O.I. Lavrik, *Poly(ADP-ribosylation) by PARP1: reaction mechanism and regulatory proteins*. Nucleic acids research, 2019. **47**(8).
241. Caputo, M., et al., *The CSB repair factor is overexpressed in cancer cells, increases apoptotic resistance, and promotes tumor growth*. DNA repair, 2013. **12**(4).
242. Sollier, J., et al., *Transcription-coupled nucleotide excision repair factors promote R-loop-induced genome instability*. Mol Cell, 2014. **56**(6): p. 777-85.
243. *Luria-Delbrück experiment*. [08/03/2021]; Available from: https://en.wikipedia.org/wiki/Luria%E2%80%93Delbr%C3%BCck_experiment.
244. Cairns, J., J. Overbaugh, and S. Miller, *The origin of mutants*. Nature, 1988. **335**(6186).
245. Cairns, J. and P.L. Foster, *Adaptive reversion of a frameshift mutation in Escherichia coli*. Genetics, 1991. **128**(4): p. 695-701.
246. Foster, P.L., *Adaptive mutation in Escherichia coli*. Journal of bacteriology, 2004. **186**(15).
247. Hastings, P.J., et al., *Adaptive amplification: an inducible chromosomal instability mechanism*. Cell, 2000. **103**(5): p. 723-31.
248. Layton, J.C. and P.L. Foster, *Error-prone DNA polymerase IV is controlled by the stress-response sigma factor, RpoS, in Escherichia coli*. Molecular microbiology, 2003. **50**(2).
249. Rousset, F., et al., *The impact of genetic diversity on gene essentiality within the Escherichia coli species*. Nature microbiology, 2021. **6**(3).
250. Ukkivi, K. and M. Kivisaar, *Involvement of transcription-coupled repair factor Mfd and DNA helicase UvrD in mutational processes in Pseudomonas putida*. DNA Repair (Amst), 2018. **72**: p. 18-27.
251. Ragheb, M.N., et al., *Inhibiting the Evolution of Antibiotic Resistance*. Mol Cell, 2019. **73**(1): p. 157-165.
252. Foster, P.L., *Stress-induced mutagenesis in bacteria*. Critical reviews in biochemistry and molecular biology, 2007. **42**(5).
253. Maslowska, K.H., K. Makiela-Dzubska, and I.J. Fijalkowska, *The SOS system: A complex and tightly regulated response to DNA damage*. Environmental and molecular mutagenesis, 2019. **60**(4).
254. Schroeder, J.W., et al., *Sources of spontaneous mutagenesis in bacteria*. Critical reviews in biochemistry and molecular biology, 2018. **53**(1).
255. Sankar, T.S., et al., *The nature of mutations induced by replication-transcription collisions*. Nature, 2016. **535**(7610).
256. Pennington, J.M. and S.M. Rosenberg, *Spontaneous DNA breakage in single living Escherichia coli cells*. Nature genetics, 2007. **39**(6).
257. Kibota, T.T. and M. Lynch, *Estimate of the genomic mutation rate deleterious to overall fitness in E. coli*. Nature, 1996. **381**(6584).
258. Imhof, M. and C. Schlotterer, *Fitness effects of advantageous mutations in evolving Escherichia coli populations*. Proceedings of the National Academy of Sciences of the United States of America, 2001. **98**(3).
259. Denamur, E. and I. Matic, *Evolution of mutation rates in bacteria*. Mol Microbiol, 2006. **60**(4): p. 820-7.
260. Matic, I., et al., *Highly variable mutation rates in commensal and pathogenic Escherichia coli*. Science (New York, N.Y.), 1997. **277**(5333).

261. Bayliss, C.D., D. Field, and E.R. Moxon, *The simple sequence contingency loci of Haemophilus influenzae and Neisseria meningitidis*. The Journal of clinical investigation, 2001. **107**(6).
262. von Wintersdorff, C.J., et al., *Dissemination of Antimicrobial Resistance in Microbial Ecosystems through Horizontal Gene Transfer*. Front Microbiol, 2016. **7**: p. 173.
263. Sun, D., *Pull in and Push Out: Mechanisms of Horizontal Gene Transfer in Bacteria*. Front Microbiol, 2018. **9**: p. 2154.
264. Husain, I., et al., *Effect of DNA polymerase I and DNA helicase II on the turnover rate of UvrABC excision nuclease*. Proc Natl Acad Sci U S A, 1985. **82**(20): p. 6774-8.
265. Westblade, L.F., et al., *Structural basis for the bacterial transcription-repair coupling factor/RNA polymerase interaction*, in *Nucleic Acids Res*. 2010. p. 8357-69.
266. Chambers, A.L., A.J. Smith, and N.J. Savery, *A DNA translocation motif in the bacterial transcription-repair coupling factor, Mfd*. Nucleic Acids Res, 2003. **31**(22): p. 6409-18.
267. Ma, J., et al., *Transcription factor regulation of RNA polymerase's torque generation capacity*. Proc Natl Acad Sci U S A, 2019. **116**(7): p. 2583-2588.
268. Hausen, P. and H. Stein, *Ribonuclease H. An enzyme degrading the RNA moiety of DNA-RNA hybrids*. Eur J Biochem, 1970. **14**(2): p. 278-83.
269. Selby, C.P. and A. Sancar, *Transcription-repair coupling and mutation frequency decline*. 1993, J Bacteriology. p. 7509-7514.
270. Schmidt, A., et al., *The quantitative and condition-dependent Escherichia coli proteome*. Nature biotechnology, 2016. **34**(1).
271. Troelstra, C., et al., *ERCC6, a member of a subfamily of putative helicases, is involved in Cockayne's syndrome and preferential repair of active genes*. Cell, 1992. **71**(6): p. 939-53.
272. Duboc, C., et al., *Preparation of DNA Substrates and Functionalized Glass Surfaces for Correlative Nanomanipulation and Colocalization (NanoCOSM) of Single Molecules*. Methods Enzymol, 2017. **582**: p. 275-296.
273. Lawson, C.L., et al., *Catabolite activator protein: DNA binding and transcription activation*. Current opinion in structural biology, 2004. **14**(1).
274. Ebright, R.H., Y.W. Ebright, and A. Gunasekera, *Consensus DNA site for the Escherichia coli catabolite gene activator protein (CAP): CAP exhibits a 450-fold higher affinity for the consensus DNA site than for the E. coli lac DNA site*. Nucleic acids research, 1989. **17**(24).
275. Zhang, X.P., et al., *Derivatives of CAP having no solvent-accessible cysteine residues, or having a unique solvent-accessible cysteine residue at amino acid 2 of the helix-turn-helix motif*. Journal of biomolecular structure & dynamics, 1991. **9**(3).
276. Burgess, R.R., et al., *Factor stimulating transcription by RNA polymerase*. Nature, 1969. **221**(5175).
277. McKinley, B.A. and M.V. Sukhodolets, *Escherichia coli RNA polymerase-associated SWI/SNF protein RapA: evidence for RNA-directed binding and remodeling activity*. Nucleic Acids Res, 2007. **35**(21): p. 7044-60.
278. Richmond, M., et al., *RapA, Escherichia coli RNA polymerase SWI/SNF subunit-dependent polyadenylation of RNA*. Biochemistry, 2011. **50**(12): p. 2298-312.
279. Tennyson, C.N., H.J. Klamut, and R.G. Worton, *The human dystrophin gene requires 16 hours to be transcribed and is cotranscriptionally spliced*. Nature genetics, 1995. **9**(2).
280. Portman, J.R., et al., *Cotranscriptional R-loop formation by Mfd involves topological partitioning of DNA*. Proceedings of the National Academy of Sciences of the United States of America, 2021. **118**(15).
281. Krummel, B. and M.J. Chamberlin, *Structural analysis of ternary complexes of Escherichia coli RNA polymerase. Deoxyribonuclease I footprinting of defined complexes*. Journal of molecular biology, 1992. **225**(2).
282. Polyakov, A., E. Severinova, and S.A. Darst, *Three-dimensional structure of E. coli core RNA polymerase: promoter binding and elongation conformations of the enzyme*. Cell, 1995. **83**(3).
283. Revyakin, A., R.H. Ebright, and T.R. Strick, *Promoter unwinding and promoter clearance by RNA polymerase: detection by single-molecule DNA nanomanipulation*. Proceedings of the National Academy of Sciences of the United States of America, 2004. **101**(14).

284. Toulmé, F., et al., *GreA and GreB proteins revive backtracked RNA polymerase in vivo by promoting transcript trimming*. The EMBO journal, 2000. **19**(24).
285. Opalka, N., et al., *Structure and Function of the Transcription Elongation Factor GreB Bound to Bacterial RNA Polymerase*. Cell, 2003. **114**(3).
286. Hawkins, M., et al., *Direct removal of RNA polymerase barriers to replication by accessory replicative helicases*. Nucleic acids research, 2019. **47**(10).
287. Tetone, L.E., et al., *Dynamics of GreB-RNA polymerase interaction allow a proofreading accessory protein to patrol for transcription complexes needing rescue*. Proceedings of the National Academy of Sciences of the United States of America, 2017. **114**(7).
288. Harden, T.T., et al., *Bacterial RNA polymerase can retain $\sigma 70$ throughout transcription*. Proceedings of the National Academy of Sciences of the United States of America, 2016. **113**(3).
289. Hiasa, H. and K.J. Marians, *Two distinct modes of strand unlinking during theta-type DNA replication*. The Journal of biological chemistry, 1996. **271**(35).
290. Cagliero, C. and D.J. Jin, *Dissociation and Re-Association of RNA Polymerase With DNA During Osmotic Stress Response in Escherichia Coli*. Nucleic acids research, 2013. **41**(1).
291. Bernhardt, T.G. and P.A. de Boer, *The Escherichia coli amidase AmiC is a periplasmic septal ring component exported via the twin-arginine transport pathway*. Molecular microbiology, 2003. **48**(5).
292. Brüning, J.G., K.K. Myka, and P. McGlynn, *Overexpression of the Replicative Helicase in Escherichia coli Inhibits Replication Initiation and Replication Fork Reloading*. Journal of molecular biology, 2016. **428**(6).

RÉSUMÉ

La formation des R-loops peut conduire à une mutagenèse et constitue pour cela une menace pour l'intégrité génomique. La mutagenèse peut entraîner la résistance aux antibiotiques chez les bactéries et la résistance à la chimiothérapie dans les cancers humains, il est donc important de comprendre comment se forment les R-loops. En utilisant une combinaison d'expérimentation sur une seule molécule *in vitro* et avec des tests de mutagenèse *in vivo*, nous avons montré comment Mfd, une protéine bactérienne connue pour son rôle dans la réparation de l'ADN, peut interagir avec l'ARN polymérase pour former des R-loops. Cette interaction génère un domaine topologique dans l'ADN qui est très enclin à la formation de la R-loop et repose sur les propriétés de Mfd qui sont partagées par de nombreuses autres protéines, suggérant qu'il s'agit d'un modèle potentiellement universel pour la formation de la R-loop.

MOTS CLÉS

Réparation d'ADN, Mutagenèse, Évolution, Molécule-Unique.

ABSTRACT

R-loop structures pose a threat to genomic integrity because their formation can lead to mutagenesis. Mutagenesis drives antibiotic resistance in bacteria and chemotherapy resistance in human cancers, therefore it is important to understand how R-loops form. Using a combination of *in vitro* single-molecule experimentation with *in vivo* mutagenesis assays we have shown how Mfd, a bacterial protein known for its role in DNA repair, can interact with RNA polymerase to form R-loops. This interaction generates a topological domain in DNA that is highly prone to R-loop formation and relies on properties of Mfd that are shared by many other proteins, hinting that this is a potentially universal model for R-loop formation.

KEYWORDS

DNA repair, Mutagenesis, Evolution, Single-Molecule.