DEFINING THE ANCESTRAL REPLICATION FORK TRAP IN TUS-DEPENDENT BACTERIA

Casey J. Toft (BSc)

College of Public Health, Medical and Veterinary Sciences

James Cook University

This dissertation is submitted for the degree of Master of Philosophy

June 2021

ACKNOWLEDGEMENTS

First and foremost I would like to express my sincere appreciation to my primary advisor A/Prof. Patrick Schaeffer for his continued support, guidance and devoted passion in his field of research, which was a true inspiration for the work I have achieved in this thesis. Patrick was constantly full of new ideas to improve my work, and these often came just as I had completed applying the advice he had just previously given. Joke aside, you were the perfect advisor for me and I look forward to working with you in the future.

I also owe a great debt to my secondary advisor Dr Alanna Sorenson who showed me the ropes in the laboratory when I had essentially no experience for even the most basic protocols. I would also like to thank Dr Sacha Jensen for taking on the role of my secondary advisor with very little notice and coating my 'final' draft in red ink before I sent it to Patrick for viewing. Additionally, I'd like to thank the rest of my candidature committee Dr Ira Cooke, Prof. Norelle Daly, A/Prof. Ellen Ariel and Prof. Maxine Whittaker for ensuring I complete all the bureaucratic hurdles a graduate student must. Ira Cooke in particular who also helped with the sequencing and *de novo* assembly of a bacterial genome required for important analyses.

Finally, a big thanks to the other Molecular and Cell biology graduate students in the third floor Science Lab, namely Rebecca Bourquin, Brittany Dewdney, Rhys Gillman and Lauren Taylor for helping tackle some of the more complicated laboratory techniques, ensuring the office was a fun and sometimes productive work environment but most importantly, providing plenty of reasons to have a coffee break.

And of course, my family who have nurtured my passion for science and have been very supportive of my career along the way.

Und natürlich du Coco Hirsch, die eine Klasse für sich ist.

STATEMENT OF THE CONTRIBUTION OF OTHERS

Nature of Assistance	Contribution	Names, Titles and Affiliations of Co-Contributors
Financial support	Stipend	JCU RTPS
	Student Support Account (SSA)	JCU
	Grant application for publication support (Article title: Defining the prototypical DNA replication fork-trap in bacteria)	JCU HDRES
	Grant application for MinION access (Project Title: *De novo* genome assembly of KRX strain for crucial *in vivo* bacterial replication termination data)	Centre for Tropical Bioinformatics and Molecular Biology (CTBMB)
Research assistance and Data collection	Production of ChIP-Seq and ChIP-qPCR data for the genome wide distribution of Tus	Morgane J. J. Moreau (PhD), James Cook University, Townsville
	Ectopic insertion of *Ter* sites and growth rates analysis	Jiri Perutka, Savitri Mandapati, Peter Enyeart, Andrew Ellington, Institute for Cell and Molecular Biology, University of Texas at Austin, Austin, TX, 78712, USA

I certify that this dissertation has not been submitted previously as part of the requirements for another degree and that it is the result of my own independent work, unless otherwise referenced and acknowledged in this thesis. This thesis contains material that is currently under review for publication (Chapter 3) and has been slightly modified to be coherent with the format of this thesis. The experiments performed and data collected by the co-authors of this publication (see table above) are included in Chapter 3 as they were essential for overall data analyses. My contribution to this work is detailed in the chapter heading and included ChIP-Seq data analysis, figure productions, synthesis of the findings and all aspects of writing.

ABSTRACT

Most of our understanding of the paradigm of bacterial chromosome replication termination has been derived from studies of *Escherichia coli*, which have revealed the existence of a highly complex "replication fork trap" – where replication forks are allowed to enter but not leave. Fork arrest is mediated by *Ter* sites where helicase-induced strand separation triggers the formation of a "locked" Tus-*Ter* complex (Mulcair et al., 2006b). Recent efforts to understand this highly debated system have focused on investigating the underlying mechanism of Tus–*Ter*-lock (TT-lock) complex formation using elaborate single-molecule-based studies, ultimately aiming to explain the need for redundant 'backup' *Ter* sites. The main conclusion of these studies has been that the formation of the 'locked' Tus-*Ter* complex is somewhat inefficient, requiring the presence of additional 'backup' *Ter* sites (Elshenawy et al., 2015). Unfortunately, none of these studies can provide answers to the following fundamental questions:

a) What is the reason for the number of *Ter* sites and their locations?
b) Are all *Ter* sites bound to Tus and how does *Ter* occupancy relate to fork arrest?
c) Is there a universal need for backup *Ter* sites in replication fork traps?
d) Is Tus really working alone or is there a possible interplay with the *dif* site and other unknown factors?

The work presented in this thesis used a combination of novel *in silico* and *in vitro* approaches to uncover the 'ancestral' *Ter* fork trap architecture in Tus-dependent bacteria and explore its potential application in Immuno-PCR diagnostics. The first high-resolution genome-wide distribution map of Tus in *E. coli* is reported comparing the state of occupancy of all previously reported *Ter* sites. To better understand the biological relevance of the high-affinity and TT-lock forming outermost *Ter* sites in *E. coli* (*TerD*, *E* and *G*), examination of replication fork traps in a variety of Enterobacteriales was undertaken, followed by the first characterisation of

an ‘ancestral’ fork trap. The binding of an evolutionarily distant *Dickeya paradisiaca* Tus to its putative *Ter* and *Ter*-lock species was examined and compared to *E. coli* Tus.

A refined minimal replication fork trap within *E. coli* comprising two clusters of three *Ter* sites is proposed with *TerB*, *C* and *G* able to block the clockwise moving replication fork, and *TerA*, *D* and *E* the anticlockwise moving replication fork. Examination into the fork trap architecture in other Tus-dependent Enterobacteriales identified a new class of fork trap, coined type I, based on its ‘ancestral’ features. The type I fork trap architecture in *Edwardsiella tarda* consists of two *Ter* sequences that are highly homologous to *TerB* including the crucial C(6) base. Furthermore, the *E. tarda* Tus protein shares 48% identity with the *E. coli* Tus sequence and the presence of a conserved C(6)-binding pocket indicates that replication fork arrest is mediated by TT-lock formation. The fork trap architecture of *D. paradisiaca* was formally characterised and is organised as a single ‘unbreachable’ *Ter* site vicinal to *tus* on the left chromosomal arm, and two *Ter* sites in opposite orientation that are able to form a TT-lock. An unexpected 50-fold lower affinity of $_{Dp}$Tus for *Ter1* compared with *Ter1*-lock was observed and it was hypothesised that $_{Dp}$Tus may be expressed at higher levels in this bacterium. This lower affinity could be attributed to a reduction in protein-DNA contacts in the $_{Dp}$Tus:*Ter1* complex. Finally, the *D. paradisiaca* Tus-*Ter*-lock interaction was developed into a $_{Dp}$Tus-based qIPCR prototype for anti-GFP detection and demonstrated its potential to be useful in multiplex applications.

CONTENTS

LIST OF TABLES

LIST OF FIGURES

List of Abbreviations and Acronyms

ApR	Ampicillin resistance
APS	Ammonium persulphate
bp	Base pair
BSA	Bovine serum albumin
CmR	Chloramphenicol resistance
Ct	Cycle threshold
Da	Dalton
$_{dd}H_2O$	Double distilled water
$_{Dp}$Tus-GFP	The Tus protein from *Dickeya paradisiaca* tagged with GFP
$_{Ec}$Tus-GFP	The Tus protein from *Escherichia coli* tagged with GFP
EDTA	Ethylenediaminetetraacetic acid
ELISA	Enzyme-linked immunosorbent assay
GFP	Green fluorescent protein
K_D	Equilibrium dissociation constant
K_{obs}	Apparent equilibrium dissociation constant
NP	Non-Permissive face
OriC	Origin of replication
P	Permissive face
q.s.	quantum satis - as much as is sufficient

qIPCR	Quantitative immuno-PCR
RBS	Ribosome binding site
RT	Room temperature
SD	Standard deviation
SDS	Sodium dodecyl sulphate
ssDNA	Single-stranded DNA
T7P	T7 promoter
TBE	Tris-borate EDTA
T_m	Denaturation midpoint temperature
Tris	Tris(hydroxymethyl)aminimethane
TT-lock	Tus-*Ter* lock
UV	Ultraviolet

1 GENERAL INTRODUCTION

Bacteria dominate every ecological niche on the planet and will continue to do so. With more major lineages of organisms than any other domain of life (Hug et al., 2016), their success can mostly be attributed to molecular mechanisms that permit fast replication and rapid evolution. This chapter reviews critical literature and synthesises our current understanding of DNA replication termination in Tus-dependent bacteria. It also explores the utility and applications of Tus in emerging biotechnologies.

1.1 Replication of the *E. coli* chromosome

Bacterial chromosome replication is often presented based on the well-established *E. coli* model although major variations to the system do exist (Wolanski et al., 2014, Lin et al., 1993, Schaeffer et al., 2005). In *E. coli*, two large multimeric protein complexes, referred to as replisomes, assemble at a single, unique replication origin (*oriC*) of the circular chromosome. Each replisome proceeds bi-directionally at approximately 60 kb/min, duplicating the leading and lagging strand until a terminus site is reached (Figure 1A) (Schaeffer et al., 2005, Neylon et al., 2005a). In favourable conditions, bacterial DNA replication can occur in an overlapping manner termed multi-fork replication whereby a second round of replication can begin prior to the first-round finishing (Figure 1A) (Cooper and Helmstetter, 1968, Fossum et al., 2007).

Multi-fork replication permits rapid growth of *E. coli* with generation times as low as ~20 min and the inheritance of partially replicated chromosomes in some daughter cells. Hence, DNA replication must be completed faithfully to guarantee the stability of the genome, in conjunction with a tightly coordinated termination system that arrests the two rapidly and oppositely progressing replisomes (Figure 1A). This ensures precious resources and energy are not wasted by over-replicating the genome. Although many of the basic components of the large replisome have distinct roles that are conserved in all three domains of life consistent with the importance of DNA replication (Benkovic et al., 2001), the systems implemented to terminate replication can differ even among bacteria of a single order (Neylon et al., 2005a). For example, the monomeric Tus protein (*i.e.* termination utilisation substance) and its cognate *Ter* DNA binding sites (Figure 1B) found in *E. coli* are not universally found across all Enterobacteriales (Galli et al., 2019). It is important to note that the gram-positive *Bacillus subtilis* utilises an alternative termination system involving the RTP (replication terminator protein) which will not be further discussed in this review (Neylon et al., 2005a, Griffiths et al., 1998).

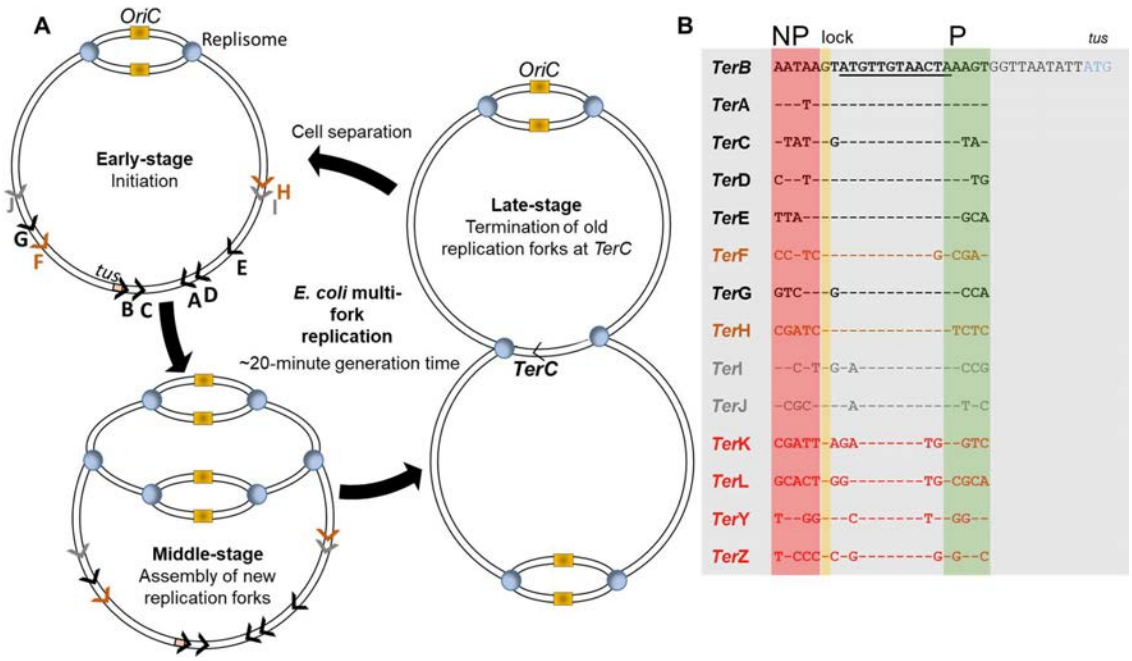


Figure 1: Replication initiation, elongation, and termination of the *E. coli* chromosome. (A) Multi-fork replication in fast growing bacteria. Multiple replisomes may load sequentially at the *oriC*, resulting in a greatly reduced generation time. Illustrated is the currently accepted replication termination fork trap architecture of *E. coli* including high affinity (black), moderate affinity (grey) and non-lock forming *Ter* sites (orange). Replication forks terminate most often at *TerC*. (B) Alignment of all 14 proposed *Ter* sequences (23 bp) including pseudo-*Ter* sites (red) highlighting the GC(6) base pair (yellow) and the strictly conserved 12 bp core sequence (underlined) in the high affinity sites. *TerB* is located 11 bp upstream of the start site (ATG) of the *tus* gene. NP: non-permissive face (red), P: permissive face (green).

The molecular mechanisms involved in bacterial DNA replication initiation and elongation have been the focus of numerous studies (Reyes-Lamothe and Sherratt, 2019). More recently, detailed insights have been gained from elaborate single-molecule studies (Trojanowski et al., 2018, Robinson and van Oijen, 2013, Hansen and Atlung, 2018, Hamdan et al., 2009, Tanner et al., 2008). Conversely, there has been a resurgence in studies to elucidate the mechanism of

replication termination which has been heatedly debated amongst researchers since the discovery of the Tus-*Ter* system (Figure 2). Many aspects of the Tus-*Ter* complex have been reviewed, including a comprehensive historical account of the discovery of the Tus-*Ter* interaction as well as a systematic comparison of the possible models of replication termination in *E. coli* in light of recent single molecule data (Neylon et al., 2005a, Berghuis et al., 2018). The next sections will focus on recent findings in DNA replication termination of Tus-dependent bacteria as well as the use of Tus in various biotechnological applications.

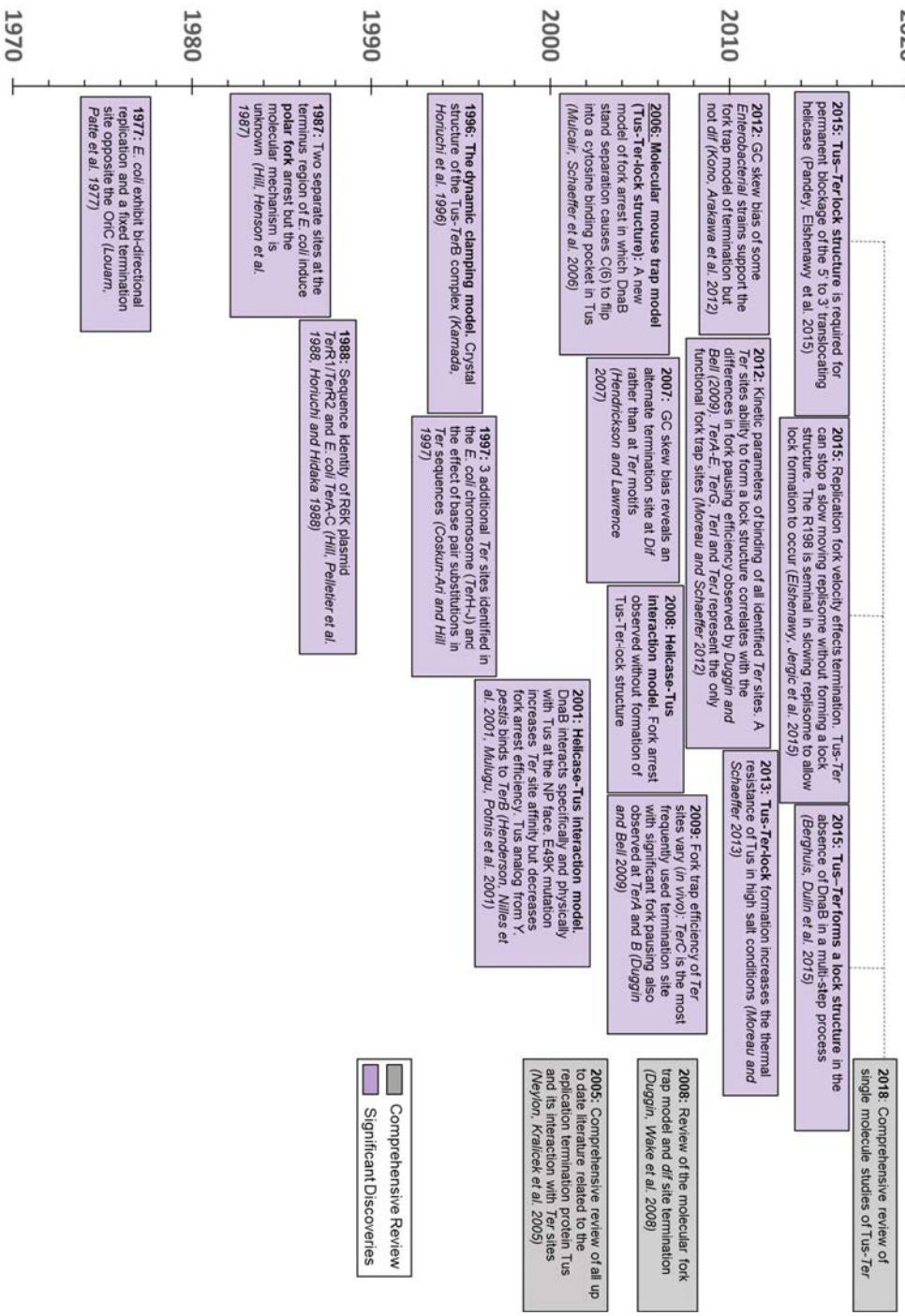


Figure 2: Literature landscape of *E. coli* replication termination studies. Historical timeline of the major studies that have aided in elucidating the complex system of bacterial replication termination in *E. coli* with a focus on studies that were performed after publication of the 'mouse trap' model (underlined) (Mulcair et al., 2006b).

1.2 The final step of DNA replication

Evidence of a fixed termination site in *E. coli* was first reported in 1977 (Louarn et al., 1977). However, it was not until a decade later that two asymmetric DNA sequences, now known as *Ter* sites (*Ter*A and *TerB*), were identified for their involvement in polar replication fork arrest (Hill et al., 1987, Hill et al., 1988). Each sequence was located on either side of the terminus region and revealed the existence of a highly complex "replication fork trap" – where replication forks were able to enter but not leave, appearing to assist accurate replication termination. Two highly homologous sequences were also identified in a 39.9 kb conjugative R6K plasmid and named *Ter*R1 and *Ter*R2, located opposite the replication origin (Horiuchi and Hidaka, 1988). Over the years, up to 14 *Ter* sites arranged throughout the *E. coli* chromosome have been proposed (*Ter*A-L, *Ter*Y and *Ter*Z), simultaneously expanding the size of the replication fork trap and increasing the perceived complexity of DNA replication termination (Figure 1B). However, only 10 of these *(TerA-TerJ)* were recognized as having replication fork arrest activity (Duggin and Bell, 2009). *Ter* sequences consist of a 16 bp consensus core sequence (5'-AGNATGTTGTAACTAA-3') (Figure 1B), and Tus binds to these different sites with varying affinity (Moreau and Schaeffer, 2012a, Duggin and Bell, 2009, Moreau and Schaeffer, 2013, Coskun-Ari and Hill, 1997, Francois et al., 1989, Hidaka et al., 1991). The chromosomal arrangement of the distal outermost *Ter* sites (*TerG-J*) suggests that they could serve as a 'backup' system if replication forks were to break through the innermost *Ter* sites (*TerA-TerD*) (Figure 1A). However, the function of distal *Ter* sites and their biological relevance is somewhat puzzling and debatable.

1.2.1 The Tus-*Ter* interaction

E. coli Ter sequences are characterised by a strictly conserved GC base pair at position six, followed by a highly conserved 12 bp core in their 23 bp sequence (Figure 1B). Each *Ter* is bound by a monomeric Tus with high specificity and affinity. The orientation of binding of Tus

to *Ter* sites results in a situation where an incoming replication fork will be unimpeded in one direction (permissive orientation), but will result in fork arrest in the opposite direction (non-permissive orientation), until the second replisome arrives (Hidaka et al., 1988, Pelletier et al., 1989). In the *E. coli* chromosome, *TerB* is located in the promoter region of the *tus* gene (11 bp upstream of the Tus coding sequence). The resulting tight association between the Tus protein and *TerB* effectively represses *tus* transcription, resulting in a tightly controlled transcription negative feedback loop (Roecklein et al., 1991, Neylon et al., 2005a). The dissociation constant (K_D) for the Tus-*Ter* complex lies in the pM to nM range depending on the ionic strength of the buffer and presents one of the strongest known DNA–protein interactions (Gottlieb et al., 1992, Elshenawy et al., 2015, Moreau and Schaeffer, 2012a). The strong binding affinity is obviously a prerequisite for such a small 36 kDa protein to be able to arrest a ~800 kDa multi-protein complex traveling at 60 kb/min. Although it was widely agreed that the Tus-*Ter* complex primarily acts as an inhibitor of the DNA unwinding activity of DnaB helicase at the forefront of the replisome, several hypotheses have been proposed over the years to explain the mechanism of polar replication fork arrest. These included a dynamic clamping model (Lee et al., 1989, Kamada et al., 1996) and a helicase interaction model (Khatri et al., 1989, Hiasa and Marians, 1992, Skokotas et al., 1995, Bastia et al., 2008). The affinity and binding mechanism of the Tus-*Ter* complex could tentatively be explained with the structural data of the Tus-*TerA* complex, showing a large number of non-specific and base-specific interactions (Kamada et al., 1996). However, the true molecular mechanism behind the polarity of fork arrest was revealed ten years later, challenging all previously proposed models (Mulcair et al., 2006b).

1.2.2 The molecular mousetrap model

At the forefront of the replication fork is the hexameric DnaB helicase which coordinates separation of the double-stranded template DNA by encircling and translocating DNA in a 5’ to 3’ direction on the lagging strand (Figure 3A). This process is coupled to continuous and

discontinuous synthesis of the leading and lagging daughter strands by several DNA polymerase III in association with the sliding clamp; and Okazaki fragment synthesis by RNA primers (DnaG primase). As DnaB helicase progresses toward the non-permissive face of *TerC* bound by a Tus protein, strand separation and breaking of the strictly conserved GC(6) base pair allows the unpaired cytosine to rotate and bind within a pre-formed pocket of Tus (Figure 3B) (Mulcair et al., 2006b). When this 'molecular mouse trap' is triggered, an extremely stable 'locked' Tus-*Ter* complex (TT-lock) which includes new base-specific interactions can form (Mulcair et al., 2006b, Moreau and Schaeffer, 2013).

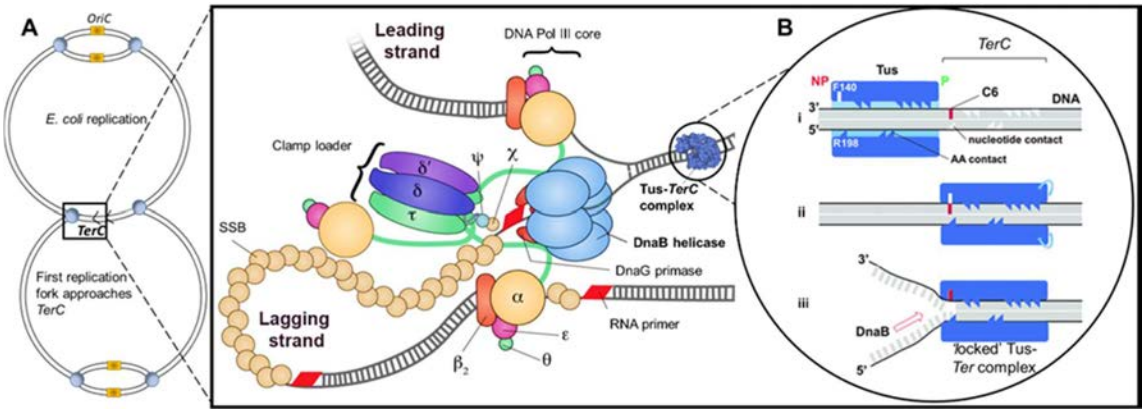


Figure 3: Termination of DNA replication in *E. coli* mediated by the formation of a 'locked' Tus-*Ter* complex. (A) Final stages of replication fork progression towards *TerC* in *E. coli*. At the forefront of the replication fork is the hexameric DnaB helicase that is arrested through its unwinding activity at the non-permissive face of the Tus-*TerC* complex. (B) Three-step model of replication fork-arrest by Tus-*TerC*: (i) binding of Tus to DNA mediated by non-specific electrostatic contacts followed by (ii) proper docking of Tus to its *Ter* site which results in a linear 'ratchet' and (iii) unzipping of double-stranded DNA by the DnaB helicase at the non-permissive (NP) end of the Tus-*Ter* complex leads to formation of the extremely stable TT-lock through the flipping of C(6) into its binding-pocket. Adapted from Lewis et al. (2017) and Moreau and Schaeffer (2012a).

The high resolution structure of Tus bound to the core *Ter* sequence with an unpaired C(6) base detailed its various specific and non-specific electrostatic protein-DNA contacts, and for the first time highlighted the important and strictly conserved amino acid residues involved in the formation of the TT-lock (Figure 4A & B). Since then, the TT-lock has been further scrutinised using single molecule techniques (Pandey et al., 2015, Elshenawy et al., 2015, Berghuis et al., 2016, Berghuis et al., 2015), designed specifically to better understand this unique mechanism of fork arrest. As such, several residues found at the non-permissive face of Tus have been shown to be important for efficient TT-lock formation to arrest incoming DNA replication forks. These include E49, I79, F140, H144, G149 and R198, which are discussed in detail in a later section. A revised multi-step model of Tus binding and docking to *Ter*, followed by TT-lock formation mediated by DnaB helicase activity has been proposed to explain the dynamic mechanism of replication fork arrest (Figure 3B i-iii). It is quite obvious how formation of a strong TT-lock can be induced by the base separation action of the DnaB helicase through the interaction of the freed C(6) with its cognate binding pocket on Tus. How Tus initially finds and docks properly onto *Ter* is less obvious (Moreau and Schaeffer, 2012a). More recently, further refinements to explain the formation and efficiency of the TT-lock have been proposed (Berghuis et al., 2015, Elshenawy et al., 2015, Berghuis et al., 2018), but none of these studies have addressed how Tus finds a *Ter* in the first place, *e.g.* if the helicase pushes Tus along the DNA?

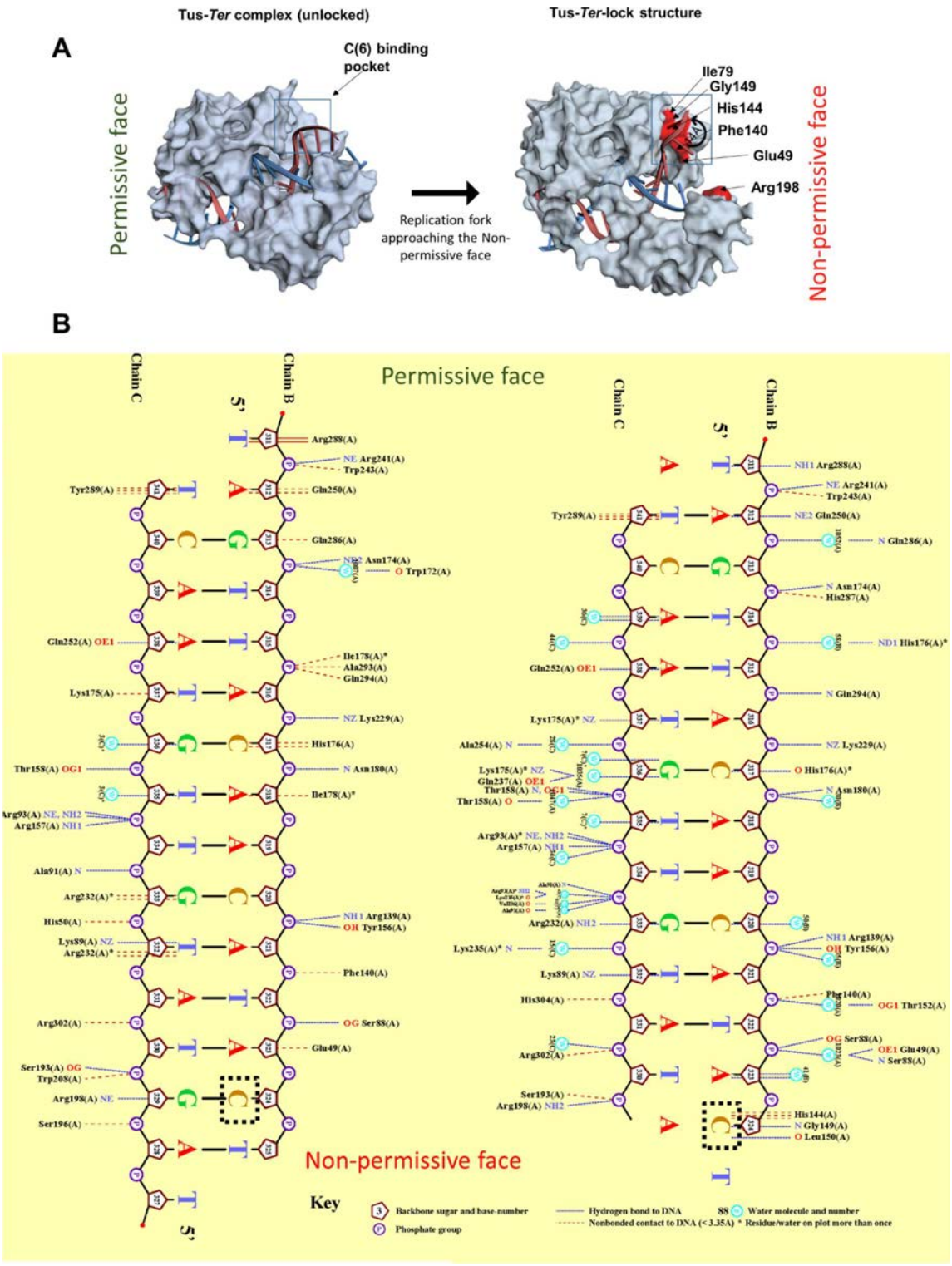


Figure 4: Tus-*Ter* lock formation. (A) Structures of Tus-*Ter* (PDB 2I05) and Tus-*Ter*-lock (PDB 2I06) illustrating the proposed fork arrest mechanism at the non-permissive face of Tus involving the cytosine binding pocket of Tus and TT-lock formation. (B) Nucplot of Tus-*Ter*

(left side) and Tus-*Ter*-lock (right side) comparing the new sequence specific interactions formed upon binding of C(6) into its binding pocket. Numerous non-specific interactions are involved in the formation of the Tus-*Ter* complex in addition to a number of new specific hydrogen bonds which play a crucial role in TT-lock stability.

1.3 Identifying the redundant '*Ter*' sites

1.3.1 Fork arrest activity

Using two-dimensional gel analysis of replication intermediates at *Ter* sites under natural growth conditions, it was shown that out of the 14 proposed *Ter* sites, *TerC* is the most frequently used termination site with significant fork pausing also observed at *TerA* and *B* (Duggin and Bell, 2009). Additional experiments in conditions where Tus was over-expressed were required to measure fork arrest activity for the remaining *Ter* sequences, which in natural conditions produced an unquantifiable signal intensity difference between paused fork and linear DNA (*TerD, G, H* and *I*) or had a negligible paused fork pattern (*TerE, F* and *J*) (Duggin and Bell, 2009). While this may have initially suggested that only *TerA, B* and *C* were required to terminate replication, the efficiency of each *Ter* site at arresting a replication fork in overproduced Tus conditions revealed a clear but variable capability for other *Ter* sites to arrest incoming replication forks (Figure 5). For this, each *Ter* sequence was cloned into a unidirectional replicating vector (pACYC184) so that fork arrest was dependent on its ability to progress through a Tus-*Ter* complex in non-permissive orientation. Of the 14 proposed *Ter* sites, 10 of them showed significant fork arrest activity (*TerA*-*J*) (Duggin and Bell, 2009) although compared to the others, the activity of *TerF* could be questioned. Of note, the innermost *Ter* sites (*TerC* and *TerB*), showed a minimal difference in efficacy at halting replication forks in overexpressed Tus conditions compared to natural cellular Tus abundance (Duggin and Bell, 2009), which has been estimated to be as few as 100 copies per cell (Natarajan et al., 1993). Accordingly, *TerK*, *L*, *Y* and *Z* were dismissed as pseudo-*Ter* (p*Ter*)

sites which may only cause fork arrest in artificial conditions (Duggin and Bell, 2009). Compared to the simple replication fork trap architecture found in some plasmids consisting of only two *Ter* sites (Horiuchi and Hidaka, 1988), many of the *Ter* sites in the *E. coli* chromosome could be considered redundant. It is not obvious whether the number or chromosomal loci of *Ter* sites in *E. coli* is common across other species as the fork trap architecture in other bacteria that carry a *tus* gene ortholog have not yet been characterised.

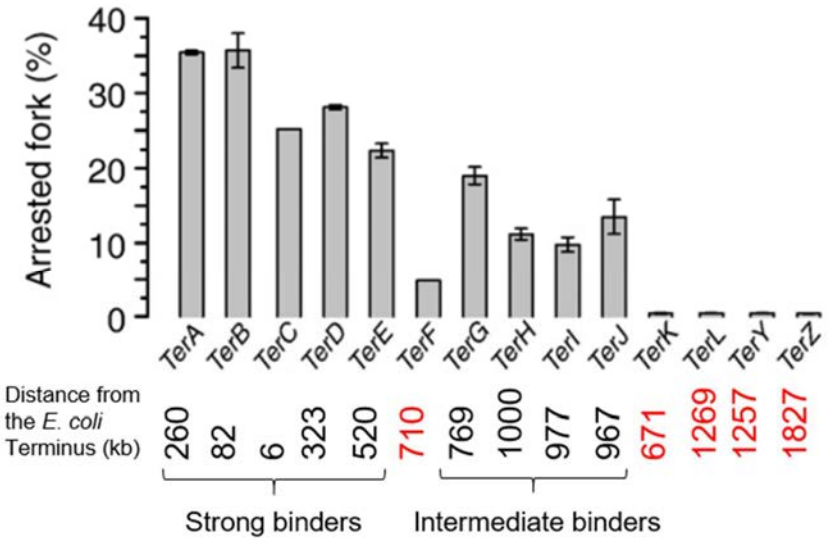


Figure 5: Fork arrest efficiency at *Ter* sites. Each *Ter* sequence was integrated into separate plasmids (pACYC184) and transformed into cells overexpressing Tus before the efficiency of fork arrest was measured using Y-fork band quantitation. Pseudo-*Ter* sites are indicated in red. The terminus in *E. coli* MG1655 (RefSeq NC_000913.3) was defined as a site located diametrically opposite to the *oriC* (chromosomal position at 1603 kb). Figure was adapted from Duggin and Bell, (2009).

1.3.2 Are all *Ter* sites able to lock?

The variable efficiency of the 10 *Ter* sites in arresting replication forks indicated that some *Ter* sequences might have a lower affinity for Tus and that they could be impaired in TT-lock formation (Duggin and Bell, 2009, Coskun-Ari and Hill, 1997). Equilibrium, thermodynamic and kinetic binding parameters for *TerA-J* sequences and their *Ter*-lock variants (unpaired C(6) *Ter* sequence) in complex with Tus were examined using various methods such as surface plasmon resonance (SPR) and differential scanning fluorimetry of GFP-tagged proteins (DSF-GTP) (Moreau and Schaeffer, 2012a, Moreau and Schaeffer, 2013). At 150 mM KCl, the dissociation kinetics obtained by SPR for Tus in complex with the strong *Ter* and *Ter*-lock sequences are immeasurably slow. However, the kinetics of the moderate *TerH-J* and weak *TerF* could be determined. Out of these, only *TerJ*-lock showed evidence of TT-lock formation yet its dissociation half-life ($t_{1/2}$ =1322 s) was similar to Tus-*TerH* and *TerI* (1244 s and 949 s respectively). Furthermore, although not directly comparable, at high 250 mM KCl, the dissociation half-lives of Tus bound to the *TerA*-lock or any of the other strong *Ter*-lock sequences are still 2-3-fold longer ($t_{1/2}$ = 4367 s for *TerB*-lock) suggesting that while *TerH*, *I* and *J* may pause replication forks, their dissociation kinetics may be too fast to completely arrest replication forks. At 250 mM KCl, *TerF* did not bind Tus and was thus proposed to be a pseudo *Ter* site.

A GFP-based thermal stability assay (GFP-Basta) was instrumental to differentiate the extreme binding affinity of Tus for the strong *TerA-E* and *G* at 150 mM KCl (Moreau and Schaeffer, 2012a). The different Tus-*Ter* binding profiles were confirmed using a polyplex qPCR DNA-binding assay (Moreau and Schaeffer, 2012b) as well as differential scanning fluorimetry of a GFP-tagged Tus (DSF-GTP) (Moreau and Schaeffer, 2013). DSF-GTP records the quenching of GFP fluorescence caused by the unfolding of Tus allowing determination of a denaturation midpoint temperature (T_m) value in free and DNA-bound states. The T_m of Tus was highly

dependent on which *Ter* or *Ter*-lock sequence it was bound to and the ionic strength of the buffer (Moreau and Schaeffer, 2013). By gradually increasing the ionic strength of the binding buffer (8.5-300 mM KCl), thereby progressively removing all non-specific electrostatic interactions between Tus and DNA, Moreau and Schaeffer were able to dissect and compare the contributions of specific interactions occurring in the Tus-*Ter* and Tus-*Ter*-lock complexes (Figure 6). The data supported previous findings that neither *TerF* nor *TerH* could form a locked complex with Tus. To explain why *TerF* and *TerH* are not able to form a TT-lock, Moreau and Schaeffer (2012a) reasoned that the substitution of a pyrimidine (T) to a bicyclic purine (G) on the leading strand at position 5, unique only to these two sites (*c.f.* Figure 1B), may cause a steric hindrance in the C(6) binding pocket of Tus. Tus-*TerB*-lock was the most salt-resistant confirming previous SPR and GFP-Basta data that had shown it was also the most stable complex, *i.e.* slowest to dissociate. Overall, the aforementioned studies provide compelling evidence that *TerA-E*, *G* and *TerJ* can form locked complexes with Tus.

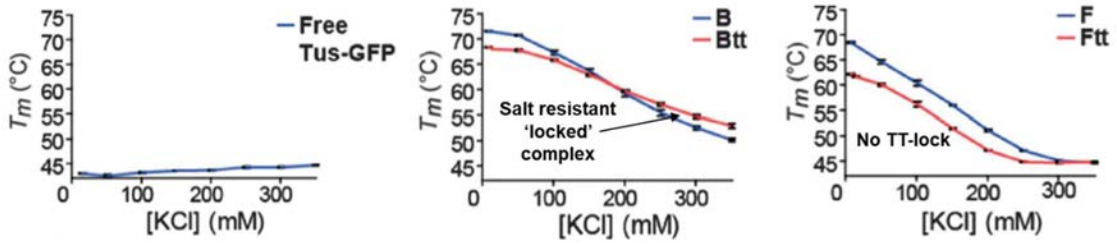


Figure 6: Salt-dependence of free and *Ter*-bound Tus-GFP. T_m values of Tus-GFP in complex with *TerB* and *TerF* (blue) and their *Ter*-lock variants (red) were measured in increasing ionic strength by DSF-GTP. As KCl concentration increases, the non-specific protein-DNA interactions reduce, allowing identification of *Ter* sequences which are capable of forming a TT-lock. This is evident for *TerB*-lock which has increased salt resistance compared to *TerB*. This pattern is not observed for the pseudo *TerF* and *TerF*-lock. Adapted from Moreau and Schaeffer, (2013).

1.4 Location of replication termination

Hendrickson and Lawrence (2007) proposed an alternate site of replication termination at the *dif* locus. The *dif* locus has been identified as the site at which the recombinases XerC and XerD bind cooperatively and act as a catalyst of site-specific recombination for resolution of dimeric chromosomes. This site is pivotal for successful segregation of a single chromosome into each resulting daughter cell during replication (Blakely and Sherratt, 1994, Bastia and Zaman, 2014). The *dif* termination hypothesis is based on the well-established asymmetric mutational pressures that occur in bidirectionally-replicated bacterial chromosomes due to the continuous and discontinuous nature of strand synthesis (Lobry, 1996). Such pressures produce an over-abundance of guanines over cytosines in the leading strand, resulting in a GC-skew footprint that can be utilised to predict the sites of replication origin and terminus (Frank and Lobry, 1999). Bioinformatic sequence analysis of the point at which the strand compositional asymmetry switches at the terminus (representative of a historical termination site) identified a site very close to the *dif* site (Hendrickson and Lawrence, 2007). It was proposed that the Tus-*Ter* system evolved to trap retrograde forks such as those involved in repair events rather than the replicative DnaB helicase. In support of their hypothesis, the authors highlighted that the *dif* locus can be found across all bacteria and that the Tus-*Ter* system has a seemingly recent ancestral evolution. Indeed, the Tus-*Ter* system has only been found in closely related Enterobacteriales and some plasmids (Henderson, 2001), although recent literature suggests that a *tus* gene may be carried by a larger group of bacteria and plasmids than previously thought (Galli et al., 2019). The notion that *dif* could be the real site of replication termination was later dismissed following 2D-gel electrophoresis experiments demonstrating unambiguous resolution of replication fork arrest at *TerC* but no significant pausing at the *dif* site located less than 30 kb away (Duggin and Bell, 2009). Although, the genomic compositional GC-skew certainly made apparent that the terminus shift point occurs closer to the chromosome dimer

resolution site (*dif*) than to the *Ter* sites in some studied species (Hendrickson and Lawrence, 2007), more recent GC-skew analyses demonstrated that this terminus shift point rather does coincide with replication fork arrest by Tus at *Ter* sites (Kono et al., 2012). The same authors further validated their findings using a novel *in vivo* accelerated laboratory evolution technique (Kono et al., 2018). Taken together, it could be speculated that *Ter* sites have rather evolved around the *dif* site in select bacteria. This is possible given that XerC and XerD that bind *dif* are unable to function to resolve dimeric chromosomes unless the two replichores are on either side of the *dif* sequence (Bastia and Zaman, 2014). However, further investigation into the fork trap architecture in other bacteria which utilise a Tus ortholog may provide a more detailed insight into the need for the multiple *Ter* sites around *dif*.

1.5 Important interactions in the *E. coli* Tus-*Ter*-lock complex

About 43 out of the 309 amino acid residues of Tus interact with a short *Ter*-lock DNA species in the reported crystal structure (Figure 4B). Mutations of a selection of these have been shown to result in partial loss or abolition of DNA fork arrest activity (Mulcair et al., 2006b, Henderson et al., 2001, Mulugu et al., 2001). Numerous DNA-interacting amino acid residues have been identified to increase fork arrest efficiency via TT-lock formation including E49, F140, H144, G149, L150 and R198 (Mulcair et al., 2006b, Berghuis et al., 2016, Pandey et al., 2015, Elshenawy et al., 2015, Berghuis et al., 2018). A phylogenetic comparison of Tus sequences from other Enterobacteriales performed prior to the discovery of the TT-lock formation, had shown that most of these amino acid residues were highly conserved in Tus (Henderson et al., 2001).

The H144A and F140A mutants were key to confirm the mechanism of TT-lock formation. In fact, a single mutation of the C(6) binding pocket residue H144 or F140 abolishes formation of a TT-lock (Mulcair et al., 2006b). This was expected as H144 forms a base-specific interaction with C(6) in the Tus-*Ter*-lock structure (Figure 4B). Surprisingly, despite its key involvement

in lock formation, it was later shown that a H144A mutation was still able, albeit less efficiently, to induce fork arrest (Berghuis et al., 2016, Elshenawy et al., 2015). Most revealing was the finding that the F140A Tus mutant is more stable in complex with *TerB* despite being unable to form a TT-lock. Mulcair et al. argued that the strictly conserved F140 residue is essential to the TT-lock mechanism and thus fork arrest in vivo, as otherwise, there would be no reason to conserve this residue (Mulcair et al., 2006b).

Interestingly, single molecule DNA unzipping experiments (Berghuis et al., 2018), revealed that a TT-lock was still able to form with the E49K Tus mutant, albeit with a dramatically decreased probability. It is important to note that E49 in Tus interacts with A(7) in the fully double-stranded *Ter* sequence, but this direct interaction is lost in the locked complex. Taken together, Berghuis et al. (2018) suggested a possible role for the residues E49 in which it guides C(6) into its binding pocket for lock formation. This is in agreement with the modest decrease in half-lives observed by Mulcair et al., (2006) for the E49A-Tus-*Ter* and *Ter*-lock complexes.

The interaction made by R198 with *Ter* and its role in fork arrest is less clear. The importance of R198 was highlighted by an R198A Tus mutant that had a 130-fold increase in K_D and was later shown to have lost fork arrest activity in a growth assay (Henderson et al., 2001, Neylon et al., 2000). Moreau and Schaeffer (2012a) first proposed a possible key role for R198 to scan the DNA in the search for a *Ter* sequence when pushed by the replisome, as the A(5) interacting with R198 plays a critical role in Tus–*Ter* complex formation prior to the unzipping of the C(6) that allows a 'locked' complex to form. R198 forms a hydrogen bond with the G(6) nucleotide (opposite C(6) lock base) in the Tus-*Ter* crystal structure (*c.f.* Figure 4B p21) although additional interaction have been proposed with the A(5) base on the same strand and T(5) on the opposite strand (Berghuis et al., 2018). These interactions rearrange in the Tus-*Ter*-lock structure to a single interaction with the phosphate backbone between A(6)-T(7) (*c.f.* Figure 4B p21). Data from single molecule flow stretching suggested that this residue, although not

involved directly in the formation of the TT-lock, has an important role in transiently stopping the replication fork and 'buying time' to enable the C(6) of *Ter* to locate its binding pocket in Tus (Elshenawy et al., 2015). The authors also showed that the R198A Tus in complex with *Ter* results in negligible fork arrest activity, but can be completely restored to wild-type efficiency if Tus is in complex with *Ter*-lock DNA (Elshenawy et al., 2015). Single molecule DNA unzipping experiments support that R198A Tus mutant can form a TT-lock similar to wild-type Tus but with a decreased lifetime (Berghuis et al., 2016, Berghuis et al., 2018). The role of this residue remains somewhat unclear, albeit its importance being evident through its numerous interactions with *Ter* and *Ter*-lock species.

1.6 The Tus protein in other bacteria

Tus has a seemingly recent ancestral evolution. A comparison of the Tus protein sequences found in other Enterobacteriaceae, including *S. typhimurium, K. ozaenae* and *Y. pestis* revealed, a ~78%, 70% and 53% identity to the *E. coli* Tus sequence respectively (Henderson et al., 2001). Despite the relatively high sequence divergence of the *Y. pestis* Tus, it was shown to bind to *E. coli Ter* sequences and arrest the DnaB helicase, albeit with somewhat lower efficiency than *E. coli* Tus (Henderson et al., 2001). It is important to note that the critical E49, F140, H144, G149, L150 and R198 residues which are involved in TT-lock formation and fork arrest activity are strictly conserved in these proteins. A more recent phylogenetic analysis of Tus revealed that it is only present among a selection of γ-Proteobacteria, within both plasmids and chromosomal DNA, and most frequently within Enterobacteriaceae (Galli et al., 2019). However, the Tus-*Ter* system is absent in species that diverged early during the evolution of the Enterobacteriales (*Plesiomonas shigelloides*). Its absence in most bacteria as well as early Enterobacteriales provides evidence of a very recent domestication of the *tus* gene, which may have occurred simultaneously with or in response to the switch to another ecological niche (Galli et al., 2019). It also strongly suggests that the domestication of the Tus-*Ter* system

occurred from a plasmid-encoded *tus* gene as a result of a gain of function (Galli et al., 2019). A possible reason for the presence of a *tus*-like gene in certain plasmids (*e.g. Pseudomonas resinovorans*) (Galli et al., 2019) may be due to the need for copy number control to ensure excessive host resources are not expended, and therefore require a tight replication arrest system. Interestingly, self-transmissible R6K plasmids have been shown to carry only two *Ter* sequences and are without a *tus* gene (Horiuchi and Hidaka, 1988), contrasting with the two clusters of multiple *Ter* sites and a *tus* gene found in *E. coli* and other Enterobacteriaceae (Horiuchi and Hidaka, 1988). This raises the questions of which of *tus* or the *Ter* sequences evolved first, and why are there so many strong *Ter* sequences that can form a 'locked' complex with Tus in *E. coli?* Finally, are all Tus proteins found in other Tus-dependent bacteria able to form a TT-lock? These questions will hopefully provide the necessary impetus for future studies outside *E. coli*.

1.7 Biotechnological applications of Tus

Over the years, a number of systematic studies have been conducted to scrutinise the affinity of Tus for DNA as well as its ability to form the ultra-stable Tus-*Ter*-lock structure. In fact, by 2007, the *E. coli* Tus protein was so well understood that it was used as a control protein to validate a sortase-mediated ligation assay for the covalent attachment of proteins to surfaces (Chan et al., 2007). Since then, a variety of biotechnological applications making use of the highly stable *E. coli* Tus-*Ter* and *Ter*-lock complexes have emerged. Interest in the applicability of the Tus-*Ter* system was largely sparked by its high-affinity (~3.4 x 10^{-13} M in 150 mM potassium glutamate buffer) and the even more stable Tus-*Ter*-*lock* complex (Mulcair et al., 2006a, Gottlieb et al., 1992).

The initial use of Tus was as a model protein to support the development of the sortase-mediated ligation assay (Chan et al., 2007). A year later, the first example of GFP-tagged Tus was reported and used as a proof-of-concept for the development of a Tus-based protein microarray

platform technology (Chatterjee et al., 2008). From that point in time onwards the intrinsic properties of Tus and *Ter* were harnessed to develop new technologies and methods. In parallel, and essentially due to its predictable properties, Tus has also been frequently used as a reference protein to validate innovative ligand binding assays such as GFP-Basta and DSF-GTP (Moreau et al., 2010, Moreau et al., 2012), a novel polyplex real-time PCR (qPCR) technique to measure protein-DNA interactions (Moreau and Schaeffer, 2012b) as well as a functional protein that can be incorporated into a self-assembling peptide hydrogel (Piluso et al., 2013). Similar to streptavidin, avidin and other high-affinity binders which are widely used in molecular science owing to their highly selective and stable interaction with biotin (Dundas et al., 2013), notable developments of *E. coli* Tus-*Ter* into molecular tools, protein delivery systems and diagnostic platforms have started to emerge (Table 1). Here, the small size and monomeric binding modality of Tus (Neylon et al., 2005a, Berghuis et al., 2018) are desirable traits compared with the possible shortcomings of the tetrameric streptavidin (Dundas et al., 2013). The next sections provide a detailed and systematic historical account of the various developments and uses of the Tus-*Ter* system.

Table 1: Biotechnological applications of Tus over the last decade

Year	Target Application	Technique	Reference
2008	The study of protein–ligand interactions	Tus based protein microarray	(Chatterjee et al., 2008, Sitaraman and Chatterjee, 2011)
2009	Irreversible protein-DNA conjugation	Photoactivatable connector between proteins and specific DNA sequences	(Dahdah et al., 2009)
2010	Simple and rapid eukaryotic protein delivery into the nucleus.	Tus nuclear localisation signal used for nuclear delivery in mammalian cells	(Kaczmarczyk et al., 2010)
2010-2017	Extremely sensitive immunoassays for biomolecular detection of antibodies and antigens	Tus-*Ter* based immuno-PCR which utilises the extreme stability and self-binding capabilities of Tus to *Ter*-lock DNA	(Morin et al., 2010, Morin et al., 2011, Johnston et al., 2014, Askin and Schaeffer, 2012, Cooper et al., 2013, Kamath et al., 2017)
2011	Quantification of the protease activity of various proteins	Tus-GFP based protease activity assay	(Askin et al., 2011)
2013	Strong DNA binding functions within a hydrogel	Self-assembling hydrogels which contain functional peptide-Tus conjugates	(Piluso et al., 2013)
2014-2018	A novel system to explore the proteins involved in the regulation of external fork pausing barriers in eukaryotes.	Tus-*Ter* induced replication fork arrest in yeasts, mice and mammalian cells.	(Willis et al., 2014, Larsen et al., 2014a, Willis et al., 2018)

1.7.1 Proteomic applications

1.7.1.1 Tus-based self-assembling protein microarray

The Tus-*Ter* interaction was key to the development of an innovative protein microarray strategy to study protein-protein interactions (Chatterjee et al., 2008, Sitaraman and Chatterjee, 2011). The core platform technology is based on an expression vector, including both a *TerB* and a *tus* cassette for the production of Tus-tagged fusion proteins that is printed onto an array slide. Thus, the printed plasmid has a dual purpose of cell-free expression allowing direct capture of the Tus-fusion protein onto the array slide through its stable interaction with *TerB*. This compares to typical microarray strategies that often directly immobilise purified proteins onto array slides, thus requiring the production of large libraries of purified proteins and the long-term maintenance of array stability (Chatterjee et al., 2008, Sutandy et al., 2013).

An initial proof of concept of this Tus-based protein array platform was reported in 2008, with a GFP-tagged Tus fusion protein (Tus-GFP). The E47Q Tus mutant which has increased affinity for *TerB* was used for the development of this technology (Henderson et al., 2001). The platform was later validated with several known protein-protein interactions such as Jun/Fos, FRB/FKBP12, p53/MDM2, and CDK4/p16 and allowed identification of p16 as a new binding partner of CDK2 (Sitaraman and Chatterjee, 2011). In principle this platform can be used to identify DNA–protein, ligand–receptor, enzyme–substrate, and drug–protein interactions (Sitaraman and Chatterjee, 2011).

1.7.1.2 Tus-based immuno-PCR diagnostics

Immuno-PCR (IPCR) is a highly sensitive immunoassay technique that combines the high affinity and specificity of interaction between an antigen and an antibody, with the extremely low copy-number detectability of DNA by PCR (Sano et al., 1992). Quantitative IPCR (qIPCR) has shown to increase analytical detection sensitivity by up to ~10^9 fold compared to the enzyme-linked immunosorbent assay (ELISA) (Chang et al., 2016). A number of linking systems exist for the conjugation of DNA to antibodies that are essential for qIPCR, many of which have been reviewed previously (Chang et al., 2016). Most protein-DNA conjugation systems produce heterogenous DNA–antibody conjugates limiting the use of qIPCR as a routine diagnostic tool in laboratories that require high batch to batch reproducibility and simplicity.

The field of qIPCR is potentially one of the most promising application of Tus where its monomeric nature enables controlled production of stoichiometric DNA-protein conjugates (Morin et al., 2010, Morin et al., 2011, Askin and Schaeffer, 2012, Johnston et al., 2014, Chang et al., 2016). In 2009, the TT-lock interaction was developed as a useful protein-DNA photo-crosslinking system (Dahdah et al., 2009). The A(7) of a *TerB*-lock sequence was substituted with a BrdU to allow for site-specific covalent binding of Tus-GFP with yields up to 65% of

protein-DNA conjugates in a few minutes. The broad utility of this photoactivatable protein-DNA conjugation system was then verified with a Tus-GFP fusion protein. The self-assembling property of the Tus-*Ter*-lock interaction presents numerous advantages, including production of irreversible stoichiometric protein-DNA conjugates with a well-defined regiospecificity.

In a further development, an extended *Ter*-lock sequence including a template sequence for PCR amplification (*Ter*-lock-T) was bound to Tus-GFP, providing the basis for the first Tus-based qIPCR assay (Figure 7) (Morin et al., 2010). Ultrasensitive detection of anti-GFP antibodies could be achieved in the femtomolar range. The system was also evaluated in a competitive format (Morin et al., 2010). Since then, there has been a steady increase of Tus-based qIPCR applications for the sensitive detection of a number of antigens and antibodies as well as protein analytes in buffer and crude complex matrices (Morin et al., 2010, Morin et al., 2011, Askin and Schaeffer, 2012, Johnston et al., 2014, Moreau and Schaeffer, 2012b, Kamath et al., 2017, Cooper et al., 2013). One of the most promising application of this system is for the detection of extremely low numbers of antibodies such as tropomyosin-specific IgE. There is also scope for quantifying target proteins in a complex biological matrix such as serum, using a 'bridging' system developed by Askin and Schaeffer (2012). With the different conjugation systems now at hand there is an obvious possibility for multiplexing qIPCR assays. Another avenue for multiplexing the Tus-based qIPCR platform could be to engineer new Tus variants or identify Tus othologs with altered DNA sequence specificities.

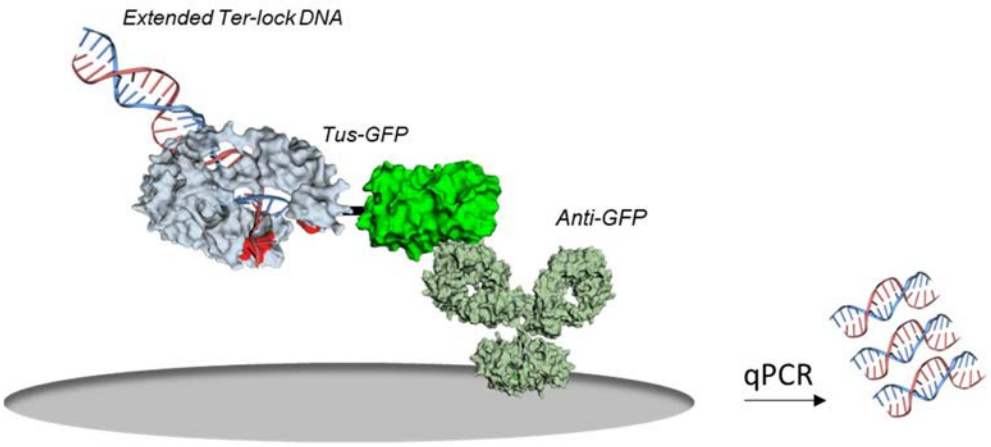


Figure 7: Tus-based qIPCR diagnostic platform. The system takes advantage of the self-assembling TT-lock interaction. Here, a Tus-GFP was linked to a *Ter*-lock-containing DNA template sequence to quantify the presence of an antibody (anti-GFP) by qPCR.

1.7.1.3 Tus-based protease assay

In 2011, a protease activity assay was developed using a Tus-GFP construct containing a protease substrate in its interdomain linker sequence. The system takes advantage of the lower T_m of Tus compared to GFP. Here, Tus-GFP is incubated with a protease with subsequent heat denaturation and centrifugation steps to remove non-proteolysed Tus-GFP. As such, the increase in GFP fluorescence in the supernatant correlates with increased protease activity. This system was successfully applied to measure the protease activities of trypsin, caspase 3, and neutrophil elastase with analytical detection sensitivities in the nanomolar range. To allow longer incubation times at a higher temperature of the Tus-GFP protease assay, a *Ter* DNA sequence can be added to increase the half-life of Tus and the analytical sensitivity of the assay. The platform should be useful in drug discovery programs for the screening and characterisation of protease inhibitors (Askin et al., 2011).

1.7.2 Cell-based applications

1.7.2.1 Protein delivery system

Interestingly, the *E. coli* Tus has been reported to contain a protein sequence that resembles a mammalian nuclear localization signal (NLS) that has been verified to have a moderate nuclear targeting efficiency (Kaczmarczyk et al., 2010, Ray et al., 2015). As such, the Tus protein has been proposed as a protein delivery system which was illustrated by the nuclear localisation by Tus-GFP. They reported that Tus-GFP can be transfected into human cells with almost exclusive nuclear localization through plasmid transfection and rapid protein transduction from culture media (Kaczmarczyk et al., 2010). Although the nuclear localisation of a bacterial protein is a unique attribute that may be a fortuitous coincidence, these findings demonstrated expression of functional Tus fusion proteins in eukaryotic cells, hinting toward its potential applications in higher organisms.

1.7.2.2 Tus-*Ter* induced replication fork arrest in eukaryotes

In 2014, Larsen et al. reported the potential use of Tus-*Ter* as a heterologous DNA replication perturbating system in eukaryotes. Here, genomic insertion of *Ter* sequences in the *Saccharomyces cerevisiae* (yeast) genome allowed them to determine the effect of site-specific heterologous DNA replication fork barriers. The authors found that the genomic insertion of *Ter* modules with concomitant Tus expression, form site-specific replication fork barriers that can be resolved by site-specific homologous recombination requiring the RecQ helicase (Sgs1) (Larsen et al., 2014b). Soon after, a similar system was applied to mouse cells, allowing Willis et al. (2014) to conclude that aberrant long-tract homologous recombination at stalled replication forks contributes to genomic instability and specifically to breast and ovarian cancer predisposition in BRCA mutant cells. In 2018, the Tus-*Ter* fork barrier system was used to investigate whether non-homologous end joining or homologous recombination was activated to repair mammalian chromosomal replication fork barriers (Willis et al., 2018). The authors

found that homologous recombination is the primary repair mechanism through the early recruitment of Rad51 at the site of the stalled replication forks.

The Tus-*Ter* system has been a useful tool to examine recombination events in eukaryotes and will no doubt be used in studies looking at the impact of replication fork barriers in other organisms (Willis et al., 2014, Willis et al., 2018).

1.7.3 Future perspective of Tus-based technologies

The Tus protein had been studied in great detail over several decades, eventually culminating in 2006 with the Tus-*Ter*-lock structure that enabled delineation of the currently accepted mechanistic model of polar fork arrest in *E. coli*. Tus had finally achieved a status of a 'model' DNA-binding protein with attractive and predictable attributes. Not surprisingly, the first application of Tus emerged soon after. While its utility as a protein-DNA conjugation system was somewhat obvious given its monomeric and extremely high-affinity binding properties, its development into eukaryotic systems as a site-specific heterologous DNA replication perturbating tool could be one of the most innovative and informative contributions to the field of DNA replication. There is obvious scope for the characterisation and development of orthologous Tus-*Ter* systems with new DNA binding properties. Finally, it is hoped that the applications of Tus presented here will provide a platform for developing new ideas into innovative biotechnologies across various disciplines.

2 GENERAL MATERIALS AND METHODS

2.1 Bioinformatic resources and protocols

This section relates mostly to my contribution to the manuscript in Chapter 3. This includes the sequencing and genome assembly of KRX *E. coli* (Chapter 3), ChIP-Seq analysis (Chapter 3), phylogenetic analysis (Chapter 3) and protein modelling (Chapter 3 & 4). Here, a list of the tools and software commonly used to complete these analyses and a brief overview of each method is provided. More detailed descriptions can be found in the Materials and Methods sections of Chapter 3.

Table 2: List of software and algorithms used for this thesis

Software and Algorithms		
MinKNOW	Oxford Nanopore Technologies	https://github.com/nanoporetech/minknow_api
Trimmomatic	(Bolger et al., 2014)	http://www.usadellab.org/cms/?page=trimmomatic
Porechop	(Wick et al., 2017)	https://github.com/rrwick/Porechop
Flye	(Kolmogorov et al., 2019)	https://github.com/fenderglass/Flye
Racon	(Vaser et al., 2017)	https://github.com/isovic/racon

Pilon	(Walker et al., 2014)	https://github.com/broadinstitute/pilon
Quast	(Gurevich et al., 2013)	http://quast.sourceforge.net/quast
Prokka	(Seemann, 2014)	https://github.com/tseemann/prokka
Bowtie2	(Langmead and Salzberg, 2012)	http://bowtie-bio.sourceforge.net/bowtie2/index.shtml
Samtools	(Li et al., 2009)	http://www.htslib.org/
Circleator	(Crabtree et al., 2014)	http://jonathancrabtree.github.io/Circleator/
genomeCoverageBed	(Quinlan and Hall, 2010)	https://bedtools.readthedocs.io/en/latest/content/tools/genomecov.html
blastn	(Altschul et al., 1990)	https://ftp.ncbi.nlm.nih.gov/blast/executables/blast+/LATEST/
Interactive Genomics Viewer (IGV)	(Thorvaldsdottir et al., 2013)	http://software.broadinstitute.org/software/igv/
EzMol Molecular display wizard	(Reynolds et al., 2018)	http://www.sbg.bio.ic.ac.uk/ezmol/

InterPro Protein Data Bank	(Mitchell et al., 2019)	https://www.ebi.ac.uk/interpro/
iTol	(Letunic and Bork, 2019)	https://itol.embl.de/
RAxML	(Stamatakis, 2014)	https://cme.h-its.org/exelixis/web/software/raxml/
MUSCLE	(Edgar, 2004)	http://www.drive5.com/muscle/downloads.htm
ImageJ	(Schneider et al., 2012)	https://imagej.nih.gov/ij/index.html
Nucplot	(Luscombe et al., 1997)	https://www.ebi.ac.uk/thornton-srv/software/NUCPLOT/
GraphPad 8	GraphPad Software	https://www.graphpad.com/scientific-software/prism/

2.1.1 Nanopore long read sequencing and genome assembly of KRX *E. coli*

DNA from a growing culture of KRX *E. coli* was prepared for sequencing on a FLO-MIN106 R9 MinION flow cell (Oxford Nanopore). DNA concentration and quality were assessed using Invitrogen Qubit 4 and agarose gel electrophoresis. The Rapid Sequencing protocol SQK-RAD004 (Oxford Nanopore) was used for library preparation and base-calling was processed using the pipeline implemented in MinKNOW software version 18.01.6 (Oxford Nanopore). After read quality filter steps, a total of 225 985 reads which had a total 146x coverage of the KRX *E. coli* genome were used for assembly. Polishing was performed with Racon iteratively

four times in combination with Pilon (version 1.23-1) using the entire Illumina ChIP-Seq data. The genome was annotated and evaluated against the K12 *E. coli* genome.

2.1.1 ChIP-Seq analysis

Illumina read quality of each biological and technical replicate fastqc file was assessed and subject to a quality control and removal of Illumina adapters. Sequenced samples were individually aligned to the polished KRX genome and the data visually assessed in terms of replicability between each replicate and for any outliers. The replicate data were pooled into a single file and aligned to the KRX *E. coli* genome. A circular annotation of the ChIP-Seq reads mapped to the KRX *E. coli* genome was created as well as the GC-skew of the chromosome. The individual positions and orientations of *Ter* sites were identified and verified using their specific 23 bp sequences and compared to the respective sites in K12 *E. coli*.

2.1.1 Fork trap characterisation in Enterobacteriales

2.1.1.1 Phylogenetic analysis of Tus

Bacterial species within the order Enterobacteriales containing a *tus* gene ortholog were identified in InterPro. Several Tus protein sequences from a selection of *Pseudoalteromonas* species were also included as outgroup. The Tus proteins from each species were aligned and a phylogenetic tree was constructed in RAxML v8.2.0 (Stamatakis, 2014). The consensus tree produced was visualised and edited in iTol (Letunic and Bork, 2019).

2.1.1.2 Characterisation of the replication fork trap

A random selection of species from different families were chosen using the phylogenetic tree previously constructed. Upon identification of a *tus* gene otholog, the adjacent *Ter* site was identified within its 50 bp 5' UTR by aligning the 23 bp *E. coli TerB* sequence. For each selected species, a BLAST search was carried out using the adjacent *Ter* sequence to locate further *Ter* sites within their genome with a refined definition of what constitutes a *Ter* site. A circular

annotation of each genome to display the architecture of the fork trap was generated alongside the GC-skew of the chromosome (5000 bp window) using Circleator (version 1.0.2).

2.1.1 Deposited Data

All ChIP-Seq data and the genome assembly of *KRX E. coli* for which it was analysed against was deposited into NCBI GEO under the accession number: GSE163680.

2.1.2 Quantification and statistical analysis

Statistics and number of biological and technical repeats are indicated in the relevant figure legends, tables and methods. Statistical analyses were performed using Graphpad Prism 7. Data are expressed as mean values ± SD, ± SE or ranges.

2.2 Materials

This section relates strictly to the general materials used to complete Chapter 4 and Chapter 5 and is completely my own work unless otherwise specified.

2.2.1 Bacterial strains and plasmids

Table 3: Genotype of bacterial strains and plasmids used

STRAIN OR PLASMID	DESCRIPTION/GENOTYPE	SOURCE
***E. coli* strains:**		
DH12s	Used for plasmid preparation. Genotype: 80d*lacZ*ΔM15 *mcrA* Δ(*mrr-hsd*RMS-*mcr*BC) *ara*D139 Δ(*ara*, *leu*)7697 Δ(*lac*X74 *gal*U *gal*K *rps*L (Str^R) *nup*G *rec*A1/F' *pro*AB^+ *lac*$I^qZ\Delta$ M15	Invitrogen
BL21(*DE3*)RIPL	Used for expression of proteins. Genotype: B F– *omp*T *hsd*S(r_B^- m_B^-) *dcm*$^+$ Tet^r *gal* *λ(DE3) endA* Hte [*argU proL* Cam^r] [*argU ileY leuW* Strep/Specr]	Stratagene
KRX	Used for expression of proteins. Genotype: [F′, *tra*D36,Δ*omp*P, *pro*A^+B^+, *lac*Iq, Δ(*lacZ*)M15]Δ*omp*T, *end*A1, *rec*A1, *gyr*A96(Nal^r), *thi*-1, *hsd*R17(r_k^-,m_k^+), e14$^-$ (McrA$^-$	Promega

	), *rel*A1, *sup*E44,Δ(*lac-pro*AB), Δ(*rha*BAD)::T7 RNA polymerase.	
Plasmids:		
pPMS1259	Ap^r, T7P, N-His_6-$_{Ec}$Tus-GFP-C	(Figure 8A) (Dahdah et al., 2009)
pIM013	Ap^r, T7P, GFP-cloning cassette	(Moreau et al., 2010)
pCT300	Ap^r, T7P, pIM013 carries $_{Dp}$Tus-GFP, N-His_6 -$_{Dp}$Tus-GFP-C	This Study (Figure 8B)

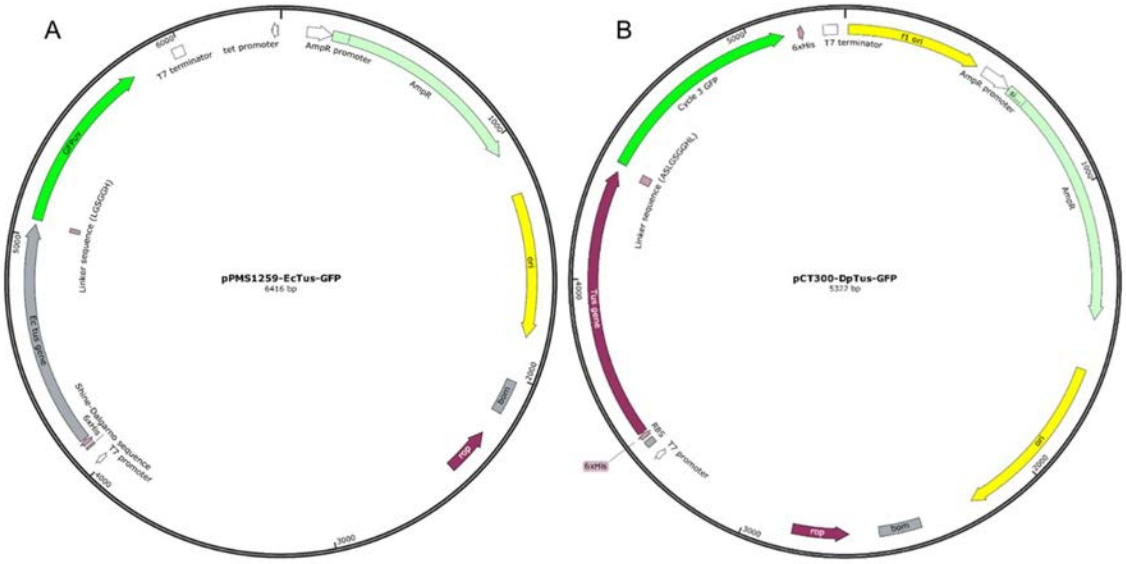


Figure 8: The pPMS1259 and pCT300 expression vectors used in this study. Details of the construction of pPMS1259 were described previously (Dahdah et al., 2009). For pCT300, the protein coding sequence for the DNA replication terminus site-binding protein (Tus) from *Dickeya paradisiaca* was codon optimized for *E. coli* and cloned into the pET-uvGFP expression vector, pIM013 (Moreau et al., 2010) with EcoRV and NheI using the Gene Synthesis Service offered by Bioneer Pacific. Other than the Tus protein, the $_{Ec}$Tus-GFP protein product only differs slightly in the shorted linker sequence (LGSGGH) compared to the $_{Dp}$Tus-GFP linker (ASLGSGGHL).

2.2.2 Key Resource Table

Table 4: Reagents and resources used in this study

REAGENT/ RESOURCE	DESCRIPTION	NOTES/ SOURCE
Transformation and Expression media		
Luria-Bertani medium (LB)	5 g of peptone, 5 g of potassium chloride and 2.5 g of yeast extract were dissolved in 500 mL $_{dd}H_2O$. For agar plates, 7.5 g of Agar was added. The appropriate antibiotics were mixed in after cooling to ~50°C: ampicillin (100 µg/ml final) and/or chloramphenicol (50 µg/ml).	
Terrific broth media (TB)	12 g of tryptone, 24 g of yeast extract and 4 mL of glycerol were dissolved in $_{dd}H_2O$ q.s. 1 L and autoclaved. Once the broth had cooled to 55°C, 10 mL of a sterile filtered phosphate buffer solution (0.17 M KH_2PO_4 and 0.72 M K_2HPO_4) was added.	
Auto-induction media	2 mL of a 1 M glucose solution (4 mM final) and 2 mL of a 0.1 M D-galactose solution (0.4 mM final) were sterile filtered into 496 mL of terrific broth.	

REAGENT/ RESOURCE	DESCRIPTION	NOTES/ SOURCE
Chemicals and Buffers		
Competent cell buffer (CCMB)	0.59 g of $CaCl_2 \cdot 2H_2O$ (80 mM final), 0.2 g of $MnCl_2 \cdot 4H_2O$ (20 mM final), 0.1 g of $MgCl_2 \cdot 6H_2O$ (10 mM final), 0.049 g of CH_3COOK (25 mM final) and 5 mL of glycerol (10% v/v final) were dissolved in $_{dd}H_2O$ q.s. 50 mL	
STET	8% sucrose (w/v), 0.1% TritonX-100 (v/v), 50 mM EDTA (pH 8) and 50 mM Tris-HCl (pH 8).	
STET++	300 µL of STET was combined with 5 µL of a 50 mg/mL lysozyme stock and 1 µL of a 10 mg/mL RNAse A stock	
Isoprop/NH4 Ac	19.3 g Ammonium acetate and 75 mL of Isopropanol was combined in $_{dd}H_2O$ q.s. 100 mL	
Lysis buffer	9 mL of a 0.5 M Na_2HPO_4 solution (45 mM final), 0.5 mL of a 1 M NaH_2PO_4 solution (5 mM final), 3.5 g of NaCl (300 mM final), 10 mL of glycerol (10% v/v final), 1 mL of a 1 M imidazole solution (10 mM final) and 14 µl of β-mercaptoethanol (2 mM final) were combined with $_{dd}H_2O$ q.s. 100 ml (pH 7.8).	
Wash buffer	Same as Lysis buffer + 20 mM imidazole (pH 7.8).	

REAGENT/ RESOURCE	DESCRIPTION	NOTES/ SOURCE
Elution buffer	Same as Lysis buffer + 200 mM imidazole (pH 7.8).	
Buffer A	4.5 mL of a 0.5 M Na_2HPO_4 solution (45 mM final), 0.25 mL of a 1 M NaH_2PO_4 solution (5 mM final), 5 mL of glycerol (10% v/v final) and 7.5 µl of β-mercaptoethanol (2 mM final) was combined with $_{dd}H_2O$ q.s. 50 mL (pH 7.8).	All purified proteins were resuspended in buffer A
Buffer B	Buffer A + 125 mM NaCl and without β-mercaptoethanol	Size exclusion chromatography buffer (Chapter 4 p126)
Buffer C	0.12 g of Tris base (20 mM final) and 0.44 g of NaCl (150 mM final) was dissolved in 45 mL of $_{dd}H_2O$, adjusted to pH 8 with a 5 M HCl solution and then topped up to 50 mL with $_{dd}H_2O$	All oligonucleotides were dissolved in buffer C
Coat buffer	Buffer A without β-mercaptoethanol	qIPCR (Chapter 5 p126)
Block buffer	Buffer C + 0.005% Tween-20 and 1% BSA	qIPCR (Chapter 5 p126)
qIPCR wash buffer	Block buffer without BSA	qIPCR (Chapter 5 p126)

REAGENT/ RESOURCE	DESCRIPTION	NOTES/ SOURCE
5xTBE buffer	54 g of Tris base, 27.4 g of boric acid and 20 mL of 0.5 M EDTA (pH 8.0) was combined with $_{dd}H_2O$ q.s. 1000 mL	
2xSDS-Sample Buffer	5 mL of a 0.5 M Tris-HCl (pH 6.8) solution (0.2 M final), 4 mL glycerol (20% v/v final), 8 mL of a 10% SDS solution (4% v/v final), 1 mL of a 0.1% Bromophenol Blue solution (0.005% v/v final), 2.8 µL of β-Mercaptoethanol (2 mM final) and 2 mL of $_{dd}H_2O$ were mixed and 1.5 mL aliquots were stored at RT.	(0.1% Bromophenol Blue solution = 10 mg in 10 mL of $_{dd}H_2O$)
Acrylamide/ Bis-acrylamide	19.3 g of acrylamide and 0.7 g of Bis-acrylamide were dissolved in 50 mL of $_{dd}H_2O$.	
Bradford reagent	Catalogue number: B6916-500ML	Sigma
Coomassie Stain (R-250)	0.5 g of R-250 Coomasie (Bio-Rad) was dissolved in 90 mL of methanol, 20 mL of glacial acetic acid and 90 mL of $_{dd}H_2O$	SDS-PAGE gels were stained for 2 hours
De-stain solution	10% acetic acid, 40% isopropanol	SDS-PAGE gels were de-stained until transparent

REAGENT/ RESOURCE	DESCRIPTION	NOTES/ SOURCE
Antibodies		
Anti-GFP IgY	Aves Labs (GFP-1010)	qIPCR (Chapter 5 p126)
Gels		
Agarose GE	Agarose gels (1-2% w/v) were used to separate digested plasmid and sheered genomic DNA and created using a standard procedure (Sambrook and Russell, 2001). Gels containing 1x gel red were run in TBE buffer (1x).	
SDS-PAGE	12% SDS-PAGE gels were used to assess protein purity and created using a standard procedure (Sambrook and Russell, 2001). Protein samples were mixed with 2X SDS sample buffer. Protein samples that required to be fully denatured were heat treated at 95°C for 5 min. All protein bands were stained with Coomasie blue staining solution (R-250) for 2 hours and destained until transparent.	
Protein quantification		
Bradford assay	The protein concentrations of purified samples were determined using Bradford reagent following the	

REAGENT/ RESOURCE	DESCRIPTION	NOTES/ SOURCE
	recommended protocol for the BMG LABTECH microplate reader (BMG LABTECH Application note 158, available at: https://www.bmglabtech.com/bradford-assay-performed-on-bmg-labtech-microplate-readers/)	
Laboratory Instruments		
French Press G-M	Used for lysis of bacterial cells (12000 p.s.i)	GlenMills
G:BOX Chemi XRQ	Used to capture agarose and SDS-PAGE gels for: • GFP Fluorescence (Blue LED Module, Filt525) • Gel red stained DNA (TLUM mid wave, UV06) • Coomasie Blue (While light)	Syngene
CFX96/ C1000 Touch Thermal Cycler	Used for DSF-GTP assay and qIPCR	Bio-Rad
BioLogic Duo- Flow	Used for analytical size exclusion chromatography and purification of proteins	Bio-Rad

2.2.3 Oligonucleotides

All oligonucleotides were purchased from Bioneer Pacific and resuspended at 400 μM in buffer C (20 mM Tris-HCl, 150 mM NaCl, pH 8). Complementary DNA strands detailed in Table 5 were annealed by heating at 82°C for 5 min followed by gradually cooling to room temperature (RT) over 1 hour.

Table 5: *Ter* and *Ter-lock* oligonucleotides

Name	Sequence	JCU#
Oligonucleotides used for DSF-GTP and GFP-EMSA		
Ter4	5′-GGCCAGAATGTTGTAACTCCCGT*GGGGGCGGGG*-3′	477
	3′-CCGGTCTTACAACATTGAGGGCA*CCCCCGCCCC*-5′	478
Ter4-lock	5′- GAATGTTGTAACTCCCGT*GGGGGCGGGG*-3′	488
	3′-CCGGTCTTACAACATTGAGGGCA*CCCCCGCCCC*-5′	478
Ter3	5′-TAAAAGTATGTTGTAACTATTTA*GGGGGCGGGG*-3′	479
	3′-ATTTTCATACAACATTGATAAAT*CCCCCGCCCC*-5′	480
Ter3-lock	5′- TATGTTGTAACTATTTA*GGGGGCGGGG*-3′	485
	3′-ATTTTCATACAACATTGATAAAT*CCCCCGCCCC*-5′	480
Ter2	5′-CATTAGTATGTTGTAACTAATTT*GGGGGCGGGG*-3′	481
	3′-GTAATCATACAACATTGATTAAA*CCCCCGCCCC*-5′	482
Ter2-lock	5′- TATGTTGTAACTAATTT*GGGGGCGGGG*-3′	486
	3′-GTAATCATACAACATTGATTAAA*CCCCCGCCCC*-5′	482
Ter1	5′-AAACAGTATGTTGTAACTAATAT*GGGGGCGGGG*-3′	483
	3′-TTTGTCATACAACATTGATTATA*CCCCCGCCCC*-5′	484
Ter1-lock	5′- TATGTTGTAACTAATAT*GGGGGCGGGG*-3′	487
	3′-TTTGTCATACAACATTGATTATA*CCCCCGCCCC*-5′	484
TerB	5'-AATAAGTATGTTGTAACTAAAGT*GGGGGCGGGG*-3'	490
	3'-TTATTCATACAACATTGATTTCA*CCCCCGCCCC*-5'	493

Name	Sequence	JCU#
TerB-lock	5′- TATGTTGTAACTAAAGT*GGGGGCGGGG*-3′ 3'-TTATTCATACAACATTGATTTCA*CCCCCGCCCC*-5'	489 493
TerF	5'-CCTTCGTATGTTGTAACGACGAT*GGGGGCGGGG*-3' 3'-GGAAGCATACAACATTGCTGCTA*CCCCCGCCCC*-5'	107 108
TerF-lock	5'- TATGTTGTAACGACGAT*GGGGGCGGGG*-3' 3'-GGAAGCATACAACATTGCTGCTA*CCCCCGCCCC*-5'	144 108

Oligonucleotides used include the 23 bp *Ter* or *Ter-lock* sequence as well as a stabilizing 10-mer GC rich region (italicized) to elevate their melting temperatures above 70°C for DSF-GTP applications.

Oligonucleotides used for real-time Immuno-PCR		
Ter1 L-barcode	CACCGCTGAGCAATAACTAGCAT[8x(A)]GAACTGGATCTCAACAGCGGTATATTAGTTACAACATACTGTTT 3′-TATAATCAATGTTGTAT-5′	496 494
Ter1-barcode	CACCGCTGAGCAATAACTAGCAT[8x(A)]GAACTGGATCTCAACAGCGGTATATTAGTTACAACATACTGTTT 3′-GGGGCGGGGGTATAATCAATGTTGTATGACAAA	496 483
ss*Ter1*-barcode	CACCGCTGAGCAATAACTAGCAT[8x(A)]GAACTGGATCTCAACAGCGGTATATTAGTTACAACATACTGTTT	496
Fwd Primer	5'-CACCGCTGAGCAATAACTAGCAT-3'	39
Rev Primer	5'-ACCGCTGTTGAGATCCAGTTC-3'	40

The *Ter1L* Barcode contains the *Ter1* site (underlined) that is free at the C(6) as well as a sequences that can be PCR amplified by the forward (blue) and reverse (red) primers. [8x(A)] represents a string of eight adenine bases. The *Ter1* complementary sequence is reused from the DSF-GTP assay and therefore has a stretch of Gs at its 3’ end that has no effect on the qIPCR platform.

2.3 Laboratory methods

2.3.1 Chemical competent BL21(DE3)RIPL

E. coli cells such as BL21(*DE3*)RIPL cells were grown on an LB agar plate supplemented with chloramphenicol (LBC) and incubated overnight at 37°C. A single colony was selected to inoculate 5 mL of LBC broth and grown overnight shaking (200 RPM) at 37°C. A 100 µL aliquot of overnight culture was used to inoculate 5 mL LBC culture which was incubated until OD_{600} reached 0.5 followed by centrifugation at 4°C and 2000*g* for 10 min. Cell pellets were washed with 3 mL ice cold CCMB buffer followed by centrifugation at 4°C and 2000*g* for 10 min. The supernatant was discarded, and the pellet resuspended in 1.5 mL CCMB buffer. Chemical competent cells were frozen in liquid nitrogen and stored at -80°C.

2.3.2 Bacterial transformation

Competent cells (25 µL) were mixed with 1 µL of plasmid DNA and incubated for 20 min on ice. Cells were heat shocked for 50 s at 42°C and immediately placed back on ice for 2 min. LB broth (500 µL) was added and the suspension was incubated at 37°C for 1 hour. Cells were very briefly centrifuged and 250 µL of supernatant removed before streaking 100 µL on LB agar plates containing the appropriate antibiotic. Plates were incubated overnight at 37°C.

2.3.3 Plasmid preparation

DH12s cells were commonly used as host strain for propagating plasmids. LBA broth (1.5 mL) was inoculated with transformed DH12s cells and incubated overnight at 37°C and 200 RPM. Plasmid DNA was extracted and purified using the Rapid-boil Mini-prep protocol available at http://www.molbi.de/protocols/miniprep_rapid_boiling_v1_0.htm. Briefly, bacteria were centrifuged at 6000*g* for 2 min and the pellet resuspended in 300 µl STET++ buffer. After 10 min incubation at RT, the cell suspension was heated for 2 min at 95°C. The cell lysate was centrifuged 15 min at 18000*g* and the supernatant transferred to a fresh 1.5 mL centrifuge tube

and mixed with 500 µl Isoprop/NH4Ac and incubated at RT for 20 min. The suspension was centrifuged for 15 min at 18000*g* and the pellet washed in 500 µl of 70% ethanol before recentrifuging for 10 min at 18000*g*. The supernatant was discarded and the pellet left to dry for 10 min before resuspending in water. DNA was quantified using Nanodrop and the quality verified by agarose gel electrophoresis.

2.3.4 Protein expression and purification

The expression and purification of *Escherichia coli* Tus-GFP (His_6-$_{Ec}$Tus-GFP) was performed as previously described by Moreau, MJ (2013) 'DNA replication in *Escherichia coli:* A comprehensive study of the Tus-*Ter* complex', PhD thesis, James Cook University, Townsville. The same protocol was used to express and purify Tus-GFP from *Dickeya paradisiaca* (His_6-$_{Dp}$Tus-GFP). Briefly, a bacterial loop retrieved from a fresh master plate of *E. coli* BL21(*DE3*)RIPL cells containing either pPMS1259 (His_6-$_{Ec}$Tus-GFP also referenced as $_{Ec}$Tus-GFP) or pCT300 (His_6-$_{Dp}$Tus-GFP also referenced as $_{Dp}$Tus-GFP) was used to inoculate 100 mL of Auto-induction medium supplemented with 100 µg/ml ampicillin and 50 µg/ml of chloramphenicol. Cultures were grown at 37 ℃ and 200 RPM until OD_{600} reached 0.8 and then at 16ºC for a further 72 hours. Cultures were centrifuged at 4℃ and 3000*g* for 40 min, resuspended at 7.5 ml/g wet cells with ice-cold lysis buffer and lysed with two successive passages in a French Press (12000 p.s.i). The lysate was centrifuged at 4℃ and 40000*g* for 40 min and the supernatant containing soluble proteins separated with Profinity IMAC nickel-charged resin (Bio-Rad) as described by Dahdah et al. (2009). The fractions containing the His_6-proteins were pooled and precipitated with an equal volume of a saturated ammonium sulfate solution at 4˚C for 1 hour followed by centrifugation at 4˚C and 40000*g* for 30 min. The supernatant was removed and the protein pellet resuspended in buffer A. Protein concentrations were determined by Bradford Assay and purity assessed by SDS-PAGE. Protein samples were thawed on ice and centrifuged at 4℃ and 40000*g* for 15 min to remove aggregates prior use.

2.3.5 Size exclusion chromatography

The detailed procedures are described in Chapter 4. Size exclusion chromatography (SEC) was used as an additional purification step for proteins and to compare the quaternary structure of $_{Dp}$Tus-GFP to $_{Ec}$Tus-GFP. Proteins were separated using a Superdex 200 10/300 GL column (GE Lifesciences) and Biologic Duoflow Chromatography System (Biorad) in buffer B.

2.3.6 GFP-EMSA

The detailed procedures are described in Chapter 4. A modified version of an electrophoretic mobility shift assay (EMSA) for GFP tagged proteins (Sorenson and Schaeffer, 2020) was used to compare the binding of Tus-GFP to *Ter* DNA species at varying NaCl concentrations. Experiments were run in duplicates.

2.3.7 DSF-GTP assay

The detailed procedures are described in Chapter 4. The Differential scanning fluorimetry of GFP-tagged proteins (DSF-GTP) assay (Moreau et al., 2012) was used to obtain melting curves for free and DNA-bound Tus-GFP at varying NaCl concentrations (*c.f.* Table 5 for oligonucleotides used).

2.3.8 Quantitative immuno-PCR (qIPCR)

The detailed procedures are described in Chapter 5. Tus-GFP in complex with an extended *Ter1*-lock sequence that can be amplified by qPCR (*Ter1*-lock-barcode) was used as a detection device to measure preadsorbed polyclonal anti-GFP IgY similarly to a previously described method (Morin et al., 2010).

3 DEFINING THE PROTOTYPICAL DNA REPLICATION FORK TRAP IN BACTERIA

This chapter is formatted as a journal manuscript that is a collaborative work (Casey J. Toft, Morgane J. J. Moreau, Jiri Perutka, Savitri Mandapati, Peter Enyeart, Alanna E. Sorenson, Andrew Ellington and Patrick M. Schaeffer). The manuscript has been slightly amended to be coherent with the rest of the thesis. My contribution to this work involved ChIP-Seq data analysis including long-read genome sequencing and assembly of KRX E. coli, phylogenetic analysis of Tus, fork trap characterisation of Enterobacteriales, production of all figures except for the ectopic insertion figure, as well as synthesis of the findings. Construction of the strains and analysis of the growth curves for ectopic insertions of Ter were completed by Jiri Perukta, Savi Mandapati and Peter Enyart as part of the Ellington lab at the institute for Cell and Molecular Biology, University of Texas at Austin. Aspects of collaborative work were drafted with help of the coauthors (cf statement of contributions of others).

3.1 Introduction

The best studied mechanism of bacterial DNA replication termination requires the presence of a DNA replication fork trap within the chromosomal terminus region. This is exemplified in *E. coli* by the presence of a cluster of five similar but distinct 23-bp *Ter* DNA sequences on each chromosomal arm, which have anti-helicase activity when they are bound by the replication termination protein Tus (Berghuis et al., 2018, Xu and Dixon, 2018, Neylon et al., 2005a). The complexity of the *E. coli* replication fork trap with respect to multiplicity and wide distribution of *Ter* sites around the chromosome is puzzling. One cluster consisting of *TerB*, *C*, *F*, *G* and *J* arrests the clockwise moving replication fork and the other oppositely-oriented cluster consisting of *TerA*, *D*, *E*, *I*, *H* arrests the anti-clockwise moving replication fork (Figure 9A). Until recently, the notions that Tus could bind to all ten of these slightly different *Ter* DNA sequences (*TerA-J*) (Figure 9B), and that these sequences all have a significant role in replication termination, have remained mostly unchallenged despite their individual binding properties for Tus being significantly different (Neylon et al., 2005b). Each *Ter* cluster consists of three high affinity, one moderate-to-low affinity and one non-lock forming *Ter* site (Figure 9A) (Moreau and Schaeffer, 2012b, Moreau and Schaeffer, 2012a). Four additional *Ter*-like sequences (*TerK*, *L*, *Y* and *Z*) can be found in the *E. coli* chromosome, one within the previously identified termination region and the other three being on the left part of the chromosome, but these were dismissed as pseudo-*Ter* sites (Duggin and Bell, 2009). Binding of Tus to the pseudo-*Ter* sites is likely to be insignificant based on their sequences (Moreau and Schaeffer, 2012a) and fork arrest efficiency (Duggin and Bell, 2009).

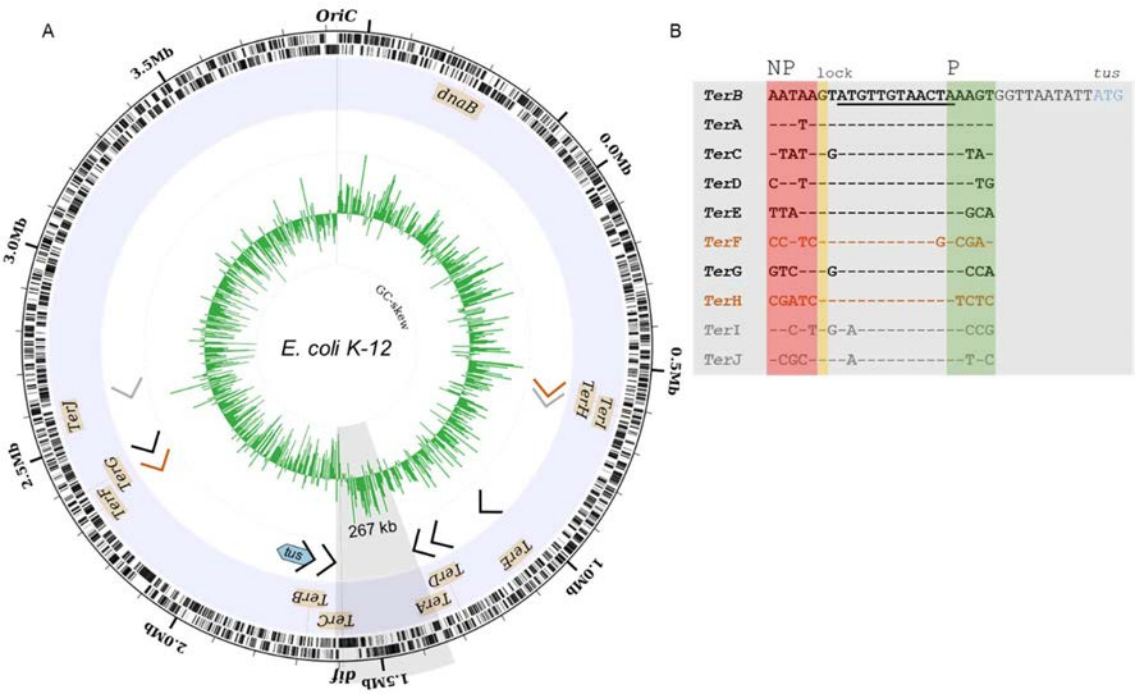


Figure 9: Chromosomal distribution and sequences of *Ter* sites in *E. coli.* (A) Circular representation of *E. coli* K12 MG1655. Illustrated from the outside to the centre of the circle: labelled forward and reverse genes; location of the ten primary *Ter*-sites (*TerA-J*) and their strand orientation; the currently accepted replication termination fork trap involving high affinity (black), moderate affinity (grey) and non-lock forming *Ter* sites (orange); GC-skew over a 5000 bp window showing a switch in polarity at the replication origin and close to *TerC* and the *dif* site. (B) *Ter* site sequences with the G(6) base complementary to C(6) highlighted in yellow and the strictly conserved 12 bp core sequence (underlined). *TerB* is located 11 bp upstream of the start codon (ATG) of the *tus* gene. NP: non-permissive face (red), P: permissive face (green).

The unique mechanism of polar DNA replication fork arrest observed in *E. coli* is due to the unusual binding mode of Tus to *Ter* and the unwinding action of the DnaB helicase at the non-permissive face of the Tus-*Ter* complex (Bastia et al., 2008, Duggin and Bell, 2009, Kaplan, 2006, Mulcair et al., 2006b, Neylon et al., 2000, Neylon et al., 2005a, Schaeffer et al., 2005).

Although a specific protein-protein interaction between the DnaB helicase and Tus had initially been proposed to have a pivotal role in polar fork arrest (Bastia et al., 2008, Mulugu et al., 2001), several studies have shown that this interaction is not necessary for polar fork arrest at the non-permissive face of the Tus-*Ter* complex (Pandey et al., 2015, Berghuis et al., 2015, Elshenawy et al., 2015, Mulcair et al., 2006b, Moreau and Schaeffer, 2012a). Tus precisely and tightly binds onto a *Ter* site, bending the DNA to prepare the molecular mouse trap that will be triggered by the 5'-3' translocation and DNA unwinding action of DnaB helicase on the lagging strand moving towards the non-permissive face of the Tus-*Ter* complex (Kamada et al., 1996, Mulcair et al., 2006b, Moreau and Schaeffer, 2012a). The progressive separation of DNA strands at the non-permissive face of the Tus-*Ter* complex ultimately breaks the GC(6) base pair (Figure 9B) in the *Ter* core sequence leading to the precise docking of the freed C(6) into a cytosine-specific binding pocket on the surface of Tus. The formation of the locked Tus-*Ter* conformation (TT-lock) slows the dissociation of Tus considerably and is believed to inhibit further DnaB helicase translocation (Kaplan, 2006, Moreau and Schaeffer, 2012a, Mulcair et al., 2006b). By using a systematic and comparative approach examining the kinetic and equilibrium parameters of all ten Tus-*Ter* and locked complexes, the binding of Tus to *Ter* and formation of the locked complex could be dissected in further detail (Moreau and Schaeffer, 2012a, Moreau and Schaeffer, 2013). The study proposed a sequential three-step model for fork arrest including initial non-specific 'sliding' of Tus on DNA mediated by weak cooperative electrostatic interactions, followed by proper ratchet-like docking of Tus onto *Ter* upon correct alignment of specific nucleotide-amino acid contacts, and finally the DnaB-induced Tus-*Ter*-lock via binding of C(6) to the cytosine binding-pocket of Tus (Moreau and Schaeffer, 2012a). The same study provided a new classification of *Ter* sites based on their kinetic and affinity parameters as well as their capacity to form a locked complex, and challenged the status quo by rejecting *TerF* and *TerH* as functional *Ter* sequences arguing that they could not induce

polar fork arrest due to their inability to form a locked complex. A substitution of the canonical T to G at position 5 in the core sequence (Figure 9B) was proposed to be the cause for this loss of function (Moreau and Schaeffer, 2012a). In addition, Coskun-Ari and Hill had previously observed a significant loss of replication fork arrest activity when *TerB* was mutated at TA(9) to CG or GC (Coskun-Ari and Hill, 1997). *TerI* and *TerJ* are the only *Ter* sites where the TA(9) is replaced by an AT(9), which taken together with the relatively fast Tus dissociation from these locked Tus-*Ter* complexes (Moreau and Schaeffer, 2012a) also raises doubts about their actual role in replication termination under natural Tus abundance conditions.

Using two-dimensional gel analysis of replication intermediates at each *Ter* site under natural growth conditions, Duggin and Bell showed that *TerC* is the most frequently used termination site with significant fork pausing also observed at *TerA* and *B* (Duggin and Bell, 2009). Additional Tus over-expression studies were required to better understand the involvement of the remaining '*Ter'* sites, which in wild-type conditions produced an unquantifiable signal intensity difference between paused fork and linear DNA (*TerD*, *G*, *H* and *I*) or had a negligible paused fork pattern (*TerE, F* and *J*) (Duggin and Bell, 2009). The sites varied significantly in their capability to arrest replication forks and this was later correlated with their respective affinity for Tus (high, moderate or weak) and their ability to form a TT-lock (Moreau and Schaeffer, 2012a). Most puzzling, however, was the observation that some pausing occurred at the outer *TerH.* To be arrested at *TerH*, the replication fork has to break through the strong *TerE* and moderate *TerI* but no fork pausing was observed at *TerE* and little at *TerI*. Although Duggin and Bell showed that pausing was abolished at *TerC* in a *tus* null strain, they did not verify if the pausing observed at the outermost *TerH-I* sites was also strictly due to Tus binding (Duggin and Bell, 2009). The low probability of the anti-clockwise fork reaching *TerH*, the absence of pausing at the strong *TerE* (Duggin and Bell, 2009) and the non-TT-lock forming characteristic of *TerH* (Moreau and Schaeffer, 2012a), suggest that the pausing observed at *TerH* could either

be due to the clockwise moving fork at the permissive face of *TerH* or to recombination events (Horiuchi et al., 1995, Mohanty et al., 2009, Rothstein et al., 2000).

The presence of the distal *Ter* sites and their involvement in DNA replication termination remains unclear. Forks most frequently meet at *TerC* and to some extent at *TerA* and *B* (Duggin and Bell, 2009). Assuming the two forks progress at equivalent rates, forks are more likely to meet at *TerC* than at *TerA* since *TerC* is almost perfectly located directly opposite to *oriC* whereas the anti-clockwise moving fork must travel an additional ~259 kb to encounter the non-permissive face of the Tus-*TerA* complex. Despite the stability of the locked Tus-*TerC* complex (Moreau and Schaeffer, 2012a) in over-expressed Tus conditions, significant pausing still occurs at *TerB* and to some extent at *TerG* (Duggin and Bell, 2009). One possible explanation for pausing at *TerB* is that in some cases, the ratchet-lock mechanism (Moreau and Schaeffer, 2012a) fails to form and the next site serves as a backup for DNA replication arrest. In support of this, single molecule DNA replication assays suggest that a replication fork approaching a non-permissive *TerB* will fail to be arrested 52% of the time as a result of an inefficient Tus-*Ter*-lock mechanism (Elshenawy et al., 2015). The authors proposed lock-formation is dependent on transient fork stoppage by an Arg198 interaction that buys time for C(6) to dock into its binding pocket (Elshenawy et al., 2015). Hence the critical need for backup *Ter* sites throughout the terminus of the genome is for replication forks that have breached the innermost *TerA* and *TerC* sites.

So far, the large number of binding, structural and single molecule studies designed to thoroughly examine the Tus-*Ter* complexes (Moreau and Schaeffer, 2012b, Moreau and Schaeffer, 2012a, Mulcair et al., 2006a, Berghuis et al., 2018, Elshenawy et al., 2015, Berghuis et al., 2015), as well as fluorescence imaging aiming at examining the progression and pausing of replication forks at natural replication barriers in live bacteria (Moolman et al., 2016), have

failed to provide a clear explanation for the need of such a large replication fork trap in *E. coli*. In fact, the most significant knowledge gap on Tus-*Ter* replication fork traps, *i.e.* the binding distribution of Tus to individual *Ter* sites in replicating bacteria, has not been addressed. As such, the function of distal *Ter* sites and their biological relevance remains unclear and calls into question as to whether or not Tus proteins (Roecklein et al., 1991, Roecklein and Kuempel, 1992) really bind these *Ter* sites, and if yes, then to what extent?

Here the effects of ectopically and precisely inserting a selection of *Ter* sites (*TerB*, *H* and *J*) into a safe chromosomal locus in both permissive and non-permissive orientations using TargeTron technology was examined, as well as the genome-wide distribution of Tus using ChIP-Seq and ChIP-qPCR to identify the functional *Ter* sites capable of halting replication forks in *E. coli*. We also characterised the fork trap architecture in closely, moderately as well as distantly related bacteria harbouring the *tus* gene. Taken together, our data enabled the delineation of a prototypical Tus-dependent DNA replication fork trap.

3.2 Materials and methods

Table 6: Key resources used in this study

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Antibodies		
Chicken anti-GFP IgY	Abcam	ab92456
HRP-conjugated goat anti-IgY (Jackson 103-035-155)	Jackson ImmunoResearch Laboratories	103-035-155
Bacterial Strains		
E. coli KRX	Promega	Cat#: L3002
BL21(*DE3*)RIPL	Stratagene	Cat#: 230280
Dickeya paradisiaca (strain Ech703)	RefSeq	NC_012880
Edwardsiella tarda (strain EIB202)	RefSeq	NC_013508
Proteus mirabilis (strain HN2p)	RefSeq	NZ_CP046048
Xenorhabdus nematophila (strain ATCC 19061)	RefSeq	NC_014228
Salmonella typhimurium (strain LT2)	RefSeq	NC_003197
Escherichia coli (strain K12 substr. MG1655)	RefSeq	U00096

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Cedecea neteri (strain ND14a)	RefSeq	NZ_CP009459
Recombinant DNA		
Plasmid: pPMS1259	Schaeffer Lab	(Dahdah et al., 2009)
Plasmid: pMM220	Schaeffer Lab	(Moreau et al., 2012)
Chemicals, Peptides, and Recombinant Proteins		
His_6-Tus-GFP	Schaeffer Lab	N/A
His_6-DnaA-GFP	Schaeffer Lab	N/A
SIGMAFAST™ 3,3′-Diaminobenzidine tablets	Sigma	d4418
SensiMix SYBR & fluorescein mastermix	Bioline	QT615-05
Critical Commercial Assays		
NEBNext Ultra DNA library preparation kit	New England BioLabs	E7370S
QuantiFluor® dsDNA System	Promega	E2670
Rapid Sequencing protocol (FLO-MIN106 R9 MinION)	Oxford Nanopore	SQK-RAD004

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Oligonucleotides		
See Moreau Appendix 2 for full list of sequences and genomic loci for amplification of *oriC* and *Ter* regions by qPCR		

3.2.1 Plasmids and protein expression

Vectors encoding His_6-Tus-GFP (pPMS1259) and His_6-DnaA-GFP (pMM220) have been described previously (Dahdah et al., 2009, Moreau et al., 2010, Moreau et al., 2012, Morin et al., 2012). *E. coli* KRX (Promega) which is a K12 derivative was used to express His_6-DnaA-GFP and His_6-Tus-GFP. His_6-Tus-GFP and His_6-DnaA-GFP proteins were expressed and affinity purified as previously described (Moreau et al., 2012). His_6-Tus-GFP was stored in a buffer containing 50 mM sodium phosphate (pH 7.8) and 10% glycerol (w/v). His_6-DnaA-GFP was in DnaA buffer (50 mM HEPES-KOH (pH 7.6), 1 mM EDTA, 1 mM β-mercaptoethanol, 20 % sucrose (w/v)).

3.2.2 Protein expression and crosslinking

The ChIP protocol was derived from previous work by Regev *et al.* and Ishikawa *et al.* (Regev et al., 2012, Ishikawa et al., 2007). Competent *E. coli* KRX bacteria were transformed with pPMS1259 (His_6-Tus-GFP) or pMM220 (His_6-DnaA-GFP), plated onto LB plates supplemented with ampicillin (100 μg/ml) and grown overnight at 37°C. For His_6-Tus-GFP expressing cells, colonies were resuspended and diluted to an OD_{600} of 0.1 in 12 mL of LB broth supplemented with ampicillin (100 μg/ml). For His_6-DnaA-GFP expressing cells, single colonies were first streaked on a master plate (LB agar plate supplemented with 100 μg/ml ampicillin and 0.4 % glucose to avoid toxicity) and incubated for 24 hours at 37°C. These

colonies were then resuspended in 12 mL of LB broth supplemented with ampicillin (100 µg/ml) at an OD_{600} of 0.1. All cultures were grown for 45 minutes at 37°C before inducing moderate levels of GFP-tagged proteins with 0.02 % Rhamnose (w/v final culture concentration). Bacteria were incubated for 2 hours at 37°C, followed by 2 hours at 16°C. Culture aliquots (9 ml) were transferred on ice for 30 minute to which 250 µl of a formaldehyde solution (36% w/v) were added (final concentration of 1%). The bacterial suspensions were reacted for 20 minutes at RT. Glycine powder was added the bacterial suspensions (0.5 M final concentration) for 5 minutes at RT followed by 5 minutes on ice. The bacterial suspensions were then centrifuged 5 minutes at 800 *g* and 4°C and washed twice with 4 mL and 10 mL of cold TCS buffer (50 mM Tris (pH 7.5), 150 mM NaCl and 2 mM KCl). The supernatants were discarded and the bacteria pellets were stored at -80°C until required. KRX bacteria (without plasmid) were subjected to the same protocol and used as control.

3.2.3 Detection and quantitation of GFP-tagged protein expression

Culture aliquots were taken prior to crosslinking, centrifuged at 1,000 g for 1 minute and the pellets were resuspended in 2X Laemmli buffer at a concentration of 7.8×10^9 bacteria.ml^{-1}. The suspensions were heated for 10 minutes at 90°C and 5 µl (corresponding to total proteins of 3.95×10^7 cells) were separated in a 10 % SDS-polyacrylamide gel alongside known amounts of purified His_6-Tus-GFP (0.5 µg) or His_6-DnaA-GFP (0.5 and 1 µg) proteins as standards for Western blot analysis. Chicken anti-GFP IgY (Abcam ab92456) and HRP-conjugated goat anti-IgY (Jackson 103-035-155) were revealed with SIGMAFAST™ 3,3′-Diaminobenzidine tablets. Protein bands were quantified using imageJ (http://rsbweb.nih.gov/ij/) and intracellular concentrations were estimated based on the intensity of bands of known protein concentration and using cell parameters determined by Volkmer and Heinemann for cell volume (4.4 fL) and cell concentration at a given OD_{600} in LB (7.8×10^8 cells.ml^{-1}.OD^{-1}) (Volkmer and Heinemann, 2011).

3.2.4 Chromosome immunoprecipitation

Bacteria pellets were resuspended in lysis buffer (10 mM Tris (pH 8), 20 % sucrose, 50 mM NaCl, 10 mM EDTA, 1 mg/ml lsozyme and 10 µg/ml RNase) in 1/10 of initial culture volume (adjusted between replicates to reach same suspension concentration). Following a 30-minute incubation period at 37°C, the lysates were diluted 5 times in IP buffer (50 mM HEPES-KOH (pH 7.5), 150 mM NaCl, 1 mM EDTA) and passed three times in a French press at 12,000 psi to ensure maximum and reproducible cell lysis and DNA shearing. The Tus-GFP lysates were heated for 10 minutes at 50°C to denature free Tus-GFP (Moreau et al., 2010, Moreau and Schaeffer, 2012a, Moreau et al., 2012). Control KRX lysates were processed identically. All lysates were centrifuged at 30,000 *g* for 20 minutes at 4°C. Supernatants were used for immunoprecipitation and as input samples. For immunoprecipitation, 96-well MaxiSorp round bottom U96 Nunc plates were coated overnight at 4°C with 50 µl of 50 mM sodium phosphate (pH 7.5) and 10 % glycerol buffer containing 0.5 µg of goat anti-GFP IgG (Abcam; Ab6673). Wells were washed once with 200 µl of TCS buffer prior to immunoprecipitation. After the wash step, 50 µl of lysate supernatant were added per well for 90 minutes at RT. Wells were washed three times with 200 µl of TCS buffer. Control immunoprecipitation experiments were performed in parallel without antibody pre-coating as background controls. Immunocaptured DNA was released by adding 50 µl of elution and de-crosslinking buffer (2 mM Tris (pH 7.5), 50 mM NaCl, 0.005 % tween and 300 µg/ml proteinase K) to each well for 1 hour at 37°C (output). Input samples were diluted 10,000 times in elution buffer (2 mM Tris (pH 7.5), 50 mM NaCl, 0.005 % Tween) and 50 µl were transferred to a tube containing proteinase K yielding a final concentration of 300 µg/ml to de-crosslink the input DNA for 1 hour at 37°C. Samples (inputs and outputs) were incubated 15 minutes at 95°C to denature proteinase K and residual crosslinked proteins. After 5 minutes incubation on ice, samples were centrifuged at 18,000 *g* for 5 minutes at 4°C and the supernatants were used for qPCR and Illumina sequencing.

3.2.5 qPCR protocol

All qPCR reactions were performed as previously described (Moreau and Schaeffer, 2012b). Oligonucleotides for amplification of *oriC* and *Ter* containing regions are listed in Additional Resources. Briefly, qPCR reactions contained 2 µl of input or output DNA sample, 8 µl of primer pairs (0.5 µM each) and 10 µl of SensiMix SYBR & fluorescein mastermix (Bioline). The protocol included a 10 minute step at 95°C followed by 40 cycles at 95°C for 10 s and 60°C for 15 s. Melt-curves were run for quality control. *Ct* values were obtained at a set threshold applied to all experiments. Standard curves were performed in triplicate with purified and serially diluted *Ter* and *oriC* amplicons in matching output buffer conditions. For each primer pair, the average slope value of three standard curves (n=3) was used to determine the primer specific amplification efficiency according to the following equation (Meijerink et al., 2001).

$$E_{amp} = 10^{\left(-\frac{1}{\text{slope}}\right)}$$

A melt-curve was performed to verify that the correct regions were amplified.

3.2.6 qPCR analysis

For all qPCR experiments, *Ct* values were determined at the same threshold value. $Ct_{(input)}$ values were corrected for the dilution factor to give ${}_cCt_{(input)}$ according to the following equation:

$${}_cCt_{(input)} = Ct_{input} - log_{E_{amp}}(dilution\ factor)$$

The immunoprecipitation efficiency of each specific target DNA region relative to a non-specific DNA region ($IP\ efficiency_{(oriC)}$) was calculated according to the following equation:

$$IP\ efficiency_{(oriC)} = \frac{E_{amp}^{\left({}_cCt_{(input)sp} - C\ _{(output)sp}\right)}}{E_{amp}^{\left({}_cCt_{(input)ns} - Ct_{(output)ns}\right)}}$$

where ${}_cCt_{(input)}$ and $Ct_{(output)}$ are values obtained for each DNA target before (input) and after ChIP (output). Specific DNA target (*i.e. Ter* sites) and non-specific control DNA region are indicated with "sp" and "ns" subscripts respectively.

3.2.7 Library preparation and Illumina sequencing

Input and output DNA samples were purified using Wizard PCR clean up and eluted in 110 µL water. Each library was prepared using the NEBNext Ultra DNA library preparation kit for Illumina. ChIP output samples (~0.25 ng) were used for library preparation. The libraries were prepared according to manufacturer instructions. Briefly, 55.5 µL of DNA suspensions were end repaired. Due to the low DNA concentrations in the suspensions, the NEB adaptors were diluted 10 fold in water to 1.5 µM for ligation as recommended. The adaptors were cleaved using uracil excision. Size selection was not recommended for sample < 50 ng. DNA was then cleaned up using Sera-Mag beads (ratio of 1.4) and eluted in 28 µL of 0.1xTE. Index primers were added by PCR using 18 cycles. (13-15 cycles were recommended for 5 ng of input material, therefore 3 cycles were added to account for the 20 fold difference in input DNA). DNA quantification was performed using the Qantifluor dsDNA system (Promega). The samples were pooled in a single library, denatured and loaded for sequencing with an Illumina MiSeq desktop sequencer (50 bp single-end sequencing). Illumina read quality was assessed using FastQC (v.0.11.8) followed by removal of Illumina adapters and leading and tailing nucleotides with a Phred score ≤ 10 over a 6 bp window using Trimmomatic (v.0.36).

3.2.8 DNA preparation for Nanopore long read sequencing

A flask containing 10 mL of LB media was inoculated with 100 µL of KRX *E. coli* overnight culture. The culture was incubated at 37°C and 150 RPM until log phase was reached (OD_{600} = 0.7) at which point chloramphenicol was added at a final concentration of 180 µg/mL to inhibit protein synthesis. The bacteria were centrifuged and resuspended in 3.5 mL lysis buffer (114

mM Tris-HCl (pH 8), 115 mM EDTA, 570 mM NaCl and 1% triton X-100). After addition of lysozyme (250 µL, 50 mg/mL), sodium dodecyl sulphate (500 µL, 10% w/v) and RNase A (4 µL, 100 mg/mL), the bacterial suspension was inverted gently 12 times and heated at 50°C for 30 min. Proteinase K (500 µL, 10 mg/mL) was added with repeated gentle mixing and heating steps. The suspension was then combined with 8 mL precipitation buffer (75% isopropanol, 2.5 mM ammonium acetate), inverted gently 15 times followed by centrifugation at 4,000 *g* at 4°C for 5 min. The supernatant was discarded. The soft DNA precipitate was transferred using a wide bore tip into a 1.5 mL tube, resuspended with 70% ethanol and stored at 4°C for 5 min, then centrifuged and the supernatant removed. The DNA pellet was air dried for 5 minutes then resuspended in 200 µL nuclease free water and heated at 50°C for 5 minutes with the tube cap open. The DNA was sheared using a syringe with a 20 gauge needle (3 times). DNA concentration and quality were assessed using Invitrogen Qubit 4 and agarose gel electrophoresis.

3.2.9 Nanopore long read sequencing and genome assembly

A Nanopore sequencing library was prepared using the Rapid Sequencing protocol SQK-RAD004 (Oxford Nanopore). As recommended by the protocol, 7.5 µl of DNA suspension (400 ng) was added to the flow cell. Sequencing was performed on a FLO-MIN106 R9 MinION flow cell. Base-calling was processed using the pipeline implemented in MinKNOW software version 18.01.6 (Oxford Nanopore). In total, 1.17 GB (253x coverage) of sequence data was generated for *E. coli* strain KRX over ~18 hours of sequencing achieved in two separate runs. Prior to assembly, all fastq files were combined and quality filtered by nanofilt version 2.5.0 (quality score ≥ 9). The remaining 225 985 reads had an average length of 3738 bp with the longest read of 179 912 bp and a total 146x coverage of the KRX *E. coli* genome. Oxford Nanopore adapters were trimmed using Porechop (version 0.2.3_seqan2.1.1) and assembled using Flye (version 2.6) with default settings for Oxford Nanopore data. Polishing was

performed with Racon iteratively four times in combination with Pilon (version 1.23-1) using the entire Illumina ChIP-Seq data (*i.e.* KRX Input, WGS, negative control and ChIP DNA). The genome was annotated using Prokka (version 1.14.0) and evaluated using Quast (version 5.0.2) against the *E. coli* K12 genome (GenBank assembly accession: GCA_000005845.2).

3.2.10 ChIP-Seq analysis

Following adapter trimming and quality control, each biological and technical replicate fastqc file of sequenced samples (*i.e.* Input, ChIP DNA and negative control) were individually aligned to the polished KRX genome using Bowtie2 (version 2.3.4.1). Samtools (version 1.9) was then used to organise each alignment file for visualisation on Interactive Genomics Viewer (IGV) to visually assess the data in terms of replicability between each replicate and for any outliers. The replicate data for Input, ChIP and negative control were pooled into three separate fastqc files and aligned to the KRX *E. coli* genome using the default settings of Bowtie2. A circular annotation of each pooled ChIP-Seq reads mapped to the KRX *E. coli* genome spanning a 23 bp window (size of an extended *Ter* site) was created as well as the GC-skew of the chromosome (5000 bp window) using Circleator (version 1.0.2). The individual positions and orientations of *Ter* sites were identified and verified using their specific 23 bp sequences and compared to the respective sites in *E. coli* (K12). The single base read counts were averaged over the 23 base *Ter* sequences for Input, ChIP and negative control DNA using genomeCoverageBed (Version: v2.26.0).

3.2.11 Genome engineering of ectopic *Ter* sites

Targetrons (mobile group II introns) were designed to insert *Ter* sites in the safe insertion region SIR.5.6 (Isaacs et al., 2011), located in the right non-structured chromosome domain (Figure 9A) of *E. coli* BL21(*DE3*) (accession number AM946981). The SIR5.6 is located about 930 kbp downstream of *oriC* (right chromosome arm, Table 2). *Ter*-targetrons (mobile group II

introns carrying *Ter* sequences) insertion was performed as described previously (Enyeart et al., 2013), replacing *lox* sites with *TerB*, *TerH* or *TerI* sequences in permissive (P) or non-permissive (NP) orientations. Insertion of *TerB* in the non-permissive orientation was also attempted using the Lambda Red recombination system but no viable colonies could be obtained. The successful insertions of the other *Ter* sites were confirmed by colony PCR and verified by sequencing (Table 2). *Ter* sites were inserted into SIR5.6 with an efficiency of 53/65 (81.5% - excluding integrations attempted for the insertion of *TerB* in the non-permissive orientation).

BL21(*DE3*) cells carrying ectopic *Ter* sites were grown in LB broth at 37°C and OD_{600} was measured every 5 minutes for 12 hours. The results were plotted as $\log_2(OD_{600})$ versus time (minute). In order to select the linear region of the curve, each point was assigned a correlation coefficient R^2 corresponding to the value of R^2 for the line consisting of that point and the five points before and after. The variance was lower when the same time window was used for all three replicates so the resulting R^2 values were averaged for all three replicates at each time point. The longest stretch in which all these averaged R^2 values were equal to or greater than 0.99 was taken as the linear range. The slope of the least-squares linear fit of the $\log_2(OD_{600})$ curve of each replicate in that time range was then taken as the growth rate and the doubling time was calculated as 1/growth rate.

3.2.12 Fork trap characterisation in Enterobacteriales

Bacterial species containing a *tus* gene ortholog were identified using InterPro entry IPR008865 including only the entries for Enterobacteriales (2518 protein sequences in total). Several Tus protein sequences from a selection of *Pseudoalteromonas* species were also included as outgroup. Ten rounds of an alignment for all sequences were generated using MUSCLE software with default settings (Edgar, 2004). The tree was constructed using the matrix of

aligned sequences in RAxML v8.2.0 (Stamatakis, 2014) which performed an ML phylogenetic analysis with 100 independent repetitions using the PROTGAMMALGI molecular evolution model in combination with an independent rapid bootstrap algorithm (--AutoMRE) to establish support for each node. The consensus tree produced by RAxML was visualised and edited in iTol (Letunic and Bork, 2019). The sequences that branched off earlier than the outgroup were removed from the final tree as these were most likely incorrectly assigned taxa or sequences that are most likely not Tus protein sequences. For clarity, clades from *Escherichia*, *Salmonella* and *Klebsiella* were collapsed due to the large number of sequences.

A random selection of species from different families were chosen including *Dickeya paradisiaca* (strain Ech703), *Edwardsiella tarda* (strain EIB202), *Yersinia pestis* (Microtus str. 91001), *Xenorhabdus nematophila* (strain ATCC 19061), *Proteus mirabilis* (strain HN2p), *Cedecea neteri* (strain ND14a) and *Salmonella typhimurium* (strain LT2). Genome assemblies that were preliminary were excluded from fork trap analysis. Upon identification of a *tus* gene otholog, the adjacent *Ter* site was identified within its 50 bp 5' UTR by aligning the 23 bp *E. coli TerB* sequence. For each selected species, a BLAST search was carried out using the adjacent *Ter* sequence to locate further *Ter* sites within their genome. Sequences were verified by inspecting each BLAST to ensure it contained a cytosine at position 6 followed by the strictly conserved 12 bp core spanning from A(8) to A(19) (Figure 9B) as well as a purine base at position 5 to ensure successful rotation of the C(6) into its binding pocket to form a locked complex with minimal steric hindrance. A circular annotation of each genome to display the architecture of the fork trap was generated alongside the GC-skew of the chromosome (5000 bp window) using Circleator (version 1.0.2).

3.2.13 Quantification and statistical analysis

Statistics and number of biological and technical repeats are indicated in the relevant figure legends, tables and methods. Statistical analyses were performed using Graphpad Prism 7. Data are expressed as mean values ± SD, ± SE or ranges.

3.3 Results

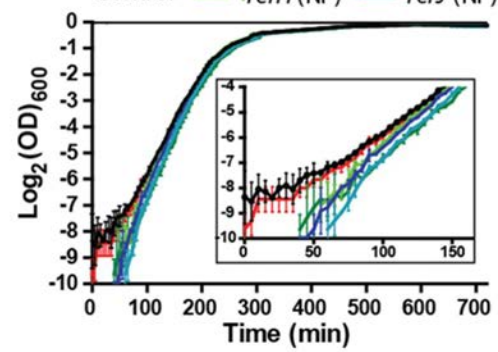
3.3.1 Effect of ectopic insertion of *TerB*, *TerH* and *TerJ* sites

The role of some of the distal *Ter* sites in replication fork arrest is questionable in light of their location, Tus binding affinity and dissociation kinetics (Moreau and Schaeffer, 2012a). *TerF* has recently been dismissed as a pseudo-*Ter* site with no possible role in replication fork arrest. While *TerH* cannot form a lock ($t_{1/2}$ = 59.1 s), the *TerI*-lock and *TerJ*-lock sequences are only moderately stabilising, $t_{1/2}$ of 196 s and 332 s respectively. Importantly, the Tus-*TerJ-lock* complex dissociates similarly to the non-locked Tus-*TerB* ($t_{1/2}$=315 s) (Moreau and Schaeffer, 2012a). These findings prompted us to examine the capacity of the most distal *Ter* sites (*i.e. TerH* and *J*) to halt DNA replication forks. For this, *TerH* (moderate non-TT-lock forming sequence), *TerJ* (moderate TT-lock forming sequence) and the strong TT-lock forming *TerB* were inserted in the right chromosome arm of *RecA*$^+$ *E. coli* strain BL21(*DE3*), 930 kbp downstream of *oriC* (right arm, SIR5.6) in both permissive (P) or non-permissive (NP) orientations using a TargeTron strategy (Enyeart et al., 2013). While ectopic *Ter* site insertions and fork trap inversions have been studied previously (Bidnenko et al., 2002, Sharma and Hill, 1995, Bidnenko et al., 2006, Esnault et al., 2007), the TargeTron technique guarantees that a 23 bp *Ter* within a short intron sequence is incorporated with minimal genomic variations. It was hypothesised that *Ter* insertions resulting in weak to moderate replication fork pausing would yield a measurable effect on bacterial growth rate while efficient fork arrest would potentially be lethal.

All sites except *TerB* could be inserted efficiently in either orientations. It is important to note that *TerB* in non-permissive orientation could not be inserted using either a TargeTron or the Lambda Red recombination system (Poteete, 2001). Growth rates were determined for viable *E. coli* cells with successful ectopic *Ter* sites insertions (Figure 10). No significant differences in growth rates were observed in the exponential growth phase between these strains.

Furthermore, all strains reached the same plateau as the control strain (Figure 10A) suggesting that *TerB* in permissive orientation as well as *TerH* and the *TerJ* in either orientation do not significantly impact replication forks which is in partial agreement with previous genomic region inversion data (Esnault et al., 2007). It was concluded that the insertion of an ectopic *TerB* site in non-permissive orientation in a strain carrying the wild type *tus* gene must be lethal to the bacteria as a result of efficient replication fork arrest 930 kbp downstream to *oriC.* The resulting DNA break induced by the non-permissive *TerB* barrier cannot be restored even in a *RecA*$^{+}$ strain *i.e.* with wild-type homologous recombination function. This is in agreement with previous direct detection of each *Ter* site's activity in wild-type cells, which found that *TerA-C* are the only substantially utilized terminators in *E. coli*, (Duggin and Bell 2009). As such it was proposed to reclassify *TerH* and *J* as pseudo-*Ter* sequences and in light of our and previous data (Moreau and Schaeffer, 2012a, Coskun-Ari and Hill, 1997) *TerI,* which forms a Tus-*Ter*-lock complex with a faster $t_{1/2}$ than *TerJ* (Moreau and Schaeffer, 2012a, Moreau and Schaeffer, 2013), can also reasonably be dismissed as a pseudo-*Ter* site.

A



B

Strain BL21(DE3)::	Growth rate $Log_2(A_{600})$/min (SD)	Doubling time T_D in min (SE)
TerB (P)	0.039 (0.0012)	25.53 (0.448)
TerH (NP)	0.041 (0.0002)	24.13 (0.087)*
TerH(P)	0.042 (0.0017)	23.69 (0.564)
TerJ (P)	0.044 (0.0011)	22.70 (0.327)
TerJ (NP)	0.042 (0.0008)	23.70 (0.259)
Control	0.039 (0.0012)	25.88 (0.469)

Figure 10: Effect of ectopic *Ter* sites on the growth rate of *E. coli* BL21(DE3). *TerB*, *TerH* and *TerJ* were inserted ~ 930 kpb downstream of *oriC* in the permissive (P) or non-permissive (NP) orientation. A culture of wild type BL21(DE3) was grown as a control. (A) Growth rates were measured in independent triplicates. Error bars represent SD. (B) Averaged growth rates

obtained from the slope of the linear regression performed between 100 and 210 minutes. Doubling times (TD) were calculated as 1/growth rates (n=3, except for *TerH* (NP), n=2). Standard deviations (SD) are shown. *One outlier was omitted. Reproduced with permission from Moreau, PhD thesis, James Cook University (2013). Thesis can be downloaded using the following link: https://researchonline.jcu.edu.au/31903/1/31903_Moreau_2013_thesis.pdf).

3.3.2 Chromosomal binding of Tus

The genome wide distribution of Tus was examined by using chromosome immunoprecipitation (ChIP)-Seq and ChIP-qPCR methods to identify the *Ter* sites that are occupied during active DNA replication. Major limitations of this study include both the low natural abundance of Tus and the unavailability of commercial antibodies. To circumvent these issues, the chromosomal distribution of GFP-tagged Tus (Tus-GFP) in exponentially growing *E. coli* (KRX) was examined (Appendix 1 Figure S1) (Moreau, PhD thesis, James Cook University (2013). It was hypothesized that expression of the chromosomal *tus* gene would be downregulated further in the presence of exogenous Tus-GFP through their binding to *TerB* within the promoter region of the chromosomal *tus* (Roecklein et al., 1991). This would have the effect of reducing endogenous Tus levels allowing Tus-GFP to efficiently compete for *Ter* sites. Due to the unique base sequences flanking each *Ter* site, ChIP samples could be sequenced using 50 bp Illumina reads, thereby ensuring the reads containing the full or partial 23 bp *Ter* sequences could be accurately mapped to the genome. Input and immunoprecipitated DNA samples were sequenced and the reads mapped back to our KRX genome assembly to generate a high-resolution genome-wide distribution map of Tus-GFP (Figure 11).

ChIP-Seq peaks were immediately apparent without the need for a peak identification workflow. Six large peaks were visible in the immunoprecipitated coverage plot, corresponding to the binding of Tus-GFP to individual *Ter* sites. A large peak at the *tus* locus was also visible

in the input DNA sample due to plasmid-encoded *tus* sequences that are identical to chromosomal *tus*. The coverage at *tus* in the input DNA indicated a plasmid copy number per chromosome of ~ 43. The read coverage in the 10 bp region between the chromosomal *tus* and *TerB* loci was inspected to ensure that the reads originating from plasmid-encoded *tus* did not bias the read count at *TerB* (Appendix 1 Figure S2). The base coverage at *TerB* is equivalent to the average chromosomal reads in the input DNA demonstrating that our method does not lead to a coverage bias in the immunoprecipitated DNA data. The average coverage values ranged from ~5 at *oriC* to ~1 in the terminus region indicating that at least three replication forks were progressing on each chromosome arm towards the terminus region. Our ChIP-seq data revealed that out of the 10 primary *Ter* sites, only the 6 high-affinity *TerA-E and G* sequences (Moreau and Schaeffer, 2012a) are significantly bound by excess Tus-GFP (Figure 11 and Figure S3).

Surprisingly, despite being the major termination site (Duggin and Bell, 2009) *TerC* was one of the least bound in this group with an average 269x read coverage compared to 430x coverage at *TerB*. The coverage at *Ter*G (410x) was similar to *TerB* suggesting that this site is almost certainly bound at normal bacterial Tus concentrations. Given the strong Tus binding and lock-forming ability of *TerG* (Moreau and Schaeffer, 2012a), our data suggest that the absence of paused fork intermediates in the fork arrest assay measured by Duggin and Bell (Duggin and Bell, 2009), is a result of the replication fork not reaching this *Ter* site. As anticipated, no binding was observed at the pseudo-*TerF* (Moreau and Schaeffer, 2012a, Moreau and Schaeffer, 2012b, Moreau et al., 2012, Duggin and Bell, 2009). Out of the three moderate *Ter* sequences, *TerJ* in its locked complex with Tus is the most stable with respect to $t_{1/2}$ (Moreau and Schaeffer, 2012a) and fork arrest activity (Duggin and Bell, 2009) yet no peak was observed strongly supporting our ectopic insertion data and that the latter is also a pseudo-*Ter* site in natural conditions. *TerH* and *TerI* sites have similar coverages (57x and 48x respectively) corresponding to only 11-13% of the coverage at *TerB*, despite the bacterial Tus-GFP

concentration being 1,000-fold higher than the normal endogenous levels of Tus, suggesting these sites would be mostly unbound at normal cellular Tus concentration. Taken together with previous affinity data, our ChIP-Seq and ectopic insertion findings support the notion that *TerH, I* and *J* do not have a role in replication fork arrest. The ChIP-Seq dataset was confirmed by ChIP-qPCR (Appendix 1 FigureS4) and allowed delineation of a refined minimal replication fork trap within *E. coli* comprising two clusters of three *Ter* sites: (a) *TerB*, *C* and *G* that can arrest a clockwise moving replication fork and (b) *TerA*, *D* and *E* that can arrest an anticlockwise moving replication fork (Figure 11).

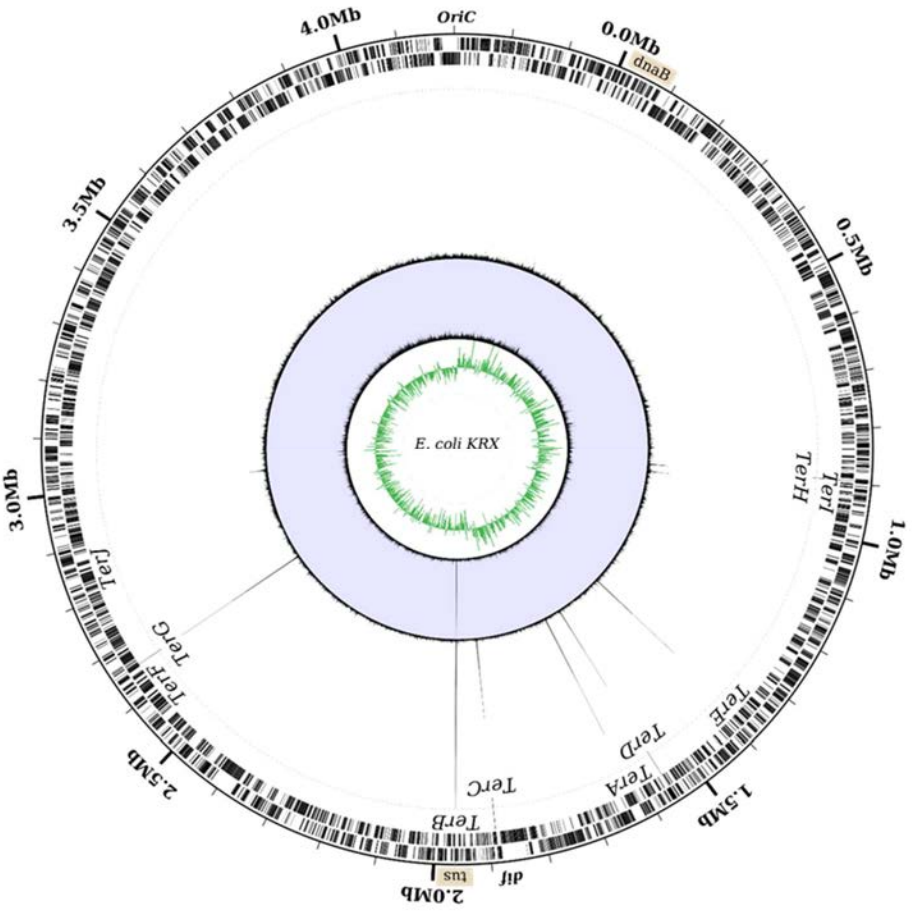
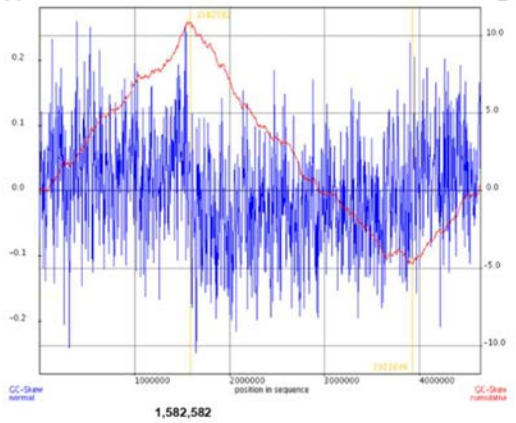


Figure 11: Circular representation of *E. coli* KRX chromosome with mapped ChIP-Seq coverage. From the outside to the centre of the circle: labelled forward and reverse genes; genomic location of sites and genes involved in DNA replication termination; combined ChIP-Seq read coverage (max = 430 reads at *TerB*), Input DNA read coverage (max = 230 reads at the *tus* gene), GC-skew over a 5,000 bp moving window. The GC-skew switches polarity at the replication origin and terminus.

3.3.3 GC-skew relative to termination site usage in *E. coli*

Although the GC-skew is a well-recognised tool to identify the origin of replication in many circular prokaryotic chromosomes (Worning et al., 2006, Touchon and Rocha, 2008, Arakawa and Tomita, 2007), the feasibility of utilising the GC-skew to predict the terminus has been debated amongst researchers due to the terminus shift point being closer to the chromosome dimer resolution site (*dif*) than to the *Ter* sites in some studied species (Hendrickson and Lawrence, 2007). However, the GC-skew has recently been shown to coincide with replication fork arrest by Tus at *Ter* sites and is not influenced by *dif* (Kono et al., 2012, Kono et al., 2018). We hypothesize that the GC-skew should correlate with the frequency of fork arrest activity at specified *Ter* sites. In other words, we propose that the GC-skew is representative of the average of the ensemble of replication forks collision loci at functional *Ter* sites. In this scenario, the inflection point should occur at the positional average between the *Ter* sites where termination occurs. Duggin and Bell showed that only *TerA, B* and *C* have significant replication arrest activity (0.19 %, 0.14 % and 0.85 % respectively) in natural Tus conditions (Duggin and Bell, 2009). We tested this scenario and found that the expected average position of replication termination (based on the positional and fractional distribution of replication fork arrest activity) almost coincided with the GC-skew inflection point, *i.e.* only 7.5 kb from the calculated inflection point derived from a sliding 1,000 bp cumulative GC-skew (Figure 12). We tested additional scenarios, but none produced a better correlation. Taken together with previously published data (Kono et al., 2012, Kono et al., 2018), the GC-skew of *E. coli* supports the involvement of *TerA, B* and *C* in replication fork arrest and provides an invaluable tool to further our understanding of replication termination in other species.

A



B

	TerA	*TerB*	*TerC*	*dif*	GC skew transition midpoint
Location in *E. coli* genome (bp)	1,341,745	1,684,226	1,609,157	1,590,777	1,582,582
In vivo replication arrest (%)	0.19	0.14	0.85	-	-
Fraction of arrest Hypothesis 1	0.161	0.119	0.72	-	-
Fraction of arrest Hypothesis 2	0.183	-	0.817	-	-
Fraction of arrest Hypothesis 3	-	-	-	1	-
					Distance from GC skew (bp)
Predicted GC-inflection Hypothesis 1 (bp)			1,575,037		-7545
Predicted GC-inflection Hypothesis 2 (bp)			1,560,221		-22361
Predicted GC-inflection Hypothesis 3			1,590,777		8195

Figure 12: The GC-skew is representative of the average of the ensemble of replication forks collision loci at functional *Ter* sites. (A) GC skew transition midpoint calculated with a 1000 bp sliding window for *E. coli* MG1655 used in the in vivo replication arrest study by Duggin and Bell, (2009). (B) Hypothetical GC skew transition midpoint loci using various fork arrest scenarios based on the ensemble and fractional distribution of replication fork collision loci at functional *Ter* sites. Only Ter sites with significant replication fork arrest activity (*TerA*, *TerB* and *TerC*) are included. *Dif* site is also shown for comparison. Locations of *TerA-C* and *dif* in *E. coli* MG1655 are indicated in the table.

Hypothesis 1: Replication fork arrest occurs with the fractional distribution of Y forks reported by Duggin and Bell, 2009.

Hypothesis 2: Replication fork arrest occurs at *TerA* and *TerC* with equal fractional distribution *i.e.* fork arrest only occurs at *TerA* and *TerC* and are never breached.

Hypothesis 3: Replication fork arrest occurs at *dif* site

Conclusion: the terminal GC-skew switch derived from hypothesis 1 (*i.e.* 1,575,037 bp) involving *TerA-C* deviates the least from the switch point derived from a sliding 1,000 bp

cumulative GC-skew (*i.e.* 1,582,582 bp) by only 7.5 kbp. It is important to note that the *dif* site is located 8 kbp from the terminal GC-skew switch point on the other chromosomal arm.

3.3.4 A narrow fork trap dyad in *Edwardsiella tarda*

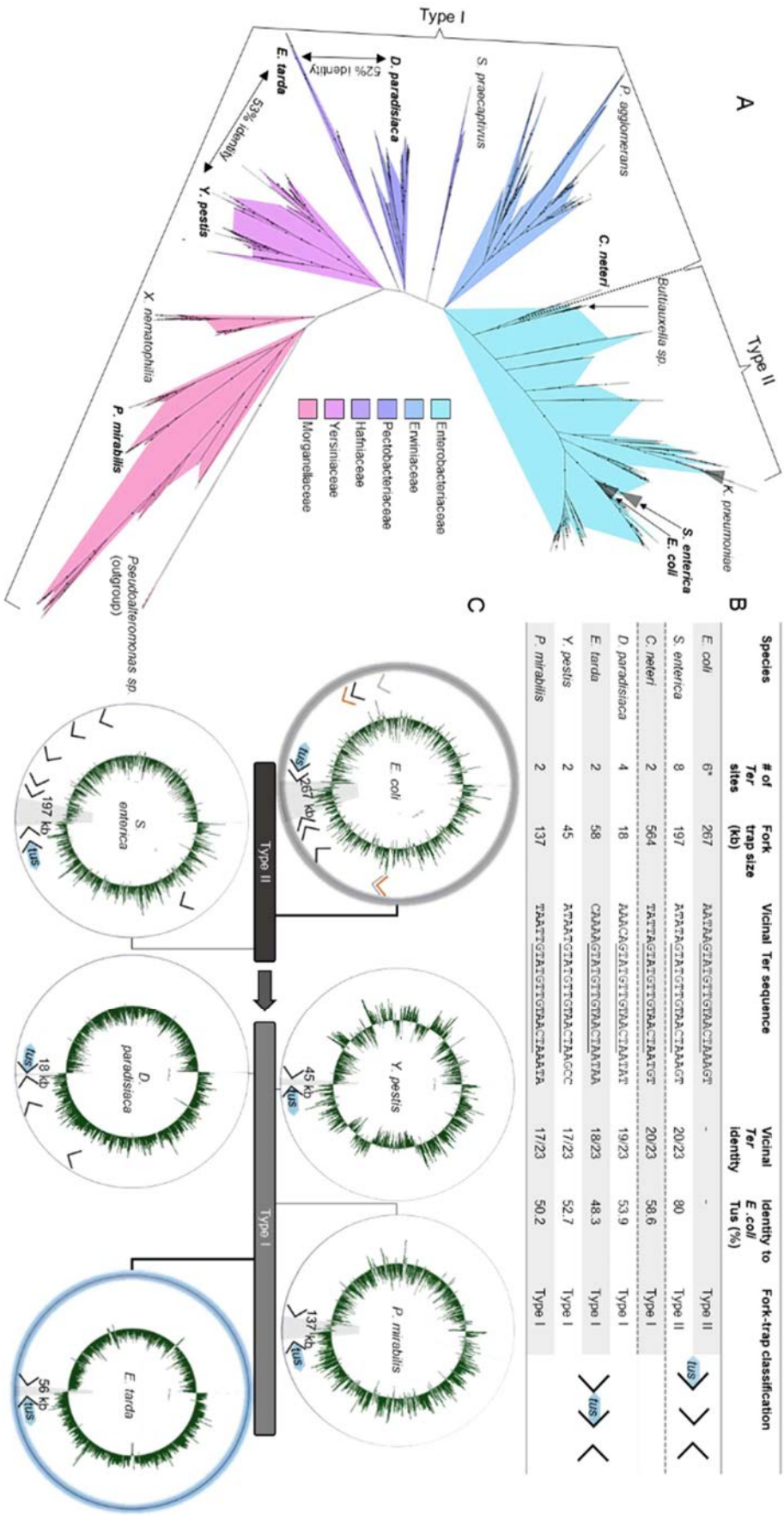
While the function of *TerA*, *B* and *C* and their replication fork arrest activity in *E. coli* is clear, the need of *TerE, D* and *G* is not, despite their high affinity for Tus and ability to form a TT-lock (Moreau and Schaeffer, 2012a). While trying to gain further insight into these seemingly redundant *Ter* sites, we examined the replication fork trap architecture in closely, moderately as well as distantly related bacteria harbouring a *tus* gene (Figure 13A). A recent phylogenetic analysis of Tus homologs in bacteria identified resident *tus* genes within the chromosomes of most Enterobacteriales (Galli et al., 2019). Using a streamlined approach, we characterised the replication fork traps in several of these species (Appendix 1 Figure S5 & Table S1). Our approach used a refined definition of what constitutes a *Ter* site: (a) a 23 bp *Ter* sequence is always located within 50 bp upstream of the *tus* gene; (b) it must contain a GC base-pair at position 6 followed by the strictly conserved 12 bp core spanning from AT(8) to AT(19) (c) it must contain a purine base at position 5 to ensure successful rotation of the C(6) into its binding pocket to form a locked complex with minimal steric hindrance (Figure 9B). Following identification of the vicinal *Ter* sequences upstream of *tus* genes in our selected bacterial genomes, BLAST searches were performed to identify other *Ter* sites within the genomes as well as their replication fork blocking orientations.

In *Salmonella enterica*, a close relative of *E. coli*, the left chromosomal arm contains five *Ter* sites, while the right chromosomal arm contains only three *Ter* sites in opposite orientation (Appendix 1 Figure S5A). The Tus protein and vicinal *Ter* sequence identities (*i.e.* corresponding to the *E. coli* Tus protein and *TerB*) were found to be 80% and 87% respectively. The distance between the innermost *Ter* sites (197 kb) is significantly reduced in the *Salmonella*

replication fork trap. In more distantly-related bacteria, such as *Dickeya paradisiaca* and *Proteus mirabilis*, despite the high sequence identity of their respective *Ter* sequences vicinal to *tus* (83% for both), a reduction in the number of *Ter* sites as well as a narrowing of the fork trap (*i.e.* the distance between the innermost *Ter* sites), was commonly seen (Figure 13B). Most striking was that the innermost *Ter* site upstream of the *tus* gene (*i.e.* corresponding to the *E. coli TerC*) was no longer present in these species (Figure 13B). To our surprise, all replication fork traps that we characterised outside the Enterobacteriaceae family lacked the innermost *Ter* site corresponding to the *TerC* in *E. coli*. In these genomes (*D. paradisiaca*, *P. mirabilis, X. nematophilia*, *E. tarda* and *Yersinia pestis*), the innermost *Ter* site is the one vicinal to *tus* gene in that cluster. We thus propose a new replication fork trap classification based on their architecture where a type I replication fork trap has one of its innermost *Ter* sites vicinal to *tus* (Figure 13B). Accordingly, the *E. coli* and *S. enterica* genomes contain type II replication fork traps. All type I replication fork traps that we identified outside of the Enterobacteriacae family are significantly narrower than the type II traps (Figure 13C). This is most evident in *D. paradisiaca* for which the innermost *Ter* inter-distance is just 18 kbp.

In *E. tarda* this pattern of simplification culminated into a narrow and perfectly symmetrical replication fork trap diametrically opposite the *oriC*, consisting of two unique *Ter* sequences. *E. tarda Ter1* and *Ter2* are only 56 kb apart and equidistantly located on either side of the hypothetical terminus site (Figure 14A). The next *Ter*-like sequence within this genome has only 65% identity to *Ter1* with a high level of degeneracy in the core sequence (see *pTer3*, Figure 14B) and would oppose an o*riC*-initiated replication fork. Most importantly, the midpoint between *Ter1* and *Ter2* (at 1,846,308 bp) coincides almost perfectly with the sharp GC-skew flip (at 1,846,910 bp) immediately suggesting that they are being used equally as replication fork barriers. The *E. tarda Ter1* sequence is 78% identical to *E. coli TerB* (Figure 13B). In contrast, the *E. coli TerC* sequence shares only 74% identity to *TerB* immediately

suggesting that the mechanism of polar DNA replication fork arrest in *E. tarda* also involves formation of a locked Tus-*Ter* complex.



Species	# of Ter sites	Fork trap size (kb)	Vicinal Ter sequence	Vicinal Ter identity	Identity to E. coli Tus (%)	Fork-trap classification
E. coli	6*	267	AATAAGTATGTTGTAACTAAAGT	-	-	Type II
S. enterica	8	197	ATATAGTATGTTGTAACTAAAGT	20/23	80	Type II
C. neteri	2	564	TATTAGTATGTTGTAACTAATGT	20/23	58.6	Type I
D. paradisiaca	4	18	AAACAGTATGTTGTAACTAATAT	19/23	53.9	Type I
E. tarda	2	58	CAAAGTATGTTGTAACTAATAA	18/23	48.3	Type I
Y. pestis	2	45	ATAATGTATGTTGTAACTAACCC	17/23	52.7	Type I
P. mirabilis	2	137	TAATTGTATGTTGTAACTAAATA	17/23	50.2	Type I

Figure 13: Phylogenetic analysis of Tus orthologs and fork trap architecture in Enterobacteriales. (A) Unrooted phylogenetic relationship of ~2500 Tus protein sequences using InterPro entries (IPR008865) highlighting the transition of a simple type I to complex type II fork trap architecture which occurs at *Cedecea*. (B) Chromosomal fork trap characteristics and classification for selected species (See Appendix 1 for their graphical representations and the complete table of species). Fork trap size (kb) corresponds to the distance between the two innermost *Ter* sites of opposite polarity. Underlined bases represent a continuous identical sequence shared between all *Ter* sequences vicinal to *tus* starting at the GC(6) base-pair. (C) The different types of replication fork trap architecture in Enterobacteriales.

While all vicinal *Ter* sequences are highly homologous (74-83% identity to *E. coli TerB)* and include the crucial C(6), and there is little doubt that the Tus orthologs from *S. enterica*, *Y. pestis* and *P. mirabilis* are able to arrest a replication fork at *TerB* (Henderson et al., 2001); the competency of *D. paradisiaca*, *X. nematophilia* and *E. tarda* Tus orthologs (only 46-54% identity to *E. coli* Tus) to form a locked complex was not clear (Figure 13B). To confirm that these Tus orthologs are fully competent in forming a locked complex (Mulcair et al., 2006a), we verified that the residues that make a critical interaction with the C(6) base are strictly conserved. It is apparent that *E. tarda* Tus despite having one of the lowest identity score with the *E. coli* Tus sequence (48%) should be fully competent in forming a locked complex with its *Ter* sequences (Figure 14C). Furthermore, a model structure of *E. tarda* Tus showed no major differences in the cytosine binding pocket when compared to *E. coli* (Figure 14D). Hence, we predict the formation of a highly efficient Tus-*Ter*-lock complex in *E. tarda,* as termination can only occur at two sites. This is also supported by the sharp GC-skew flip midway between *Ter1* and *Ter2* suggesting that replication forks rarely break-through the replication fork trap dyad in *E. tarda.*

In our efforts to investigate the type I/II fork trap evolutionary divergence, we identified a unique group of *Cedecea* species within the Enterobacteriaceae family that uses a type I replication fork trap system with only two oppositely-oriented *Ter* sites for *C. neteri* (Figure 13B & Appendix 1 Figure S5B). Against our expectations, the fork traps within this rare genus of bacteria were the widest (507-564 kb) of all investigated bacteria and the GC-skew switch in *C. neteri* fitted unambiguously with the *dif* location.

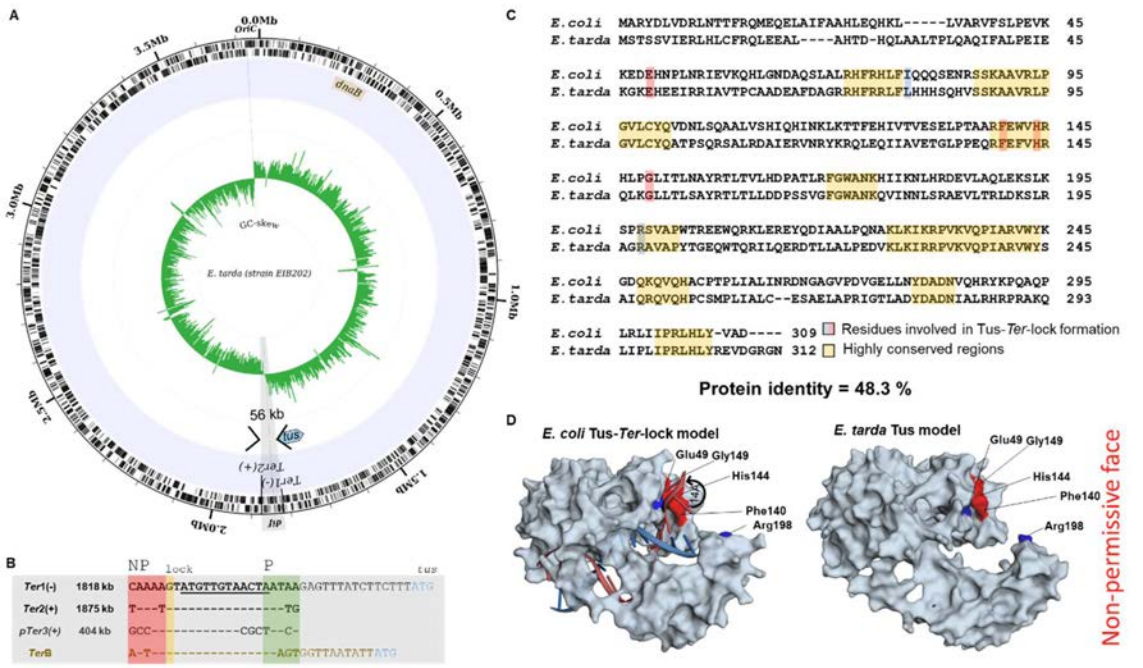


Figure 14: Prototypical type I replication fork trap. (A) Circular representation of *E. tarda (strain EIB202)* chromosome. Illustrated from the outside to the centre of the circle: forward and reverse genes, labelled genomic location of identified *Ter* sites involved in DNA replication termination, simplified annotation of the termination fork trap utilised, GC-skew over a 5000 bp moving window. The sharp GC-skew switches polarity at the replication origin and between the two identified *Ter* sites near *dif*. (B) Sequence alignment and genomic locations of the *E. tarda Ter* sites including a pseudo-*Ter* (*pTer*) close to *oriC* and *TerB* from *E. coli*. *Ter1* is located slightly upstream of the start site (ATG) of the *tus* gene similar to *TerB* in *E. coli.* The strictly conserved 12 bp core sequence is underlined and the G(6) base complementary to C(6) is highlighted in yellow. *NP: non-permissive face (red), P: permissive face (green).* (C) Tus protein sequence alignment with highlighted conserved residues. (D) Comparison of the *E. coli* Tus-*Ter*-lock complex 3D structure (PDB 2I06) and the modelled structure of *E. tarda* Tus protein using SWISS-MODEL. The essential amino acid residues in the cytosine binding pocket are indicated.

3.4 Discussion

3.4.1 A simplified type II replication fork trap in *E. coli*

Since the discovery of the first *Ter* sites and Tus coding sequence in *E. coli*, additional *Ter* sites were identified simultaneously expanding the size of the replication fork trap and increasing the perceived complexity of DNA replication termination. The systematic analyses of individual *Ter* sites both *in vitro* and *in vivo* with respect to their affinity and kinetics for Tus, ability of forming a TT-lock structure as well as their position and orientation within the bacterial genome have provided a wealth of information as to how this seemingly simple protein-DNA interaction impedes replication forks. In fact, Tus-*Ter* has become one of the best-understood protein-DNA complexes, leading to the development of a variety of biotechnological applications (Willis et al., 2014, Larsen et al., 2014b). Yet, we are only just starting to understand the modus operandi of Tus *in vivo*. Duggin and Bell showed evidence of a simple replication fork trap involving just *TerA, B and C* under normal bacterial concentrations of Tus (Duggin and Bell, 2009). We found that the observed distribution of fork arrest events at these sites fits with the terminus GC-skew flip in the *E. coli* genome.

Taken together, our findings allow us to propose a simplified replication fork trap *in E. coli* consisting of just six *Ter* sites (three in each cluster) and support the notions that: (i) Tus binds preferentially to the high affinity *Ter* sites *in vivo*; (ii) Tus-bound *TerC* and *TerA* are sufficient to block replication forks progressing towards their non-permissive face; (iii) *TerB* is most likely only used when a replication fork passes through an unbound *TerC*; (iv) replication forks are unlikely to reach the outer *Ter* sites; (v) *TerH, I and J* are unlikely to be bound at natural Tus concentrations, are unable to block replication forks and thus cannot be considered as functional *Ter* sites.

While the roles of *TerA*, *B*, and *C* are now clear, the need for *TerD*, and particularly the distant *TerE* and *TerG* in the *E. coli* genome still remains somewhat enigmatic. As such, if we consider that a single genomic insertion of *TerB* in the non-permissive orientation at SIR5.6 was not viable despite its distant position from *tus* and the low natural abundance of Tus (Natarajan et al., 1993), this would support the notion that *TerD, E* and *G* although probably bound by Tus are rarely used to arrest replication forks.

3.4.2 The prototypical type I DNA replication fork trap

The *E. coli* Tus-*Ter* mediated replication fork arrest mechanism has been intensely scrutinized in an attempt to better understand the final step in bacterial DNA replication. However, it now appears that the type II replication fork trap which is mostly found in Enterobacteriaceae is more of an exception or even an anomaly with respect to its many redundant *Ter* sites and their wide spread around the chromosome. Ironically, *E. coli* within this family, was probably the worst choice to examine the replication fork trap mechanism and its wider role in DNA replication termination. It seems that the complexity of the type II fork trap in *E. coli* has merely distracted scientists from capturing the elegance and simplicity of the type I system in other Enterobacteriales. Nevertheless, the work on the *E. coli* Tus-*Ter* complex was instrumental to decipher the unique TT-lock mechanism (Mulcair et al., 2006b) which still stands true and is seemingly conserved in all *tus*-harboring bacteria.

The architecture and complexity of replication fork traps vary significantly across *tus*-harbouring bacteria. Yet, the two distinct classes of fork traps contain highly conserved *Ter* sequences despite moderate identity scores between Tus sequences. In the narrow type I fork trap, the *Ter1* vicinal to *tus* acts as a primary *Ter* site to arrest an incoming DnaB helicase travelling toward its non-permissive face (Figure 13B). In type II fork traps, the *Ter* vicinal to *tus* is rarely used as it is in second position from the terminus (*c.f. TerB* and *TerC* in *E. coli*,

Figure 9A). We initially suspected that the large distance between *TerB* and *TerA* could have been the selective driver for acquiring an additional *Ter* site closer to the terminus and *dif* site in *E. coli* rather than the inherent need for a backup system due to inefficient TT-lock formation that has previously been reported (Pandey et al., 2015, Elshenawy et al., 2015, Berghuis et al., 2015). However, the discovery of an unusually wide type I fork trap with a GC-skew switch fitting with the *dif* in *C. neteri* site advocates against this possibility. It is unclear whether replication forks would ever reach a blocking Tus-*Ter* complex during DNA replication in these bacteria as the GC-switch occurs diametrically opposite to the origin at the *dif* locus.

The sharp terminal GC-skew switch observed in several bacterial chromosomes with a prototypical type I system such as *E. tarda*, strongly suggests that replication forks never break through the *Ter* sites in these narrow replication fork traps advocating against the need for backup *Ter* sites. As such the essential requirements to efficiently trap DNA replication forks in the terminus region of Tus-dependent bacteria are two *Ter* sites that are diametrically opposite of the *oriC* and fully competent to form a TT-lock, representative of the prototypical type I replication fork trap.

3.5 Conclusion and perspective

Our *in silico* data support the notion that all bacteria harbouring a type I fork trap use a Tus-*Ter* interaction that is competent in arresting an incoming replication fork by producing a TT-lock. We have discovered several Tus-dependent bacteria that do not require redundant *Ter* sites. The Tus-*Ter* interactions in these bacteria should have increased binding affinities and lock strengths due to the absence of backup *Ter* sites in the fork trap. As such, it would be interesting to examine what the kinetic and thermodynamic parameters in these orthologous Tus-*Ter* systems are? Further examination of these prototypical fork trap systems will certainly help dissect the essential features and requirements of the unique Tus-*Ter* interaction.

Moving forward, a deeper understanding of the overall prevalence of the type I fork trap in bacteria is warranted to gain insight into the biological drivers that require or eliminate the need for redundant *Ter* sites in type II replication fork traps. Initially, our data might have suggested that in *E. coli* the distance of *TerB* from the terminus is not optimal for efficient replication termination and thus an additional *TerC* site with increased fork arrest activity was required to narrow the fork trap near the *dif* site. However, further data mining and the discovery of the extremely wide type I fork traps found in *Cedecea* species do not seem to support that hypothesis. Comparative examination of narrow versus wide type I fork traps could thus be key to shed further light on the evolutionary drivers for this system.

It is clear that the prototypical type I replication fork trap system will provide a great pedagogical tool for teaching DNA replication termination in curricula dealing with the central dogma in molecular biology. Furthermore, the development of a number of high-throughput proteomic and bioinformatic tools prompted by investigations into the *E. coli* Tus-*Ter* interaction, will no doubt facilitate further studies of orthologous systems. Indeed, Tus orthologs with different *Ter* binding-affinities will be very useful to develop finely tuneable assays to study DNA replication and transcription perturbation effects (Willis et al., 2014) and other biotechnologies (Morin et al., 2010, Moreau and Schaeffer, 2012b).

While the prototypical type I system clearly demonstrates that there is no inherent requirement for a backup system to trap replication forks in the terminus region, the existence of both very narrow and very wide replication fork traps is puzzling. The wide type I fork traps found in *Cedecea* suggest that here, replication stalling activity may not be a primary purpose. As such, further comparative studies will be critical to fully decipher the mechanism of DNA replication termination and particularly the intersection between *dif* sites and fork traps as well as possible additional roles of Tus-*Ter e.g.* in chromosomal segregation (Moolman et al., 2016). Finally,

the diversity of type I and type II fork traps with respect to the number of *Ter* sites and their narrow or wide distribution begs the question as to what the evolutionary drivers for such variety are? Further examination of both wide and narrow type I replication fork traps will undoubtedly be instrumental to fully understand the replication fork trap and its possible interactions with other factors essential to DNA replication termination.

4 CHARACTERISATION OF A TYPE I REPLICATION FORK TRAP IN *DICKEYA PARADISIACA*

This chapter describes the first characterisation of an 'ancestral' type I fork trap. The binding of D. paradisiaca Tus to putative Ter and Ter-lock species was examined and compared to E. coli Tus. This chapter is part of a manuscript in preparation for publication (Toft, Sorenson and Schaeffer).

4.1 Introduction:

DNA replication termination in bacteria has been the subject of a long and complex debate within the scientific community. Most of our understanding of this system originated from the biochemical characterisation of the replisomal proteins of Gram-positive *Bacillus subtilis* and Gram-negative *Escherichia coli*. Both model species carry a replication fork trap system situated in the terminus region ensuring that replication forks are allowed to enter but not leave, thus preventing over-replication (Midgley-Smith et al., 2018). Initially, the *E. coli* replication fork trap system involved in DNA replication termination was assumed to be conserved across bacteria like replication initiation (Louarn et al., 1977, Hill et al., 1988). It was later reported that the *E. coli* Tus-*Ter* fork trap system shared no sequence, structural or binding homology to the RTP-*Ter* of *B. subtilis* (Bussiere et al., 1995, Wilce et al., 2001, Neylon et al., 2005a). The RTP protein is only conserved in a subset of *Bacillus* species (Griffiths et al., 1998) and evolved independently of Tus. The notion that a replication fork trap is essential was later dismissed by the observation that deletion of chromosomal *tus* in *E. coli* has little effect on overall cell

viability (Sharma and Hill, 1995, Gottlieb et al., 1992, Roecklein et al., 1991). The rapid growth of genomic data supported this notion, *e.g. Vibrio cholerae*, a bacterium closely related to *E. coli,* is devoid of any replication fork trap system (Galli et al., 2019). Nevertheless, the *tus* gene is commonly found in Enterobacteriales (Galli et al., 2019), suggesting an ancestral termination system with a strong evolutionary drive. The Tus-*Ter* system therefore presents an ideal model to study the biological significance of a replication fork trap in bacteria.

The *E. coli* Tus-*Ter* system has been thoroughly scrutinized, with the discovery of the Tus-*Ter*-lock mechanism of polar fork arrest in 2006 driving a resurgence of investigations into the system (Mulcair et al., 2006b). The high complexity of the replication fork trap architecture in terms of the number and spatial distancing of *Ter* sites within the *E. coli* chromosome has remained mostly undisputed, with up to 14 proposed *Ter* sites, although these have been reduced to 6 (Chapter 3 this thesis, unpublished data). Investigations into the *Ter* fork trap architecture in other Tus-dependent bacteria, revealed that the large number of *Ter* sites in *E. coli* are more of an exception and that there are two distinct classes of replication fork traps (Figure 15) which contain almost identical *Ter* sequences: (type I) a prototypical fork trap architecture in which the innermost *Ter,* acting as a repressor of the *tus* gene, is the primary fork arrest site; and (type II) the *Ter* vicinal to *tus* is less frequently used as it is in the second position from the terminus (*E. coli*) (Figure 15). Of importance, the new type I fork trap architecture is almost exclusively found in bacteria outside of the Enterobacteriacae family and for many of these species, such as *Edwardsiella tarda,* it contains only two closely positioned *Ter* sequences in opposite orientation, close to the *dif* site. Interestingly, the amino acid residues of *E. coli* Tus that make critical interactions with the C(6) base of *Ter* and are part of the C(6)-binding pocket, are mostly conserved in *E. tarda* Tus despite having a moderate protein sequence identity with the *E. coli* protein (48%). Thus, it was hypothesised that in the 'ancestral' bacterial species containing a *tus* gene, only two *Ter* sites were required to sequester replication

forks and as such the ancestral Tus protein must have been capable of forming a very strong TT-lock structure. This is supported by the terminus GC-skew shift observed for these species (Figure 15) which strongly suggest replication forks never break through these *Ter* sites and that there may be some interplay between the *dif* site and a fork trap. One of the most intriguing type I fork trap architecture, with respect to its large size and *Ter* sites as well as its *dif* and GC skew loci, was identified in *Cedecea* species (Figure 15) and suggested that replication stalling activity may not be a primary function of Tus here. Consequently, these *in silico* observations raise questions about the *Ter*-binding properties of orthologous Tus and their TT-lock forming efficiencies in type I replication fork trap systems.

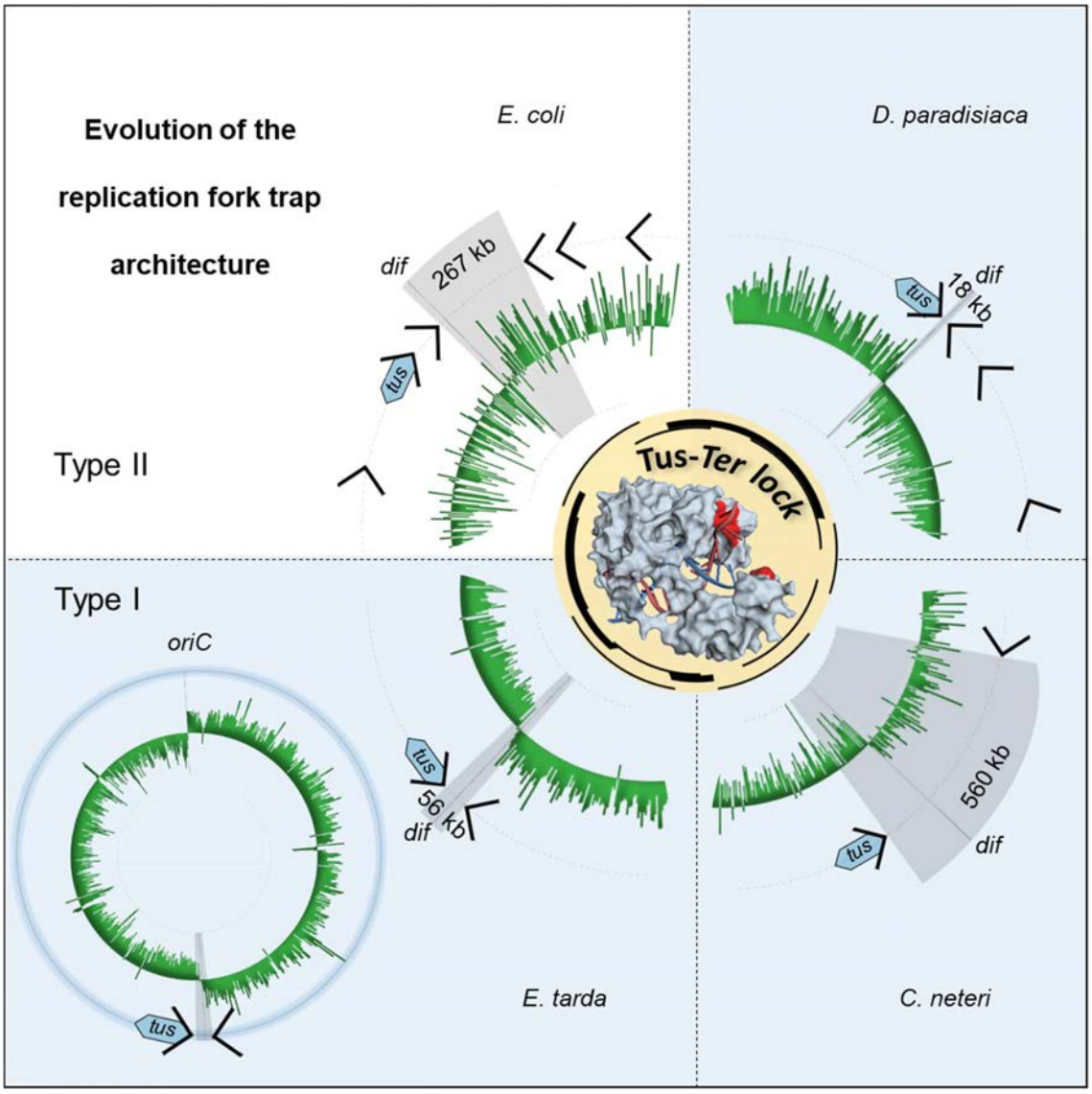


Figure 15: The different replication fork trap architectures and distribution of innermost *Ter* sites in *Escherichia coli*, *Dickeya paradisiaca*, *Cedecea neteri* and *Edwardsiella tarda*. The newly described type I 'prototypical' replication fork traps (shaded in light blue) are compared to the complex 'degenerate' *E. coli* type II fork trap architecture.

The Tus proteins from *Salmonella typhimurium*, *Klebsiella ozaenae*, *Yersinia pestis* and *Proteus mirabilis* have been studied prior to the elucidation of the TT-lock mechanism (Henderson et al., 2001). There were two main conclusions emanating from this study that are relevant today. Firstly, the whole-cell extracts containing over-expressed Tus from each species had similar *TerB* binding activity as an *E. coli* extract (Henderson et al., 2001). Secondly, except for *P. mirabilis* Tus which was not included in the replication fork arrest assay, all other heterologously expressed Tus proteins were capable of arresting replication forks within a mini-chromosome carrying two *E. coli TerB* sites, one on each chromosomal arm with their non-permissive end pointing toward the *oriC*. However, a slightly reduced fork arrest efficiency was observed for *Y. pestis* Tus at *TerB* (Henderson et al., 2001). Interestingly, the evolutionarily distant *Y. pestis* (53% sequence identity between *E. coli* and *Y. pestis* Tus) has been shown to carry only two *Ter* sites within a narrow type I replication fork trap (Appendix 1 Figure S5D). As such, no backup *Ter* sites are present suggesting that *Y. pestis* Tus may have increased binding affinity for its *Ter* sequences and form a highly efficient TT-lock.

'Ancestral' type I replication fork trap systems have not been characterised formally (Figure 16A). This study examines for the first time the *Dickeya paradisiaca* type I replication fork trap system (Figure 16B-D) and more specifically the *Ter*-binding and TT-lock forming ability of *D. paradisiaca* Tus ($_{Dp}$Tus). The *D. paradisiaca* type I replication fork trap system was selected for the following reasons:

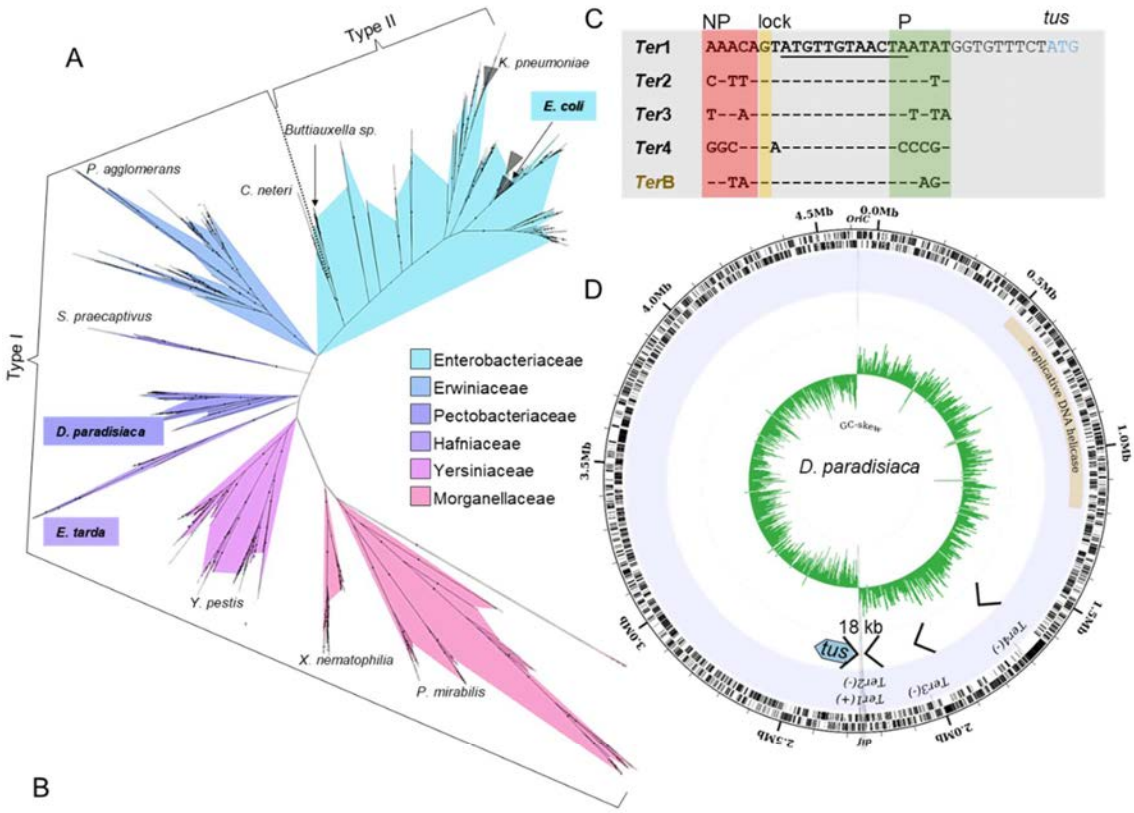
a) A total of four putative *Ter* sites were identified in the *D. paradisiaca* chromosome (Figure 16C) with the left chromosomal arm containing a single *Ter* site vicinal to *tus*, while the right chromosomal arm contains three putative *Ter* sites in opposite orientation (Figure 16D).
b) The distance between the innermost *Ter* sites is just ~18 kbp (Figure 16D) which is in extreme contrast with the wide *E. coli* fork trap.

c) The extremely sharp terminal GC-skew switch suggests that historically, the vicinal *Ter* sequence (*Ter1*) is not breached by replication forks (Figure 16D) and therefore it could be hypothesized that this site is extremely efficient at arresting replication forks.

d) Examining the binding of $_{Dp}$Tus to the *Ter* sequences of the right chromosomal arm (containing three *Ter* sites) may provide valuable insight regarding the need for backup *Ter* sites, *i.e.* are additional *Ter* sites required due to lower Tus-*Ter* complex stability or replication fork arrest activity on this arm?

e) The $_{Dp}$Tus sequence shares ~54% identity to $_{Ec}$Tus and 52% identity to *E. tarda* Tus ($_{Et}$Tus) presenting a valuable model protein sitting at the intersection of two extremely different replication fork trap systems (Figure 16B).

A combination of DSF-GTP assay (Moreau et al., 2012) and GFP-EMSA (Sorenson and Schaeffer, 2020) were applied to characterise and compare the binding of $_{Dp}$Tus and $_{Ec}$Tus to *Ter and Ter*-lock species.



B

Species	# of *Ter* sites	Fork trap size (kb)	Vicinal Ter sequence	Vicinal *Ter* identity	Identity to *E .coli* Tus (%)	Fork-trap classification
E. coli	6*	267	AATAAGTATGTTGTAACTAAAGT	-	-	Type II
D. paradisiaca	**4**	**18**	**AAACAGTATGTTGTAACTAATAT**	**19/23**	**53.9**	**Type I**
E. tarda	2	58	CAAAAGTATGTTGTAACTAATAA	18/23	48.3	Type I

52

Figure 16: Phylogenetic analysis of Tus in Enterobacteriales and the chromosomal fork trap architecture in *Dickeya paradisiaca* (this thesis Chapter 3, unpublished data). (A) Phylogenetic relationship of ~2500 Tus protein sequences using InterPro entries (IPR008865) highlighting the evolutionary distance of *D. paradisiaca* from *E. coli* and *E. tarda* (B) Chromosomal fork trap characteristics and classification of *D. paradisiaca* compared to *Escherichia coli* and *Edwardsiella tarda*. Fork trap size (kb) corresponds to the distance between the two innermost *Ter* sites of opposite polarity. Underlined bases represent a continuous identical sequence shared between *Ter* sequences vicinal to *Tus*. (C) Sequence alignment of the *D. paradisiaca Ter* sites and *E. coli TerB*. *Ter1* is located slightly upstream of the start site (ATG) of the *tus* gene similar to *TerB* in *E. coli.* The conserved 12 bp core sequence is underlined and the G(6) base complementary to C(6) is highlighted in yellow. *NP: non-*

permissive face (red), P: permissive face (green). (D) Circular representation of *D. paradisiaca* chromosome. Illustrated from the outside to the centre of the circle: forward and reverse genes, labelled genomic locations of identified *Ter* sites involved in DNA replication termination, simplified annotation of the termination fork trap utilised, GC-skew over a 5000 bp moving window. The sharp GC-skew switches polarity at the replication origin and between the *Ter1* and *Ter2* sites near *dif*.

4.2 Materials and Methods

4.2.1 Expression and purification of Tus-GFP proteins

$_{Ec}$Tus-GFP was expressed and Ni-affinity purified as previously described (Moreau et al., 2012). The $_{Dp}$Tus coding sequence (Dd703_1900) was synthesised in a codon optimized form for expression in *E. coli* (Bioneer), and cloned into pIM013 (pET-uvGFP) (Moreau et al., 2010) to create pCT300 (pET-N-6His-$_{Dp}$Tus-GFP-C). The pCT300 expression vector carrying the coding sequence of $_{Dp}$Tus-GFP was expressed in *E. coli* BL21(DE3)RIPL using Overnight Express TB Medium, containing 100 µg/ml ampicillin and 50 µg/ml chloramphenicol. In a 1 L flask, 100 mL of TB expression medium was inoculated with a bacterial loop sourced from a fresh overnight master plate culture and incubated at 37°C and 200 RPM until the optical density reached 0.7. Proteins were expressed over 72 hours at 16°C. Lysis and purification procedures were performed as for $_{Ec}$Tus-GFP (Moreau et al., 2012). Protein concentrations were determined by Bradford Assay and purity assessed by SDS-PAGE. Purified Tus-GFP samples were stored in buffer A.

4.2.2 Size exclusion chromotography

Purified $_{Dp}$Tus-GFP, $_{Ec}$Tus-GFP and GFP in buffer A were loaded individually on a Superdex 200 10/300 GL column connected to a BioLogic Duo-Flow system and BioLogic QuadTec UV-

Vis detector (Bio-Rad). Size exclusion chromatography (SEC) was performed at 4°C, in buffer B (buffer A + 125 mM NaCl and without β-mercaptoethanol) and a flow rate of 0.5 mL/min. Absorbance was recorded at 280 nm using BioLogic QuadTec UV-Vis detector (Bio-Rad) and plotted as a function of the retention time. GFP Fluorescence in each fraction (0.5 mL) of $_{Dp}$Tus-GFP was measured in the Bio-Rad CFX96 C1000 Touch Thermal Cycler (FAM channel) and superimposed onto the SEC elution profile obtained at 280 nm absorbance. Purity of fractions containing purified proteins were evaluated by SDS-PAGE and Coomassie Blue staining.

4.2.3 GFP-EMSA

The binding of *Ter* species to Tus-GFP proteins was examined using a modified electrophoretic mobility shift assay (GFP-EMSA) (Sorenson and Schaeffer, 2020). Briefly, 3 µL of a working stock of $_{Dp}$Tus-GFP or $_{Ec}$Tus-GFP (4 µM) in buffer A were mixed with 3 µL of *Ter* or *Ter*-lock DNA (4.5 µM) in buffer C and 6 µL of either $_{dd}H_2O$ (yielding 37.5 mM final NaCl) or a 425 mM NaCl solution (yielding 250 mM final NaCl). Reactions were left at RT for 10 min and loaded onto a 1% agarose gel in TBE and run at 80 V for 40 min. Protein and DNA bands were visualised using a G:BOX Chemi XRQ. GFP fluorescence was first captured (Blue LED module, Filt525), followed by Gelred staining (in 3XGelred solution for 30 min) and DNA fluorescence (TLUM mid-wave, UV06).

4.2.4 DSF-GTP

All DSF-GTP (Moreau et al., 2012) reactions were run in Hard-Shell 96-Well PCR Plates (Bio-Rad) sealed with Microseal B Adhesive sealer (Bio-Rad). Briefly, 50 µl reactions were equilibrated for 10 min at RT prior to melt curve analysis in a real-time thermal cycler (Bio-Rad CFX96 C1000 Touch Thermal Cycler). Temperature range was set from 25 to 90°C, increasing in 0.5°C increments every 30 s and a stabilization phase of 30 s between increments.

GFP fluorescence was recorded using the FAM channel. Denaturation midpoint temperature (T_m) data were analysed in GraphPad Prism (version 8.3.1).

4.2.5 Thermal stability of Tus-GFP:*Ter* and *Ter-lock* complexes

The T_m of free Tus-GFP and Tus-GFP:*Ter* complexes were analysed in 37.5 mM, 144 mM, 250 mM and 400 mM of NaCl by DSF-GTP. For this, 12.5 µL of $_{Dp}$Tus-GFP or $_{Ec}$Tus-GFP (4 µM in buffer A) were mixed with 12.5 µL of *Ter* or *Ter*-lock DNA (4.5 µM in buffer C) and 25 µL of water or NaCl solutions (212.5, 425 or 725 mM). Reactions were left at RT for 10 min and subject to the DSF-GTP protocol described above. All reactions were performed at least in triplicates.

Additionally, the T_m of the Tus-GFP proteins in the presence of increasing *Ter1* or *Ter1*-lock species (ranging from 1 - 10 µM final concentration) was analysed in 37.5 mM and 144 mM of NaCl by DSF-GTP to estimate their apparent dissociation constant (K_{obs}). Here, reactions were performed in duplicates.

4.2.6 Mathematical and Statistical Analysis

Statistics and number of biological and technical repeats are indicated in the relevant figure legends and methods. Statistical analyses were performed using Graphpad Prism 7. Data are expressed as mean values ± SD. To determine if Tus-GFP proteins are affected equally by the ionic strength of the buffer, a simple linear regression analysis was performed for the $_{Dp}$ and $_{Ec}$ Tus-GFP slopes (DFd = 27, P = 0.039).

4.3 Results:

4.3.1 *In silico* analyses of $_{Dp}$Tus

The formation of a 'locked' Tus-*Ter* complex (TT-lock) is crucial to arrest DNA replication in *E. coli*. Through rigorous independent investigations, several residues (E49, I79, F140, H144, G149 and R198) found at the non-permissive face of $_{Ec}$Tus have been shown to be important for *Ter* binding (Kamada et al., 1996) and efficient TT-lock formation to arrest incoming DNA replication forks (Mulcair et al., 2006b, Elshenawy et al., 2015). Triangular comparison of Tus protein sequences from the evolutionarily distant *D. paradisiaca* (Pectobacteriaceae) and *E. tarda* (Hafniaceae) with *E. coli,* revealed that only ~1/3 of the amino acid residues (124) are conserved between all three species and they are often clustered in highly conserved uninterrupted sequences (Figure 17A-B). Amongst these shared residues, E49, I79, F140, H144, G149 and R198 that have been shown to make critical interactions with *Ter* and *Ter*-lock species are almost fully conserved in all three Tus sequences with the exception of $_{Et}$Tus (Figure 17A). Indeed, in $_{Et}$Tus the corresponding $_{Ec}$Tus I79 residue is conservatively substituted to L79. I79 is a key residue within the cytosine binding pocket of $_{Ec}$Tus that makes a critical interaction with the C(6) base in the TT-lock complex (Mulcair et al., 2006b, Elshenawy et al., 2015). Of note, *E. coli* (type II, 'degenerate trap') and *E. tarda* (type I, 'ancestral trap') use two extremely different replication fork trap systems, *i.e.* six and two *Ter* sites respectively (*c.f.* Figure 16B) and the Tus proteins have the highest evolutionary divergence sharing just 27 distinct residues in addition to the 124 residues shared between the three Tus species examined. Interestingly, $_{Dp}$Tus (type I, 'slightly degenerate trap') shares an additional 40 residues exclusively with $_{Et}$Tus, and another 48 distinct residues with $_{Ec}$Tus, seemingly sitting at the intersection between the two extremes of the fork trap systems (Figure 15 & Figure 17B).

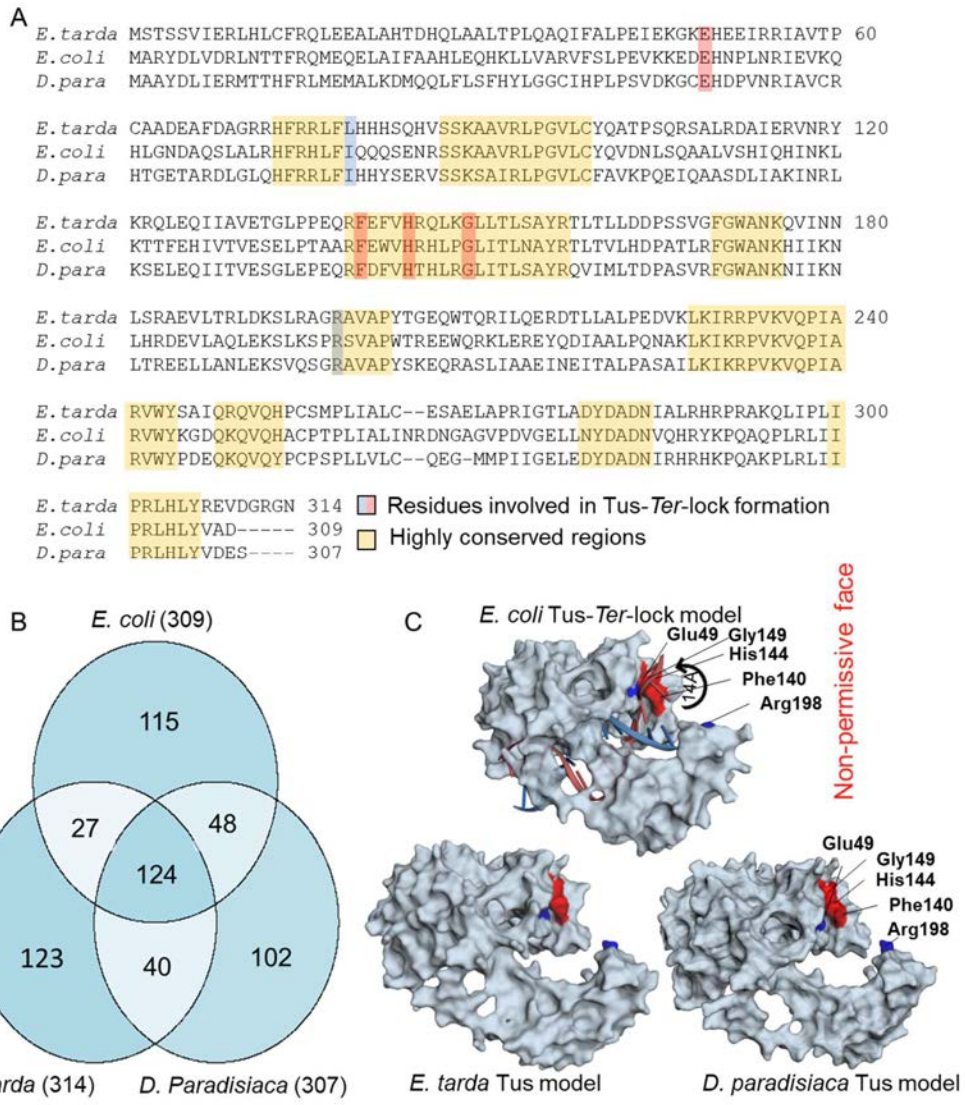


Figure 17: *In silico* analyses of $_{Dp}$Tus. (A) Multiple sequence alignment of Tus protein sequences from *E. tarda* (NC_013508), *E. coli* (U00096) and *D. paradisiaca* (NC_012880) with highlighted conserved residues. (B) Venn diagram describing the shared and distinct amino acid residues between the selected species. The number next to the species name represents the total number of amino acid residues of that Tus protein. $_{Dp}$Tus shares a similar number of residues with both $_{Et}$Tus and $_{Ec}$Tus species. (C) Comparison of the *E. coli* Tus-*Ter*-lock complex 3D structure (PDB 2I06) and the modelled structure of $_{Dp}$Tus (QMEAN = -0.79)

and $_{Et}$Tus (QMEAN = -1.89) proteins using SWISS-MODEL. The essential amino acid residues in the cytosine binding pocket are indicated.

While there is no structural data for free $_{Ec}$Tus, the structural coordinates of $_{Ec}$Tus in complex with *Ter* and *Ter*-lock species are available (Kamada et al., 1996, Mulcair et al., 2006b, Elshenawy et al., 2015). Using these datasets, structural homology modelling was performed on *D. paradisiaca* and *E. tarda* Tus. The modelled structures (based on PDB 2I06) showed no major differences in their C(6) binding pocket when compared to $_{Ec}$Tus, nor in their overall tertiary structure (Figure 17C). Taken together, the data support the notion that $_{Dp}$Tus can form a Tus-*Ter* complex that is competent in arresting an incoming replication fork by producing an extremely stable TT-lock.

4.3.2 Thermal stability and quaternary structure analyses of $_{Dp}$Tus-GFP

A pET vector system was used to express GFP-tagged $_{Ec}$Tus ($_{Ec}$Tus-GFP) and $_{Dp}$Tus ($_{Dp}$Tus-GFP) (*c.f.* Figure 8 p45). The two fusion constructs containing a slightly different linker sequence (Figure 18A) were expressed at similar levels and proteins were purified in identical fashion using a streamlined cell lysis and Ni-affinity chromatography process (Moreau and Schaeffer, 2012a). $_{Dp}$Tus-GFP and $_{Ec}$Tus-GFP showed no significant difference in their electrophoretic mobility (Figure 18B). The fusion proteins were in identical buffer conditions allowing direct thermal stability comparison by DSF-GTP (Moreau et al., 2012). DSF-GTP has been successfully applied to examine $_{Ec}$Tus-GFP binding to *Ter* sites (Moreau and Schaeffer, 2013). Here, the denaturation midpoint (T_m) of Tus-GFP can be accurately measured in the presence or absence of DNA. Consequently, it was important to ascertain that $_{Dp}$Tus-GFP also produced a single, unique T_m peak. A distinct T_m peak was observed for $_{Dp}$Tus-GFP (39.4°C±0.7) that was lower than $_{Ec}$Tus-GFP (45.4°C±0.3) in the same low salt (37.5 mM NaCl)

buffer conditions (Figure 18C). Neither $_{Dp}$Tus-GFP nor $_{Ec}$Tus-GFP generated a T_m peak in a second run (i.e. 25–50 °C), demonstrating that none of these Tus proteins can refold after heat denaturation.

$_{Ec}$Tus is a monomeric protein (Coskun-Ari et al., 1994). To examine the quaternary structure of $_{Dp}$Tus-GFP, size exclusion chromatography (SEC) was used. For this, $_{Dp}$Tus-GFP (63.8 kDa) and $_{Ec}$Tus-GFP (64.6 kDa) samples were run on a Superdex 200 10/300 GL column. The elution profiles of both protein samples were similar (Figure 18D). The retention time (Rt) of $_{Dp}$Tus-GFP peak was 34.0 min. The Rt of $_{Ec}$Tus-GFP was 34.5 min despite its similar MW to $_{Dp}$Tus-GFP. GFP was run as a proteolysis control yielding a single peak with a Rt of 35.6 min. The elution curves indicate that $_{Dp}$Tus and $_{Ec}$Tus-GFP were not proteolyzed, and this was confirmed by SDS-PAGE (Figure 18E). The difference in Rt observed for $_{Ec}$Tus-GFP and $_{Dp}$Tus-GFP is probably due to a difference in their Stokes radius (Horiike et al., 1983), and would suggest that $_{Dp}$Tus-GFP might be larger. SDS-PAGE analysis (Figure 18E) clearly shows the removal of a protein band contaminant at 80 kDa that could be seen prior to SEC (Figure 18B). Taken together, the SEC data show that $_{Dp}$Tus just like $_{Ec}$Tus is a monomer and does not significantly aggregate in the buffer conditions despite its lower thermal stability.

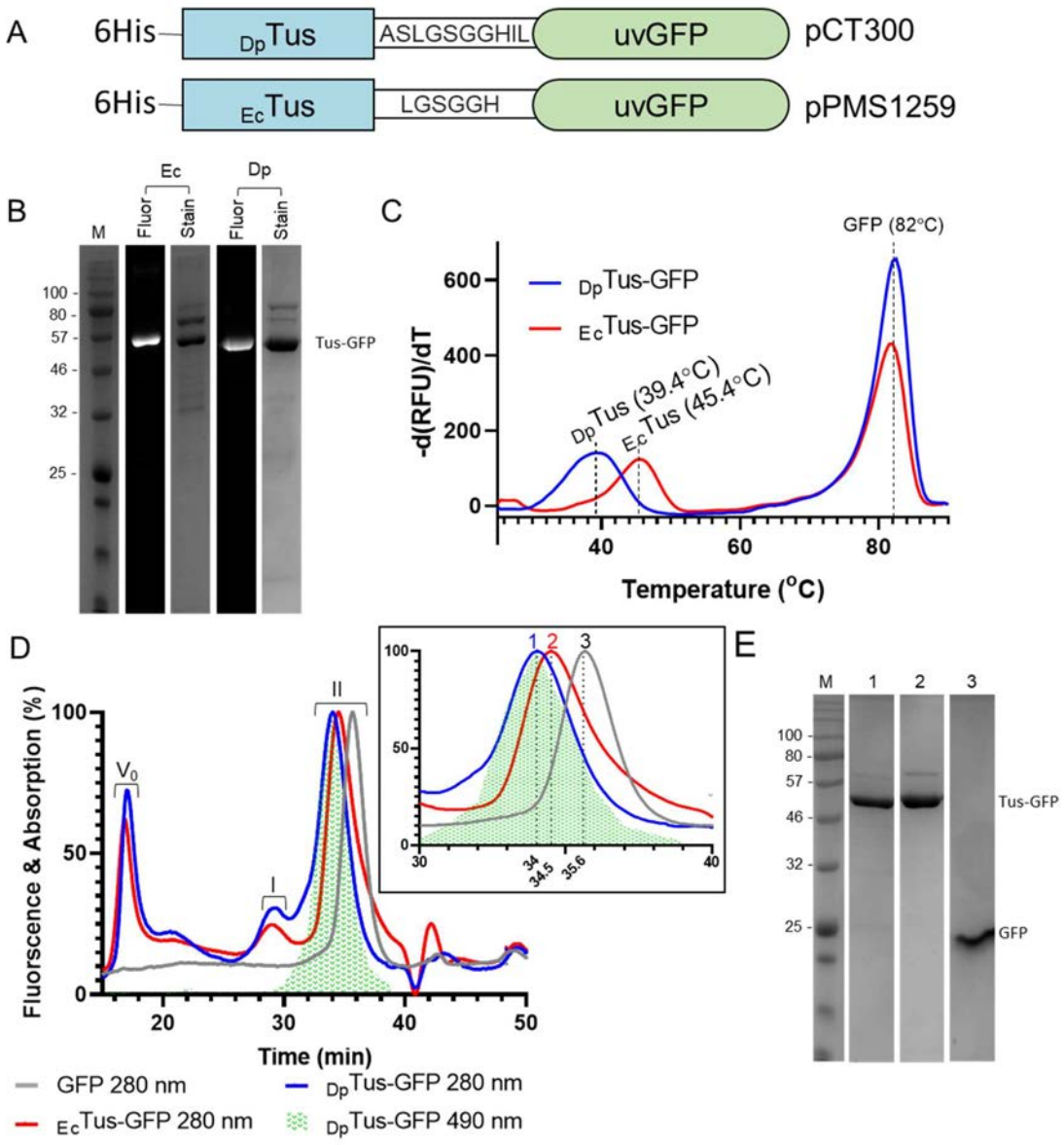


Figure 18: Comparison of $_{Dp}$Tus-GFP and $_{Ec}$Tus-GFP. (A) Tus-GFP constructs used in this study. Although different backbone vectors are used to express Tus-GFP species, the protein products only differ slightly in the GFP-linker. (B) SDS-PAGE analysis of the purified Tus-GFP samples. M: protein marker, lane 1: $_{Ec}$Tus-GFP visualised with GFP Fluorescence (Blue LED module, Filt 525), lane 2: the same $_{Ec}$Tus-GFP gel stained with Coomasie blue R-250, lane 3: $_{Dp}$Tus-GFP visualised via GFP Fluorescence and lane 4: the same $_{Dp}$Tus-GFP gel stained with Coomasie blue R-250 (C) T_m of $_{Dp}$Tus-GFP (39.4°C) and $_{Ec}$Tus-GFP (45.4°C) measured

by DSF-GTP (D) SEC elution profiles of GFP (grey line, Rt = 35.6 min), $_{Ec}$Tus-GFP (red line, Rt = 34.5 min) and $_{Dp}$Tus-GFP (blue line, Rt = 34 min) measured at 280 nm. SEC was performed with a flow rate of 0.5 mL/min. The GFP fluorescence of each fraction of $_{Dp}$Tus-GFP was measured using CFX96 Touch Real-Time PCR (FAM channel) to identify the peak corresponding to the $_{Dp}$Tus-GFP protein. (E) SDS-PAGE analysis of peak fractions from gel filtration chromatography. 1: $_{Dp}$Tus-GFP, 2: $_{Ec}$Tus-GFP and 3: GFP. Note: SDS-PAGE samples were not heat treated before loading to allow assessment of GFP fluorescence. As a result the Tus-GFP bands appear at a lower molecular weight than expected (MW of 63.8 kDa for $_{Dp}$Tus-GFP and 64.6 kDa for $_{Ec}$Tus-GFP).

4.3.3 Comparison of $_{Dp}$Tus-GFP and $_{Ec}$Tus-GFP binding to *Ter* DNA

A GFP based electrophoretic mobility shift assay (GFP-EMSA) was used to evaluate and compare the binding of $_{Dp}$Tus-GFP and $_{Ec}$Tus-GFP to different *Ter and Ter*-lock DNA species. The *Ter*-lock oligonucleotides used in this study are single-stranded at their non-permissive end to free the C(6) allowing it to interact with the cytosine-binding pocket of Tus (Figure 19A) (Moreau and Schaeffer, 2013). The *D. paradisiaca Ter1* and *Ter4* as well as the strong TT-lock forming *E. coli TerB* and weak non-TT-lock forming *TerF* were chosen for cross comparison (Figure 19A). *Ter1* and *Ter4* were selected as they were hypothesised to be the strongest and weakest $_{Dp}$Tus binders respectively, based on their loci in the type I fork trap architecture of *D. paradisiaca* (Figure 16D), and their sequence identities to *TerB* (Figure 19A). All protein-DNA complexes were assembled in low-salt conditions to increase their stability and maximise the chances of observing a band shift for weak *Ter* species as well as in high-salt conditions to weaken non-specific protein-DNA interactions and make apparent any potential binding preference of $_{Dp}$Tus-GFP for the selected *Ter* or *Ter*-lock species.

$_{Ec}$Tus-GFP produced similar discrete shifted bands with all *Ter* and *Ter*-lock DNA species in both low and high salt conditions demonstrating that its binding to weak or strong *Ter* species leads to formation of stable protein-DNA complexes that cannot be distinguished using GFP-EMSA. This was in stark contrast with $_{Dp}$Tus-GFP which only produced discrete shifted bands with *Ter1* and *Ter1*-lock as well as *TerB* and *TerB*-lock in both salt conditions (Figure 19B). Interestingly, $_{Dp}$Tus-GFP in the presence of *Ter4, Ter4*-lock and *TerF* only produced smeared protein and DNA band shifts, that are almost completely lost in high salt (Figure 19B) immediately suggesting weak and labile protein-DNA complexes with these *Ter* species. The very weak $_{Dp}$Tus-GFP binding observed for the *Ter4* and *Ter4*-lock species at 1 μM even in low-salt conditions may suggest that this site is indeed a pseudo-*Ter* in *D. paradisiaca*.

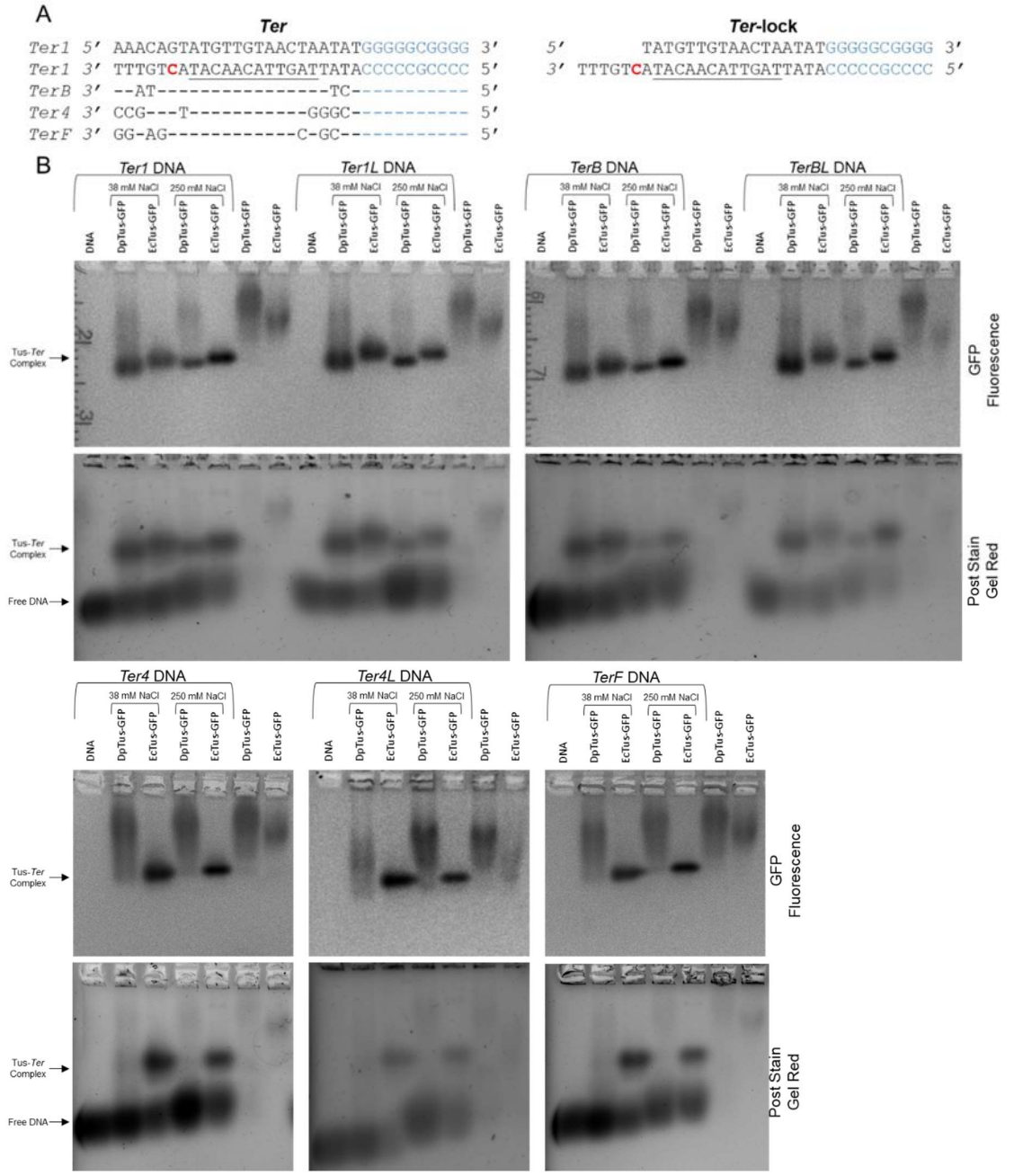


Figure 19: DNA binding of Tus-GFP species to selected *Ter* and *Ter*-lock containing sequences. (A) Oligonucleotides used in this study with an example of the ds*Ter1* compared to *Ter1*-lock that ensures the C(6) base (red) is free. The core *Ter* sequence is underlined. The GC rich sequence (blue) was added to increase the T_m. Base-changes between *Ter* species are displayed by the bottom strand (i.e. in the 3' – 5' direction) and are annealed to their complimentary *Ter* or *Ter*-lock top strand. (B) An electrophoretic mobility shift assay for GFP-

tagged proteins (Sorenson and Schaeffer, 2020), was used to compare the DNA-binding of $_{Dp}$Tus-GFP to strong (*TerB*, *Ter1*), weak (*TerF*, *Ter4*) *Ter* sequences and their respective *Ter*-lock analogues. The binding of $_{Ec}$Tus-GFP to these *Ter* sequences was also evaluated for comparison. Tus-GFP at a concentration of 1 μM were combined with *Ter* sequences at a concentration of 1.1 μM in low and high-salt conditions and analysed by agarose gel electrophoresis and GelRed staining. Gel migration patterns were quantified on a G:BOX Chemi XRQ. Top gels represent GFP fluorescence of Tus-GFP bands before Gelred staining (Blue LED module, Filt525). Bottom gels represent Gelred fluorescence of DNA bands (TLUM mid-wave, UV06). Strong binding partners are identified by complete retardation of the Tus-GFP bands that co-migrate with the DNA while weak or negligible binding is shown by smearing of DNA that do not result in a discrete band. The GFP-EMSA was repeated twice for each reaction with no difference observed between gels.

4.3.4 Salt resistance of Tus–*Ter* and Tus–*Ter*-lock using DSF-GTP

DSF-GTP has previously been applied to examine the salt-dependence of $_{Ec}$Tus-GFP in complex with various *Ter* and *Ter*-lock sequences (Moreau and Schaeffer, 2013). The authors found that the major factor that increased the salt resistance of $_{Ec}$Tus is the formation of the TT-lock complex due to the increased number of base-specific interactions over non-specific interactions (Moreau and Schaeffer, 2013). Using a similar protocol, the salt resistance of $_{Dp}$Tus-GFP in complex with *Ter* and *Ter*-lock species was examined. Here, DSF-GTP was performed in high (250 mM), moderate (144 mM) and low (37.5 mM) NaCl to examine and compare the binding of $_{Dp}$Tus-GFP and $_{Ec}$Tus-GFP to various *Ter* and *Ter*-lock DNA.

The thermal stability of $_{Dp}$Tus-GFP increased at a linear rate of 6°C/M NaCl (Figure 20) which is twice the rate observed for $_{Ec}$Tus-GFP (3°C/M NaCl). This indicates that more stabilising interactions occur with salt ions for $_{Dp}$Tus-GFP compared to $_{Ec}$Tus-GFP.

It is important to note that no significant increase in T_m was observed for $_{Dp}$Tus-GFP in the presence of *TerF*, *Ter4* or *Ter4*-lock, at all NaCl concentrations (Figure 20) confirming the previously observed band shift data (Figure 19B) and that *Ter4* is a pseudo-*Ter* site. However, the T_m value of $_{Ec}$Tus-GFP in the presence of *Ter4* was nearly identical to that with *TerF* (Figure 20). In fact, as previously reported for *TerF*, the salt-dependence of $_{Ec}$Tus-GFP in complex with with *Ter4* and *Ter4*-lock was essentially the same (i.e. curves do not cross), suggesting that *Ter4* cannot form a TT-lock with $_{Ec}$Tus and the protein-DNA interactions in these complexes are mostly non-specific (Moreau and Schaeffer, 2013). The data clearly demonstrate that *Ter4* can be reasonably dismissed as a pseudo-*Ter* site that should not be included in the *D. paradisiaca* fork trap.

Surprisingly, all other *Ter*-lock species led systematically to the largest increase in T_m of $_{Dp}$Tus-GFP at any NaCl concentration compared to the corresponding *Ter* species (*e.g.* $\Delta T_{m(37.5\ mM\ NaCl)}$ = +20.6°C ±0.3 for *TerB*-lock and +15.3°C ±0.5 for *TerB*). This is in complete contrast to $_{Ec}$Tus-GFP, where *Ter* variants are more stabilising than their corresponding *Ter*-lock species in low salt (*e.g.* $\Delta T_{m(37.5\ mM\ NaCl)}$ = +20.4°C ±0.4 for *TerB*-lock and +23.4°C ±0.4 for *TerB*) (Figure 20). In fact, for $_{Ec}$Tus-GFP, this trend inverts at 250 mM NaCl where more base-specific protein-DNA interactions subside with *TerB*-lock than with *TerB*. Indeed, *TerB*, *Ter1* and their respective *Ter*-lock species are almost interchangeable. In fact, at 144 mM NaCl, the ΔT_m of $_{Dp}$Tus-GFP in complex with *Ter1* and *Ter1*-lock ($\Delta T_{m(144\ mM\ NaCl)}$ = +1.8°C ±1.2 and +11.4°C ±1.2 respectively) are indistinguishable with *TerB* and *TerB*-lock ($\Delta T_{m(144\ mM\ NaCl)}$ = +2.1°C ±1.0 and +11.1°C ±1.1 respectively). The salt-resistance trends obtained with *Ter2* and *Ter2*-lock also suggest a strong *Ter* capable of forming a TT-lock. Indeed, $_{Dp}$Tus-GFP in complex with *Ter2*-lock was just slightly more sensitive to salt ($\Delta T_{m(144\ mM\ NaCl)}$ = +10.8°C±0.6) than with *Ter1*-lock.

TerH was used as moderate reference *Ter* that fails to form a TT-lock and hypothetically could allow identification of further *D. paradiasiaca* pseudo-*Ter* sites. Indeed, *TerH* was reclassified as a pseudo-*Ter* sequence due to its inability to arrest nor pause replication forks *in vivo* (*c.f.* Figure 10 p77), in addition to low coverage corresponding to only 13% of the coverage at *TerB* in the ChIP-Seq dataset (*c.f.* Figure 11 p81). Higher T_m values were obtained for $_{Ec}$Tus-GFP in complex with *Ter3* as with *TerH* ($\Delta T_{m(144\ mM\ NaCl)}$ = +12.0°C±0.6 and +9.8°C±0.4 respectively). For both Tus-GFP proteins, *TerH* and *TerH*-lock and *Ter3 and Ter3-lock* are quite similar with respect to their curve profiles and trends, however *Ter3* is still slightly more stabilising at all NaCl concentrations. Of note, the same AT base-pair substitution is present in *Ter3* and *TerH* at position 20. In *E. coli*, it has previously been shown that this A20T substitution in *TerB* significantly reduces fork arrest activity (33% compared to *TerB*) despite only a modest increase in the K_{obs} (90 pM for *TerB* to 290 pM for A20T *TerB*) (Coskun-Ari and Hill, 1997). This calls into question as to whether *Ter3* would be capable of arresting a replication fork in either bacterial species. Taken together, there is a clear ranking of *D. paradisiaca Ter* sites with respect to their binding to $_{Dp}$Tus, *i.e. Ter1 = Ter2 > Ter3* and *Ter4* is a pseudo-*Ter* that cannot be bound in normal cellular Tus concentrations.

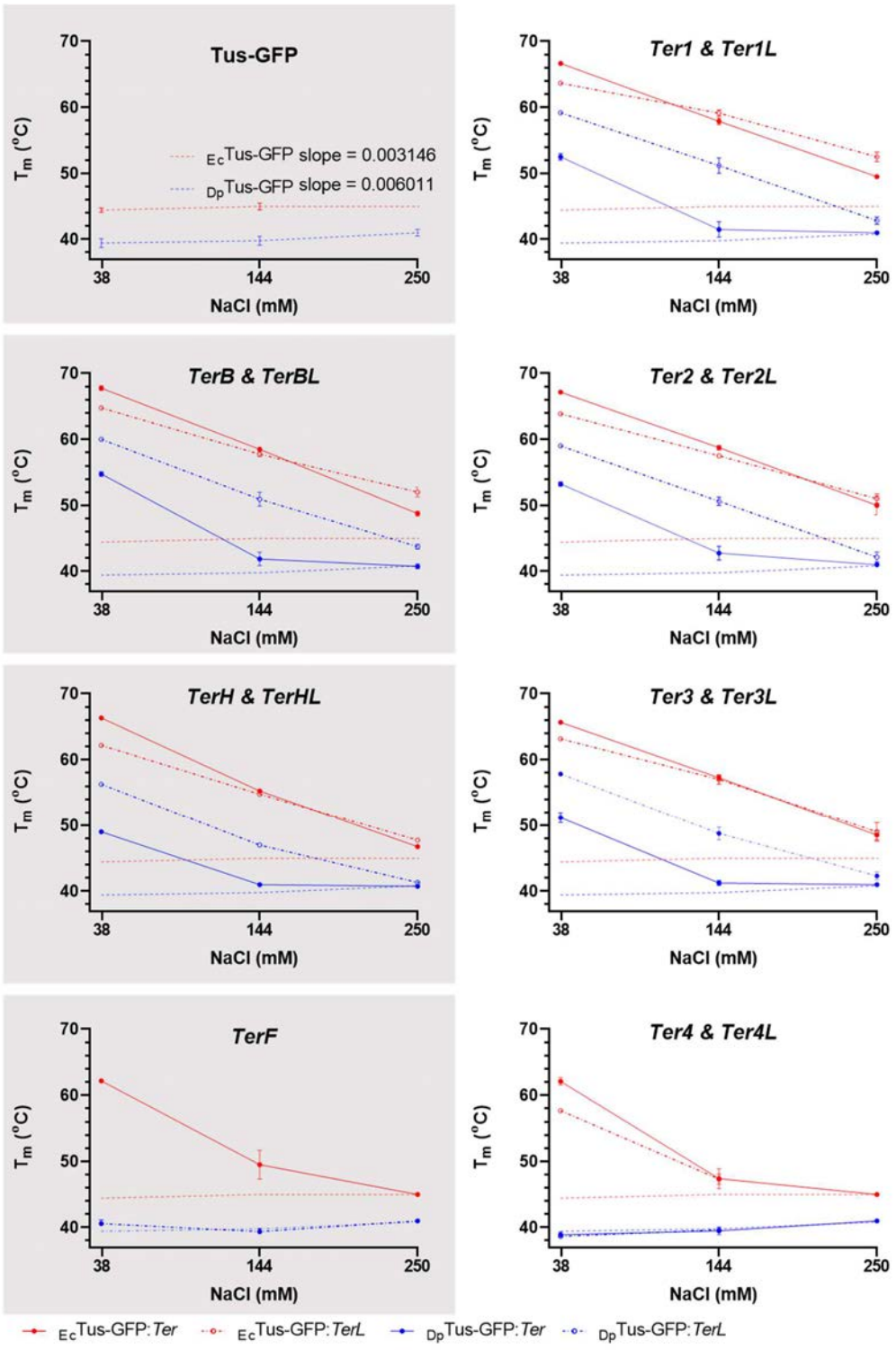


Figure 20: Thermal profiles of $_{Dp}$ and $_{Ec}$Tus-GFP bound to *Ter* and *Ter*-lock sequences in low, moderate and high ionic strength. Thermal stability curves represent $_{Dp}$Tus-GFP (blue) and $_{Ec}$Tus-GFP (red) in complex with *Ter* (continuous lines) and *Ter*-lock (dotted lines)

sequences. Reactions were performed at 1 μM protein and 1.1 μM DNA concentrations, in the presence of low (38 mM), moderate (144 mM) and high (250 mM) NaCl concentrations ($n \geq 3$). A linear regression model was used to determine the significant difference for the salt dependence slopes for free Tus-GFP ($p=0.0393$, DFd=27).

4.3.5 Dissecting the cumulative binding contributions of the Tus-*Ter* interaction

The DSF-GTP data revealed major differences between $_{Dp}$Tus-GFP and $_{Ec}$Tus-GFP in complex with *Ter* and *Ter*-lock species. The significantly reduced thermal stabilization of $_{Dp}$Tus-GFP with *Ter1* compared with *Ter1*-lock is puzzling. In the conditions tested, *Ter4,* was not significantly bound by $_{Dp}$Tus-GFP. As such, *Ter4* was used as a baseline DNA species to differentiate the non-specific and base specific DNA binding contributions that are driving Tus-*Ter* and Tus–*Ter*-lock complex formation for each Tus species. The thermal stability profiles obtained with *Ter4* were used as a proxy to visualize the thermal stabilization contributions of non-specific DNA binding to the proteins which are negligible in the case of $_{Dp}$Tus-GFP. Here, the ΔT_m contributions of *Ter4* at different NaCl concentrations were simply overlaid on the ΔT_m profiles obtained for each Tus-GFP with *TerB* and *Ter1* as well as their respective *Ter*-lock species (Figure 21). These were chosen as they are the strongest *Ter* sites in their respective bacterial replication fork traps.

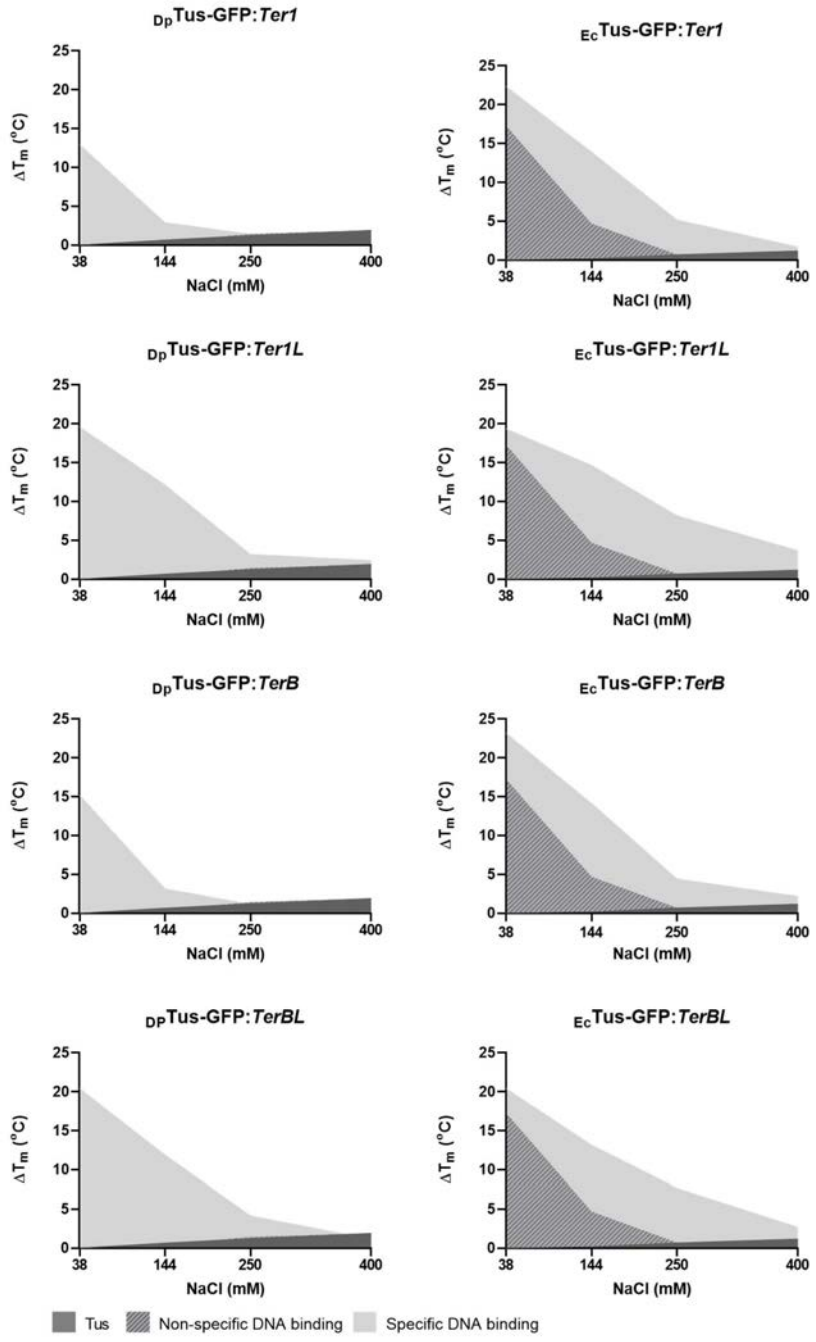


Figure 21: Differences in the specific and non-specific DNA binding properties of $_{Dp}$ and $_{Ec}$Tus-GFP in complex with strong *Ter* and *Ter*-lock sequences. The thermal stability profile of *Ter4* was used as a proxy for the thermal stabilization effects from non-specific DNA binding to the proteins which are negligible in the case of $_{Dp}$Tus-GFP. ΔT_m values were calculated as $T_{m(\text{free Tus-GFP or Tus-GFP:}Ter)} - T_{m(\text{Tus-GFP at 37.5 mM NaCl})}$.

The T_m of $_{Dp}$Tus-GFP:*Ter1* was significantly lower than with *Ter1*-lock at all NaCl concentrations. The data suggest that a larger number of protein-DNA interactions occur with *Ter1*-lock resulting in increasing the affinity of the complex. If these interactions are base specific, increasing the ionic strength of the buffer should lead to a less severe reduction in T_m for the $_{Dp}$Tus-GFP:*Ter1*-lock as base-specific interactions should be impacted less than electrostatic and non-specific interactions. The data support this hypothesis and similar trends can also be seen with *TerB* and *TerB*-lock species. This implies that $_{Dp}$Tus makes considerably more specific-contacts with *Ter*-lock than *Ter* and reinforces the importance of TT-lock formation to arrest replication forks that is less obvious for $_{Ec}$Tus-GFP. Indeed, in complete contrast to what is observed for *Ter1*-lock, the stabilising effect of *Ter1* is significantly lower for $_{Dp}$Tus-GFP compared to $_{Ec}$Tus-GFP ($\Delta T_{m(144\ mM\ NaCl)}$ = + 1.7°C and + 15°C respectively). It is clear that $_{Dp}$Tus has a substantially weaker affinity for double stranded *Ter* sites compared to $_{Ec}$Tus as a result of a lesser number of protein-DNA contacts. In fact, the thermal stabilization contributions of *Ter1* for $_{Dp}$Tus-GFP are completely lost at 250 mM NaCl. This is in stark contrast with $_{Ec}$Tus-GFP (+4.5°C). Interestingly, *Ter1*-lock is still slightly stabilising $_{Dp}$Tus-GFP at 400 mM NaCl (+2.4°C) demonstrating that this sequence is able to successfully compete with higher ion concentrations.

Most surprisingly, the binding of *Ter1*-lock to $_{Dp}$Tus-GFP or $_{Ec}$Tus-GFP results in the same thermal stabilization at 37.5 mM NaCl ($\Delta T_{m(37.5\ mM\ NaCl)}$ = +19.7°C and +19.5°C respectively). In these conditions, electrostatic interactions are expected to have the strongest influence on the stability of a protein-DNA complex. Taken together, the large increase in thermal stabilization of $_{Dp}$Tus-GFP with the *Ter*-lock species hints toward the notion that a greater number of protein-DNA contacts occur within its C(6)-binding pocket compared to $_{Ec}$Tus-GFP.

4.3.6 Affinity of $_{Dp}$Tus-GFP for *Ter1* and *Ter1L*

The observed dissociation constant (K_{obs}) of *TerB* and *TerB*-lock in complex with $_{Ec}$Tus-GFP have previously been determined by DSF-GTP (Moreau and Schaeffer, 2013). The K_{obs} obtained, reflected the K_D measured by surface plasmon resonance (SPR). Here, the effect of *Ter1* and *Ter1*-lock at concentrations ranging from 1 μM to 10 μM on the T_m of $_{Dp}$Tus-GFP was compared to $_{Ec}$Tus-GFP in both low and moderate salt conditions (Figure 22 and Table 7).

At 37.5 mM NaCl, the K_{obs} for $_{Dp}$Tus-GFP:*Ter1* was substantially higher compared to the *Ter1*-lock complex ($K_{obs(37.5\ mM\ NaCl)} = 10^{-9.4}$ M and 10^{-15} M respectively) (Figure 22 & Table 7). This significant increase in affinity for the *Ter*-lock complex is obviously fundamental to fork arrest activity. In the same conditions, the K_{obs} for $_{Ec}$Tus-GFP:*Ter1* and *Ter1*-lock could not be determined ($<10^{-15}$ M) but *Ter1* is clearly more stabilizing.

At 144 NaCl, K_{obs} values could be obtained and compared between Tus-GFP species (Figure 22 and Table 7). It is clear that both $_{Dp}$ and $_{Ec}$Tus-GFP form high-affinity complexes with *Ter1*-lock ($K_{obs(144\ mM\ NaCl)} = 10^{-8.1}$ M and $10^{-9.3}$ M respectively). As expected, $_{Ec}$Tus-GFP also has an equivalently high affinity for *Ter1* ($K_{obs(144\ mM\ NaCl)} = 10^{-8.7}$ M), obscuring the specific contributions to TT-lock formation in these conditions. This is in stark contrast with the affinity of $_{Dp}$Tus-GFP for *Ter1* ($K_{obs(144\ mM\ NaCl)} = 10^{-6.4}$ M) which is ~ 200-fold lower. Considering that Tus must initially bind a fully double stranded *Ter* sequence and is only able to form a locked complex after the GC(6) base-pair has been unzipped by a translocating DNA helicase, the significantly lower affinity of $_{Dp}$Tus for *Ter1* raises the question as to how $_{Dp}$Tus finds its *Ter* sites. This is particularly puzzling, when considering the number of $_{Ec}$Tus is estimated to be as low as 100 copies per *E. coli* cell (Gottlieb et al., 1992).

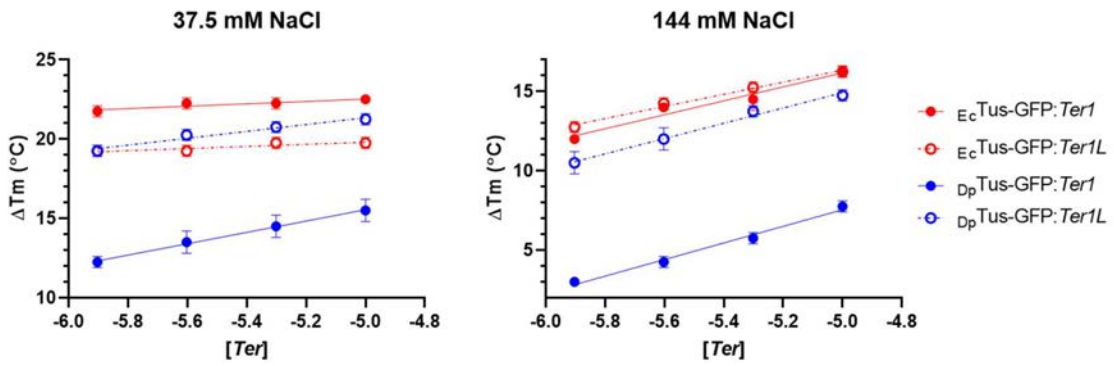


Figure 22: Affinity of $_{Dp}$ and $_{Ec}$Tus-GFP for *Ter1* and *Ter1*-lock in low and moderate salt conditions. DSF-GTP data were obtained with increasing concentrations of *Ter* species ranging from 1 μM to 10 μM with 1 μM constant Tus-GFP, and expressed with ΔT_m as a function of *Ter* concentration. K_{obs}, slope and R-square values are given in Table 7. Error bars represent the SD (N = 2).

Table 7: *Ter1* and *Ter1*-lock-induced thermal stabilization of $_{Dp}$ and $_{Ec}$Tus-GFP

Complex	NaCl (mM)	X-intercept	K_{obs} (M)	R squared	Slope
$_{Dp}$Tus-GFP:Ter1	37.5	-9.354	$10^{-9.4}$	0.9968	3.571
	144	-6.443	$10^{-6.4}$	0.9885	5.232
$_{Dp}$Tus-GFP:Ter1L	37.5	-14.89	10^{-15}	0.9657	2.159
	144	-8.099	$10^{-8.1}$	0.9894	4.402
$_{Ec}$Tus-GFP:Ter1	37.5	-35.14	$<10^{-15}$	0.8526	0.7474
	144	-8.675	$10^{-8.7}$	0.9571	4.817
$_{Ec}$Tus-GFP:Ter1L	37.5	-34.80	$<10^{-15}$	0.8000	0.6644
	144	-9.280	$10^{-9.3}$	0.9888	3.820

4.4 Discussion

In silico analysis of the fork trap architecture of *D. paradisiaca* identified four *Ter* sites, organised as a single unbreachable *Ter* site vicinal to *tus* on the left chromosomal arm to arrest clockwise travelling replication forks, and three *Ter* sites in opposite orientation on the right chromosomal arm with the outermost sites acting as backup systems. The binding of $_{Dp}$Tus to its four putative *Ter* and *Ter*-lock sequences was systematically compared to $_{Ec}$Tus. The data unambiguously demonstrate that $_{Dp}$Tus can bind and form a TT-lock complex with *Ter1-3* but not with the predicted *Ter4* which was dismissed as a pseudo-*Ter* site. The DSF-GTP data obtained at different ionic strengths allowed us to rank each *Ter* sequence with respect to their strength of binding to $_{Dp}$Tus, which also fitted with their individual chromosomal locations and sequence identities to *TerB.* The innermost *Ter1* and *Ter2* were found to be the strongest and most stabilising $_{Dp}$Tus binding sites followed by a weaker *Ter3*. The obvious trend of weakening from *Ter1 to Ter4* suggests that there are minimal evolutionary driving forces at play for outer *Ter* sites to remain functional in *D. paradisiaca*. While the need for six high-affinity *Ter* sites (*TerA-E* and *G*) in *E. coli* that can form strong locked complexes with Tus is still unclear, the $_{Dp}$Tus data possibly negate a requirement for backup *Ter* sites to arrest replication forks in *D. paradisiaca* as the strongest Tus-*Ter*-lock complexes are formed with the two innermost *Ter* sites.

While *Ter4* binding was somewhat expected with $_{Ec}$Tus due to its significant non-specific DNA binding, the 50-fold reduction in $_{Dp}$Tus binding affinity seen for double-stranded *Ter1* compared with the *Ter1*-lock in 144 mM NaCl conditions was unexpected and most intriguing. Indeed, this *Ter* binding behaviour is in complete contrast to what was observed with $_{Ec}$Tus in this as well as previous studies (Moreau and Schaeffer, 2012a, Mulcair et al., 2006b). It is important to note that $_{Dp}$Tus would probably not bind to its *Ter* sites if expressed in similar low numbers as observed for $_{Ec}$Tus(Gottlieb et al., 1992). As such, given its lower affinity to *Ter*, it

could reasonably be argued that $_{Dp}$Tus would require to be expressed at much higher levels in *D. paradisiaca* to be able to act as a repressor and bind its other *Ter* sites. The integrated Tus-*Ter* negative feedback loop system essentially allows the affinity of Tus for its vicinal *Ter* site to evolve without affecting its ability to form a locked complex. Most of the decreased affinity of $_{Dp}$Tus for its vicinal *Ter1* can be attributed to a reduction in protein-DNA contacts for the double-stranded DNA sequence. This is in sharp contrast with the $_{Ec}$Tus protein, which binds with comparable high affinity to both *Ter* and *Ter*-lock species, and where the differences in dissociation rates between these species are mostly masked at 150 mM KCl (Moreau and Schaeffer, 2012a). In fact, for $_{Ec}$Tus, formation of the TT-lock can only be observed at 250 mM KCl by a ~40-fold decrease in dissociation rate for the *Ter*-lock species (Mulcair et al., 2006b, Moreau and Schaeffer, 2012a). Taken together, the data could indicate that a greater number of protein-DNA contacts occur within the C(6)-binding pocket of $_{Dp}$Tus. Thus, it would be highly informative to examine the rate of dissociation of the *Ter* and *Ter*-lock species with $_{Dp}$Tus as well as preforming a mutational analysis on the permissive and non-permissive ends of *Ter* species to delineate the critical base interactions.

5 EVALUATION OF A *D. PARADISIACA* TUS-BASED IMMUNO-PCR PLATFORM TECHNOLOGY

This chapter describes the preliminary development of the D. paradisiaca Tus-Ter1-lock interaction into a quantitative immuno-PCR assay for the sensitive detection of an antibody.

Classic immunoassays can suffer from a lack of analytical sensitivity. This can be circumvented by transforming an immunoassay into an immuno-PCR (IPCR) assay which combines the high affinity and specificity of interaction between an antigen and an antibody, with the extremely low copy-number detectability of DNA by PCR (Sano et al., 1992). This requires the conjugation of a nucleic acid molecule to an antibody (Wiener et al., 2020, Chang et al., 2016). Moreover, quantitative IPCR (qIPCR) has been shown to improve analytical detection sensitivity by up to ~10^9 fold compared to the traditional enzyme-linked immunosorbent assay (ELISA) with many different assay formats available (Chang et al., 2016). The reproducibility of qIPCR that is essential for its diagnostic use, is largely dependent on the procedure used to form the DNA–antibody conjugate.

One of the most recent platforms that has gained significant traction in the field of qIPCR is the Tus-based protein-DNA 'linker' system. The monomeric nature and high affinity of the self-assembling Tus-*Ter*-lock complex enables controlled production of near-irreversible and stoichiometric DNA-protein conjugates (Morin et al., 2010, Morin et al., 2011, Askin and Schaeffer, 2012, Johnston et al., 2014, Chang et al., 2016). The $_{Ec}$Tus-based qIPCR system has

been applied for the ultrasensitive detection of GFP antibodies in the femtomolar range as well as detection of extremely low numbers of antibodies such as tropomyosin-specific IgE and melioidosis-specific IgG in crude complex matrices (Morin et al., 2010, Morin et al., 2011, Askin and Schaeffer, 2012, Johnston et al., 2014, Moreau and Schaeffer, 2012b, Kamath et al., 2017, Cooper et al., 2013). However, the relatively high affinity of $_{Ec}$Tus for non-*Ter* DNA (K_D = 360 nM in 100 mM KCl) (Neylon et al., 2000) is a limitation of the use of Tus-based qIPCR platform in a multiplex configuration. The GFP-EMSA and DSF-GTP data for $_{Dp}$Tus-GFP:*Ter4* showed an almost absence of non-specific DNA-binding at a 1 μM protein-DNA concentration albeit forming an high-affinity complex with *Ter1*-lock similar to what is observed with $_{Ec}$Tus-GFP at 144 mM NaCl (Chapter 4, this thesis). As such, it was hypothesised that this could be an advantageous attribute for its use in multiplex qIPCR. Here, the $_{Dp}$Tus-*Ter1*-lock interaction was investigated for its potential application as a sensitive antibody detection qIPCR assay and compared to the well-established $_{Ec}$Tus-based system (Morin et al., 2010).

5.1 Materials and methods

5.1.1 Expression and purification of Tus-GFP proteins

$_{Dp}$Tus-GFP and $_{Ec}$Tus-GFP were expressed and Ni-affinity purified as previously described (Section 4.2.1 p103).

5.1.2 qIPCR

Polyclonal chicken anti-GFP IgY (Aves Labs GFP-1010, MW ~180 kDa) were adsorbed (IgY suspension concentrations ranging from 56 nM to 560 fM in coat buffer in 50 μL aliquots) overnight at RT onto the surface of wells of a Nunc clear polystyrene (PS) U96 MicroWell™ plate with a MaxiSorp™ (Nunc, 449824). Microplates were blocked for 1 hr at RT with 100 μL block buffer. The $_{Dp}$ or $_{Ec}$Tus-GFP (2.5 μl at 4 μM in buffer A) were assembled with *Ter1*-lock-barcode (1 μl at 10 μM in buffer C) and block buffer (6.5 μl) for 15 min at RT. Each

assembled complex was further diluted to 5 nM in block buffer of which a volume of 50 µL was applied to each well for 30 min at RT. The microplates were washed four times with 200 µL qIPCR wash buffer. A solution containing PCR primers (50 µL, 0.5 µM each in $_{dd}H_2O$) was used to release the barcode DNA for 1 hr at RT. A volume of 10 µL was combined with 10 µL iTaq™ Universal SYBR® Green Supermix (Bio-Rad) in 96-well hard-shell full-skirted PCR plates (Bio-Rad) for PCR amplification. Real-time PCR (qPCR) was performed in a CFX96 C1000 Touch Thermal Cycler with 40 cycles at 95°C for 10 s and 60°C for 20 s. No antibody controls (No Ab) were performed in parallel to measure the background signal (indirect and direct non-specific *Ter*-barcode binding to well surface). No template controls (NTC) were also performed and used as a baseline to set Ct threshold values between separate qPCR runs. Additional experiments were performed with *Ter1*-barcode and single-stranded *Ter1*-barcode (*i.e.* only the top strand that can be amplified) using the same protocol. ΔCt values are defined as: $\Delta Ct = Ct_{(\text{No Ab control})} - Ct_{(\text{sample of interest})}$. All reactions were completed as duplicates but have not been repeated independently.

5.2 Results

As a proof of principle, $_{Dp}$Tus-GFP in complex with an extended *Ter1*-lock sequence that can be amplified by qPCR (*Ter1*-lock-barcode) (Figure 23A), was used as a detection device to measure a range of known concentrations of preadsorbed polyclonal anti-GFP IgY and compared with $_{Ec}$Tus-GFP. For this, $_{Dp}$ and $_{Ec}$Tus-GFP were pre-assembled with the *Ter1*-lock-barcode (Figure 23B) at 1 µM and diluted to 5 nM for anti-GFP (56 nM to 560 fM) detection. After several wash steps the *Ter1*-lock-barcode was allowed to dissociate and quantified by qPCR (Figure 23C). Cycle threshold (Ct) values were obtained and used to evaluate the analytical detection limit and background Ct values of each qIPCR system represented by the non-specific adsorption and dissociation of the *Ter1*-lock-barcode to and from the surface of the wells (Figure 23D). The data show that both Tus-GFP:*Ter1*-lock-barcode qIPCR systems

perform similarly, with anti-GFP detection becoming apparent in the low picomolar range (56 pM) (Figure 23D).

A *Ter1*-barcode as well as a single-stranded barcode species, ss*Ter1*-barcode, were also evaluated with both systems. The main reason to examine the qIPCR systems with these species was to compare the non-specific DNA binding of the detection devices. Here, *Ter1*-barcode and ss*Ter1*-barcode (Figure 23A), which form moderate and negligible interactions respectively with $_{Dp}$Tus-GFP, were assembled with $_{Dp}$ and $_{Ec}$ Tus-GFP in identical conditions as for the *Ter1*-lock-barcode (Figure 23C). The Ct values of the background control wells (*i.e.* No Ab) vary between the different barcode assemblies (Figure 23D) despite the same barcode template being amplified (Figure 23A). Accordingly, ΔCt values for each sample were determined ($\Delta Ct = -1(Ct_{(sample\ of\ interest)} - Ct_{(No\ Ab\ control)})$) to allow direct comparison of the Ct values obtained for each type of qIPCR system (Figure 23E). The $_{Ec}$Tus qIPCR systems with the *Ter1*-barcode has a similar antibody detection profile as observed with *Ter1*-lock-barcode (Figure 23E). However, there is no significant increase in ΔCt value at 56 pM anti-GFP visible for the *Ter1*-barcode with the $_{Dp}$Tus system despite a clear increase in ΔCt value at 560 pM which is similar to *Ter1*-lock-barcode (Figure 23E).

The Ct values for the ss*Ter1*-barcode in the different systems significantly rise above the background at 560 pM anti-GFP (Figure 23E). The $_{Ec}$Tus-GFP:ss*Ter1*-barcode data show a steeper increase in Ct values at higher anti-GFP concentrations (ΔCt = 1.2 and 4.6 at 560 pM and 56 nM anti-GFP respectively) compared to the $_{Dp}$Tus system (ΔCt = 1.5 and 2.9 at 560 pM and 56 nM anti-GFP respectively). It is evident that less ss*Ter1*-barcode remained bound than the *Ter1* and *Ter1*-lock-barcode species to $_{Dp}$ and $_{Ec}$Tus-GFP. This was most apparent with $_{Dp}$Tus-GFP in the 56 nM anti-GFP wells.

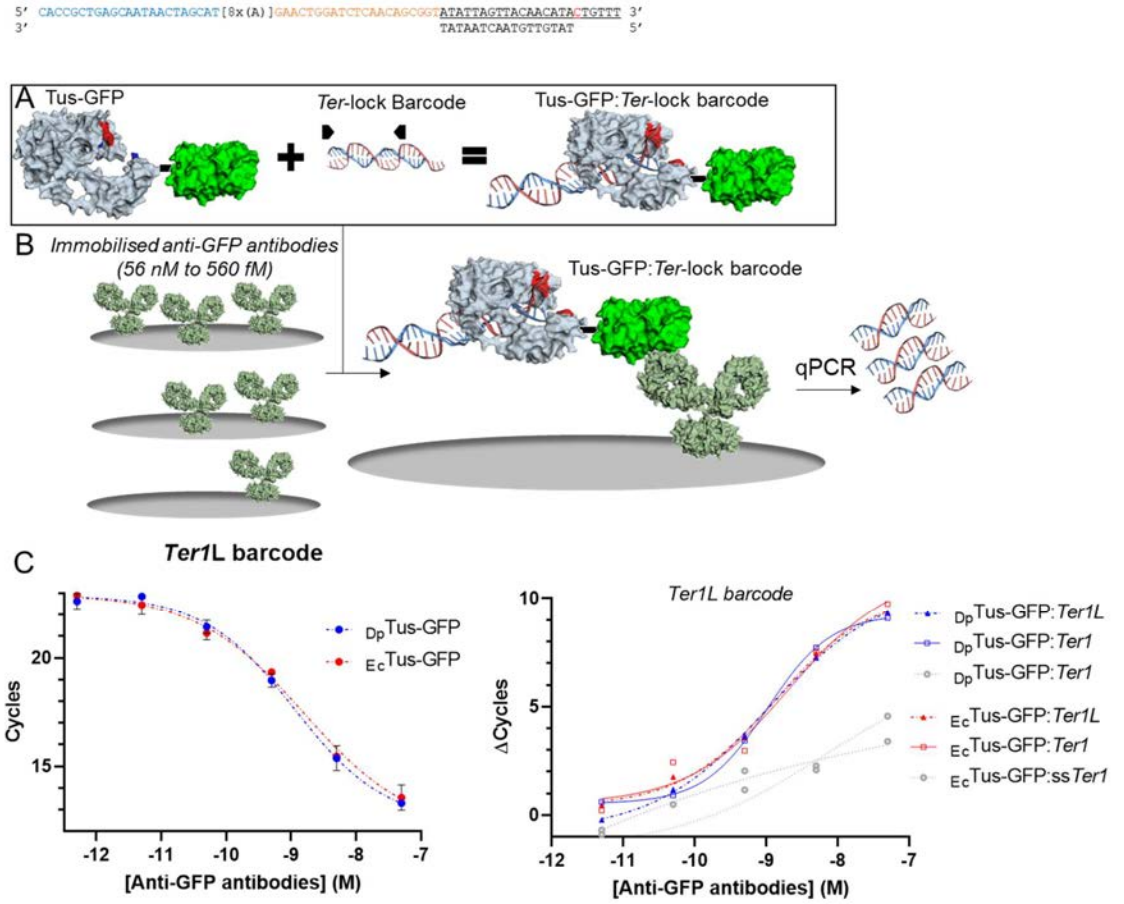


Figure 23: Tus-based quantitative immuno-PCR (qIPCR) platform using a $_{Dp}$Tus-*Ter* detection device. (A) Details of the *Ter1*- and *Ter1*-lock-barcodes used. Forward and reverse primer binding sites are in blue and orange respectively. The top barcode strand contains a *Ter1* site (underlined) and is annealed to either a complementary *Ter1* with a free C(6) (red) that can form a 'locked' Tus-*Ter* complex (*Ter1L*) or a complete *Ter1* complementary sequence. It should be noted that the *Ter1* complementary sequence has a large stretch of G bases at its 3' end that has no effect on the qPCR (B) Schematic representation of the $_{Dp}$Tus-GFP self-assembled with the *Ter1*-lock-barcode (5 nM) and (C) detection of preadsorbed polyclonal anti-GFP IgY from suspensions ranging between 56 nM and 560 fM. After the wash and dissociation steps, qPCR detects the *Ter1*-lock-barcode. (D) Comparison of anti-GFP detection with $_{Dp}$ and $_{Ec}$Tus-GFP assembled to *Ter1*-lock-, *Ter1*- or single stranded *Ter1*-barcode species by qIPCR

(Plot of mean values and SD, N=2 technical replicates). It is important to note that each experiment used identical assembly conditions, dilution and wash steps. The background Ct values (No Ab control) can be used to normalise and directly compare the different detection devices (E). The ΔCt values were calculated as $\Delta Ct = Ct_{(No\ Ab\ control)} - Ct_{(sample\ of\ interest)}$.

5.3 Discussion

The $_{Dp}$Tus was successfully developed into a qIPCR system for the detection of polyclonal anti-GFP IgY and was comparable to $_{Ec}$Tus-GFP in its analytical sensitivity (Morin et al., 2010). This was consistent with the similar affinity of these Tus-GFP species for *Ter1*-lock at 144 mM NaCl (*c.f.* Table 7, p122). The *Ter1*-lock-barcode could be replaced by a fully double-stranded *Ter1*-barcode resulting in a slight decrease in the analytical detection sensitivity of the $_{Dp}$Tus qIPCR system somewhat consistent with the lower affinity of $_{Dp}$Tus for *Ter1* (*c.f.* Table 7 p122). Unfortunately, more datapoints are needed to allow a better differentiation between these barcode species and systems. Here, more subtle differences in the analytical sensitivities of the $_{Dp}$ and $_{Ec}$Tus-based qIPCR systems might also become more apparent.

As expected, the ss*Ter1*-barcode lead to a significant reduction in analytical sensitivity compared to the *Ter1*- and *Ter1*-lock-barcode species. Although, the non-specific DNA binding of $_{Ec}$Tus was most pronounced, especially in the high anti-GFP wells. This is consistent with the *Ter4*-binding data indicating a very weak, non-specific DNA binding quantum for $_{Dp}$Tus-GFP compared with $_{Ec}$Tus-GFP (Chapter 4, this thesis).

In this proof-of-principle assay, the Tus-GFP:*Ter*-barcode detection devices were used at a relatively high concentration (i.e. 5 nM) compared with the original $_{Ec}$Tus-based qIPCR platform (Morin et al., 2010). To capture the true potential of $_{Dp}$Tus for qIPCR diagnostics, it is important that optimal conditions are identified *i.e.* the concentration of $_{Dp}$Tus-GFP:*Ter1*-lock-barcode that produces the largest ΔCt window between the No Ab control and the plateau

of anti-GFP detection. This would allow a better comparison of these qIPCR systems with respect to their analytical sensitivity. Nevertheless, the data clearly demonstrate that the $_{Dp}$Tus-*Ter1*-lock interaction has potential as a qIPCR assay for antibody detection and its low affinity for other non-specific DNA species could be a positive attribute for its use in multiplex applications.

6 CONCLUSIONS AND PERSPECTIVES

Many of the basic components of the large replisome have distinct roles and interactions that are conserved in all three domains of life consistent with the importance of DNA replication (Benkovic et al., 2001). However the systems implemented to terminate replication can differ even among bacteria of a single order (Neylon et al., 2005a). In *E. coli,* the first two termination sites (*TerA* and *TerB*) and the Tus coding sequence, were discovered over 30 years ago (Hill et al., 1987). Since then, the *E. coli* Tus-*Ter* interaction has been intensely scrutinized both *in vitro* and *in vivo* eventually culminating with the discovery of the Tus-*Ter*-lock structure that enabled delineation of the currently accepted mechanistic model of polar replication fork arrest in *E. coli* (Mulcair et al., 2006b). 'En voyage' to decipher this highly debated system, additional *Ter* sites were identified and some dismissed upon further examination, expanding the size of the replication fork trap, and increasing the perceived complexity of DNA replication termination in *E. coli*. Systematic analyses of individual *Ter* sites in *E. coli* with respect to their affinity and kinetics for Tus, ability of forming a TT-lock and replication fork arrest activity, provided a new ranking for these (Moreau and Schaeffer, 2012a, Moreau and Schaeffer, 2013, Duggin and Bell, 2009). Each *Ter* cluster consisted of three high affinity (*TerA*-*E* or *G*), one moderate-to-low affinity (*TerI* or *J*) and one non-lock forming *Ter* site (*TerF* or *H*) (Duggin and Bell, 2009, Moreau and Schaeffer, 2012a). The existence of non-lock forming *TerF* and *TerH* as well as weak *TerI* and *TerJ*, raised doubts about their actual role in the replication fork trap (Moreau and Schaeffer, 2012a). The chromosomal arrangement of these *Ter* sites suggests that they may serve as a 'backup' system if replication forks were to break through the innermost *Ter* sites

whether bound or not by Tus (*TerA-TerD*). Recent efforts to explain the biological relevance of backup *Ter* sites, have mainly focused on investigating the efficiency of TT-lock formation in *E. coli* using elaborate single-molecule-based studies (Berghuis et al., 2018). The main conclusion being that formation of a 'locked' Tus-*Ter* is somewhat inefficient, requiring the presence of additional 'backup' *Ter* sites. Unfortunately, none of these studies could provide answers to the following fundamental questions:

e) What is the reason for the number of *Ter* sites and their locations?
f) Are all *Ter* sites bound to Tus and how does *Ter* occupancy relate to fork arrest?
g) Is there a universal need for backup *Ter* sites in replication fork traps?
h) Is Tus really working alone or is there a possible interplay with the *dif* site and other unknown factors?

A combination of novel *in silico* and *in vitro* approaches was instrumental to address some of these questions that are fundamental to the replication fork trap system in Tus-dependent bacteria.

A simplified replication fork trap in E. coli:

A first of its kind ChIP-Seq analysis allowed to map the genome-wide binding profile of Tus in high-resolution. Taken together with previous affinity data (Moreau and Schaeffer, 2012a) and ectopic insertion findings, it is clear that *TerH*, *I* and *J* are negligibly bound by Tus and as such can be dismissed as pseudo-*Ter* sites. A refined minimal replication fork trap within *E. coli* comprising two clusters of three *Ter* sites is proposed with *TerB*, *C* and *G* blocking clockwise moving replication forks, and *TerA*, *D* and *E*, the anticlockwise moving replication fork. The coverage at *TerG* was similar to *TerB* and suggested that this site is almost certainly bound at normal bacterial Tus concentrations. Taken together with the strong Tus binding and lock-forming ability of *TerG* (Moreau and Schaeffer, 2012a), the ChIP-Seq data suggest that

the absence of paused fork intermediates observed by Duggin and Bell (2009), is a result of the replication fork not reaching this *Ter* site. In fact, our data demonstrate the termination GC-skew flip in *E. coli* coincides with a simple replication fork trap involving just *TerA, B* and *C* based on their fractional use in replication fork arrest activity (Duggin and Bell, 2009). As such, the need for *TerD*, and particularly the distant *TerE* and *TerG* in the *E. coli* chromosome, although probably bound by Tus, still remains enigmatic (see below).

There is no inherent requirement for backup Ter sites to trap replication forks:

To better understand the biological relevance of the outermost *Ter* sites (*TerD, E and G*), the fork trap architectures of a diverse subset of Tus-dependent Enterobacteriales were examined. A prototypical fork trap architecture was identified and classed as type I based on its 'ancestral' features. In the prototypical type I fork trap, the *Ter1* vicinal to *tus* acts as a primary *Ter* site to arrest an incoming DnaB helicase travelling toward its non-permissive face. The type I fork trap architecture can be found in all families of bacteria within Enterobacteriales. Indeed, while investigating the evolutionary shift between type I (mostly found outside the Enterobacteriaceae) and type II fork traps, an unusually wide type I fork trap consisting of only two oppositely-oriented *Ter* sites (564 kb apart) was discovered in *C. neteri* within Enterobacteriacae. It is unclear whether replication forks would ever reach the *C. neteri Ter* sites as the GC-skew flip coincides with the *dif* locus (diametrically opposite to the origin) which is asymmetrically located ~104 kb from the nearest *Ter* site.

The fork trap architecture of *D. paradisiaca* sits at the intersection between two extreme fork trap systems found in *E. coli* and *E. tarda.* Indeed, the *D. paradisiaca* type I fork trap is organised as a single 'unbreachable' *Ter* site vicinal to *tus* on the left chromosomal arm, and two confirmed *Ter* sites in opposite orientation. It is important to note that the amino acid residues that make critical interactions with the C(6) base in the TT-lock (Mulcair et al., 2006a)

are conserved in $_{Dp}$Tus despite having one of the lowest overall identity score with the *E. coli* Tus sequence (54%). The data unambiguously demonstrate that $_{Dp}$Tus can bind and form a TT-lock complex with *Ter1-3* but not with the predicted *Ter4* which was dismissed as a pseudo-*Ter* site. In fact, the innermost *Ter1* and *Ter2* are the strongest and most stabilising $_{Dp}$Tus binding sites followed by a weaker *Ter3* suggesting that there are minimal evolutionary driving forces at play for outer *Ter* sites to remain functional in *D. paradisiaca*. On the opposite spectrum of type I fork trap architectures, very narrow systems can be found such as in *Edwardsiella tarda* where a 56 kbp fork trap is located diametrically opposite to *oriC*, with termination seemingly occurring between its two unique *Ter* sites and coinciding with *dif*. The sharp terminal GC-skew switch in the *E. tarda* chromosome, strongly suggests that replication forks rarely break through its *Ter* sites, advocating against an inherent need for backup *Ter* sites in 'ancestral' type I fork traps.

Different evolutionary forces at play in the Tus-Ter system:

Most intriguing and unexpected, was the 50-fold reduction in $_{Dp}$Tus binding affinity for double-stranded *Ter1* compared with the *Ter1*-lock. Indeed, this *Ter* binding behaviour is in complete contrast to what was observed with $_{Ec}$Tus in this thesis as well as previous studies (Moreau and Schaeffer, 2012a, Mulcair et al., 2006b). Of note, $_{Dp}$Tus would probably not bind to its *Ter* sites if expressed in similar low numbers as observed for $_{Ec}$Tus (Gottlieb et al., 1992). Given its lower affinity to *Ter*, it was argued that $_{Dp}$Tus would require to be expressed at much higher levels in *D. paradisiaca* to be able to act as a repressor and bind to its other *Ter* sites. This leads to the possibility that the integrated Tus-*Ter* negative feedback loop system essentially allows the affinity of Tus for its vicinal *Ter* site to evolve without affecting its ability to form a locked complex. As such, most of the decreased affinity of $_{Dp}$Tus for its vicinal *Ter1* can be attributed to a reduction in protein-DNA contacts for the double-stranded DNA sequence. The data strongly suggest that a greater number of protein-DNA contacts occur within the C(6)-binding

pocket of $_{Dp}$Tus compared to $_{Ec}$Tus. Finally, despite the relatively high protein sequence variability seen in Tus (except for the C(6) binding pocket) the *Ter* sequences have very little plasticity.

Final remarks and perspectives:

The findings in this thesis highlight the true simplicity of replication termination in Tus-dependent bacteria harbouring an 'ancestral' type I fork trap. The biological drivers that require redundant *Ter* sites as well as possible interactions of Tus with other factors remain to be identified. Moving forward, it would be of high interest to show that $_{Dp}$Tus is expressed at a significantly higher level in *D. paradisiaca* compared to $_{Ec}$Tus as a result of its reduced affinity for the vicinal *Ter* site as well as exploring its protein-DNA kinetics and dynamics. The same goes for the *C. neteri* and *E. tarda* systems which should be highly informative. Finally, the proof-of-concept $_{Dp}$Tus-based qIPCR system has potential to be useful in multiplex applications as it is significantly less burdened by non-specific protein-DNA interactions.

7 REFERENCES

ALTSCHUL, S. F., GISH, W., MILLER, W., MYERS, E. W. & LIPMAN, D. J. 1990. Basic local alignment search tool. *J Mol Biol,* 215**,** 403-10.

ARAKAWA, K. & TOMITA, M. 2007. The GC skew index: a measure of genomic compositional asymmetry and the degree of replicational selection. *Evol Bioinform Online,* 3**,** 159-68.

ASKIN, S. P., MORIN, I. & SCHAEFFER, P. M. 2011. Development of a protease activity assay using heat-sensitive Tus-GFP fusion protein substrates. *Anal Biochem,* 415**,** 126-33.

ASKIN, S. P. & SCHAEFFER, P. M. 2012. A universal immuno-PCR platform for comparative and ultrasensitive quantification of dual affinity-tagged proteins in complex matrices. *Analyst,* 137**,** 5193-5196.

BASTIA, D. & ZAMAN, S. 2014. Mechanism and physiological significance of programmed replication termination. *Semin Cell Dev Biol,* 30**,** 165-73.

BASTIA, D., ZZAMAN, S., KRINGS, G., SAXENA, M., PENG, X. & GREENBERG, M. M. 2008. Replication termination mechanism as revealed by Tus-mediated polar arrest of a sliding helicase. *Proc Natl Acad Sci U S A,* 105**,** 12831-6.

BENKOVIC, S. J., VALENTINE, A. M. & SALINAS, F. 2001. Replisome-mediated DNA replication. *Annu Rev Biochem,* 70**,** 181-208.

BERGHUIS, B. A., DULIN, D., XU, Z. Q., VAN LAAR, T., CROSS, B., JANISSEN, R., JERGIC, S., DIXON, N. E., DEPKEN, M. & DEKKER, N. H. 2015. Strand separation establishes a sustained lock at the Tus-Ter replication fork barrier. *Nat Chem Biol,* 11**,** 579-85.

BERGHUIS, B. A., KOBER, M., VAN LAAR, T. & DEKKER, N. H. 2016. High-throughput, high-force probing of DNA-protein interactions with magnetic tweezers. *Methods,* 105**,** 90-8.

BERGHUIS, B. A., RADUCANU, V. S., ELSHENAWY, M. M., JERGIC, S., DEPKEN, M., DIXON, N. E., HAMDAN, S. M. & DEKKER, N. H. 2018. What is all this fuss about Tus? Comparison of recent findings from biophysical and biochemical experiments. *Crit Rev Biochem Mol Biol,* 53**,** 49-63.

BIDNENKO, V., EHRLICH, S. D. & MICHEL, B. 2002. Replication fork collapse at replication terminator sequences. *EMBO J,* 21**,** 3898-907.

BIDNENKO, V., LESTINI, R. & MICHEL, B. 2006. The Escherichia coli UvrD helicase is essential for Tus removal during recombination-dependent replication restart from Ter sites. *Mol Microbiol,* 62**,** 382-96.

BLAKELY, G. W. & SHERRATT, D. J. 1994. Interactions of the site-specific recombinases XerC and XerD with the recombination site dif. *Nucleic Acids Res,* 22**,** 5613-20.

BOLGER, A. M., LOHSE, M. & USADEL, B. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics,* 30**,** 2114-20.

BUSSIERE, D. E., BASTIA, D. & WHITE, S. W. 1995. Crystal structure of the replication terminator protein from B. subtilis at 2.6 A. *Cell,* 80**,** 651-60.

CHAN, L., CROSS, H. F., SHE, J. K., CAVALLI, G., MARTINS, H. F. & NEYLON, C. 2007. Covalent attachment of proteins to solid supports and surfaces via Sortase-mediated ligation. *PLoS One,* 2**,** e1164.

CHANG, L., LI, J. & WANG, L. 2016. Immuno-PCR: An ultrasensitive immunoassay for biomolecular detection. *Anal Chim Acta,* 910**,** 12-24.

CHATTERJEE, D. K., SITARAMAN, K., BAPTISTA, C., HARTLEY, J., HILL, T. M. & MUNROE, D. J. 2008. Protein microarray on-demand: a novel protein microarray system. *PLoS One,* 3**,** e3265.

COOPER, A., WILLIAMS, N. L., MORRIS, J. L., NORTON, R. E., KETHEESAN, N. & SCHAEFFER, P. M. 2013. ELISA and immuno-polymerase chain reaction assays for the sensitive detection of melioidosis. *Diagn Microbiol Infect Dis,* 75**,** 135-8.

COOPER, S. & HELMSTETTER, C. E. 1968. Chromosome replication and the division cycle of Escherichia coli B/r. *J Mol Biol,* 31**,** 519-40.

COSKUN-ARI, F. F. & HILL, T. M. 1997. Sequence-specific interactions in the Tus-Ter complex and the effect of base pair substitutions on arrest of DNA replication in Escherichia coli. *J Biol Chem,* 272**,** 26448-56.

COSKUN-ARI, F. F., SKOKOTAS, A., MOE, G. R. & HILL, T. M. 1994. Biophysical characteristics of Tus, the replication arrest protein of Escherichia coli. *J Biol Chem,* 269**,** 4027-34.

CRABTREE, J., AGRAWAL, S., MAHURKAR, A., MYERS, G. S., RASKO, D. A. & WHITE, O. 2014. Circleator: flexible circular visualization of genome-associated data with BioPerl and SVG. *Bioinformatics,* 30**,** 3125-7.

DAHDAH, D. B., MORIN, I., MOREAU, M. J., DIXON, N. E. & SCHAEFFER, P. M. 2009. Site-specific covalent attachment of DNA to proteins using a photoactivatable Tus-Ter complex. *Chem Commun (Camb)***,** 3050-2.

DUGGIN, I. G. & BELL, S. D. 2009. Termination structures in the Escherichia coli chromosome replication fork trap. *J Mol Biol,* 387**,** 532-9.

DUNDAS, C. M., DEMONTE, D. & PARK, S. 2013. Streptavidin-biotin technology: improvements and innovations in chemical and biological applications. *Appl Microbiol Biotechnol,* 97**,** 9343-53.

EDGAR, R. C. 2004. MUSCLE: a multiple sequence alignment method with reduced time and space complexity. *BMC Bioinformatics,* 5**,** 113.

ELSHENAWY, M. M., JERGIC, S., XU, Z. Q., SOBHY, M. A., TAKAHASHI, M., OAKLEY, A. J., DIXON, N. E. & HAMDAN, S. M. 2015. Replisome speed determines the efficiency of the Tus-Ter replication termination barrier. *Nature,* 525**,** 394-8.

ENYEART, P. J., CHIRIELEISON, S. M., DAO, M. N., PERUTKA, J., QUANDT, E. M., YAO, J., WHITT, J. T., KEATINGE-CLAY, A. T., LAMBOWITZ, A. M. & ELLINGTON, A. D. 2013. Generalized bacterial genome editing using mobile group II introns and Cre-lox. *Mol Syst Biol,* 9**,** 685.

ESNAULT, E., VALENS, M., ESPELI, O. & BOCCARD, F. 2007. Chromosome structuring limits genome plasticity in Escherichia coli. *PLoS Genet,* 3**,** e226.

FOSSUM, S., CROOKE, E. & SKARSTAD, K. 2007. Organization of sister origins and replisomes during multifork DNA replication in Escherichia coli. *EMBO J,* 26**,** 4514-22.

FRANCOIS, V., LOUARN, J. & LOUARN, J. M. 1989. The terminus of the Escherichia coli chromosome is flanked by several polar replication pause sites. *Mol Microbiol,* 3**,** 995-1002.

FRANK, A. C. & LOBRY, J. R. 1999. Asymmetric substitution patterns: a review of possible underlying mutational or selective mechanisms. *Gene,* 238**,** 65-77.

GALLI, E., FERAT, J. L., DESFONTAINES, J. M., VAL, M. E., SKOVGAARD, O., BARRE, F. X. & POSSOZ, C. 2019. Replication termination without a replication fork trap. *Sci Rep,* 9**,** 8315.

GOTTLIEB, P. A., WU, S., ZHANG, X., TECKLENBURG, M., KUEMPEL, P. & HILL, T. M. 1992. Equilibrium, kinetic, and footprinting studies of the Tus-Ter protein-DNA interaction. *J Biol Chem,* 267**,** 7434-43.

GRIFFITHS, A. A., ANDERSEN, P. A. & WAKE, R. G. 1998. Replication terminator protein-based replication fork-arrest systems in various Bacillus species. *J Bacteriol,* 180**,** 3360-7.

GUREVICH, A., SAVELIEV, V., VYAHHI, N. & TESLER, G. 2013. QUAST: quality assessment tool for genome assemblies. *Bioinformatics,* 29**,** 1072-5.

HAMDAN, S. M., LOPARO, J. J., TAKAHASHI, M., RICHARDSON, C. C. & VAN OIJEN, A. M. 2009. Dynamics of DNA replication loops reveal temporal control of lagging-strand synthesis. *Nature,* 457**,** 336-9.

HANSEN, F. G. & ATLUNG, T. 2018. The DnaA Tale. *Frontiers in Microbiology,* 9.

HENDERSON, T. A., NILLES, A. F., VALJAVEC-GRATIAN, M. & HILL, T. M. 2001. Site-directed mutagenesis and phylogenetic comparisons of the Escherichia coli Tus protein: DNA-protein interactions alone can not account for Tus activity. *Mol Genet Genomics,* 265**,** 941-53.

HENDRICKSON, H. & LAWRENCE, J. G. 2007. Mutational bias suggests that replication termination occurs near the dif site, not at Ter sites. *Mol Microbiol,* 64**,** 42-56.

HIASA, H. & MARIANS, K. J. 1992. Differential inhibition of the DNA translocation and DNA unwinding activities of DNA helicases by the Escherichia coli Tus protein. *J Biol Chem,* 267**,** 11379-85.

HIDAKA, M., AKIYAMA, M. & HORIUCHI, T. 1988. A consensus sequence of three DNA replication terminus sites on the E. coli chromosome is highly homologous to the terR sites of the R6K plasmid. *Cell,* 55**,** 467-75.

HIDAKA, M., KOBAYASHI, T. & HORIUCHI, T. 1991. A newly identified DNA replication terminus site, TerE, on the Escherichia coli chromosome. *J Bacteriol,* 173**,** 391-3.

HILL, T. M., HENSON, J. M. & KUEMPEL, P. L. 1987. The terminus region of the Escherichia coli chromosome contains two separate loci that exhibit polar inhibition of replication. *Proc Natl Acad Sci U S A,* 84**,** 1754-8.

HILL, T. M., PELLETIER, A. J., TECKLENBURG, M. L. & KUEMPEL, P. L. 1988. Identification of the DNA sequence from the E. coli terminus region that halts replication forks. *Cell,* 55**,** 459-66.

HORIIKE, K., TOJO, H., YAMANO, T. & NOZAKI, M. 1983. Interpretation of the stokes radius of macromolecules determined by gel filtration chromatography. *J Biochem,* 93**,** 99-106.

HORIUCHI, T. & HIDAKA, M. 1988. Core sequence of two separable terminus sites of the R6K plasmid that exhibit polar inhibition of replication is a 20 bp inverted repeat. *Cell,* 54**,** 515-23.

HORIUCHI, T., NISHITANI, H. & KOBAYASHI, T. 1995. A new type of E. coli recombinational hotspot which requires for activity both DNA replication termination events and the Chi sequence. *Adv Biophys,* 31**,** 133-47.

HUG, L. A., BAKER, B. J., ANANTHARAMAN, K., BROWN, C. T., PROBST, A. J., CASTELLE, C. J., BUTTERFIELD, C. N., HERNSDORF, A. W., AMANO, Y., ISE, K., SUZUKI, Y., DUDEK, N., RELMAN, D. A., FINSTAD, K. M., AMUNDSON, R., THOMAS, B. C. & BANFIELD, J. F. 2016. A new view of the tree of life. *Nat Microbiol,* 1**,** 16048.

ISAACS, F. J., CARR, P. A., WANG, H. H., LAJOIE, M. J., STERLING, B., KRAAL, L., TOLONEN, A. C., GIANOULIS, T. A., GOODMAN, D. B., REPPAS, N. B., EMIG, C. J., BANG, D., HWANG, S. J., JEWETT, M. C., JACOBSON, J. M. & CHURCH, G. M. 2011. Precise manipulation of chromosomes in vivo enables genome-wide codon replacement. *Science,* 333**,** 348-53.

ISHIKAWA, S., OGURA, Y., YOSHIMURA, M., OKUMURA, H., CHO, E., KAWAI, Y., KUROKAWA, K., OSHIMA, T. & OGASAWARA, N. 2007. Distribution of stable DnaA-binding sites on the Bacillus subtilis genome detected using a modified ChIP-chip method. *DNA Res,* 14**,** 155-68.

JOHNSTON, E. B., KAMATH, S. D., LOPATA, A. L. & SCHAEFFER, P. M. 2014. Tus-Ter-lock immuno-PCR assays for the sensitive detection of tropomyosin-specific IgE antibodies. *Bioanalysis,* 6**,** 465-476.

KACZMARCZYK, S. J., SITARAMAN, K., HILL, T., HARTLEY, J. L. & CHATTERJEE, D. K. 2010. Tus, an E. coli protein, contains mammalian nuclear targeting and exporting signals. *PLoS One,* 5**,** e8889.

KAMADA, K., HORIUCHI, T., OHSUMI, K., SHIMAMOTO, N. & MORIKAWA, K. 1996. Structure of a replication-terminator protein complexed with DNA. *Nature,* 383**,** 598-603.

KAMATH, S. D., JOHNSTON, E. B., IYER, S., SCHAEFFER, P. M., KOPLIN, J., ALLEN, K. & LOPATA, A. L. 2017. IgE reactivity to shrimp allergens in infants and their cross-reactivity to house dust mite. *Pediatr Allergy Immunol,* 28**,** 703-707.

KAPLAN, D. L. 2006. Replication termination: mechanism of polar arrest revealed. *Curr Biol,* 16**,** R684-6.

KHATRI, G. S., MACALLISTER, T., SISTA, P. R. & BASTIA, D. 1989. The replication terminator protein of E. coli is a DNA sequence-specific contra-helicase. *Cell,* 59**,** 667-74.

KOLMOGOROV, M., YUAN, J., LIN, Y. & PEVZNER, P. A. 2019. Assembly of long, error-prone reads using repeat graphs. *Nat Biotechnol,* 37**,** 540-546.

KONO, N., ARAKAWA, K. & TOMITA, M. 2012. Validation of bacterial replication termination models using simulation of genomic mutations. *PLoS One,* 7**,** e34526.

KONO, N., TOMITA, M. & ARAKAWA, K. 2018. Accelerated Laboratory Evolution Reveals the Influence of Replication on the GC Skew in Escherichia coli. *Genome Biol Evol,* 10**,** 3110-3117.

LANGMEAD, B. & SALZBERG, S. L. 2012. Fast gapped-read alignment with Bowtie 2. *Nat Methods,* 9**,** 357-9.

LARSEN, N. B., HICKSON, I. D. & MANKOURI, H. W. 2014a. Tus-Ter as a tool to study site-specific DNA replication perturbation in eukaryotes. *Cell Cycle,* 13**,** 2994-8.

LARSEN, N. B., SASS, E., SUSKI, C., MANKOURI, H. W. & HICKSON, I. D. 2014b. The Escherichia coli Tus-Ter replication fork barrier causes site-specific DNA replication perturbation in yeast. *Nat Commun,* 5**,** 3574.

LEE, E. H., KORNBERG, A., HIDAKA, M., KOBAYASHI, T. & HORIUCHI, T. 1989. Escherichia coli replication termination protein impedes the action of helicases. *Proc Natl Acad Sci U S A,* 86**,** 9104-8.

LETUNIC, I. & BORK, P. 2019. Interactive Tree Of Life (iTOL) v4: recent updates and new developments. *Nucleic Acids Res,* 47**,** W256-W259.

LEWIS, J. S., SPENKELINK, L. M., JERGIC, S., WOOD, E. A., MONACHINO, E., HORAN, N. P., DUDERSTADT, K. E., COX, M. M., ROBINSON, A., DIXON, N. E. & VAN OIJEN, A. M. 2017. Single-molecule visualization of fast polymerase turnover in the bacterial replisome. *Elife,* 6.

LI, H., HANDSAKER, B., WYSOKER, A., FENNELL, T., RUAN, J., HOMER, N., MARTH, G., ABECASIS, G., DURBIN, R. & GENOME PROJECT DATA PROCESSING, S. 2009. The Sequence Alignment/Map format and SAMtools. *Bioinformatics,* 25**,** 2078-9.

LIN, Y. S., KIESER, H. M., HOPWOOD, D. A. & CHEN, C. W. 1993. The chromosomal DNA of Streptomyces lividans 66 is linear. *Mol Microbiol,* 10**,** 923-33.

LOBRY, J. R. 1996. Asymmetric substitution patterns in the two DNA strands of bacteria. *Mol Biol Evol,* 13**,** 660-5.

LOUARN, J., PATTE, J. & LOUARN, J. M. 1977. Evidence for a fixed termination site of chromosome replication in Escherichia coli K12. *J Mol Biol,* 115**,** 295-314.

LUSCOMBE, N. M., LASKOWSKI, R. A. & THORNTON, J. M. 1997. NUCPLOT: a program to generate schematic diagrams of protein-nucleic acid interactions. *Nucleic Acids Res,* 25**,** 4940-5.

MEIJERINK, J., MANDIGERS, C., VAN DE LOCHT, L., TONNISSEN, E., GOODSAID, F. & RAEMAEKERS, J. 2001. A novel method to compensate for different amplification efficiencies between patient DNA samples in quantitative real-time PCR. *J Mol Diagn,* 3**,** 55-61.

MIDGLEY-SMITH, S. L., DIMUDE, J. U., TAYLOR, T., FORRESTER, N. M., UPTON, A. L., LLOYD, R. G. & RUDOLPH, C. J. 2018. Chromosomal over-replication in Escherichia coli recG cells is triggered by replication fork fusion and amplified if replichore symmetry is disturbed. *Nucleic Acids Res,* 46**,** 7701-7715.

MITCHELL, A. L., ATTWOOD, T. K., BABBITT, P. C., BLUM, M., BORK, P., BRIDGE, A., BROWN, S. D., CHANG, H. Y., EL-GEBALI, S., FRASER, M. I., GOUGH, J.,

HAFT, D. R., HUANG, H., LETUNIC, I., LOPEZ, R., LUCIANI, A., MADEIRA, F., MARCHLER-BAUER, A., MI, H., NATALE, D. A., NECCI, M., NUKA, G., ORENGO, C., PANDURANGAN, A. P., PAYSAN-LAFOSSE, T., PESSEAT, S., POTTER, S. C., QURESHI, M. A., RAWLINGS, N. D., REDASCHI, N., RICHARDSON, L. J., RIVOIRE, C., SALAZAR, G. A., SANGRADOR-VEGAS, A., SIGRIST, C. J. A., SILLITOE, I., SUTTON, G. G., THANKI, N., THOMAS, P. D., TOSATTO, S. C. E., YONG, S. Y. & FINN, R. D. 2019. InterPro in 2019: improving coverage, classification and access to protein sequence annotations. *Nucleic Acids Res,* 47**,** D351-D360.

MOHANTY, B. K., BAIRWA, N. K. & BASTIA, D. 2009. Contrasting roles of checkpoint proteins as recombination modulators at Fob1-Ter complexes with or without fork arrest. *Eukaryot Cell,* 8**,** 487-95.

MOOLMAN, M. C., TIRUVADI KRISHNAN, S., KERSSEMAKERS, J. W., DE LEEUW, R., LORENT, V., SHERRATT, D. J. & DEKKER, N. H. 2016. The progression of replication forks at natural replication barriers in live bacteria. *Nucleic Acids Res,* 44**,** 6262-73.

MOREAU, M. J., MORIN, I. & SCHAEFFER, P. M. 2010. Quantitative determination of protein stability and ligand binding using a green fluorescent protein reporter system. *Mol Biosyst,* 6**,** 1285-92.

MOREAU, M. J. & SCHAEFFER, P. M. 2012a. Differential Tus-Ter binding and lock formation: implications for DNA replication termination in Escherichia coli. *Mol Biosyst,* 8**,** 2783-91.

MOREAU, M. J. & SCHAEFFER, P. M. 2012b. A polyplex qPCR-based binding assay for protein-DNA interactions. *Analyst,* 137**,** 4111-3.

MOREAU, M. J. & SCHAEFFER, P. M. 2013. Dissecting the salt dependence of the Tus-Ter protein-DNA complexes by high-throughput differential scanning fluorimetry of a GFP-tagged Tus. *Mol Biosyst,* 9**,** 3146-54.

MOREAU, M. J. J., MORIN, I., ASKIN, S. P., COOPER, A., MORELAND, N. J., VASUDEVAN, S. G. & SCHAEFFER, P. M. 2012. Rapid determination of protein stability and ligand binding by differential scanning fluorimetry of GFP-tagged proteins. *RSC Advances,* 2**,** 11892-11900.

MORIN, I., ASKIN, S. P. & SCHAEFFER, P. M. 2011. IgG-detection devices for the Tus-Ter-lock immuno-PCR diagnostic platform. *Analyst,* 136**,** 4815-4821.

MORIN, I., DIXON, N. E. & SCHAEFFER, P. M. 2010. Ultrasensitive detection of antibodies using a new Tus-Ter-lock immunoPCR system. *Mol Biosyst,* 6**,** 1173-5.

MORIN, I., SCHAEFFER, P. M., ASKIN, S. P. & DIXON, N. E. 2012. Combining RNA-DNA swapping and quantitative polymerase chain reaction for the detection of influenza A nucleoprotein. *Anal Biochem,* 420**,** 121-6.

MULCAIR, M. D., SCHAEFFER, P. M., OAKLEY, A. J., CROSS, H. F., NEYLON, C., HILL, T. M. & DIXON, N. E. 2006a. A molecular mousetrap determines polarity of termination of DNA replication in E-coli. *Cell,* 125**,** 1309-1319.

MULCAIR, M. D., SCHAEFFER, P. M., OAKLEY, A. J., CROSS, H. F., NEYLON, C., HILL, T. M. & DIXON, N. E. 2006b. A molecular mousetrap determines polarity of termination of DNA replication in E. coli. *Cell,* 125**,** 1309-19.

MULUGU, S., POTNIS, A., SHAMSUZZAMAN, TAYLOR, J., ALEXANDER, K. & BASTIA, D. 2001. Mechanism of termination of DNA replication of Escherichia coli involves helicase-contrahelicase interaction. *Proc Natl Acad Sci U S A,* 98**,** 9569-74.

NATARAJAN, S., KAUL, S., MIRON, A. & BASTIA, D. 1993. A 27 kd protein of E. coli promotes antitermination of replication in vitro at a sequence-specific replication terminus. *Cell,* 72**,** 113-20.

NEYLON, C., BROWN, S. E., KRALICEK, A. V., MILES, C. S., LOVE, C. A. & DIXON, N. E. 2000. Interaction of the Escherichia coli replication terminator protein (Tus) with DNA: a model derived from DNA-binding studies of mutant proteins by surface plasmon resonance. *Biochemistry,* 39**,** 11989-99.

NEYLON, C., KRALICEK, A. V., HILL, T. M. & DIXON, N. E. 2005a. Replication termination in Escherichia coli: structure and antihelicase activity of the Tus-Ter complex. *Microbiol Mol Biol Rev,* 69**,** 501-26.

NEYLON, C., KRALICEK, A. V., HILL, T. M. & DIXON, N. E. 2005b. Replication termination in Escherichia coli: structure and antihelicase activity of the Tus-Ter complex. *Microbiol. Mol. Biol. Rev.,* 69**,** 501-26.

PANDEY, M., ELSHENAWY, M. M., JERGIC, S., TAKAHASHI, M., DIXON, N. E., HAMDAN, S. M. & PATEL, S. S. 2015. Two mechanisms coordinate replication termination by the Escherichia coli Tus-Ter complex. *Nucleic Acids Res,* 43**,** 5924-35.

PELLETIER, A. J., HILL, T. M. & KUEMPEL, P. L. 1989. Termination sites T1 and T2 from the Escherichia coli chromosome inhibit DNA replication in ColE1-derived plasmids. *J Bacteriol,* 171**,** 1739-41.

PILUSO, S., CASSELL, H. C., GIBBONS, J. L., WALLER, T. E., PLANT, N. J., MILLER, A. F. & CAVALLI, G. 2013. Site-specific, covalent incorporation of Tus, a DNA-binding protein, on ionic-complementary self-assembling peptide hydrogels using transpeptidase Sortase A as a conjugation tooldaggerDedicated to the memory of Joachim H. G. Steinke.double daggerElectronic supplementary information (ESI) available: Further experimental data. See DOI: 10.1039/c3sm00131hClick here for additional data file. *Soft Matter,* 9**,** 6752-6756.

POTEETE, A. R. 2001. What makes the bacteriophage lambda Red system useful for genetic engineering: molecular mechanism and biological function. *FEMS Microbiol Lett,* 201**,** 9-14.

QUINLAN, A. R. & HALL, I. M. 2010. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics,* 26**,** 841-2.

RAY, M., TANG, R., JIANG, Z. & ROTELLO, V. M. 2015. Quantitative tracking of protein trafficking to the nucleus using cytosolic protein delivery by nanoparticle-stabilized nanocapsules. *Bioconjug Chem,* 26**,** 1004-7.

REGEV, T., MYERS, N., ZARIVACH, R. & FISHOV, I. 2012. Association of the chromosome replication initiator DnaA with the Escherichia coli inner membrane in vivo: quantity and mode of binding. *PLoS One,* 7**,** e36441.

REYES-LAMOTHE, R. & SHERRATT, D. J. 2019. The bacterial cell cycle, chromosome inheritance and cell growth. *Nat Rev Microbiol,* 17**,** 467-478.

REYNOLDS, C. R., ISLAM, S. A. & STERNBERG, M. J. E. 2018. EzMol: A Web Server Wizard for the Rapid Visualization and Image Production of Protein and Nucleic Acid Structures. *J Mol Biol,* 430**,** 2244-2248.

ROBINSO N, A. & VAN OIJEN, A. M. 2013. Bacterial replication, transcription and translation: mechanistic insights from single-molecule biochemical studies. *Nat Rev Microbiol,* 11**,** 303-15.

ROECKLEIN, B., PELLETIER, A. & KUEMPEL, P. 1991. The tus gene of Escherichia coli: autoregulation, analysis of flanking sequences and identification of a complementary system in Salmonella typhimurium. *Res Microbiol,* 142**,** 169-75.

ROECKLEIN, B. A. & KUEMPEL, P. L. 1992. In vivo characterization of tus gene expression in Escherichia coli. *Mol Microbiol,* 6**,** 1655-61.

ROTHSTEIN, R., MICHEL, B. & GANGLOFF, S. 2000. Replication fork pausing and recombination or "gimme a break". *Genes Dev,* 14**,** 1-10.

SAMBROOK, J. & RUSSELL, D. W. 2001. *Molecular cloning : a laboratory manual,* Cold Spring Harbor, N.Y., Cold Spring Harbor Laboratory Press.

SANO, T., SMITH, C. L. & CANTOR, C. R. 1992. Immuno-PCR: very sensitive antigen detection by means of specific antibody-DNA conjugates. *Science,* 258**,** 120-2.

SCHAEFFER, P. M., HEADLAM, M. J. & DIXON, N. E. 2005. Protein--protein interactions in the eubacterial replisome. *IUBMB Life,* 57**,** 5-12.

SCHNEIDER, C. A., RASBAND, W. S. & ELICEIRI, K. W. 2012. NIH Image to ImageJ: 25 years of image analysis. *Nat Methods,* 9**,** 671-5.

SEEMANN, T. 2014. Prokka: rapid prokaryotic genome annotation. *Bioinformatics,* 30**,** 2068-9.

SHARMA, B. & HILL, T. M. 1995. Insertion of inverted Ter sites into the terminus region of the Escherichia coli chromosome delays completion of DNA replication and disrupts the cell cycle. *Mol Microbiol,* 18**,** 45-61.

SITARAMAN, K. & CHATTERJEE, D. K. 2011. Protein-protein interactions: an application of Tus-Ter mediated protein microarray system. *Methods Mol Biol,* 723**,** 185-200.

SKOKOTAS, A., HIASA, H., MARIANS, K. J., O'DONNELL, L. & HILL, T. M. 1995. Mutations in the Escherichia coli Tus protein define a domain positioned close to the DNA in the Tus-Ter complex. *J Biol Chem,* 270**,** 30941-8.

SORENSON, A. E. & SCHAEFFER, P. M. 2020. Electrophoretic Mobility Shift Assays with GFP-Tagged Proteins (GFP-EMSA). *Methods Mol Biol,* 2089**,** 159-166.

STAMATAKIS, A. 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics,* 30**,** 1312-3.

SUTANDY, F. X., QIAN, J., CHEN, C. S. & ZHU, H. 2013. Overview of protein microarrays. *Curr Protoc Protein Sci,* Chapter 27**,** Unit 27 1.

TANNER, N. A., HAMDAN, S. M., JERGIC, S., LOSCHA, K. V., SCHAEFFER, P. M., DIXON, N. E. & VAN OIJEN, A. M. 2008. Single-molecule studies of fork dynamics in Escherichia coli DNA replication. *Nat Struct Mol Biol,* 15**,** 998.

THORVALDSDOTTIR, H., ROBINSON, J. T. & MESIROV, J. P. 2013. Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. *Brief Bioinform,* 14**,** 178-92.

TOUCHON, M. & ROCHA, E. P. 2008. From GC skews to wavelets: a gentle guide to the analysis of compositional asymmetries in genomic data. *Biochimie,* 90**,** 648-59.

TROJANOWSKI, D., HOLOWKA, J. & ZAKRZEWSKA-CZERWINSKA, J. 2018. Where and When Bacterial Chromosome Replication Starts: A Single Cell Perspective. *Front Microbiol,* 9**,** 2819.

VASER, R., SOVIC, I., NAGARAJAN, N. & SIKIC, M. 2017. Fast and accurate de novo genome assembly from long uncorrected reads. *Genome Res,* 27**,** 737-746.

VOLKMER, B. & HEINEMANN, M. 2011. Condition-dependent cell volume and concentration of Escherichia coli to facilitate data conversion for systems biology modeling. *PLoS One,* 6**,** e23126.

WALKER, B. J., ABEEL, T., SHEA, T., PRIEST, M., ABOUELLIEL, A., SAKTHIKUMAR, S., CUOMO, C. A., ZENG, Q., WORTMAN, J., YOUNG, S. K. & EARL, A. M. 2014. Pilon: an integrated tool for comprehensive microbial variant detection and genome assembly improvement. *PLoS One,* 9**,** e112963.

WICK, R. R., JUDD, L. M., GORRIE, C. L. & HOLT, K. E. 2017. Completing bacterial genome assemblies with multiplex MinION sequencing. *Microb Genom,* 3**,** e000132.

WIENER, J., KOKOTEK, D., ROSOWSKI, S., LICKERT, H. & MEIER, M. 2020. Preparation of single- and double-oligonucleotide antibody conjugates and their application for protein analytics. *Sci Rep,* 10**,** 1457.

WILCE, J. A., VIVIAN, J. P., HASTINGS, A. F., OTTING, G., FOLMER, R. H., DUGGIN, I. G., WAKE, R. G. & WILCE, M. C. 2001. Structure of the RTP-DNA complex and the mechanism of polar replication fork arrest. *Nat Struct Biol,* 8**,** 206-10.

WILLIS, N. A., CHANDRAMOULY, G., HUANG, B., KWOK, A., FOLLONIER, C., DENG, C. & SCULLY, R. 2014. BRCA1 controls homologous recombination at Tus/Ter-stalled mammalian replication forks. *Nature,* 510**,** 556-9.

WILLIS, N. A., PANDAY, A., DUFFEY, E. E. & SCULLY, R. 2018. Rad51 recruitment and exclusion of non-homologous end joining during homologous recombination at a Tus/Ter mammalian replication fork barrier. *PLoS Genet,* 14**,** e1007486.

WOLANSKI, M., DONCZEW, R., ZAWILAK-PAWLIK, A. & ZAKRZEWSKA-CZERWINSKA, J. 2014. oriC-encoded instructions for the initiation of bacterial chromosome replication. *Front Microbiol,* 5**,** 735.

WORNING, P., JENSEN, L. J., HALLIN, P. F., STAERFELDT, H. H. & USSERY, D. W. 2006. Origin of replication in circular prokaryotic chromosomes. *Environ Microbiol,* 8**,** 353-61.

XU, Z. Q. & DIXON, N. E. 2018. Bacterial replisomes. *Curr Opin Struct Biol,* 53**,** 159-168.

8 APPENDIX

Appendix 1: Supplementary data for chapter 3 - Defining the prototypical DNA replication fork trap in bacteria

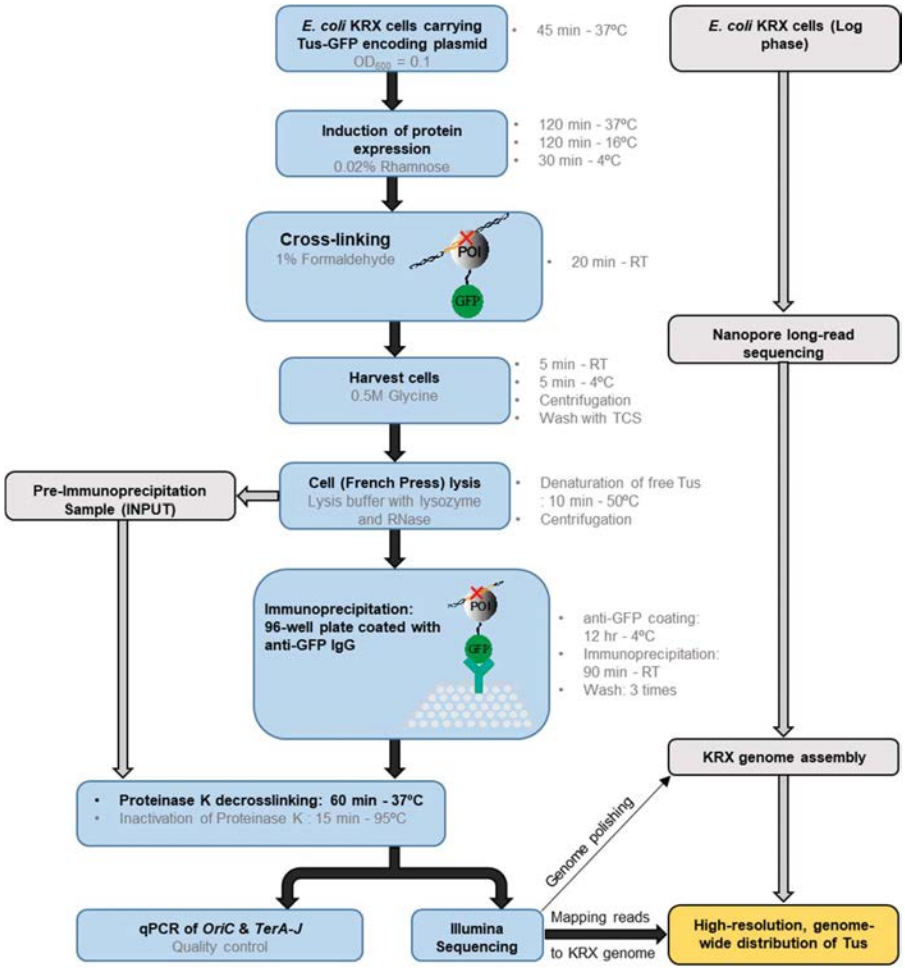


Figure S1: ChIP-qPCR and ChIP-Seq process using a 96-well plate format coated with anti-GFP IgG, and genome assembly for KRX *E. coli* strain. See 3.2 Materials and methods, p64 for detailed procedures.

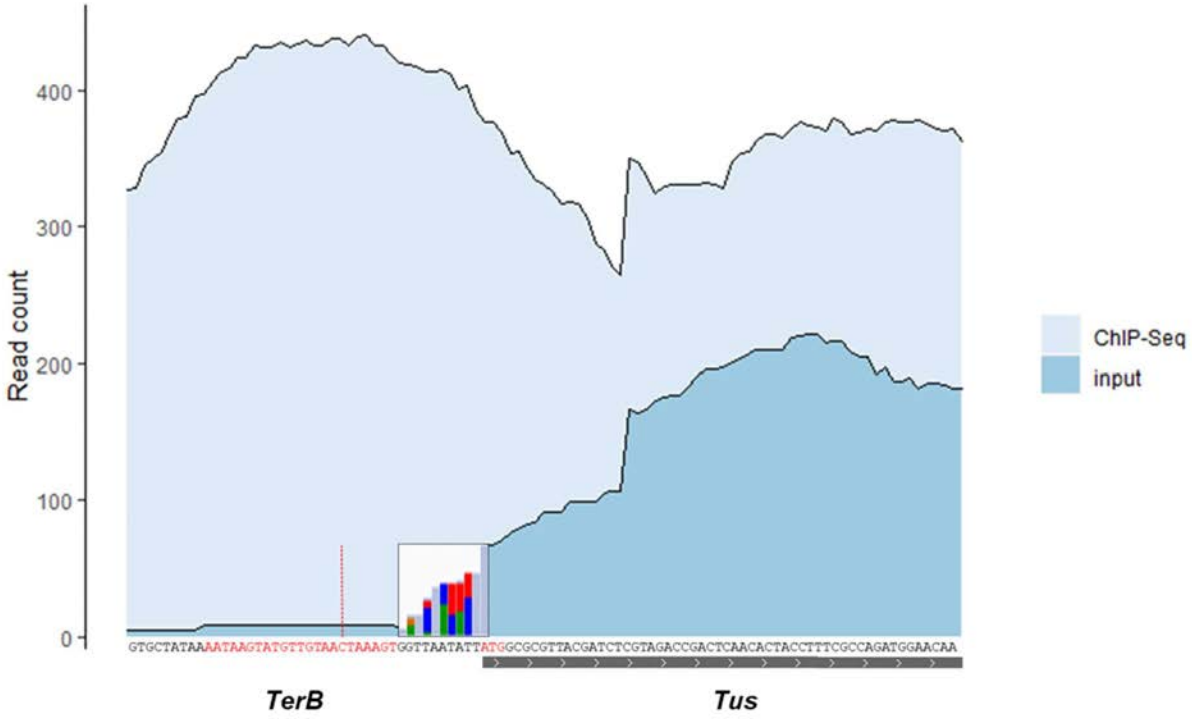


Figure S2: Nucleotide read count at genomic *TerB* and *tus* gene loci for immunoprecipitated DNA (ChIP-Seq) and non-immunoprecititated DNA (Input). The boxed bar chart between *TerB* and *tus* shows an ambiguous 10 nucleotide sequence with partial identity between the plasmid and genome sequences upstream of the start codon (ATG) of the *tus* sequence. The data show that the high read count originating from the plasmid *tus* sequence (i.e. misaligned to the genomic *tus* locus) does not bias the read count at *TerB*. Based on the average *tus* read count the plasmid number is ~100-200/bacteria. The dashed red line indicates the C(6) position in *TerB* critical for the formation of the Tus-*Ter*-lock structure.

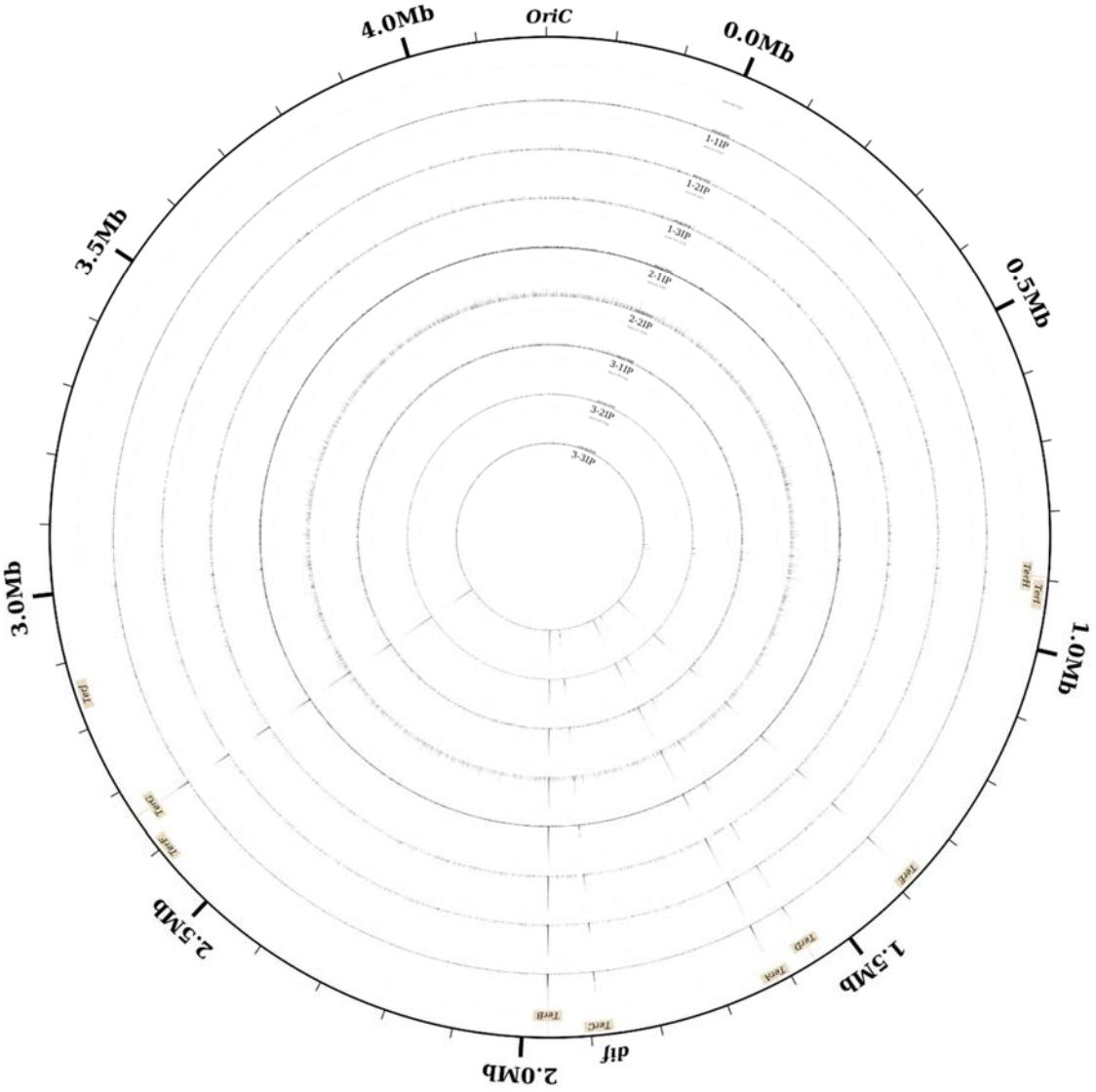


Figure S3: Individually mapped ChIP-Seq coverage for experimental and biological replicates (n = 8). Three biological replicates of immunoprecipitated DNA (ChIP) are shown each consisting of 2-3 technical replicates as indicated. The difference in peak height between different Ter sites is consistent despite varying depths of sequencing between replicates. See Figure 11, p81 for pooled data.

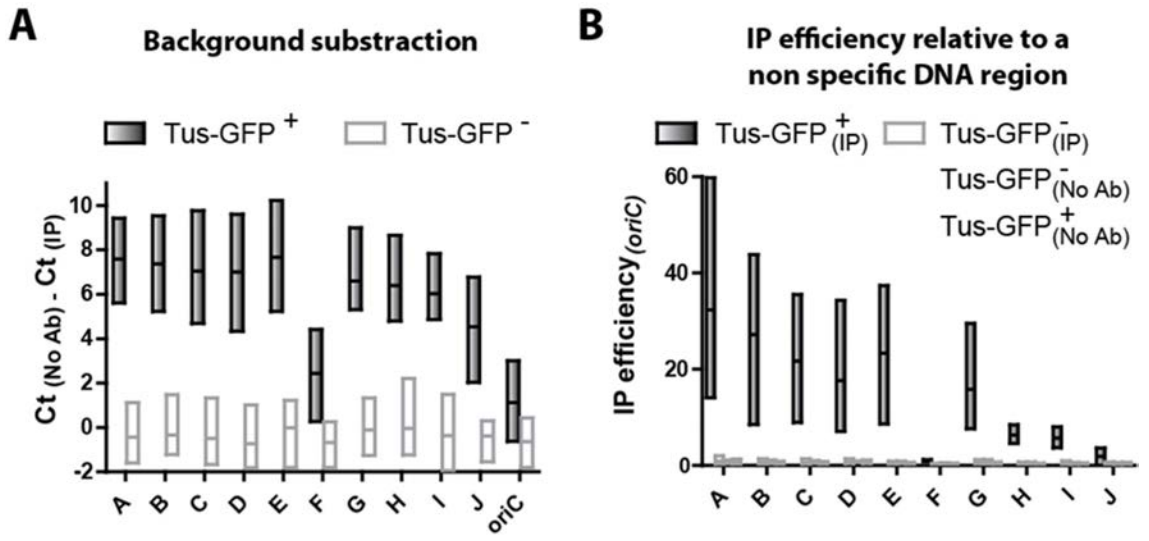
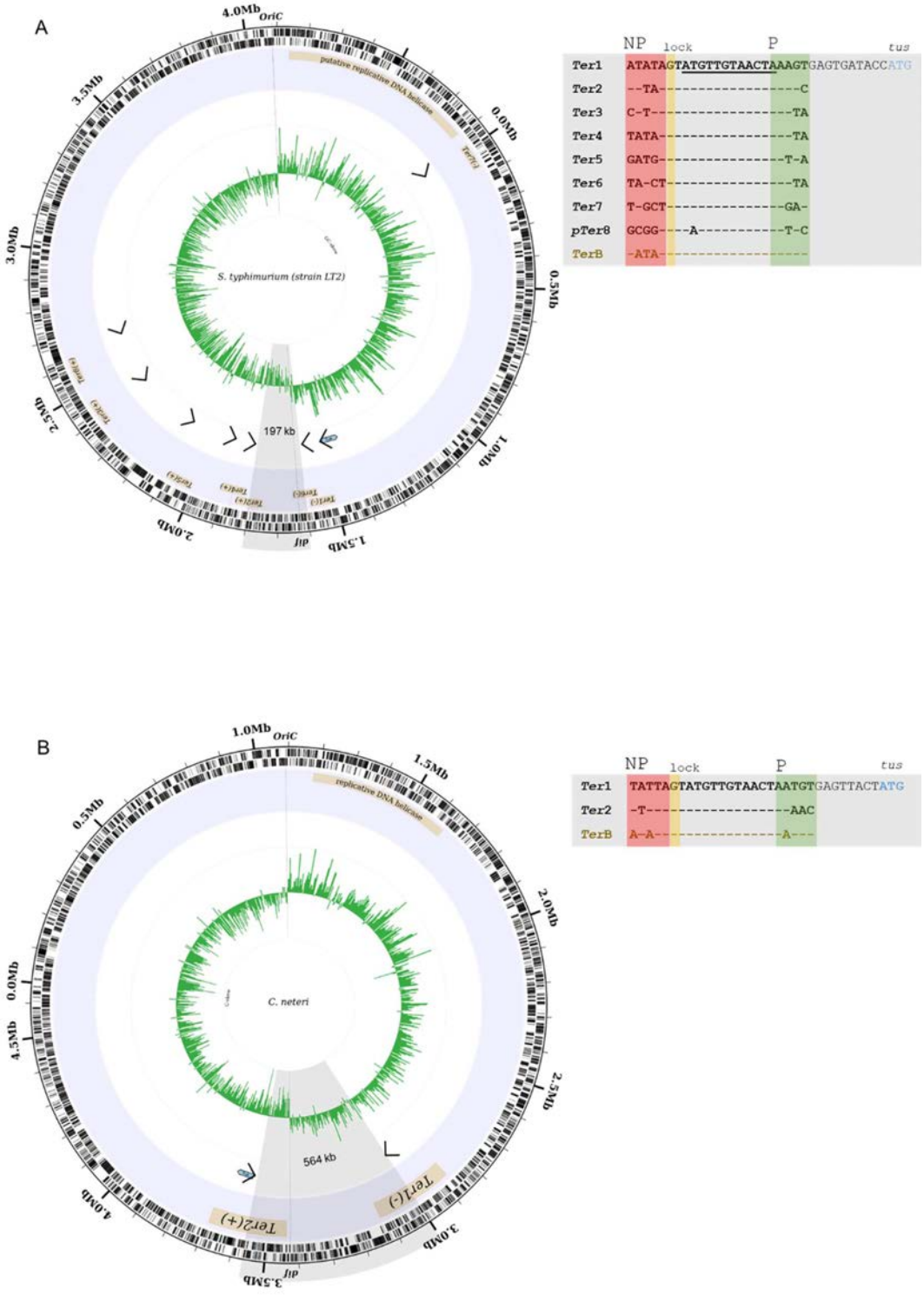
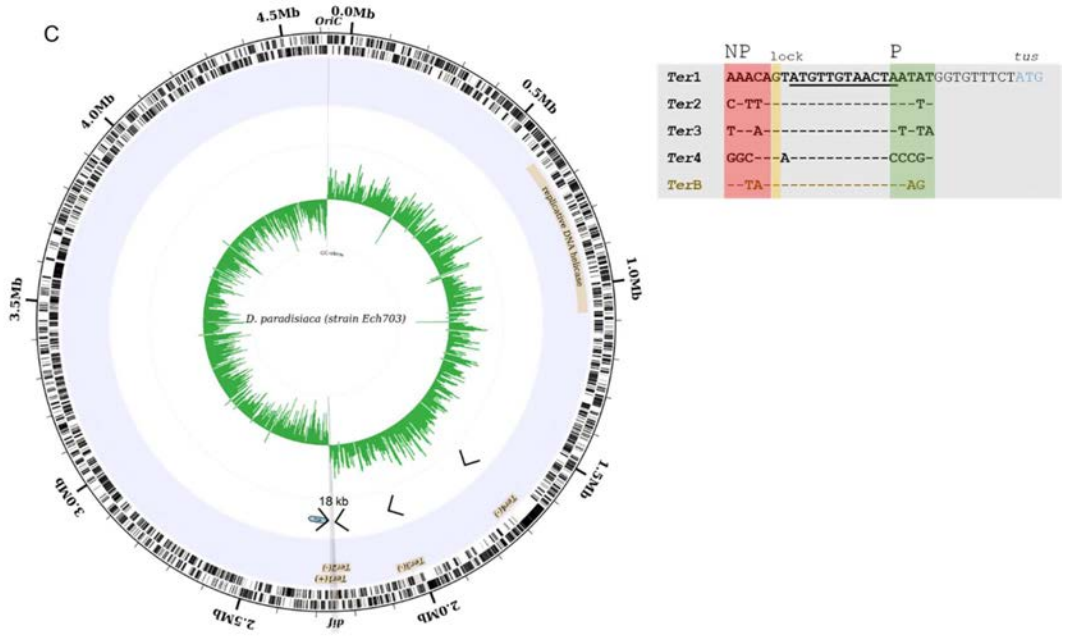
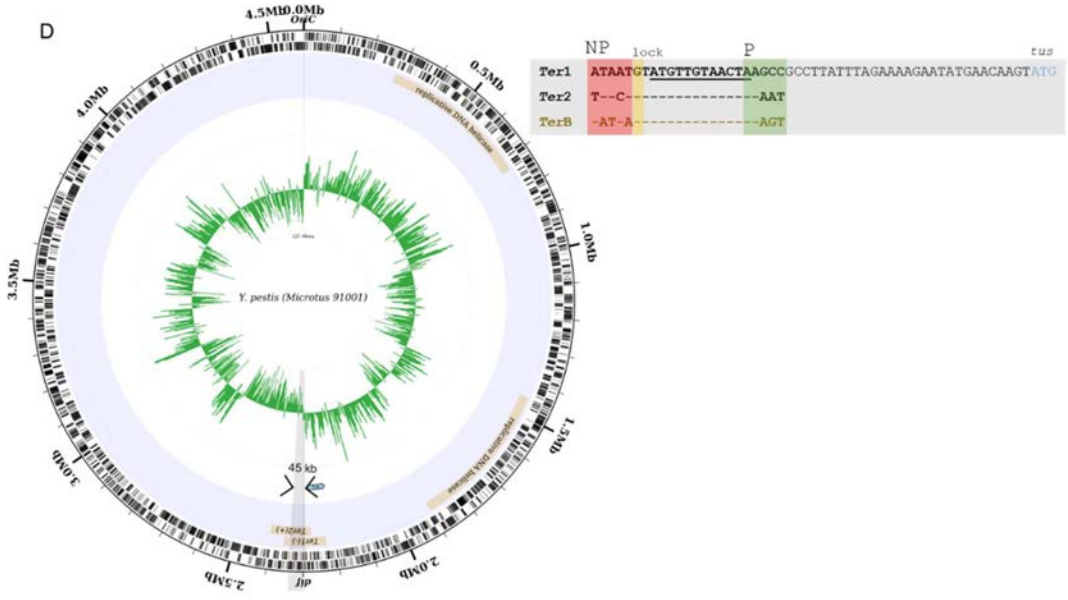


Figure S4: Distribution of Tus-GFP on Ter sites in KRX *E. coli* cells by ChIP-qPCR. (A) Difference in Ct-values between immunoprecipitated DNA (IP) and background control experiments in absence of anti-GFP (No Ab) obtained for Tus-GFP+ and Tus-GFP¬- control KRX cells. (B) IP efficiency of Ter sites relative to a non-specific oriC region obtained for Tus-GFP+ and Tus-GFP- control KRXcells in the presence (IP) or absence of anti-GFP IgG antibody (No Ab). Floating bars represent minimum, maximum and mean values. Reproduced with permission from Moreau, PhD thesis, James Cook University (2013). Thesis can be downloaded at: https://researchonline.jcu.edu.au/31903/1/31903_Moreau_2013_thesis.pdf.


A
4.0Mb
OriC
3.5Mb
0.0Mb
3.0Mb
0.5Mb
2.5Mb
1.0Mb
2.0Mb
1.5Mb
dif
putative replicative DNA helicase
S. typhimurium (strain LT2)
197 kb
NP
lock
P
tus
Ter1 ATATAGTATGTTGTAACTAAAGTGAGTGATACCATG
Ter2 --TA----------C
Ter3 C-T----------TA
Ter4 TATA----------TA
Ter5 GATG----------T-A
Ter6 TA-CT----------TA
Ter7 T-GCT----------GA-
pTer8 GCGG----A----------T-C
TerB -ATA----------
B
1.0Mb
OriC
1.5Mb
0.5Mb
2.0Mb
0.0Mb
4.5Mb
2.5Mb
4.0Mb
3.0Mb
3.5Mb
dif
replicative DNA helicase
C. neteri
564 kb
Ter2(+)
TerI(-)
NP
lock
P
tus
Ter1 TATTAGTATGTTGTAACTAATGTGAGTTACTATG
Ter2 -T----------AAC
TerB A-A----------A---


C
4.5Mb
0.0Mb
OriC
0.5Mb
1.0Mb
1.5Mb
2.0Mb
2.5Mb
3.0Mb
3.5Mb
4.0Mb
D. paradisiaca (strain Ech703)
18 kb
dif
NP
lock
P
tus
Ter1 AAACAGTATGTTGTAACTAATATGGTGTTTCTATG
Ter2 C-TT----------T-
Ter3 T--A----------T-TA
Ter4 GGC---A----------CCCG-
TerB --TA----------AG


D
4.5Mb
0.0Mb
OriC
0.5Mb
1.0Mb
1.5Mb
2.0Mb
2.5Mb
3.0Mb
3.5Mb
4.0Mb
Y. pestis (Microtus 91001)
45 kb
dif
NP
lock
P
tus
Ter1 ATAATGTATGTTGTAACTAAGCCGCCTTATTTAGAAAAGAATATGAACAAGTATG
Ter2 T--C----------AAT
TerB -AT-A----------AGT

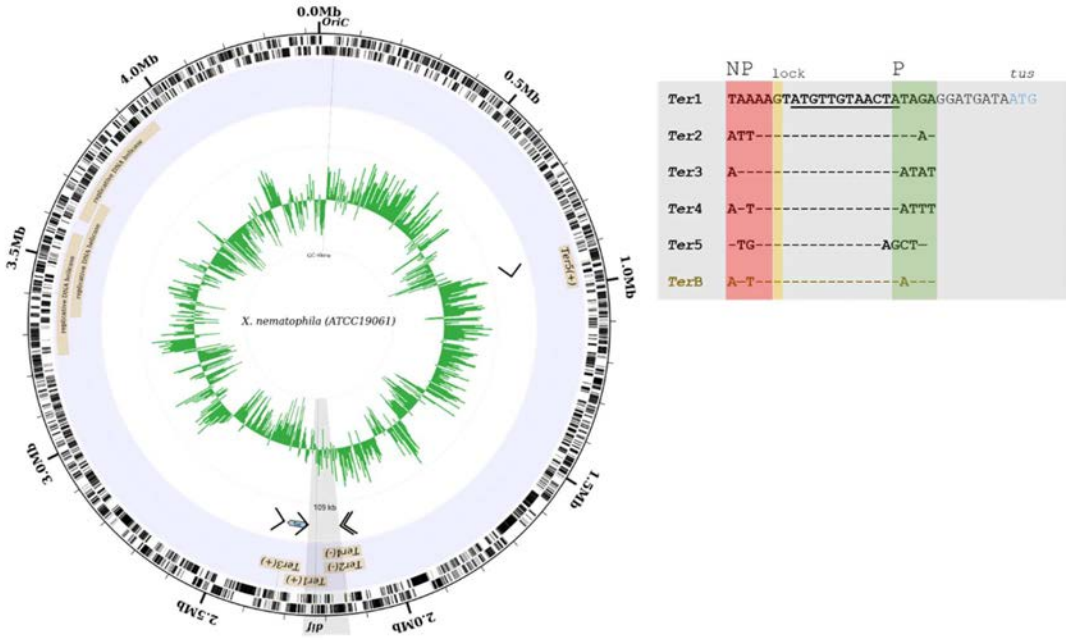


Figure S5: Circular representation of (A) *Salmonella typhimurium, (B) Cedecea neteri, (C) Dickeya paradisiaca, (D) Yersinia pestis* and *(E) Xenorhabdus nematophila* chromosomes and their *Ter* sequences. From the outside of the circle: Forward and reverse genes; genomic locations of identified *Ter* sites involved in DNA replication termination; and GC-skew with a 5000 bp sliding window. In each alignment, *Ter1* represents the *Ter* sequence adjacent to *tus* in the indicated chromosome followed by the RBS and start codon of *tus*. The G(6) complementary to C(6) is highlighted in yellow and the strictly conserved 12 bp core sequence is underlined. The non-permissive face (NP) is highlighted in red and the permissive face (P) is highlighted in green.

Data Table S1: Chromosomal fork trap architecture and classification for selected bacteria.

Family	Species	# of *Ter* sites	Fork trap size (kb)	Vicinal *Ter* sequence	Vicinal *Ter* identity	Identity to *E .coli* Tus (%)	Fork-trap classification	
Enterobacteriacae	*Escherichia coli*	6*	267	AATAAGTATGTTGTAACTAAAGT	-	-	Type II	
	Salmonella typhimurium	8	197	ATATAGTATGTTGTAACTAAAGT	20/23	80	Type II	
	Cronobacter dublinensis	9	47	ATAAAGTATGTTGTAACTAATGT	20/23	65	Type II	
	Atlantibacter hermannii	8	43	AAATAGTATGTTGTAACTAAAGG	20/23	61.8	Type II	
	Shimwellia blattae	5	72	AATAAGCATGTTGTAACTAAAGA	21/23	60	Type II	
	Buttiauxella agrestis	5	297	CTTTAGTATGTTGTAACTAATGG	18/23	60.5	Type II	
	Cedecea netari str. FDAARGOS	3	507	CATTAGTATGTTGTAACTAAAGT	21/23	59.2	Type I	
	Cedecea neteri str. ND14a	2	564	TATTAGTATGTTGTAACTAATGT	20/23	58.6	Type I	
Erwiniaceae	*Pantoea agglomerans*	4	246	TTATAGTATGTTGTAACTATAAA	16/23	55.4	Type I	
Pectobacteriaceae	*Sodalis praecaptivus*	2	111	GTATAGTATGTTGTAACTAATAG	16/23	50.2	Type I	
	Dickeya Paradisiaca	4	18	AAACAGTATGTTGTAACTAATAT	19/23	53.9	Type I	
Hafniaceae	*Edwardsiella tarda*	2	58	CAAAAGTATGTTGTAACTAATAA	18/23	48.3	Type I	
Yersiniaceae	*Yersinia pestis*	2	45	ATAATGTATGTTGTAACTAAGCC	17/23	52.7	Type I	
Morganellaceae	*Xenorhabdus nematophila*	4	109	TAAAAGTATGTTGTAACTATAGA	19/23	46.3	Type I	
	Proteus mirabilis	2	137	TAATTGTATGTTGTAACTAAATA	17/23	50.8	Type I	

Fork trap size corresponds to the distance between the two innermost *Ter* sites of opposite polarity expressed in kb. Underlined bases represent a continuous identical sequence shared between all *Ter* sequences vicinal to *tus* starting at the GC(6) base-pair.

Appendix 2: List of sequences and genomic loci for amplification of *oriC* and *Ter* regions by qPCR (chapter 3)

```
BLASTN 2.10.1+: Matrix: blastn matrix 1 -3,
Gap Penalties: Existence: 5, Extension: 2

Database: final_high_quality_krx_assembly.fasta
           1 sequences; 4,491,350 total letters
```

Query= TerAF

```
Length=24

>contig_2_pilon
Length=4491350

 Score = 48.1 bits (24),  Expect = 2e-07
 Identities = 24/24 (100%), Gaps = 0/24 (0%)
 Strand=Plus/Plus

Query  1        CAACCATTAAACCGATTCGCGGTC  24
                ||||||||||||||||||||||||
Sbjct  1619469  CAACCATTAAACCGATTCGCGGTC  1619492
```

Query= TerAR

```
Length=20

>contig_2_pilon
Length=4491350

 Score = 40.1 bits (20),  Expect = 3e-05
 Identities = 20/20 (100%), Gaps = 0/20 (0%)
 Strand=Plus/Minus

Query  1        AGTTGCGATTTCTCCCCTGG  20
                ||||||||||||||||||||
Sbjct  1619613  AGTTGCGATTTCTCCCCTGG  1619594
```

Query= TerBF

```
Length=22
```

```
>contig_2_pilon
Length=4491350

 Score = 44.1 bits (22),  Expect = 2e-06
 Identities = 22/22 (100%), Gaps = 0/22 (0%)
 Strand=Plus/Plus


Query  1        TTACCTCTGCCTGACACTACGC  22
                ||||||||||||||||||||||
Sbjct  1962934  TTACCTCTGCCTGACACTACGC  1962955
```

Query= TerBR

```
Length=23

>contig_2_pilon
Length=4491350
 Score = 46.1 bits (23),  Expect = 6e-07
 Identities = 23/23 (100%), Gaps = 0/23 (0%)
 Strand=Plus/Minus

Query  1        TGTTGAGTCGGTCTACGAGATCG  23
                |||||||||||||||||||||||
Sbjct  1963056  TGTTGAGTCGGTCTACGAGATCG  1963034
```

Query= TerCF

```
>contig_2_pilon
Length=4491350

 Score = 46.1 bits (23),  Expect = 6e-07
 Identities = 23/23 (100%), Gaps = 0/23 (0%)
 Strand=Plus/Plus

Query  1        CTGCATGTGGCACCTGTTAATGA  23
                |||||||||||||||||||||||
Sbjct  1887884  CTGCATGTGGCACCTGTTAATGA  1887906
```

Query= TerCR

```
Length=20

>contig_2_pilon
Length=4491350

 Score = 40.1 bits (20),  Expect = 3e-05
 Identities = 20/20 (100%), Gaps = 0/20 (0%)
 Strand=Plus/Minus

Query  1        GCTGTACGTCCGTTGTGCTA  20
                ||||||||||||||||||||
Sbjct  1888006  GCTGTACGTCCGTTGTGCTA  1887987
```

Query= TerDF

```
Length=25

>contig_2_pilon
Length=4491350

 Score = 50.1 bits (25),  Expect = 5e-08
 Identities = 25/25 (100%), Gaps = 0/25 (0%)
 Strand=Plus/Plus

Query  1        GGCATGATGTCGCGCtttttttATG  25
                |||||||||||||||||||||||||
Sbjct  1558486  GGCATGATGTCGCGCTTTTTTTATG  1558510
```

Query= TerDR

```
Length=25

>contig_2_pilon
Length=4491350

 Score = 50.1 bits (25),  Expect = 5e-08
 Identities = 25/25 (100%), Gaps = 0/25 (0%)
 Strand=Plus/Minus

Query  1        GGGTATTAAGGAGTATTCCCCATGG  25
                |||||||||||||||||||||||||
Sbjct  1558610  GGGTATTAAGGAGTATTCCCCATGG  1558586
```

Query= TerEF

Length=20

>contig_2_pilon
Length=4491350

 Score = 40.1 bits (20), Expect = 3e-05
 Identities = 20/20 (100%), Gaps = 0/20 (0%)
 Strand=Plus/Plus

```
Query  1        GAAGTCGCCGTCTGGTTTAT  20
                ||||||||||||||||||||
Sbjct  1377410  GAAGTCGCCGTCTGGTTTAT  1377429
```

Query= TerER

Length=20

>contig_2_pilon
Length=4491350

 Score = 40.1 bits (20), Expect = 3e-05
 Identities = 20/20 (100%), Gaps = 0/20 (0%)
 Strand=Plus/Minus

```
Query  1        TACGGCGGAAGTTAATGGTC  20
                ||||||||||||||||||||
Sbjct  1377581  TACGGCGGAAGTTAATGGTC  1377562
```

Query= TerFF

Length=21

>contig_2_pilon
Length=4491350

 Score = 42.1 bits (21), Expect = 8e-06
 Identities = 21/21 (100%), Gaps = 0/21 (0%)

```
 Strand=Plus/Plus

Query  1        CACATCTTCGGGAGTCGGTTC  21
                |||||||||||||||||||||
Sbjct  2596624  CACATCTTCGGGAGTCGGTTC  2596644
```

Query= TerFR

```
Length=22

>contig_2_pilon
Length=4491350

 Score = 44.1 bits (22),  Expect = 2e-06
 Identities = 22/22 (100%), Gaps = 0/22 (0%)
 Strand=Plus/Minus

Query  1        GGTTGAGTGGTAAACGCTGCTG  22
                ||||||||||||||||||||||
Sbjct  2596754  GGTTGAGTGGTAAACGCTGCTG  2596733
```

Query= TerGF

```
Length=20

>contig_2_pilon
Length=4491350

 Score = 40.1 bits (20),  Expect = 3e-05
 Identities = 20/20 (100%), Gaps = 0/20 (0%)
 Strand=Plus/Plus

Query  1        CCAAGCGAGTACCCCACCAG  20
                ||||||||||||||||||||
Sbjct  2656294  CCAAGCGAGTACCCCACCAG  2656313
```

Query= TerGR

```
Length=23

>contig_2_pilon
```

```
Length=4491350

 Score = 46.1 bits (23),  Expect = 6e-07
 Identities = 23/23 (100%), Gaps = 0/23 (0%)
 Strand=Plus/Minus

Query  1        CACGGTTGTATGTTGATCTCCCA  23
                |||||||||||||||||||||||
Sbjct  2656435  CACGGTTGTATGTTGATCTCCCA  2656413
```

Query= TerHF

```
Length=24

>contig_2_pilon
Length=4491350

 Score = 48.1 bits (24),  Expect = 2e-07
 Identities = 24/24 (100%), Gaps = 0/24 (0%)
 Strand=Plus/Plus

Query  1       TGAAGGACAAACTGGAAACGCTGA  24
               ||||||||||||||||||||||||
Sbjct  895054  TGAAGGACAAACTGGAAACGCTGA  895077
```

Query= TerHR

```
Length=20

>contig_2_pilon
Length=4491350

 Score = 40.1 bits (20),  Expect = 3e-05
 Identities = 20/20 (100%), Gaps = 0/20 (0%)
 Strand=Plus/Minus

Query  1       CAGACTACCGCCACCACAAT  20
               ||||||||||||||||||||
Sbjct  895201  CAGACTACCGCCACCACAAT  895182
```

Query= TerIF

```
Length=22
```

```
>contig_2_pilon
Length=4491350

 Score = 44.1 bits (22),  Expect = 2e-06
 Identities = 22/22 (100%), Gaps = 0/22 (0%)
 Strand=Plus/Plus

Query  1       ATTGCTGGAACGGTTGATTGCG  22
               ||||||||||||||||||||||
Sbjct  920804  ATTGCTGGAACGGTTGATTGCG  920825
```

Query= TerIR

```
Length=20
>contig_2_pilon
Length=4491350

 Score = 40.1 bits (20),  Expect = 3e-05
 Identities = 20/20 (100%), Gaps = 0/20 (0%)
 Strand=Plus/Minus

Query  1       CTCGCCGTCTTTACGTAGCA  20
               ||||||||||||||||||||
Sbjct  920921  CTCGCCGTCTTTACGTAGCA  920902
```

Query= TerJF

```
Length=20

>contig_2_pilon
Length=4491350

 Score = 40.1 bits (20),  Expect = 3e-05
 Identities = 20/20 (100%), Gaps = 0/20 (0%)
 Strand=Plus/Plus

Query  1        GACGATACGACGCACCGATG  20
                ||||||||||||||||||||
Sbjct  2849492  GACGATACGACGCACCGATG  2849511
```

Query= TerJR

```
Length=22
>contig_2_pilon
Length=4491350

 Score = 44.1 bits (22),  Expect = 2e-06
 Identities = 22/22 (100%), Gaps = 0/22 (0%)
 Strand=Plus/Minus

Query  1        CTGGTGATGCCGAACATGGAAG  22
                ||||||||||||||||||||||
Sbjct  2849641  CTGGTGATGCCGAACATGGAAG  2849620
```

Query= OriCF

```
Length=22
>contig_2_pilon
Length=4491350

 Score = 44.1 bits (22),  Expect = 2e-06
 Identities = 22/22 (100%), Gaps = 0/22 (0%)
 Strand=Plus/Plus

Query  1        CGCACTGCCCTGTGGATAACAA  22
                ||||||||||||||||||||||
Sbjct  4205199  CGCACTGCCCTGTGGATAACAA  4205220
```

Query= OriCR

```
Length=22
>contig_2_pilon
Length=4491350

 Score = 44.1 bits (22),  Expect = 2e-06
 Identities = 22/22 (100%), Gaps = 0/22 (0%)
 Strand=Plus/Minus

Query  1        CCCTCATTCTGATCCCAGCTTA  22
                ||||||||||||||||||||||
Sbjct  4205313  CCCTCATTCTGATCCCAGCTTA  4205292
```