Live cell studies of bacterial DNA replication, recombination, and degradation

Wiktor, Jakub

DOI
10.4233/uuid:1f384cf1-01c3-4484-a4ca-1ea1227e8bc0

Publication date
2017

Document Version
Final published version

Citation (APA)
Wiktor, J. (2017). *Live cell studies of bacterial DNA replication, recombination, and degradation*. [Dissertation (TU Delft), Delft University of Technology]. https://doi.org/10.4233/uuid:1f384cf1-01c3-4484-a4ca-1ea1227e8bc0

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

Live cell studies of bacterial DNA replication, recombination, and degradation

Jakub M. Wiktor

Propositions

accompanying the dissertation

LIVE CELL STUDIES OF BACTERIAL DNA REPLICATION, RECOMBINATION, AND DEGRADATION

by

Jakub Marek WIKTOR

1. The biological reality of homologous recombination is less romantic than the biophysical picture. *(Chapter 3 of this thesis)*
2. The lack of replication arrest due to DSB damage shows that bacteria grow too fast to reflect on the state of their chromosome. *(Chapter 2 of this thesis)*
3. Because DNA repair is studied in a very artificial way, we don't know much about its real nature. *(Chapter 2 and 3 of this thesis)*
4. Politicians should be obliged to provide a reference section for every statement they make concerning climate change, evolution, and economy.
5. Big data ruined casual web browsing.
6. The way in which molecular biology is taught is more suited for accountants than scientists.
7. The reason why threats of synthetic biology are discussed by professors and not by students, is that the first ones don't go to the lab anymore.
8. The fact that UvrABC complex was discovered in organisms that live in the dark illustrates how misleading scientific nomenclature can be.
9. Unfortunately, breakthrough discoveries are only incidental.
10. Science is great for experiencing adventures, but terrible for establishing a stable career.

These propositions are regarded as opposable and defendable, and have been approved as such by the promotor prof. dr. C. Dekker.

Stellingen

behorende bij het proefschrift

LIVE CELL STUDIES OF BACTERIAL DNA REPLICATION, RECOMBINATION, AND DEGRADATION

door

Jakub Marek WIKTOR

1. De biologische realiteit van homologe recombinatie is minder romantisch dan het biofysische plaatje. *(Hoofdstuk 3, dit proefschrift)*
2. Dat dubbelstrengs DNA breuken geen replicatie blokkade veroorzaken, laat zien dat bacteriën te snel groeien om te reflecteren op de status van hun chromosoom. *(Hoofdstuk 2, dit proefschrift)*
3. Omdat DNA reparatie bestudeerd wordt op een kunstmatige manier weten we niet veel over de echte natuur. *(Hoofdstuk 2 en 3, dit proefschrift)*
4. Politici zouden verplicht moeten zijn om een referentie sectie aan te leveren voor elke bewering over klimaatverandering, evolutie, en economie.
5. Big data heeft ongedwongen surfen op het web verpest.
6. De manier waarop moleculaire biologie wordt onderwezen is meer geschikt voor accountants dan voor biologen.
7. De reden waarom gevaren van synthetische biologie door professoren en niet door studenten worden besproken, is dat professoren niet meer in het lab werken.
8. Het feit dat het UvrABC complex is ontdekt in organismen die in het donker leven illustreert hoe misleidend wetenschappelijke naamgeving kan zijn.
9. Helaas zijn wetenschappelijke doorbraken maar incidenteel.
10. Wetenschap is heel avontuurlijk, maar verschrikkelijk voor het tot stand brengen van een stabiele carrière.

Deze stellingen worden opponeerbaar en verdedigbaar geacht en zijn als zodanig goedgekeurd door de promotor prof. dr. C. Dekker.

Live cell studies of bacterial DNA replication, recombination, and degradation

Jakub Marek Wiktor

LIVE CELL STUDIES OF BACTERIAL DNA REPLICATION, RECOMBINATION, AND DEGRADATION

Proefschrift

ter verkrijging van de graad van doctor
aan de Technische Universiteit Delft,
op gezag van de Rector Magnificus prof. ir. K.C.A.M. Luyben,
voorzitter van het College voor Promoties,
in het openbaar te verdedigen op donderdag 22 juni 2017 om 15:00 uur

door

Jakub Marek WIKTOR

Master of Science in Biotechnology,
University of Gdańsk, Polen
geboren te Gdańsk, Polen.

Dit proefschrift is goedgekeurd door de

promotor: prof. dr. C. Dekker

Samenstelling promotiecommissie:

Rector Magnificus,	voorzitter
Prof. dr. C. Dekker,	Technische Universiteit Delft
Onafhankelijke leden:	
Prof. dr. ir. S. Tans,	Technische Universiteit Delft
Prof. dr. J.H.J. Hoeijmakers,	Erasmus Medical Center Rotterdam
dr. ir. S. J. J. Brouns,	Technische Universiteit Delft
dr. O. Espeli,	Collège de France
Prof. dr. N. H. Dekker,	Technische Universiteit Delft, reservelid
Overige leden:	
Prof. dr. D.J. Sherratt,	University of Oxford
dr. C. Lesterlin,	Université Lyon

Keywords: Recombination, end-resection, replication, double stranded breaks, DNA repair

Printed by: Gildeprint

Front & Back: J. M. Wiktor

Casimir PhD series, Delft-Leiden 2017-22

ISBN 978-90-8593-306-9

An electronic version of this dissertation is available at
http://repository.tudelft.nl/.

CONTENTS

1

INTRODUCTION

Life depends on the integrity of cellular DNA, while the dynamic environment of the cytoplasm and the cellular metabolism puts the genetic material under a constant threat of damage. The extreme accuracy with which the DNA is replicated is not only due to the efficiency of the copying process, but also due to post replicative error-correction mechanisms. Cells can restore the integrity of the genome even after the formation of double stranded breaks (DSB), using a homologous recombination (HR) reaction that prepares the ends of a broken strand, localizes an intact repair template, and copies the information lost during the break. Replication itself can cause DSBs and possibly the most important function of HR is to guard the integrity of DNA during its synthesis. In this chapter, we first review some basic principles of DNA replication and recombination metabolism. Next, we look into the relation between the replication and the formation of DSBs.

1

1.1. INTRODUCTION

Deoxyribonucleic acid, or shortly DNA, is a polymer built out of four bases: adenosines (A), cytosines (C), thymines (T), and guanines (G). The information necessary for the existence and the propagation of life is encoded in the sequence of those nucleotides. The relation between the DNA and genetic information was for long a subject of a scientific dispute, which was eventually solved by the efforts of many brilliant minds. First convincing experimental evidence linking DNA with genes came from laboratory of Frederic Griffith in the early 20th century. Griffith demonstrated that he could isolate, what he called the 'transforming principle', from virulent, dead cells and use that material to transform a batch of living *Streptococcus pneumoniae* bacteria. Transformed bacteria, previously harmless, became virulent and mice infected with them died [1]. Those findings were soon after extended and refined by Oswald Avery, Colin MacLeod, and Maclyn McCarty, who isolated the 'transformed principle' and characterized it as the DNA [2]. Sometime later Alfred Hershey together with Martha Chase used T2 phages with radioactive-labelled DNA to show that the genetic information of the phages is stored within the strands of DNA [3]. The final confirmation came when James Watson and Francis Crick, based on X-ray crystallography work of Rosalind Franklin, described the structure of the DNA double helix in one of the most influential discoveries of modern science [4]. The X-ray structure showed two twisted strands of DNA connected by hydrogen bonds between opposite nucleotides. In the DNA double helix, every adenine is facing a thymine and every cytosine is facing a guanine. Such an arrangement of basepairs assures that both strands contain the same biological message and allows for an elegant copying mechanism, where each of the strands is equally suitable to serve as a replication template for a new duplex. The state of cellular DNA is very dynamic, as instructions encoded within the sequence of basepairs are constantly interpreted, processed, replicated, broken and repaired.

1.2. FLOW OF GENETIC INFORMATION

The principle of the flow of genetic information in the cell was described by Francis Crick as the 'central dogma of molecular biology' [5]. According to the 'central dogma' the information encoded on the DNA serves as a template to produce strands of another four-nucleotide polymer called ribonucleic acid (RNA). The reading of DNA and synthesis of RNA is referred to as the process of transcription and is done by RNA polymerases. The majority of the RNA present in a cell is not coding any protein, but instead is incorporated into complex molecular machines called ribosomes, which are involved in the process of translation: their function is to translate the information encoded in the sequence of the messenger RNA into chains of amino acids that form proteins. In its original form, the central dogma forbids for the information stored within the protein to be translated back into the sequence of RNA or DNA and this hypothesis still largely holds true. In special condition, a class of viral enzymes called reverse transcriptases can synthesize DNA based on the sequence of RNA. Reverse translation does not exist, or is yet to be discovered. In order to pass the genetic information from the old generation to the new generation, DNA has to be replicated and distributed between the new cells. During the replication, each of the strands of DNA serves as a template to synthesize a

brand new strand.

1.3. REPLICATION

For any organism, even those as tiny as an *Escherichia coli* bacterium, replication of the cellular DNA is a tremendous effort. In order to divide, each *E. coli* has to copy its circular, 4.6 million basepairs long chromosome with an extraordinary precision. If failed to do so, errors made during replication would result in catastrophic outcomes, even in the cell death. In the favorable conditions the entire *E. coli* chromosome is copied in about 40-minutes, with an average speed of about 1000 base pairs per second[1]. Yet the error rate of the replication is as low as 1 mistake in 10^9 copied base pairs [6], which means that an average *E. coli* can undergo more 200 divisions before acquiring a single mutation in its genome. This astonishing precision is made possible due to several mechanisms that complement each other and that are acting during and also after the synthesis of new strands of DNA. Many molecular details of those mechanisms have been studied and are well known and characterized today. However, the precise description of the relation between the replication in the busy environment of the cell and the accuracy of the synthesis remains one of the most exciting questions in the field of biology. In the next sections we will describe the most important principles of DNA replication and error correction.

INITIATION OF REPLICATION

The location at which the synthesis of new DNA begins is called the 'origin of replication', or *oriC* in *E. coli*. The double helix of DNA is stabilized by numerous hydrogen bonds and has to be opened to expose stretches of single-stranded DNA, which are the substrates for the synthesis. DNA synthesis is catalyzed by a class of enzymes called DNA polymerases, but before polymerases can begin their action, a replication complex has to be loaded onto the DNA. During the course of evolution multiple mechanisms for opening the DNA at origin did arise, but in this thesis we focus on events specific to the *E. coli* bacteria. *E. coli* has only one origin of replication on its circular chromosome that fires once per replication cycle[2]. Replication machineries established at the *oriC* subsequently travel in opposite directions until they meet again after replicating a similar amount of DNA. The opening of the replication bubble and assembly of replication factories is a well-orchestrated process. In the first step, the initiator proteins, called DnaA, recognize and bind to target sequences, DnaA boxes, spread within the sequence of the *oriC* region. Multiple DnaA proteins thus oligomerize into a protein filament on the DNA Fig. 1.1. Next, using ATP hydrolysis, the filament exerts a torsional force on the DNA that causes an adjacent AT-rich region to melt and the replication bubble opens. Single-stranded DNA of the replication bubble is a substrate onto which a DnaC hexamer helps to loads a replicative helicase, called DnaB. Once helicases are loaded onto the *oriC*, replication of the chromosome begins [8]. It is possible to perturb the early steps of the initiation of replication by targeting a dCas9 protein onto the *oriC* region.

[1] The circular chromosome is replicated by 2 replications factories, each synthesizing 2.3 Mbps.

[2] For bacteria, the time between cell divisions can be shorter than the replication time, because a new round of replication can be initiated before the completion of the previous one. That way daughter cells inherit a chromosome that is already replicating [7].

Once a dCas9 is tightly bound to the origin, interactions between DnaA and *oriC* are inhibited and initiation of replication is stopped. This is the subject of **Chapter 5**.

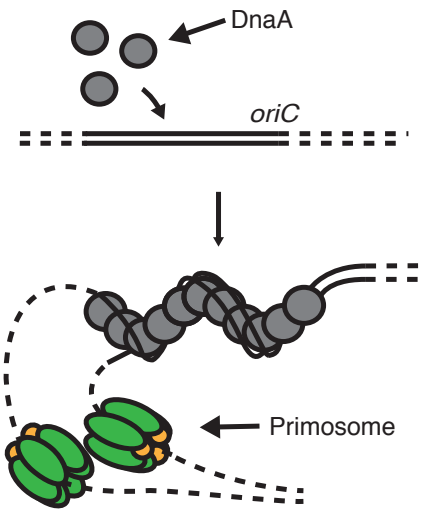


Figure 1.1: **Bacterial initiation of replication.** DnaA proteins bind to DnaA boxes within the *oriC* region and form a filament. This induces a region of ssDNA to be exposed, on which a primosome complex is formed. Adapted from **Chapter 5**.

ELONGATION OF REPLICATION

The location at which the new strands of DNA are created is limited to a specific structure within the inner volume of the cell, where multi-enzyme complexes, called replisomes, synthesize DNA. The synthesis requires the unzipping of the double helix by the replicative helicase, exposing two single strands of the DNA template. This Y-shaped structure is called the replication fork. All known DNA polymerases catalyze the elongation of the growing DNA strand only in the 5' to 3' direction [9] and this limitation had a decisive consequence for the architecture of the replication fork in all kingdoms of life [9]. The synthesis of DNA is limited to 5'-3' direction for pragmatic reasons: the energy used to create the bond between the nucleotide and the growing strand is stored on the triphosphate group. In case of spontaneous dephosphorylation of the energetic bond, the replication machinery can simply wait for another nucleotide. In the opposite case, when the energetic phosphorous group would be placed directly on the strand, the hydrolyzed bond would require to be recharged. The replicative polymerase can sense a mistake in the incorporation of a mispaired nucleotide and remove the error using its 3' to 5' exonuclease activity. A correct nucleotide can be added without the need for reactivation of 3' end of the growing strand. Two strands of the DNA duplex are of opposite polarity, and therefore synthesis of only one of two strands, called the 'leading strand',

can be continuous. The opposite strand, called the 'lagging strand', is synthesized in fragments requiring RNA primers. In *E. coli*, DnaB helicase travels on the lagging strand and forms a complex with three DnaG primases [10]. Once the sufficient amount of ss-DNA is exposed by the helicase, DnaG synthesizes a RNA primer, which allows for DNA polymerase to replicate a stretch of DNA. In *E. coli*, the length of lagging-strand fragments is in the range of 1000 to 2000 nucleotides[3] and those fragments are called the '*Okazaki* fragments'. Bases of the single stranded DNA in the lagging strand are not immediately embedded inside of the DNA double helix, and hence are more vulnerable to DNA damaging agents. This fact may be important for the genetic variability and mutation rates [11].

ERROR CORRECTION

The frequency of errors made by a polymerase during the replication of a new strand is equal to 1 incorrect nucleotide in 10^5 steps, but as we mentioned above, the final error rate measured in genetic studies is only 1 in 10^9. This 4 order-of-magnitude improvement in accuracy is due to action of two additional mechanisms that act during and right after the synthesis of DNA. The first reaction happens during the elongation step and is performed directly by the DNA polymerase. For a replicative DNA polymerase, incorporation of a new nucleotide is inhibited when the previous nucleotide is mis-parried with the template strand. In the case of incorrect base-pairing, the polymerase can proofread the wrong nucleotide, excise it and incorporate a correct one. The dependence of the function of the DNA polymerase on the correct pairing of the previous nucleotide is the reason why the lagging strand synthesis is dependent on a RNA primer[4]. This proofreading activity contributea to 100-fold increase in the fidelity of the replication. After the replication fork has passed, the residual errors on the daughter strand can be removed and corrected by a mismatch repair mechanism (MMR). In *E. coli*, MMR is initiated when a dimer of MutS proteins recognizes, a mismatched base, binds to it, and recruits a dimer of MutL proteins. MutL stimulates MutS to activate MutH, which searches for a near hemi-methylated GATC motif. Once the correct motif is found, MutH nicks the daughter strand whereupon downstream processes remove and correct the mismatch [9]. In *E. coli*, Dam methylase adds a methyl group to the adenine in the GATC motif after the replication. This modification is not instantaneous, and the newly synthesized strand remains unmethylated for a period of time after the replication. The methylation pattern distinguishes the daughter from the mother strand [9]. MMR is the final contribution to reaching the extreme high accuracy of the DNA synthesis.

1.4. DOUBLE STRANDED BREAK REPAIR

Possibly the most dramatic event that can happen to the DNA is a double stranded break (DSB). DSBs arise when the integrity of both DNA strands is interrupted at a single locus. If left unrepaired properly, it can lead to the chromosomal translocations, deletions, or cell death. *E. coli* bacteria are entirely dependent on the homologous recombination (HR) [12] pathway for repair of DSBs, while some organisms also have an alternative

[3] But only 100–200 nucleotides in eukaryotic cells.

[4] Unlike the replicative DNA polymerase, RNA polymerase subunit of the DnaG lacks the proofreading activity, and can synthesize a RNA primer *de novo*.

pathway of nonhomologous end joining (NHEJ). HR uses a homologous template to recover the genetic integrity, whereas NHEJ does not.

NONHOMOLOGOUS END JOINING

During NHEJ, two separated ends of a DSB are connected and religated to each other. In yeast, NHEJ is initiated by binding of a Ku protein heterodimer to the ends DSB. Ku stabilizes the DNA ends, protects them from non-specific degradation and recruits other downstream NHEJ components [13]. Before ligation by the DNA ligase IV complex, some breaks require initial processing to remove nucleotides damaged during the break, or to expose short regions of microhomologies, later used to connect two ends of the DSB. Because some genetic information can be irreversibly lost during the end processing, or added during DNA synthesis, NHEJ is commonly considered an error-prone mechanism [14]. NHEJ seems to provide a rapid, minute-scale response to DSBs [15], but it comes with a threat of genomic rearrangements. In mammalian cells, formation of DSBs on different chromosomes can stimulate frequent chromosomal translocations [16, 17]. Interestingly NHEJ is present in organisms that spend a portion of their lives in quiescent stage [18]. Within bacteria NHEJ is found for instance in *Bacillus subtilis*, that can form haploid spores [19], or in *Mycobacteria* which can enter into non-replicating, dormant states [20, 21]. *E. coli* however lacks genes involved in NHEJ repair and relies solely on homologous recombination to repair DSBs.

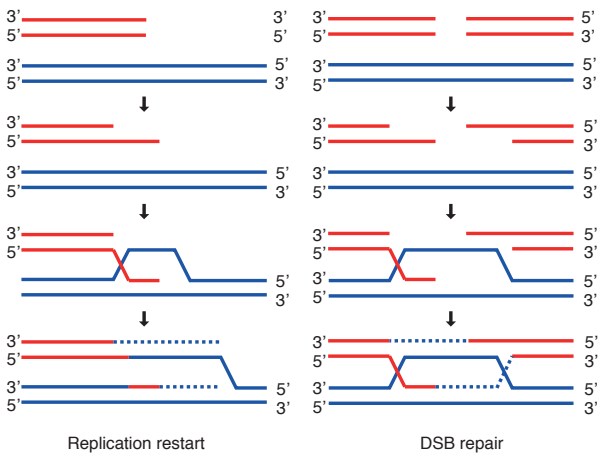


Figure 1.2: **Homologous recombination involvement, in the restart of replication forks (left) and the repair of double stranded breaks (right).** In the first step of HR, 5' ends are resected by RecBCD and 3' overhangs are coated with RecA protein,. Subsequently, ssDNA-RecA filament localises the homologous template. Pairing between the homologies results in the replication and recovery of the information lost during the formation of DSBs (dashed lines).

HOMOLOGOUS RECOMBINATION

Homologous recombination (HR) is more common than NHEJ and can be found in all kingdoms of life [22]. HR uses a homologous sequence to recover the integrity of DNA

duplex that is lost during the formation of a DSB. Typically, the replicated sister chromosome serves as a repair template. HR is also involved in genetic recombination during the crossing-over, which also involves formation of DSBs, but in a programmed manner [23]. Historically, first reports describing HR were linked to sexual reproduction in eukaryotic cells [24], and with bacterial conjugation, which in some way also resembles sexual reproduction [25, 26]. The fact that the same HR mechanism can play a role in DNA repair was realized only later [27–29]. While there are major differences in the repertoire of enzymes involved in the HR, the most fundamental principles of the reaction remain conserved and analogous amongst organisms. HR is initiated by recognition and resection of DNA ends by DNA exonucleases (the RecBCD complex in *E. coli*). End resection produces stretches of 3' ends of single stranded (ss)DNA, which is subsequently coated with recombination proteins (RecA in *E. coli*). 3' ssDNA-RecA can localize and form a complex with the repair template. Next, upon displacing one strand of the repair template duplex, DNA synthesis restores the genetic material lost during DSB and end resection. After the synthesis, the template and the repaired strands are mutually connected by Holliday junctions (HJ), which are resolved by resolvase complexes whereupon recombination is finished (see Fig. 1.2).

END RESECTION BY RECBCD

In *E. coli* the three component RecBCD complex is responsible for detection and processing of 99% of all DSBs [30]. Both RecB and RecD are helicases, with opposite polarities, RecB is 3' to 5' helicase, and RecD is 5' to 3' helicase. RecB is also a nuclease. RecC subunit holds the complex togather. The list of activities that RecBCD can execute makes it the molecular equivalent of a Swiss army knife. It can catalyze somewhat contrasting reactions, which may seem to contradict its involvement in the DNA repair. First, RecBCD can recognize and bind to blunt DNA ends with a very high affinity (K_m ~1nM [31]). Next, *in vitro* experiments show that the RecBCD complex is strong helicase/nuclease, capable of unwinding and degrading even up to 30,000 base pairs during a single binding event [32] with a high speed of 1000 bps^{-1} [33]. It is also a DNA-dependent ATPase, ssDNA endo- and exonuclease and dsDNA exonuclease [34]. All those activities are orderly controlled by an 8-base DNA motif called the χ site (Chi, 5'-GCTGGTGG-3'). When RecBCD approaches the χ from the 3' direction, it drastically changes its biochemical nature. Prior the activation, RecB and RecD are helicases that translocate on the opposite strands of the duplex, and the complex degrades both strands of the DNA, cleaving the 3' end more often than the 5' end. After the recognition of χ, the nuclease activity towards the 3' end is attenuated and RecBCD initiates loading of RecA on the growing 3' ssDNA tail [34]. Steps of end-resection catalysed by the RecBCD are shown in Fig. 1.3. The details of the change happening during χ recognition are not fully understood. Data gathered in genetic experiments showed that when the RecD subunit is absent, recombination is as efficient as for the wild-type complex, but it is not dependent on χ, which suggested that after the χ site recognition RecD is ejected from the complex. However, single-molecule experiments directly showed that the RecD subunit remains associated with the complex [35]. It is possible that the recognition of χ causes structural rearrangements within the complex that are reflected in changes in activities [36]. The reason why bacterial recombination is dependent on χ is still not fully understood. Most probably, χ sites can mark *E. coli* chromosome as a substrate for

recombination and at the same time allowing for RecBCD nuclease activity to degrade foreign DNA. Fragments released during RecBCD resection were shown to be used during acquisition of protospacers used in CRISPR resistance [37].

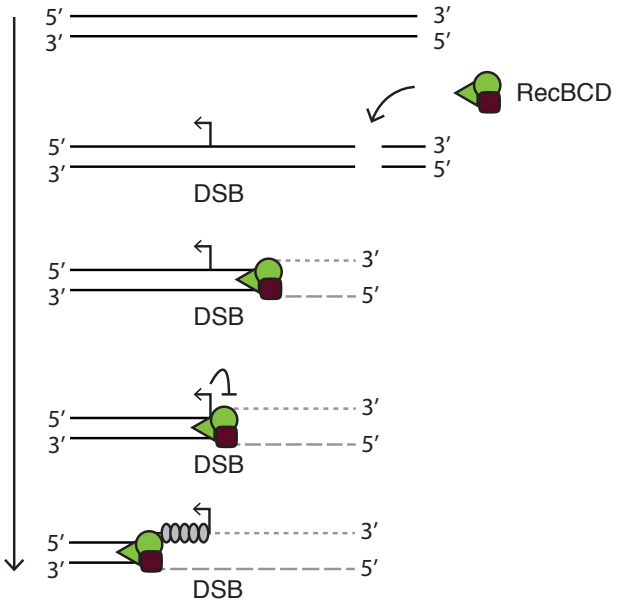


Figure 1.3: **Model of RecBCD end-resection.** The RecBCD complex binds to the ends of a DSB. Initially it acts as a helicase/nuclease that degrades both strands of DNA duplex, with a preference for the 3' end (dashed lines). Recognition of a χ site (shown as an arrow) results in activation of the recombination activity of RecBCD. After χ recognition the, 3' end nuclease activity is attenuated and the RecBCD loads RecA protein on the growing ssDNA 3' tail.

HOMOLOGY SEARCH AND STRAND INVASION

During HR, the repair of broken strand is absolutely depend on the ability to find and interact with an intact repair template elswere in the cells, most likely present on the sister chromosome[5]. This is a task of unquestionable difficulty. From the perspective of the broken locus, the overwhelming majority of the genomic sequence is heterologous and RecA-ssDNA has to discard most of the genome in order to satisfy the accuracy of repair. The search of homology remains a key question that has yet to be answered to fully understand the biology of HR. RecA filament formed on ssDNA can induce donwstream DNA repair processes. RecA bound to ssDNA can cleave LexA transcription repressor, what activates the SOS response [38], inducing the expression of DNA-repair genes, activating enzymes already present in the cell and postponing cell division by the action of

[5]Bacteria are generally considered as haploid, but constant replication provides abundant repair templates.

SulA proteins. Most of our understanding of the search process comes from much simplified *in vitro* models done with purified components. In solution RecA-ssDNA complex can localize and invade a double stranded DNA template. Several mechanisms can potentially facilitate the search process. A filament of RecA-ssDNA can sample multiple sites simultaneously by 'intersegmental contact sampling' [39], engaging in short-lived, weak interaction before establishing a proper interaction. DNA supercoiling increases the local concentration of DNA and has positive effects on homologous pairing [40, 41], possibly because the intersegmental contacts are more frequent when the concentration of template is higher. Once RecA-ssDNA is weakly bound to heterologous sequence, it can slide on the target DNA to probe proximal regions for adequate pairing, which may decrease the search time [42]. The minimal length of ssDNA needed to establish a long-lived complex with the homology is 8 bp [43]. Disregarding shorter homologies highly reduces the complexity of the search. For instance, an average 8-nt motif occurs only 70 times in the *E. coli* genome. It has been recently shown that inside of a cell, fluorescent RecA form very long bundles, spanning the length of almost the entire cell [44]. Such a bundle, if loaded with ssDNA, can probe a very large number of sites in parallel and locate a homology is a reasonably short time. Remarkably, purified RecA binds, but does not hydrolyze ATP in order to locate and invade the homologous strand. The hydrolysis is necessary to dissociate a previously bound strand and is rather used to reset the search [45, 46], confronting the idea that the RecA is a molecular motor that searches for homology. Other factors may be involved in the search process. For instance, in yeast, DSB damage initiates rearrangement in chromatin structure and degradation of a fraction of the histones [47, 48]. Without histones, chromatin in the nucleus is more mobile and the DSB can scan a larger volume during the search. In *E. coli*, RecN, a protein essential for DSB repair, is involved in the cohesion of sister strands [49], but the details of the process are still to be discovered.

RESOLUTION OF JOINT MOLECULES

Ultimately the search and invasion of the repair template leads to the synthesis of DNA and recovery of the genetic information that was damaged due to the DSB. After the synthesis and ligation of DNA, the repaired and template strands are connected by Holliday junctions (HJ), and this final intermediate state of HR has to be resolved. The resolution of these HJs is performed in *E. coli* by the RuvABC complex. RuvA specifically binds to the HJ and recruit RuvB that forms a hexameric helicase that can translocate the junction. RuvC is the enzyme responsible for cutting of the strands of HJ, but its activity is most pronounced at preferred sequence motifs. Translocation of HJ by RuvB helps to expose this preferred sequence and facilitate the resolution [46, 50]. In *E. coli* also RecG protein can translocate HJ [51] and can be involved in the resolution in cells that lack RuvABC [52].

1.5. SOURCES OF DOUBLE STRANDED BREAKS

Cellular physiology is well prepared to respond to damage of the DNA in a timely and accurate manner. It is estimated that each human cell is a subject to about 70,000 different lesions per day [11], 75% of which constitute of relatively easy-to-repair single-stranded breaks. The rest include methylated or oxidased bases that are repaired by the

base excision repair (BER) [53], thymine-dimers and bulky adducts are removed by the nucleotide excision repair (NER) [54], or mispaired bases, corrected by the mismatch repair (MMR)[6] [55]. Moreover, ssDNA breaks can be converted into the highly toxic DSBs when the replication fork attempts to synthesize DNA on such broken template. Most of the studies of DNA repair processes are performed with the use of exogenous sources of breaks, such as UV[7] [56] and gamma [57] irradiation, or reactive chemicals [58]. Use of site-specific endonucleases, like I-SceI [44, 59], or CRISPR-Cas9 [60] is also a common approach to generate localized DSBs within cellular DNA. Recent technological improvements in high-throughput sequencing and single-cell methodology revealed that endogenous DNA damage is very common and cells are subjected to inevitable and frequent assaults to the DNA even while growing in favorable conditions [6, 61, 62]. Formation of DSBs by endogenous sources is illustrated by the phenotype of *E. coli* deficient in components of HR pathway. Cells deprived of *recA* or *recBCD* genes are characterised with growth defects (the growth of *recA* deletion cells is reduced to only 50% when compared to wild-type cells [46, 63]), and show the characteristic hallmarks of DSBs, such as accumulation of linear DNA [63]. Fluorescent reporters coupled to the activation of the SOS response allowed for the direct visualization of cells undergoing DSB repair. As much as 2.1% cells per generation were experiencing those breaks [61]. In next paragraphs we discuss how the replication can contribute to the formation of DSBs during the normal growth.

SINGLE STRANDED BREAKS

A single stranded break, if not repaired within a right window of time, can be converted into a DSB, for example, when encountered by a passing replication fork. Note also that DNA repair mechanisms, such as BER can leave a ssDNA gap as an intermediate of the repair step, which is only later sealed by a ligase [64]. Abortive action of topoisomerase I enzyme, reactive oxygen species, and spontaneous disintegration of deoxyriboses can also contribute to the formation of ssDNA breaks [65]. The lagging strand is synthesized with short *Okazaki* fragments which in *E. coli* are ligated by a ligase Lig. Inactivation of a thermo-sensitive variant of *lig* by the increase of temperature leads to the RecBCD-dependent degradation of the DNA [29]. An elegant example illustrating the formation of a DSB by the replication fork passing through a single stranded break is the use of Flp nickase in the study of yeast DSB response. A modified Flp recombinase that cleaves only one strand of the *FRT* site introduces a site specific single-stranded break in the genome. The growth of yeast cells carrying a Flp-nickase system was shown to be dependent on homologous recombination due to the formation of a DSB by the replication fork passing over a ssDNA break (see Fig. 1.4) [47].

CONFLICTS BETWEEN REPLICATION AND TRANSCRIPTION

Transcription is as crucial for the survival of life as is replication. Both of those processes require to access the genetic template to function, and therefore physical conflicts between them are inevitable. The rate at which RNA polymerase (RNAP) synthesizes RNA is

[6] Mispaired bases are not DNA lesions, but correction by MMR can lead to the formation of DSBs when the replcation machinery attempts to replicate over a MMR repair intermediate.

[7] UV radiation does not induce DSBs directly, but breaks can be created by repair processes, or when the replication machinery encounter region demaged by the UV.

10–20 times lower than the speed at which the replication machinery proceeds along the DNA. Therefore, collisions in both orientations, head-on and head-to-tail, are threathening the stability of the genome. During a head-on collision, the DnaB helicase (which in *E. coli* travels on the lagging strand) can occasionally encounter a RNAP transcribing the same strand in the opposite direction and may lead to an arrest of the replication fork. In bacteria, to reduce the chances of such events the orientation of transcription of highly expressed genes, such as genes coding ribosomal RNA, is aligned with the direction of replication [66]. In such an orientation, the replication fork may just slow down and eventually pass the RNAP barrier without major consequences. Interestingly, in cells in which one of the highly expressed operons was inverted to induce frequent collisions between RNAP and the replication fork, high levels of DNA damage were observed [67]. In *Bacillus subtilis*, head-on collisions increase the mutation rate within the affected region, which possibly can explain why 87% of essential genes are oriented to match the direction of replication [68]. Initially, the head-to-tail collisions were believed to constitute a lesser threat to the genomic stability, but in the light of recent studies also have to be considered as a genome-damaging factor [69, 70]. Replicative helicases and replication restart factors accumulate at the sites of co-directional transcription-translation conflicts in *B. subtilis* cells [69]. Accumulation of those factors is a marker of replication pathologies at the sites of potential conflicts. Particularly dangerous are backtracked RNA polymerases [70, 71] that not only form a stable binding with the DNA, but also may leave a stable DNA:RNA hybrid. If such a hybrid is formed on a leading strand, polymerase can begin extension at the 3' end of the RNA remaining after transcription. The 5' end of the same RNA molecule cannot be ligated to the DNA, leading to a stable ssDNA nick, which is converted to a DSB by the next passage of the replication fork.

ARRESTED REPLICATION FORKS

The arrest and collapse of the replication fork is yet another event leading to the formation of DSBs during physiological growth conditions. Multiple factors may interfere with action of the replisome, causing abnormalities in the fork progression and leading to stalling, arrest, or disassembly of the replication machinery. Starvation [72, 73], nucleotide-depletion [74], and polymerase inhibitors that do not stop helicases [75, 76], DNA lesions, and proteins bound to DNA are examples of factors that may inactivate the replication fork. A lot of our understanding of the physiology of the arrested replication fork comes from the studies of Tus-*Ter* system in *E. coli*. Replication forks on a circular bacterial chromosome face each other at the location opposite to the *oriC*. The progression of replisomes over that region is blocked by a protein Ter, which binds to one of *ter* sequences forming a stable impassable barrier [77, 78]. When a Tus-*Ter* barrier is inserted into an ectopic position on the chromosome, bacteria accumulate DSBs, which is reflected in the induction of the SOS response [79] and accumulation of linear fragments of DNA [80, 81].

Other chromosomal pathologies at the terminus region are evident in the case of *recG* deletion. Chromosome marker frequency analysis shows an increased concentration of DNA at the sites of replication fork collisions [82]. Fork collisions can leave the 3' DNA overhangs that signal to assemble of a replication fork, similarly to what happend durig HR reaction (Fig. 1.2). In wild-type cells RecG removes those 3' DNA ends and inhibits the possible initiation of replication at that region. Monitoring of RecA binding using

CHIP-seq technique shows that RecA protein binding is enriched in the region of terminus, possibly because of frequent conflicts between replisomes [83]. Some evidence suggests that a stalled replication fork, for instance blocked by an array of TetR-*tetO* or LacI-*lacO* is very stable and will not disassemble spontaneously [84, 85], and rather has to be converted to a DSB by a collision with another replication fork approaching from the back [79, 86].

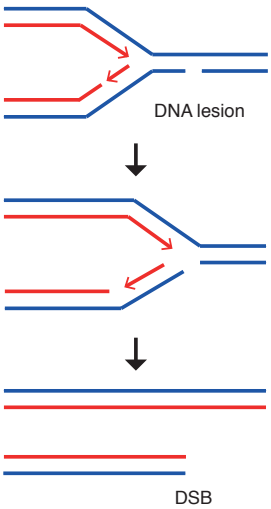


Figure 1.4: **DNA lesions can be a source of double stranded breaks.** In this example, the replication over a ssDNA nick causes DSB.

1.6. IN THIS THESIS

It seems that the evolution of life on this planet repeatedly acknowledges the value of the genetic information, by providing an abundant variety of elegant mechanisms for the repair of damage occurring to DNA. The need for such mechanisms tells us that the state of cellular DNA is under constant threat of disintegration and decay. This may be obvious now, but was previously neglected and it took almost two decades after the discovery of the double-helix structure of DNA to realize that DNA is subject to a range of distinct forms of damage. Double-stranded breaks (DSBs) are a particularly dangerous damage occurring when both of the DNA strands are broken at the same position along the DNA. In order to recover the integrity of the genome, the broken strands must undergo an elaborate process to find a repair template elsewhere in the cell where the same genetic code is imprinted.

In this thesis we focus on aspects of the repair of such lesions and we approach this problem from multiple angles to obtain insights into the processing of breaks and the relation between the homology search and the spatial organization of bacterial genome. Specifically, this PhD thesis addresses the following:

In **Chapter 1** we look into the literature and discuss the connection between replication and formation of double stranded breaks. The spatial organization of the replication forks keeps sister chromosomes in close proximity within the cell. Given the fact that the majority of DSBs occur due to the DNA replication, we propose that this shapes the mechanisms of homology search. We search for evidence for our hypothesis in the arrangement of χ sites on the *E. coli* chromosome, in sister condensation after a period of genotoxic stress, and in interchanges between HR (homologous recombination) and NHEJ (non homologous end joining) as preferred repair pathways in eukaryotic cells in response to the presence of sister templates.

In **Chapter 2** we describe our efforts to dissect the spatiotemporal dynamics of end resection after site-specific induction of DSBs on the chromosome. With a variety of methods, we follow the progression of the RecBCD complexes as they prepare the way for ssDNA-RecA nucleofilament formation that searches for the repair template. With a two color 'fluorescent ruler', we do observe the fast rate of end resection, as it happens at a single DSB site. Differences in processing of the two broken ends resulting from a DSB follow the asymmetry of the χ sites distribution. We find that reduction of end resection, by deletion of the *recD* gene, results in a promiscuous recombination phenotype that allows chromosomal DSBs to be repaired with a plasmid template.

Chapter 3 confronts the popular 'needle in the haystack' model for homology search. Drawing conclusions from multiple failed experiments, we propose that homologous recombination, in spite of previous suggestions, is not undergoing a genome-wide search, but is preferably directed towards the sister. The global search is inhibited by the collective effects of genome organization factors that reduce the mobility of genomic loci. We show that fluorescently labeled DSB regions condense into the middle of the cell in a process that likely assists the search for homology. Our explanation aligns well with the replication being the main cause of chromosomal double stranded breaks.

In **Chapter 4** we use our knowledge gained from studying end resection to establish a novel molecular biology tool for genomic editing. Cells bereft of the *recD* gene are found to be able to integrate long linear fragments of DNA, if those fragments are flanked by regions of homology. We demonstrate the superiority of our system for integration of long fragments. Next we show that *recD* cells can be transformed with genomic DNA isolated from a different strain, which has great technical advantages over the use of P1 phage and phage transductions as well as has implications for horizontal gene transfer among bacteria.

Finally, in **Chapter 5**, we move away from the breaks and take control over the initiation of replication. Using the now renowned CRISPR/dCas9 system, we prevent interactions between the origin of replication and DnaA proteins, which arrests cells in a non-replicating stage of the cell cycle. Utilizing fluorescence microscopy and flow cytometry, we show that arrested cells maintain their growth and elongation. Remarkably, when arrested cells are shifted from 37°C to 42°C, dCas9 binding is inactivated and the previously arrested culture can thrive again.

Homologous recombination is truly a remarkable process and despite a century of scientific studies, some aspects of it are still far from understood. Specifically, the process of the homology search has proven to be notably deceiving. The main goal of this thesis is to convince the reader that the search for homology has to be considered within

the context of the organized cellular genome, where the majority of breaks are spatially localized in close proximity to the replication machinery and where the mobility of DNA is limited by the nuclear organization.

REFERENCES

[1] F. Griffith, *The Significance of Pneumococcal Types,* The Journal of Hygiene **27**, 113 (1928).

[2] O. T. Avery, C. M. MacLeod, and M. McCarty, *Studies on the chemical nature of the substance inducing transformation of pneumococcal types,* The Journal of Experimental Medicine **79**, 137 LP (1944).

[3] A. D. Hershey and M. Chase, *Independent functions of viral protein and nucleic acid in growth of bacteriophage,* The Journal of General Physiology **36**, 39 LP (1952).

[4] J. D. WATSON and F. H. C. CRICK, *Molecular Structure of Nucleic Acids: A Structure for Deoxyribose Nucleic Acid,* Nature **171**, 737 (1953).

[5] F. CRICK, *Central Dogma of Molecular Biology,* Nature **227**, 561 (1970).

[6] J. Jee, A. Rasouly, I. Shamovsky, Y. Akivis, S. R. Steinman, B. Mishra, and E. Nudler, *Rates and mechanisms of bacterial mutagenesis from maximum-depth sequencing,* Nature **534**, 693 (2016), arXiv:NIHMS150003 .

[7] S. Cooper and C. E. Helmstetter, *Chromosome replication and the division cycle of Escherichia coli Br,* Journal of Molecular Biology **31**, 519 (1968).

[8] A. Costa, I. V. Hood, and J. M. Berger, *Mechanisms for Initiating Cellular DNA Replication,* Annual Review of Biochemistry **82**, 25 (2013), arXiv:15334406 .

[9] B. Alberts, A. Johnson, J. Lewis, M. Raff, K. Roberts, and P. Walter, *Molecular Biology of the Cell*, 4th ed. (Gerland Science, 2002).

[10] A. V. Mitkova, S. M. Khopde, and S. B. Biswas, *Mechanism and Stoichiometry of Interaction of DnaG Primase with DnaB Helicase of Escherichia coli in RNA Primer Synthesis,* Journal of Biological Chemistry **278**, 52253 (2003).

[11] A. Tubbs and A. Nussenzweig, *Endogenous DNA Damage as a Source of Genomic Instability in Cancer,* Cell **168**, 644 (2017).

[12] L. Cui and D. Bikard, *Consequences of Cas9 cleavage in the chromosome of Escherichia coli.* Nucleic acids research **44**, 4243 (2016).

[13] A. J. Davis and D. J. Chen, *DNA double strand break repair via non-homologous end-joining,* Translational cancer research **2**, 130 (2013).

[14] J. M. Daley, P. L. Palmbos, D. Wu, and T. E. Wilson, *Nonhomologous end joining in yeast.* Annual review of genetics **39**, 431 (2005).

[15] T. Graham, J. Walter, and J. Loparo, *Two-Stage Synapsis of DNA Ends during Nonhomologous End Joining,* Molecular Cell **61**, 850 (2017).

[16] C. Richardson and M. Jasin, *Frequent chromosomal translocations induced by DNA double-strand breaks,* Nature **405**, 697 (2000).

[17] V. Roukos, T. C. Voss, C. K. Schmidt, S. Lee, D. Wangsa, and T. Misteli, *Spatial dynamics of chromosome translocations in living cells,* Science (New York, N.Y.) **341**, 660 (2013).

[18] S. Shuman and M. S. Glickman, *Bacterial DNA repair by non-homologous end joining,* Nat Rev Micro **5**, 852 (2007).

[19] G. R. Weller, B. Kysela, R. Roy, L. M. Tonkin, E. Scanlan, M. Della, S. K. Devine, J. P. Day, A. Wilkinson, F. d'Adda di Fagagna, K. M. Devine, R. P. Bowater, P. A. Jeggo, S. P. Jackson, and A. J. Doherty, *Identification of a DNA nonhomologous end-joining complex in bacteria.* Science (New York, N.Y.) **297**, 1686 (2002).

[20] E. A. Lamont, J. P. Bannantine, A. Armién, D. S. Ariyakumar, and S. Sreevatsan, *Identification and Characterization of a Spore-Like Morphotype in Chronically Starved Mycobacterium avium Subsp. Paratuberculosis Cultures,* PLOS ONE **7**, e30648 (2012).

[21] N. C. Stephanou, F. Gao, P. Bongiorno, S. Ehrt, D. Schnappinger, S. Shuman, and M. S. Glickman, *Mycobacterial nonhomologous end joining mediates mutagenic repair of chromosomal double-strand DNA breaks.* Journal of bacteriology **189**, 5237 (2007).

[22] G. A. Cromie, J. C. Connelly, and D. R. F. Leach, *Recombination at double-strand breaks and DNA ends: Conserved mechanisms from phage to humans,* Molecular Cell **8**, 1163 (2001).

[23] J. W. Szostak, T. L. Orr-Weaver, R. J. Rothstein, and F. W. Stahl, *The double-strand-break repair model for recombination,* Cell **33**, 25 (1983).

[24] T. H. Morgan, *RANDOM SEGREGATION VERSUS COUPLING IN MENDELIAN INHERITANCE.* Science (New York, N.Y.) **34**, 384 (1911).

[25] J. LEDERBERG and E. L. TATUM, *Gene recombination in Escherichia coli.* Nature **158**, 558 (1946).

[26] E. L. Tatum and J. Lederberg, *Gene Recombination in the Bacterium Escherichia coli,* Journal of Bacteriology **53**, 673 (1947).

[27] M. A. Resnick, *The repair of double-strand breaks in DNA: A model involving recombination,* Journal of Theoretical Biology **59**, 97 (1976).

[28] N. P. Higgins, *The bacterial chromosome* (ASM Press, Washington, D.C, 2005).

[29] A. Kuzminov, *Collapse and repair of replication forks in Escherichia coli.* Molecular microbiology **16**, 373 (1995).

[30] S. C. Kowalczykowski, D. A. Dixon, A. K. Eggleston, S. D. Lauder, and W. M. Rehrauer, *Biochemistry of homologous recombination in Escherichia coli.* Microbiological reviews **58**, 401 (1994).

[31] L. J. Roman and S. C. Kowalczykowski, *Characterization of the helicase activity of the Escherichia coli RecBCD enzyme using a novel helicase assay.* Biochemistry **28**, 2863 (1989).

[32] L. J. Roman, A. K. Eggleston, and S. C. Kowalczykowski, *Processivity of the DNA helicase activity of Escherichia coli recBCD enzyme.* The Journal of biological chemistry **267**, 4207 (1992).

[33] B. Liu, R. J. Baskin, and S. C. Kowalczykowski, *DNA unwinding heterogeneity by RecBCD results from static molecules able to equilibrate,* Nature **500**, 482 (2013).

[34] M. S. Dillingham and S. C. Kowalczykowski, *RecBCD enzyme and the repair of double-stranded DNA breaks.* Microbiology and molecular biology reviews : MMBR **72**, 642 (2008).

[35] K. M. Dohoney and J. Gelles, *[chi]-Sequence recognition and DNA translocation by single RecBCD helicase/nuclease molecules,* Nature **409**, 370 (2001).

[36] M. Wilkinson, Y. Chaban, and D. B. Wigley, *Mechanism for nuclease regulation in RecBCD,* eLife **5**, e18227 (2016).

[37] A. Levy, M. G. Goren, I. Yosef, O. Auster, M. Manor, G. Amitai, R. Edgar, U. Qimron, and R. Sorek, *CRISPR adaptation biases explain preference for acquisition of foreign DNA,* Nature **520**, 505 (2015).

[38] M. Radman, *SOS repair hypothesis: phenomenology of an inducible DNA repair which is accompanied by mutagenesis.* Basic life sciences **5A**, 355 (1975).

[39] A. L. Forget and S. C. Kowalczykowski, *Single-molecule imaging of DNA pairing by RecA reveals a three-dimensional homology search,* Nature **482**, 423 (2012), arXiv:NIHMS150003 .

[40] I. De Vlaminck, M. T. J. van Loenhout, L. Zweifel, J. den Blanken, K. Hooning, S. Hage, J. Kerssemakers, and C. Dekker, *Mechanism of Homology Recognition in DNA Recombination from Dual-Molecule Experiments,* Molecular Cell **46**, 616 (2012).

[41] T. Shibata, C. DasGupta, R. P. Cunningham, and C. M. Radding, *Purified Escherichia coli recA protein catalyzes homologous pairing of superhelical DNA and single-stranded fragments.* Proceedings of the National Academy of Sciences of the United States of America **76**, 1638 (1979).

[42] K. Ragunathan, C. Liu, and T. Ha, *RecA filament sliding on DNA facilitates homology search,* eLife **1**, e00067 (2012).

[43] Z. Qi, S. Redding, J. Y. Lee, B. Gibb, Y. Kwon, H. Niu, W. A. Gaines, P. Sung, and E. C. Greene, *DNA sequence alignment by microhomology sampling during homologous recombination,* Cell **160**, 856 (2015), arXiv:NIHMS150003 .

[44] C. Lesterlin, G. Ball, L. Schermelleh, and D. J. Sherratt, *RecA bundles mediate homology pairing between distant sisters during DNA break repair,* Nature **506**, 249 (2014).

[45] S. C. Kowalczykowski, *Biochemistry of genetic recombination: energetics and mechanism of DNA strand exchange.* Annual review of biophysics and biophysical chemistry **20**, 539 (1991).

[46] S. C. Kowalczykowski, *Initiation of genetic recombination and recombination-dependent replication.* Trends in biochemical sciences **25**, 156 (2000).

[47] V. Dion, V. Kalck, C. Horigome, B. D. Towbin, and S. M. Gasser, *Increased mobility of double-strand breaks requires Mec1, Rad9 and the homologous recombination machinery.* Nature cell biology **14**, 502 (2012).

[48] M. H. Hauer, A. Seeber, V. Singh, R. Thierry, R. Sack, A. Amitai, M. Kryzhanovska, J. Eglinger, D. Holcman, T. Owen-Hughes, and S. M. Gasser, *Histone degradation in response to DNA damage enhances chromatin dynamics and recombination rates,* Nature Structural & Molecular Biology (2017), 10.1038/nsmb.3347.

[49] E. Vickridge, C. Planchenault, C. Cockram, I. G. Junceda, and O. Espéli, *Management of E. coli sister chromatid cohesion in response to genotoxic stress,* Nature Communications **8**, 14618 (2017).

[50] A. J. van Gool, N. M. Hajibagheri, A. Stasiak, and S. C. West, *Assembly of the Escherichia coli RuvABC resolvasome directs the orientation of holliday junction resolution.* Genes & development **13**, 1861 (1999).

[51] C. J. Rudolph, A. L. Upton, G. S. Briggs, and R. G. Lloyd, *Is RecG a general guardian of the bacterial genome?* DNA repair **9**, 210 (2010).

[52] A. A. Mahdi, G. J. Sharples, T. N. Mandal, and R. G. Lloyd, *Holliday junction resolvases encoded by homologous rusA genes in Escherichia coli K-12 and phage 82.* Journal of molecular biology **257**, 561 (1996).

[53] T. Lindahl, *An N-Glycosidase from Escherichia coli That Releases Free Uracil from DNA Containing Deaminated Cytosine Residues,* Proceedings of the National Academy of Sciences of the United States of America **71**, 3649 (1974).

[54] A. Sancar and W. Rupp, *A novel repair enzyme: UVRABC excision nuclease of Escherichia coli cuts a DNA strand on both sides of the damaged region,* Cell **33**, 249 (1983).

[55] R. S. Lahue, K. G. Au, and P. Modrich, *DNA mismatch correction in a defined system,* Science **245**, 160 LP (1989).

[56] C. J. Rudolph, A. L. Upton, and R. G. Lloyd, *Replication fork stalling and cell cycle arrest in UV-irradiated Escherichia coli,* Genes & Development **21**, 668 (2007).

[57] P. M. Krawczyk, T. Borovski, J. Stap, T. Cijsouw, R. ten Cate, J. P. Medema, R. Kanaar, N. a. P. Franken, and J. A. Aten, *Chromatin mobility is increased at sites of DNA double-strand breaks,* Journal of Cell Science **125**, 2127 (2012).

[58] K. Hanada, M. Budzowska, M. Modesti, A. Maas, C. Wyman, J. Essers, and R. Kanaar, *The structure-specific endonuclease Mus81–Eme1 promotes conversion of interstrand DNA crosslinks into double-strands breaks,* The EMBO Journal **25**, 4921 (2006).

[59] M. T. Certo, B. Y. Ryu, J. E. Annis, M. Garibov, J. Jarjour, D. J. Rawlings, and A. M. Scharenberg, *Tracking genome engineering outcome at individual DNA breakpoints,* Nat Meth **8**, 671 (2011).

[60] P. Mali, J. Aach, P. B. Stranges, K. M. Esvelt, M. Moosburner, S. Kosuri, L. Yang, and G. M. Church, *CAS9 transcriptional activators for target specificity screening and paired nickases for cooperative genome engineering,* Nat Biotech **31**, 833 (2013).

[61] J. M. Pennington and S. M. Rosenberg, *Spontaneous DNA breakage in single living Escherichia coli cells,* Nature Genetics **39**, 797 (2007).

[62] A. Canela, S. Sridharan, N. Sciascia, A. Tubbs, P. Meltzer, B. P. Sleckman, and A. Nussenzweig, *DNA Breaks and End Resection Measured Genome-wide by End Sequencing,* Molecular Cell **63**, 898 (2016).

[63] M. M. Cox, M. F. Goodman, K. N. Kreuzer, D. J. Sherratt, S. J. Sandler, and K. J. Marians, *The importance of repairing stalled replication forks,* Nature **404**, 37 (2000).

[64] S. Uphoff, R. Reyes-Lamothe, F. G. de Leon, D. J. Sherratt, and A. N. Kapanidis, *Single-molecule DNA repair in live bacteria,* Proceedings of the National Academy of Sciences **110**, 8063 (2013).

[65] B. S. Abrahams and D. H. Geschwind, *Advances in autism genetics: on the threshold of a new neurobiology.* Nature reviews. Genetics **9**, 341 (2008).

[66] A. Srivatsan, A. Tehranchi, D. M. MacAlpine, and J. D. Wang, *Co-Orientation of Replication and Transcription Preserves Genome Integrity,* PLOS Genetics **6**, e1000810 (2010).

[67] J. D. Wang, M. B. Berkmen, and A. D. Grossman, *Genome-wide coorientation of replication and transcription reduces adverse effects on replication in Bacillus subtilis,* Pnas **104**, 5608 (2007).

[68] S. Paul, S. Million-Weaver, S. Chattopadhyay, E. Sokurenko, and H. Merrikh, *Accelerated gene evolution through replication-transcription conflicts,* Nature **495**, 512 (2013).

[69] H. Merrikh, C. Machon, W. H. Grainger, A. D. Grossman, and P. Soultanas, *Co-directional replication-transcription conflicts lead to replication restart,* Nature **470**, 554 (2011).

[70] D. Dutta, K. Shatalin, V. Epshtein, M. E. Gottesman, and E. Nudler, *Linking RNA polymerase backtracking to genome instability in E. coli,* Cell **146**, 533 (2011), arXiv:NIHMS150003 .

[71] A. Helmrich, M. Ballarino, E. Nudler, and L. Tora, *Transcription-replication encounters, consequences and genomic instability,* Nat Struct Mol Biol **20**, 412 (2013).

[72] Y. Katou, Y. Kanoh, M. Bando, H. Noguchi, H. Tanaka, T. Ashikari, K. Sugimoto, and K. Shirahige, *S-phase checkpoint proteins Tof1 and Mrc1 form a stable replication-pausing complex,* Nature **424**, 1078 (2003).

[73] J. A. Tercero and J. F. X. Diffley, *Regulation of DNA replication fork progression through damaged DNA by the Mec1/Rad53 checkpoint,* Nature **412**, 553 (2001).

[74] K. Labib and B. Hodgson, *Replication fork barriers: pausing for a break or stalling for time?* EMBO reports **8**, 346 LP (2007).

[75] M. Pacek, A. V. Tutter, Y. Kubota, H. Takisawa, and J. C. Walter, *Localization of MCM2-7, Cdc45, and GINS to the Site of DNA Unwinding during Eukaryotic DNA Replication,* Molecular Cell **21**, 581 (2006).

[76] J. Walter and J. Newport, *Initiation of Eukaryotic DNA Replication: Origin Unwinding and Sequential Chromatin Association of Cdc45, RPA, and DNA Polymerase α,* Molecular Cell **5**, 617 (2000).

[77] M. D. Mulcair, P. M. Schaeffer, A. J. Oakley, H. F. Cross, C. Neylon, T. M. Hill, and N. E. Dixon, *A Molecular Mousetrap Determines Polarity of Termination of DNA Replication in E. coli,* Cell **125**, 1309 (2006).

[78] T. M. Hill, J. M. Henson, and P. L. Kuempel, *The terminus region of the Escherichia coli chromosome contains two separate loci that exhibit polar inhibition of replication.* Proceedings of the National Academy of Sciences of the United States of America **84**, 1754 (1987).

[79] B. Sharma and T. M. Hill, *Insertion of inverted Ter sites into the terminus region of the Escherichia coli chromosome delays completion of DNA replication and disrupts the cell cycle,* Molecular Microbiology **18**, 45 (1995).

[80] H. Bierne, S. D. Ehrlich, and B. Michel, *Deletions at stalled replication forks occur by two different pathways.* The EMBO Journal **16**, 3332 (1997).

[81] B. Michel, S. D. Ehrlich, and M. Uzest, *DNA double-strand breaks caused by replication arrest.* The EMBO Journal **16**, 430 (1997).

[82] C. J. Rudolph, A. L. Upton, A. Stockum, C. A. Nieduszynski, and R. G. Lloyd, *Avoiding chromosome pathology when replication forks collide,* Nature **500**, 608 (2013).

[83] C. A. Cockram, M. Filatenkova, V. Danos, M. El Karoui, and D. R. F. Leach, *Quantitative genomic analysis of RecA protein binding during DNA double-strand break repair reveals RecBCD action in vivo.* Proceedings of the National Academy of Sciences of the United States of America **112**, E4735 (2015).

[84] B. T. I. Payne, I. C. van Knippenberg, H. Bell, S. R. Filipe, D. J. Sherratt, and P. McGlynn, *Replication fork blockage by transcription factor-DNA complexes in Escherichia coli.* Nucleic acids research **34**, 5194 (2006).

[85] C. Possoz, S. R. Filipe, I. Grainge, and D. J. Sherratt, *Tracking of controlled Escherichia coli replication fork stalling and restart at repressor-bound DNA in vivo.* The EMBO journal **25**, 2596 (2006).

[86] V. Bidnenko, S. Ehrlich, and B. Michel, *Replication fork collapse at replication terminator sequences,* The EMBO Journal **21**, 3898 (2002).

2

EXTENSIVE AND ASYMMETRIC END-RESECTION BY RECBCD IN VIVO DURING DOUBLE STRANDED BREAK REPAIR

The formation of 3' single-stranded DNA overhands is a first and essential step during homology-directed double stranded break (DSB) repair. In Escherichia coli this step is primarily performed by a complex molecular machine called RecBCD. To address the question how the process of resection is happening within a living-cell we developed and used a fluorescent reporter to directly observe resection of individual, inducible DSBs. With this assay we demonstrated that during the end-resection RecBCD degrades tens of kilobases of DNA with the approximate speed of 1000 bps $^{-1}$, preferentially in the direction of terminus of replication. Next, we confirmed our finding with a qPCR assay. The deletion or recD gene drastically reduced the length of resection. Shortened length of the resection allowed for recombination with short, ectopic homologies during DSB repair and increased efficiency of horizontal gene transfer between strains. Altogether, in this works we visualised and measured the end-resection by the RecBCD in live cells and established a relation between the length of end-resection and the choice of homologous recombination template.

2.1. INTRODUCTION

Double stranded breaks (DSB) are exceptionally toxic DNA damage events due to the simultaneous loss of integrity of both strands of the DNA duplex. If left unrepaired, or when repaired incorrectly, can have dramatic consequences such as cell death, deleterious mutations, chromosomal translocations, or genomic rearrangements [1]. Double stranded breaks can be repaired with a template-independent non-homologous end joining (NHEJ) reaction, or by the homologous recombination (HR), where the information is recovered from an intact, homologous template. While NHEJ is present only in a limited number of organisms, belonging to eukaryotes, archaea, and some phyla of bacteria [2], HR seems to be common in all domains of life [3]. The most fundamental steps in HR share a great resemblance among mechanisms found in different organisms. Findings established in simple model organism, such as budding yeast, or bacteria, have been successfully extrapolated and helped to understand processes found in mammalian cells [4].

In *Escherichia coli*, the DSB repair is initiated by binding of the complex molecular machine: RecBCD nuclease/helicase to the ends of a broken double stranded DNA (dsDNA). Upon this binding two molecular motors RecD, and RecB translocate on the DNA with opposite polarity, where RecD is the dominant motor. The activity of the RecBCD is regulated by an asymmetric octamer sequence called χ (Chi = *c*rossover *h*otspot *i*nstigator = 5'-GCTGGTGG-3', Fig. 2.1a) [5]. After recognition of a χ RecD motor is inactivated and the complex initiates recombination, by producing a 3' ssDNA overhangs coated with RecA recombination protein [6]. The nuclease activity of RecBCD *in vitro* is dependent on the ratio between the concentration of Mg^{2+} and ATP. When Mg^{2+} is in excess over ATP, RecBCD degrades both strands of the DNA duplex, with the preference for 3' end, and the recognition of χ attenuates degradation of 3' end, but not of 5' end [7]. Alternatively, when the ATP is in excess over Mg^{2+}, RecBCD acts as a simple helicase and only after recognition of χ produces a single nick within the 3' strand [8]. The question which reaction occurs in vivo remained unanswered [9]. Next, the single-stranded DNA/RecA nucleofilament created by the RecBCD undergoes a search for suitable repair template and the repair process continues [10]. The main function of HR in *E. coli* seems to be reactivation of collapsed replication forks using sister chromosome as a template [11, 12]. The majority of chromosomal χ sites is oriented to activate RecBCD molecules translocating in the direction of *oriC* (Fig. 2.1a). Interestingly, it was also shown that the search for repair homology during DSB in *E. coli* is directed towards replicated sister chromosome [13].

While great efforts were made to elucidate the mechanism of RecBCD activity, the direct observation of the end-resection process in vivo is still lacking. It is of particular interest to measure the dynamics and lengths of the resection in live cells. *In vitro* RecBCD complex extraordinarily fast and processive, but it is unknown whether it also holds true in its natural environment of the cell. Most of chromosomal χ sites are oriented to be recognised by a RecBCD translocating towards the direction of *oriC*, therefore one could expect to observe uneven processing of two of the ends of a DSB (Fig. 2.1a). Here we observe the end-resection as it happens in live cells with fluorescent microscopy and validate our findings with qPCR assay. Our study shows that the resection *in vivo* is very fast, broad, and asymmetric, with the preference for *terminus*-end of the DSB. When

the RecD motor is removed we observed that the length of the resection was drastically reduced and we show that this shortened resection has implications on the choice of template homology during the repair. We also find that inactivation of the long resection may influence horizontal gene transfer.

2.2. RESULTS

FLUORESCENT REPORTERS OF END-RESECTION IN LIVING CELLS.

To study the dynamics of end-resection after DSB formation in live *E. coli* cells, we developed an *in vivo* fluorescent reporter system consisting of three components: an I-SceI inducible cut site [14] integrated at the *codA* locus, a *parSMt1*/ParBMt1-YFP [15] marker placed 1.5 kb from the I-SceI recognition site, and a *parSP1*/ParBP1-mCherry [16] marker placed at variable distances from the break (Fig. 2.1a). The I-SceI cut site was flanked on each end by 2 χ sites (Fig. 2.6) to induce RecBCD recombination activity [17]. The I-SceI coding gene [18] fused to a LAA degradation tag [19] was placed on a low-copy pSC101-origin plasmid under the control of an anhydrotetracycline (aTc) inducible promoter. Genetic integrations and expression of ParB proteins, did not affect chromosome organization and segregation, as the fluorescent microscopy analysis showed native positioning of the integrated sites, comparable to previously published studies (Fig. 2.7) [20].

Induction of the DSB created free dsDNA ends, which are recognized and processed by the RecBCD complex [17]. RecBCD is a helicase-nuclease that can process very large stretches of DNA [21, 22]. We hypothesized that the degradation of a DNA fragment containing the *parS* site will cause the dissociation of the ParB proteins and consequently a loss of a focus Fig. 2.1b), and that the loss of focus can be used as an *in vivo* reporter of the dynamics of end-resection.

Induction of I-SceI by a 15-minute pulse of aTc, followed by a wash in a media lacking aTc led to formation and processing of DSBs. Analysis of fluorescent microscopy images showed growing proportion of cells without any YFP foci. The proportion of induced cells increased rapidly after the induction of DSBs and after 60 minutes remained stable at about 30% for next consecutive time-points Fig. 2.1c, red squares). In the non-induced culture almost all cells contained at least one YFP focus throughout the time of an experiment, showing no traces of DSB damage. A short pulse of induction, together with a LAA-degradation tag linked to I-SceI allowed us to trigger formation of DSBs in a limited time-window, in a fraction of cells, allowing for observation of advancement of end-resection in a synchronized culture.

I-SCEI CREATES STABLE DSBS.

Formation of a DSB induces a SOS response in bacteria, which leads to inhibition of cell division and the increase of cell lengths [23]. Size analysis of an unperturbed *E. coli* culture shows time-stable distributions, with a mean cell length distribution of 3.7 ± 1.7 μm (pooled data from all time points, n = 12575 cells, Fig. 2.1d, characteristic to cells grown in M9 supplemented medium [24]. In the induced culture elongation of cells was visible and followed the dynamics of disappearance of YFP foci. Growing fraction of SOS-induced cells was pronounced 90 minutes after the induction of DSB and slowly increased in consecutive time-points (Fig. 2.1e). Fitting log-normal distribution to pooled

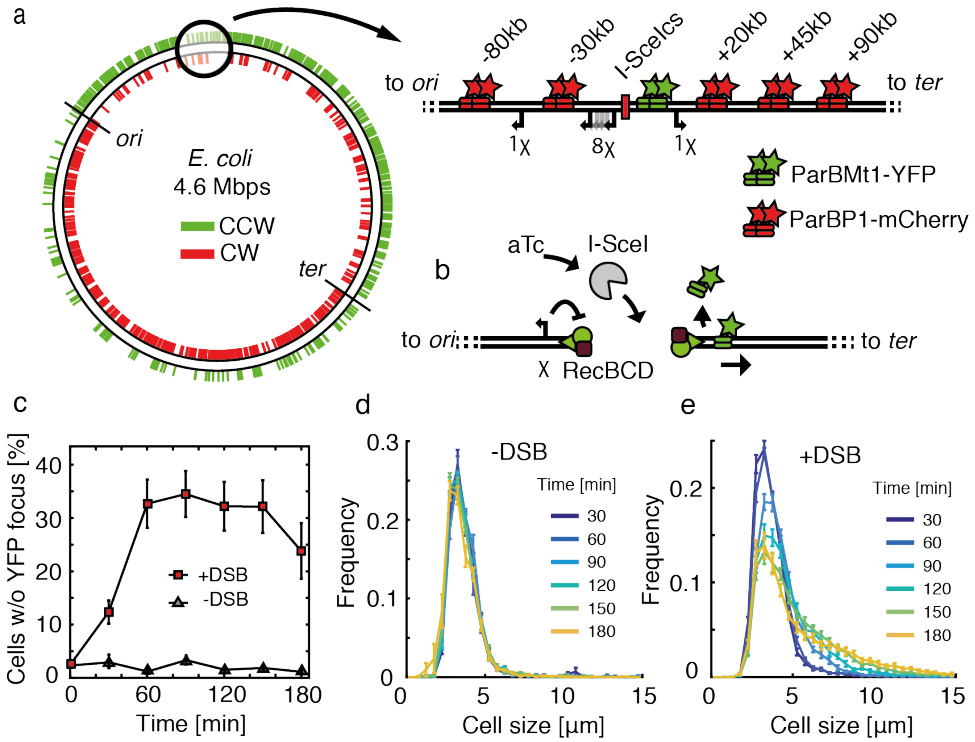


Figure 2.1: **Double-stranded DNA break processing in E. coli.** (a) Schematic showing the asymmetric distribution of χ sites plotted on a circular representation of E. coli chromosome. Sites oriented to activate RecBCD translocating in the counterclockwise direction are shown in green; in clockwise direction are shown in red. Positions of *oriC* and terminus region are indicated. The top right panel shows a zoomed in schematic of region into which the I-SceI cut site (I-SceIcs) was integrated. Green stars indicate position of parSPMT1 to which ParBMT1-YFP binds, red stars show the position of parSP1 to which ParBP1-mCherry binds. Distance from the I-SceIcs is shown on top. Arrows show positions of chromosomal χ sites that are oriented to activate RecBCD during translocation from the I-SceIcs. (b) Schematic representing the experimental procedure. aTc is inducing the expression of I-SceI from an aTc-inducible promoter. I-SceI induces the site-specific DSB that is recognised and processed by RecBCD complexes. RecBCD end-resection is controlled by a correctly oriented χ site (shown on left). Translocation of RecBCD and degradation of parS sites results in displacement of fluorescent ParB proteins. (c) Percentage of cells in the population that lost YFP focus after induction of DSBs (red squares, mean ± SEM, n = 15) and in the control culture (green triangles, mean ± SEM, n = 4). (d) Distribution of cells sizes (mean ± SEM, n = 4) in the population in which DSBs were not induced. (e) Distribution of cells sizes (mean ± SEM, n = 15) in the population in which DSBs were induced.

cell lengths data for each of the time-points showed that induced cultures consisted of two populations with different mean lengths, showing that those cultures consisted of both induced and un-induced cells (Fig. 2.8). Surprisingly we did not observe the recovery of induced cultures to their initial state, that would tell us that the site-specific DSBs were repaired during the experiment. Moreover, when we traced the fate of induced single cells over time, we observed that the displacement of foci is irreversible, strongly suggesting that *parS* sequences were degraded (Fig. 2.9). The presence of repair sister-homology is necessary for the repair by homologous recombination and in the case when I-SceI enzyme targeted all of the cut sites the completion of the repair would not be possible. The idea that this happens in case of our experiment is supported by the formation of a stable population of elongated cells, together with rapid increase, followed by a stabilization of cells lacking YFP foci (Fig. 2.1c). Expression of I-SceI even by low concentrations of aTc can lead to induction of smaller proportion of cell, but with levels sufficient to cut all integrated cut sites, because of high diffusivity of small molecules in bacterial cytoplasm. Freely diffusing protein can scan the entire volume of bacterial cell in a matter of seconds [25]. This all-or-nothing phenotype makes it troublesome to study the completion of homologous recombination, but provides a robust framework to look into early steps, and in our case we focused on end-resection by RecBCD.

SINGLE DSB ANALYSIS SHOWS χ SITE RECOGNITION WITHIN LIVING CELLS.
Fast (30s/frame) time lapse imaging of cells undergoing end-resection allowed us to study the fate of single DSBs in the native environment of a cell. We could distinguish YFP and mCherry foci related to the same DSB site even in cells with multiple genome copies, because in *E. coli* markers positioned closely on the chromosome are separated by a short physical distance in a cell [26]. We observed that the loss of a YFP focus is followed by the disappearance of mCherry focus only in some of the cells (Fig. 2.2a). In cells carrying +20 kb mCherry marker that have lost YFP focus, the mCherry focus was lost in 59 ± 2% (n = 46 DSBs) of cases, and in the remaining 41 ± 2% localized mCherry signal was maintained during the course of the time-lapse experiment (Fig. 2.2b). The cells with -30 kb mCherry marker showed significantly different behavior and the disappearance of YPF focus almost never led to loss of mCherry signal (3 ± 3%, n = 61 DSBs, Fig. 2.2b). Events of foci loss were very abrupt and happened between two consecutive time frames, much faster that what could be expected from a photo-bleaching of fluorophores, when compared to the spots that were not yet processed by the end-resection machinery (Fig. 2.10). The number of χ sites between the I-SceI cut-site and the mCherry marker was different between -30 kb (10 χ sites) and +20 kb (3 χ sites) strains, what most in our opinion explains the observed difference. Assuming that the RecBCD complex encounters 3 χ sites in the direction of +20 kb marker (2 integrated into DSB cassette and one genomic) the calculated efficiency of χ site recognition is 26% (see supplementary information for details), which lies in good agreement with previously measured values of about 20% - 40% [27]. Our fluorescent marker approach provides can allow us to estimate the cellular speed at which RecBCD processes ends of DSBs. In case of +20 kb DSB sites that were characterized by the loss of both YFP and mCherry markers, foci disappeared within two time-points (n = 16 DSBs). Degradation of 2 × 10 kb [4] in 30s, or less, results in the estimated speed in the range of 1000 bps^{-1}, that stays in a good

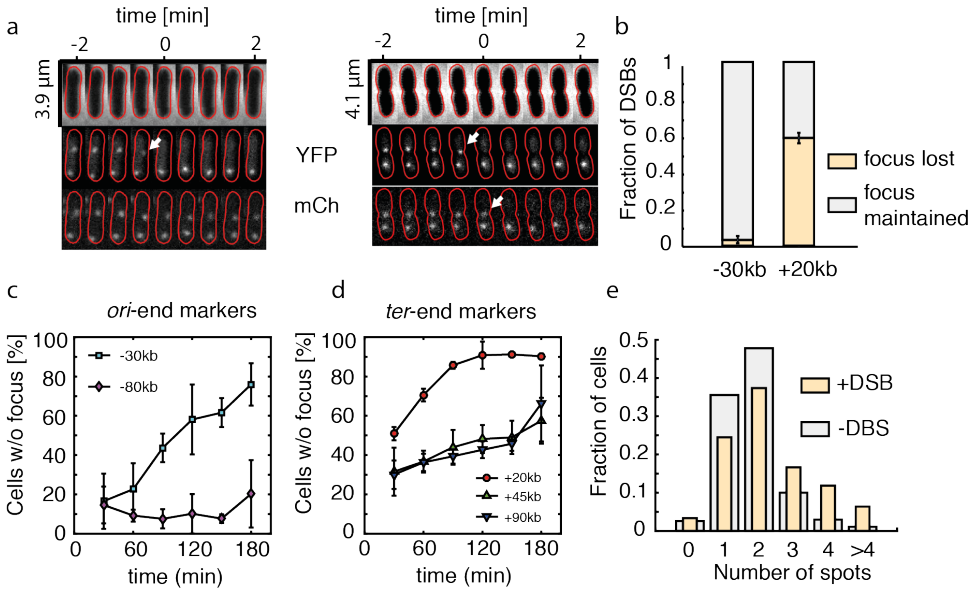


Figure 2.2: **End-resection in live cells.** (a) Individual cells undergoing end-resection of DSBs formed by I-SceI. Top: phase contrast channel; middle: +1.5 kb marker in YFP channel; bottom: +20 kb marker in mCherry channel. Red lines show cell outlines obtained using Oufti software. Arrows point at events of loss of foci. (b) Fraction (mean ± SD; -30 kb: n = 61; +20 kb: n = 46) of DSBs in which mCherry focus was lost, or maintained. (c) Percentage (mean ± SD, n = 3) of induced cells that lost mCherry at different time-points after the induction of DSBs. The markers positioned on the ori-end of the DSB are shown. (d) Percentage (mean ± SD, n = 3) of induced cells that lost mCherry at different time-points after the induction of DSBs. The markers positioned on the ter-end of the DSB are shown. (e) mCherry foci number distribution in the -80 kb strain 3h after induction of DSBs (yellow; n = 271 cells), and in uninduced culture (grey; n = 1327 cells).

agreement with previously measured *in vitro* data using optical tweezer, or DNA curtains methods [21, 28]. The speed of RecBCD in cellular environment was not measured previously due to the limitation of genetic approaches, that focused on the outcomes of the RecBCD reaction and not the reaction itself. Taken together these data show that using our system we can detect and study the events of end-resection done by the RecBCD complex in live cells. We directly measured that the RecBCD processes ends of DSBs with very high speed in cells.

ASYMMETRIC LARGE-SCALE RESECTION OF DSB ENDS.

Having showed that we can trace the fate of DSB sites in live *E. coli* cells we next set to characterize the spatiotemporal dynamics of end-resection using different mCherry markers and 30-minute snapshot analysis. We induced cells with a 15-minute pulse of aTc, followed by the wash, and for further analysis we selected cells which lost the YFP focus, that is cells which induced DSB and initiated the resection. We observed differences in processing of *ori*-end and *ter*-end ends of DSBs. The +20 kb marker was lost in 51 ± 3% of cells 30 minutes after the induction of DSB, what was consistent with fast time-lapse measurement (Fig. 2.2d). Distant markers, +45 kb and +90 kb showed slower resection,

that was progressing monotonously at a similar rate over the course of the experiment (Fig. 2.2d). The resection of the closest *ori*-end marker, -30 kb, was very low and at 30 minutes after the induction only 16 ± 14% of cells that lost YFP focus also did not contain mCherry marker, what was again consistent with fast time-lapse experiments (Fig. 2.2c). Interestingly, the most distant *ori*-end marker, -80 kb, was not degraded even in late time-points (Fig. 2.2c), but instead its copy number increased. Before the induction cells carrying that marker showed typical distribution of copy number for that *locus*, but 3h after induction we found large fraction of elongated cells with unusually high number of +80 kb foci (Fig. 2.2e and 2.11). Corresponding marker placed on the *ter*-end of the DSB was mostly resected (Fig. 2.2d and 2.11). This result shows that the formation and DSB response is not inhibiting initiation of replication in *E. coli* and each round of initiation has a potential to copy DNA proceeding the DSB. Differences in the resection of *ori*- and *ter*- end markers can be well explained by the different number of χ sites at each end, making *ori*-end more resistant to RecBCD degradation and allowing for a large resection of the *ter*-end during the repair.

qPCR ASSAY SHOWS DNA DEGRADATION DURING END-RESECTION.

Next, in order to verify if the large scale, dynamics and asymmetry was not caused, or influenced by the imaging conditions, or by the use of fluorescent ParB proteins we developed a qPCR based method to monitor the progression of RecBCD in synchronously resecting *E. coli* cells population, similar to the one used in yeast cells [29]. Using qPCR probes we measured concentrations of chromosomal markers in the proximity of the DSB as they were processed by the end-resection machinery (Fig. 2.3b). We induced the formation of DSBs by addition of aTc to the rich LB medium and compared the frequency of chromosomal markers to a non-induced culture. Similarly, to the microscopy assay, we observed long-distance and asymmetric resection of DNA flanking the DSB, at levels nearly identical to the microscopy assay (Fig. 2.3b and 2.12). Again, *ter*-end of the DSB was resected in a greater extent than the *ori*-end, and reached markers as far as +250 kb. In the qPCR assay the signal coming from a chromosomal marker can only be reduced when the region of interest is degraded. If the RecBCD complex was acting as a helicase that unzips the DNA without the degradation and then creates ssDNA nick at χ, as it was proposed in one of the previous models [8], we would not detect resection with qPCR assay. Based on our data we favour a model where RecBCD is a helicase/nuclease and the degradation of significant amounts of DNA takes place during end-resection. Moreover, because the end-resection profile obtained with the qPCR assay matches very well the microscopy results, we concluded that the fluorescent ParB fusions had no significant influence on the DSB processing.

SHORT END-RESECTION IN THE ABSENCE OF RECD.

E. coli strains lacking the gene coding for RecD are recombination proficient and can integrate homologous linear DNA into chromosome [30]. RecBC complex *in vitro* lack the activity for degradation of DNA and can load RecA onto ssDNA in a χ independent manner [31]. We asked whether the profile of end-resection in the Δ*recD* cells is significantly different from the wild-type cell. To test this, we deleted the *recD* gene from bacterial chromosome and used such strain to assay the end-resection with the same qPCR assay as for wild-type bacteria. Indeed, the resection in the mutant cells was highly reduced

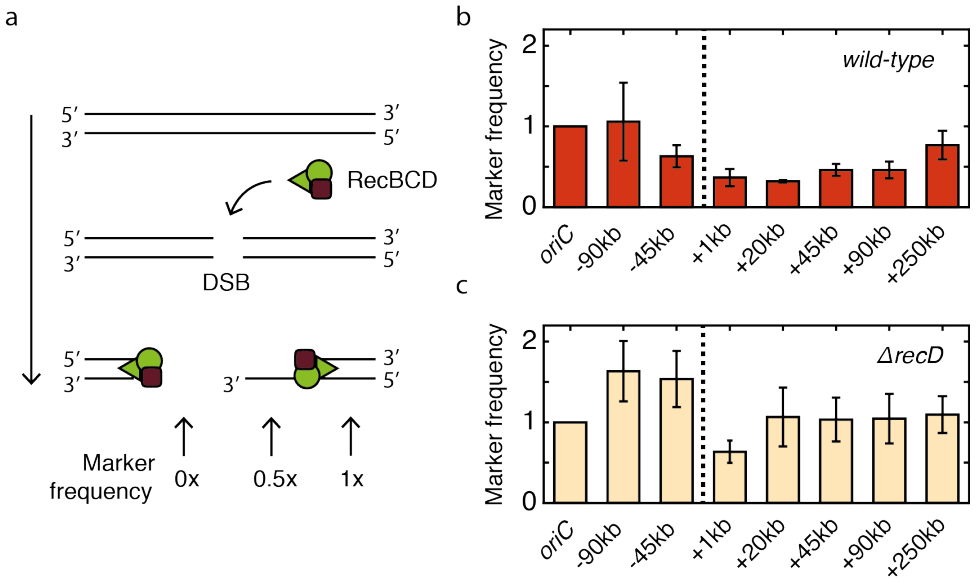


Figure 2.3: **End-resection profile studied with qPCR.** (a) Influence of the end-resection on the frequency of occurrence of the chromosomal marker. (b) The profile of wild-type end-resection measured 1h after the induction of DSBs (mean ± SD, n = 3). Dashed line indicates the position of the DSB. (c) The profile of end-resection in Δ*recD* cells measured 1h after the induction of DSBs (mean ± SD, n = 3).

and only the closest +1 kb marker showed 50% resection while the resection of remaining markers was not present (Fig. 2.3c). Reduction in the half of the signal can be explanation by an unwinding of the dsDNA by RecBC complex and loading of RecA on the 3'ss-DNA tail. Additional cellular exonucleases could have been involved in the degradation of 5' ssDNA tail formed during RecBC resection, which was not bound and protected by RecA [32]. This result directly shows that the end-resection in Δ*recD* mutants *in vivo* is highly reduced.

SHORTENED END-RESECTION AND RECOMBINATION TEMPLATES.

Next we wondered how the length of resection at the DSB site influences the selection of the repair template in *E. coli*. To study the efficiency of the HR we developed a reporter system consisting of a low-copy number plasmid carrying an inducible I-SceI cut site flanked by homologous arms. We also integrated an ectopic, I-SceI-resistant repair template with same homologous arms into the chromosome. The DSB site on the plasmid was flanked by an array of 3 χ sites at each end in the χ+ plasmid, and not flanked in the case of χ- plasmid (Fig. 2.4a). DSBs formed on plasmids can be repaired either by an intact copy of a plasmid present in the cell, or by the recombination with an ectopic chromosomal site. The recombination with the chromosome alters the sequence of the plasmid, therefore can be detected with a repair-specific primer pair after isolation of plasmids from cells. After induction of DSBs and purification of plasmids we quantified the ratio of modified plasmids to a reference sequence of I-SceI coding plasmid using

qPCR (Fig. 2.4b). After 4h of induction we detected increased signal from recombinant plasmid for all samples, but the difference was significant only in case of Δ*recD* cells (Fig. 2.4c, *t*-test p > 0.01). Interestingly, recombination with the chromosome was not detected in the RecBCD cells even for the constructs with χ sites. It is very likely that the probability of χ recognition by RecBCD complex is too low to initiate the recombination of a broken plasmid at the regions of homologous arms and directs the repair to use an intact plasmid as a donor of homology.

The repair of a χ DSB plasmid was monitored in a strain that carried an ectopic repair template with χ sites, therefore introducing a mismatched region between a plasmid and chromosome. In the Δ*recD* background, for the χ- constructs efficiency of recombination was lower than for the χ+ plasmid. The recombination between sequences with multiple mismatches was recently shown to be possible, but less efficient when assayed in yeast model system [33]. Similar effect of mismatches on efficiency of reaction could explain the reduced recombination of χ plasmid observed in our experiments. Our data demonstrate that ectopic homologies can be efficiently used for the homologous recombination during DSB repair when the length of end-resection is reduced.

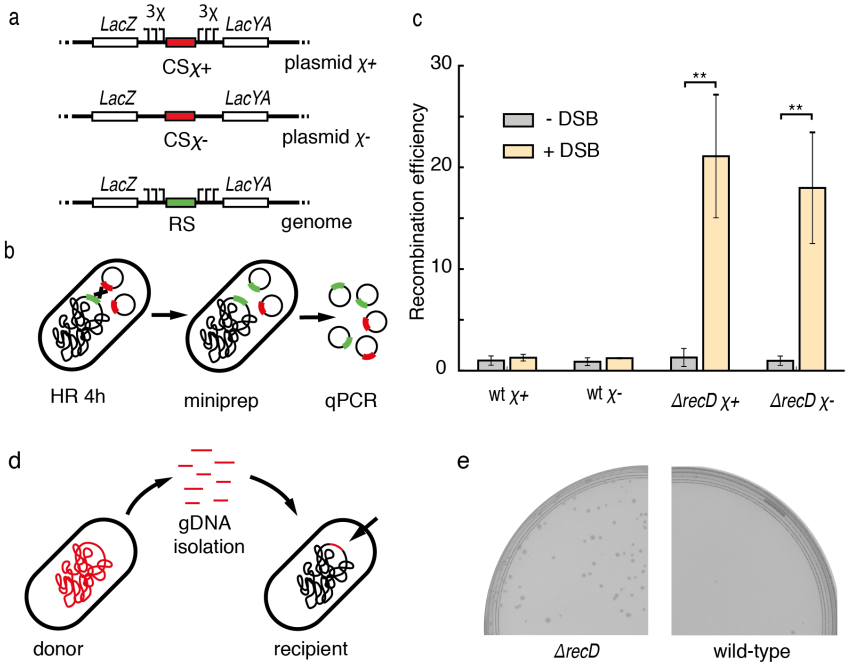
DELETION OF *recD* INFLUENCES HORIZONTAL GENE TRANSFER.

Next we wondered whether the shortened end-resection can have an influence on the process of horizontal gene transfer (HGT). In the natural environment the main factors contributing to lateral gene transfer are transformation, conjugation, and phage transduction [34]. To test how the Δ*recD* mutation influences the HGT between different *E. coli* cultures we designed a system where we first isolate a library of fragments of the chromosome from a donor carrying chloramphenicol (Cm) resistance marker then used it to electroporate into recipient Cm-sensitive strain, and screened for Cm resistant colonies (Fig. 2.4d). We found colonies when the Δ*recD*, but not wild-type cells were used as the recipients of DNA (Fig. 2.4e). PCR genotyping confirmed that Cm-resistant Δ*recD* colonies contained the Cm-resistance marker as well as Δ*recD* specific genomic watermark, confirming integration into genome (Fig. 2.13). End-resection of linear DNA ends by RecBCD was previously demonstrated to degrade incoming DNA [35]. Here we show that the HGT of collection of linear chromosomal fragments was possible when the long end-resection was disabled by Δ*recD* mutation. Next, we also asked if the mutations of *recD* gene are common within sequenced *E. coli* genomes. When we screened *E. coli* genomes deposited in the RefSeq Database we found that about 67% of genomes (n = 285 sequences) carried at least 1 non-synonymous substitution in *recD* gene (Fig. 2.14). On the other hand, equivalent analysis of *recA* gene showed variability in only 4.6% screened sequences. This result suggests that the end-resection in naturally growing *E. coli* may be occasionally modified by a mutation. Together, this data shows that the loss of long end-resection is potentially an important factor promoting the HGT in *E. coli* population, considering that mutations within the *recD* gene are common within *E. coli* strains.

2.3. DISCUSSION

HIGH SPEED OF *E. coli* END-RESECTION

In this work we directly observe the process of end-resection during DSB repair in individual *E. coli* cells. With the fluorescence microscopy we demonstrated that the re-



[H]

Figure 2.4: **Length of end-resection influences the homologous recombination.** (a) Genetic maps of the components of homologous recombination reporter. The I-SceI cut site in red was placed on low copy number plasmid. Repair template was integrated into genome. Arrows represent χ sites. (b) Recombination was induced by formation of DSBs. After 4h, reporter plasmids were isolated from cells and the concentration of recombinant plasmids was measured with qPCR. (c) The efficiency of recombination (mean ± SD; n = 3) in wild-type and in δ*recD* backgrounds. p values were obtained using the two-tailed t-test ($** \leq 0.01$). (d) Chromosomal DNA was isolated from a donor strain in a form of short linear fragments. After delivery of fragments into the recipient strain, antibiotic-resistance marker was integrated into the genome. (e) Growth of recombinant colonies in presence of chloramphenicol.

section in bacterial is unusually fast and large, reaching tens of kilobases in matter of seconds and removing a significant portion of the chromosome. Interestingly, the speed of resection measured here is orders of magnitude faster than the eukaryotic resection, which can progress at the rate of 1 bps^{-1} in vegetative yeast cells [36], or 10 bps^{-1} during meiotic resection [37]. The single-molecule experiments of RecBCD activity showed the speed of the resection, prior to χ activation of about 1000 bps^{-1} and was conducted with a naked DNA as a substrate [28, 38]. Here, we show that also *in vivo*, when the DNA is frequently bound by the DNA-associated elements, the high rate of end-resection is maintained. The high speed and processivity of RecBCD is believed to constitute a first line of the defence against some phages, and indeed, inactivation of RecBCD yields reduced resistance to phage infection [39].

DEGRADATION DURING END-RESECTION

In vitro, the degradation of DNA RecBCD complex prior to the recognition of χ is dependent on the ration between the concentration of Mg^{2+} and ATP. When the Mg^{2+} is in excess the RecBCD degrades both strands of DNA during translocation and after the recognition of a χ the degradation is limited to 5' ssDNA end only [7]. Alternatively, when the concentration of ATP is higher than Mg^{2+} the RecBCD unwinds the DNA without degradation, and produces a single nick after the activation by χ [8]. The question which of those models is correct in case of the reaction occurring in a cell was difficult to address, due to technical limitations of previously used assays. The first model, including the degradation of both strands of DNA, is more suited to explain the results of our microscopy and qPCR data. RecBCDs helicase activity alone, that is unwinding without degradation would not result in the reduction of qPCR signal, because the concentration of the template material in that case would be constant, but we observe clear loss of the DNA at regions surrounding the DSB. Moreover, the loss of ParB foci was stable in our experiments, again suggesting that the *parS* sequences, to which ParB protein bound were degraded by the action of RecBCD. We conclude that in cell, the degradation of DNA is an inseparable feature of the RecBCD end-resection before χ (Fig. 2.5a).

CONSEQUENCES OF SHORT RESECTION IN Δ*recD*

The length of end-resection was shown to have notable consequences on the selection of the homologous template in eukaryotic cells. Shortened resection of DSBs in budding yeast resulted in the repair with additional templates, thus a normal resection was beneficial for a cell because it assured the fidelity of repair [40]. Russell et al, demonstrated that in Δ*recD E. coli* cells linearized plasmid, containing regions of chromosomal homologies is efficiently integrated into chromosome by HR. In this study we experimentally demonstrated that in vivo the resection by RecBC complex is drastically shortened, what, together with χ independent RecA loading (Fig. 2.5), allows for the recombination with short, ectopic homologies. In wild-type cells RecBCD activity removes so much DNA before initiating the recombination that such ectopic homologies are inefficient for repair, even if contain χ sites. The length of resection may also have an impact on HGT. Despite *E. coli* is deficient in natural transformation it encodes components for uptake of DNA from the environment [41], what together with Δ*recD* phenotype may promote transfer of genetic markers between cells. We suspect that RecBCD may also supress the

transfer of genetic markers in cases when short DNA molecules, originating from fragmented chromosomes of dead cells are used. On the other hands, cells with mutations in *recBCD* (Fig. 2.14) can be a subject of HGT.

END-RESECTION *in vivo*

Arguably, the main function of HR in *E. coli* is the repair of collapsed replication forks, and DSBs formed by the replication in general [11, 12, 39]. When such break is formed, the RecBCD processes, and initiates recombination only with the *ori*-end of DSB, because the *ter*-end did not have chance to be synthetized. Such *ori*-direction bias is a reasonable explanation for the asymmetry in the orientation of χ sites on the genome and also explains why in our experiments the *ter*-end was excessively degraded by RecBCD. Very high rates of resection can support recombination with the replicated sister, and at the same time inhibit the recombination withshort, ectopic homologies (Fig. 2.5).

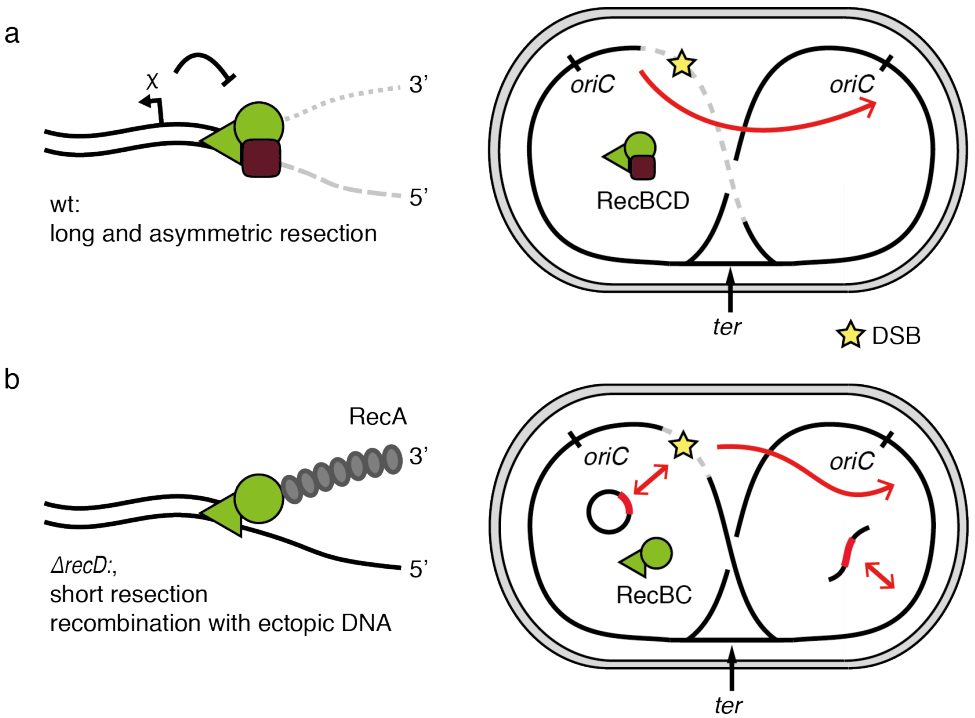


Figure 2.5: **Models of end-resection in *E. coli*.** (a) Homologous recombination in wild-type cells. During resection RecBCD degrades both strand of DNA, and recombination is activated by recognition of χ site. After formation of a DSB, large fragments of the chromosome are rapidly degraded, preferentially in the direction of terminus. Such end-resection drives recombination to use intact sister chromosome as a repair template. (b) Deletion of recD gene removes the influence of χ sites and greatly shortens the resection. As a consequence of shortening the resection, short homologies, such as plasmids, or short linear fragments can be used as templates for recombination.

2.4. MATERIALS AND METHODS

STRAINS

Cut site, fluorescent markers, and repair templates were integrated into *E. coli* K12 TB28 [42] (MG1655, Δ*lacIZYA*) cells. Cloning reactions were transformed into Top10 cells (Thermo Fisher Scientific). To construct the fluorescent reporter of end-resection, we first integrated the construct containing the I-SceI recognition cut site, flanked by 2 χ sites on each side and a *parS-MT1* sequence [15] (amplified from pBlueDSBarms-parSMT1-BglII-frtCmR) using the lambda-red method [43] into TB28 cells. Distal *parS-P1* [15] sequences were integrated into TB28 cells (amplified from pBlueHTarms-parSMT1-BglII-frtCmR) and then combined with the cut site strain using P1 phage transduction. MG1655 deletions of *recD* was done with a lambda-red protocol in a MG1655 background, followed by P1 transduction. Resistances were removed using the pCP20 plasmid [43].

Due to the length of lambda-red PCR products required to integrate homologous recombination reporter (RS) from one template plasmid two fragments from pLS6 were amplified using with Phusion polymerase (Thermo Scientific) with primer pairs Jw246 and Jw358, and Jw245 and Jw359. Resulting fragments shared 20 bp of homology and the full length fragment was assembled using Phusion polymerase. The full-length fragment was integrated into genome of TB28 cells using lambda-red protocol.

PLASMIDS

All plasmids were constructed using the Gibson assembly protocol (NEB) with the exception of pBlueDSBarms-parSMT1-BglII-frtCmR and pBlueHTarms-parSMT1-BglII-frtCmR, which were constructed using restriction digestion and ligation. To clone these plasmids pBlueHTarms and pBlueDSBarms (both were gifts from Hugo Snippert) were linearised using BglIL restriction enzyme (New England Biolabs). Fragments containing *parS-P1* and *parS-Mt1* follower by chloramphenicol resistance cassettes were amplified with primer Jw024 and JW025 and cut with BglII from plasmid pGBKD3-*ParSP1* and pGBKD3-*ParSPMT1* [44]. Next, fragments were ligated with linearized backbone.

Plasmid p15aSceI_deg was cloned using Gibson assembly by fusing fragment amplified from the plasmid pdCas9degRNA [24] with Jw201 and Jw202, then a fragment from a plasmid pDL2655 [45] amplified with primers Jw205 and Jw206 to create p15aSceI_deg_kan. Resulting plasmid was kanamycin resistant, and to change the resistance to chloramphenicol we performed additional round of Gibson assembly. A fragment amplified from that plasmid with primers Jw210 and Jw211 and fused with a fragment amplified from pKD13 [43] with primers Jw212 and Jw213 to create a plasmid p15aSceI.

Cloning of pSC101SceI_deg was similar to the previous plasmid, but in the first step we amplified a backbone from pHippACMR (gift from Helena Shomar Monges) using primers Jw203 and Jw204 and fused it with a fragment amplified from pDL2655 with primers Jw205 and Jw207 to create pSC101SceI_deg_cm. Kanamycin resistance gene was exchanged with chloramphenicol resistance gene by fusing a fragment amplified from pSC101SceI_deg_cm with primers Jw208 and Jw209 with a fragment amplified from pKD13 with primers Jw212 and Jw213.

Mix & Go E. coli transformation kit (Zymo Research) was used to transform plasmids

into the cells. List of plasmids and primers used in this study can be found in the Supplementary Table 2.1–2.4.

GROWTH CONDITIONS AND MEDIA

Bacteria used for microscopy experiments were grown in M9 minimal medium supplemented with 0.2% glucose at 37°C. qPCR experiments were carried out on bacteria grown in Lysogeny Broth (LB) at 37°C with addition of antibiotics. The overnight LB culture was refreshed on the following day 1 to 500 and incubated until reached OD ≈ 0.1, then I-SceI expression was induced with addition of aTc to the media (2ng/ml) for 1h. For microscopy experiments with snapshots the over-night culture was grown in M9-glucose media in 37°C and refreshed on the following day 1 to 100. When the culture reached OD ≈ 0.1 the I-SceI expression was induced by the addition of aTc (2 ng/ml) for 15 minutes. Then the culture was centrifuged and cells were resuspended in a fresh M9-glucose media without aTc and grown in 37°C. For timelapse experiment cells were grown as in snapshot experiment, but the imaging started after initial 15 minutes induction without the washing step. Ampicillin (100 μg/ml), chloramphenicol (34 μg/ml) and kanamycin (50 μg/ml) were added to cultures when required.

MICROSCOPY AND IMAGE ANALYSIS

Microscopy experiments were performed using a Nikon Ti-E microscope with a 100x oil-immersion phase-contrast objective (CFI Plan Apochromoat λ DM 100x). Fluorescence was excited using a Lumencor Spectra X LED light source and images were acquired using an Andor Zyla 4.2 CMOS camera. Fluorescent image exposure was selected to provide a good signal-to-noise ratio maintaining minimal exposure time (for snapshot experiments exposure times were: YFP: 400 ms, mCherry: 400 ms, and for 30 s time-lapse experiment YFP: 200 ms, mCherry: 200 ms time-lapse experiments were imaged in a custom-made environmental chamber held at 37°C. Cells were imaged on M9 media pads containing 1% agarose. Shortly, 2.5 μl of the culture was pipetted on the pad and once cells were absorbed onto the M9-agarose the pad was placed on the microscope slide and imaged.

Bacterial cells in phase-contrast images were segmented using MATLAB and Oufti package [46]. Foci were detected using a Crocker and Grier routine (adapted for MATLAB by Blair and Dufresne, http://site.physics.georgetown.edu/matlab/) embedded in a custom, automated image-processing pipeline. For each detected cell, an according region of interest was cropped from fluorescent images. In the next step each cell was validated based on fluorescent signal and only cells which mean fluorescent intensity above global background level were selected. After initial foci detection, only foci which intensity exceed the cellular background noise level were selected. Degradation events in fast time-lapse experiments were selected manually in FIJI software [47].

QUANTITATIVE PCR

Induction of the DSB for quantitative PCR measurements of degradation was done as described previously [48], but with induction with aTc. Genomic DNA was isolated with the Wizard Genomic DNA purification kit (Promega). DNA concentration was measured using Quant-iT^{TM} (ThermoFisher) and for each qPCR reaction equal amounts of template DNA were used. For end-resection experiments C_t value of each marker was com-

pared to C_t value of the *oriC* primer pair, used as an integral frequency normalization marker. Normalized values were then compared to corresponding values obtained for the un-induced strain. We used dye-based GoTaq qPCR Master Mix (Promega) and Eco Real-Time PCR system (Illumina) for all degradation qPCR reactions. Multiple primer pairs were tested for each chromosomal marker were tested and only the pairs which resulted in efficiency close to 100% were used for final experiments.

For the recombination-efficiency experiment, cells were induced as for the qPCR degradation experiment, but after the addition of aTc (2 ng/ml) the culture was incubated for 4h and then plasmid DNA was isolated using Wizard Plus SV Miniprep (Promega). Plasmid DNA was further purified with ethanol precipitation. 30 μl of plasmid DNA was mixed with 300 μl of 100% ethanol and 30 μl of 3M sodium acetate and incubated in -20°C for 1h and next, centrifuged at 4°C for 45 minutes at maximum speed (Eppendorf 5418R). Supernatant was discarded and the DNA pellet was suspended in 300 μl of 70% ethanol solution, upon which the centrifugation step was repeated. Supernatant was discarded, the DNA pellet was dried and re-suspended in 100 μl of milliQ water. DNA concentration was measured using Quant-iT^{TM} (ThermoFisher) and for each qPCR reaction, equal amounts of template DNA was used. To calculate the recombination efficiency, we first calculated C_t values for primer pair specific to the recombination product (with primers *HR product f.* and *HR product r.*) and compared this value to the C_t values for primer pair specific to the p15a_SceIdeg plasmid (with primers *p15a f.*, and *p15a r.*). We assumed that the concentration of p15a_SceIdeg was not affected by the recombination, therefore we could use it as a reference to measure the concentration of the product of recombination. Finally, normalised C_t values were compared to the normalised values obtained with uninduced culture of wildtype cell with χ+ plasmid.

PURIFICATION OF GENOMIC DNA FOR HGT EXPERIMENT

Genomic DNA of a strain carrying the Cm^R gene was prepared using Promega Wizard Genomic DNA Kit following the manufacturers protocol. Shortly, 1 ml of overnight culture of Jx097 and Jx098 cells was used to isolate the DNA. After the isolation, the DNA sample was further purified using ethanol precipitation protocol. Electrocompetent cells were prepared as follows: Cells were grown in LB in 37°C until the culture reached OD ≈ 0.6. Then 1ml of culture was centrifuged (10000RPM, 4°C, Eppendorf 5418R). and washed 3 times in 1 ml of milliQ water. After final wash cell pellet was resuspended in 50 μl of milliQ water. All steps were done in 4°C. 100 ng of purified genomic DNA was used for electroporation into electrocompetent cells.

PROBABILITY OF χ SITE RECOGNITION

The probability of degradation of a DNA marker after n χ sites can be written as $p_d = 1 - (1 - p_\chi)^n$, where $(1 - p_\chi)$ is the probability of not recognizing a χ site. From the measured value of $p_d = 0.586$ obtained in our experiments for 3 χ sites, $(1 - p_\chi) = 0.745$, and therefore probability of recognizing single site is $p_\chi = 0.255$.

BIOINFORMATICS

Mutation rates in *recABCD* genes was calculated on genomes downloaded RefSeq release 82 database. 330 genomes were screened for *recABCD* genes using custom MATLAB scripts, resulting in 285 genomes where all 4 genes were present. Next, each gene

was translated into amino-acid sequence and aligned with canonical sequence (from K12 MG1655 strain) using BLASTP algorithm. Alignments were processed using custom MATLAB scripts. We calculated number of sequences with at least one non-synonymous mutation for each gene. For calculating mutation rates we counted the mean number of unique non-synonymous mutations in each amono-acid sequence and divided it by the the total number of amino-acids.

ACKNOWLEDGMENTS

We thank Felicia Tijen-Fooh and Amelie Erben for valuable contributions in initial experiments leading to this work; Theo van Laar and Jaco van der Torre for the experimental advices; Hugo Snippert for the gift of the I-SceI cut site plasmids; Helena Shomar for the gift of the pSC101 origin plasmid; Jacob Kerssemakers for Matlab discussions; and Christian Lesterlin for I-SceI, TB28 and DSB constructs, discussions and help with initial experiments and encouragement. We thank Jorine Eeftens and Hyun Youk for critical reading of the manuscript. This work was supported by European Research Council NanoforBio No. 247072 (CD) and SynDiv 16 669598 (CD); Wellcome Trust [SIA099204/Z/12Z]; Leverhulme Trust [RP2013-K-017]. J.W. work in D.J.S. laboratory was supported by EMBO ASTF 393-2013.

2.5. SUPPLEMENTARY INFORMATION

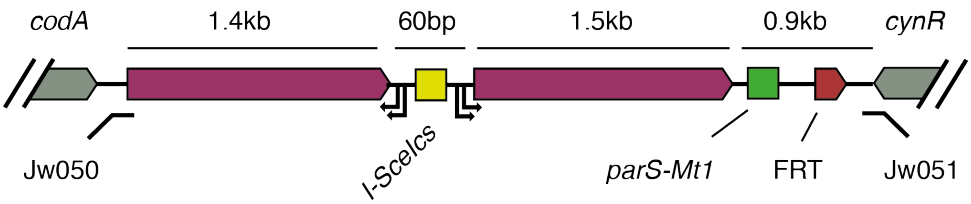


Figure 2.6: Genetic map showing the cassette with I-SceI cut site (I-SceIcs, yellow) flanked by 2χ sites at each end. Position of 1.5 kb parSPMT1 (green) and FRT (red) sequence is shown. Primers Jw050 and Jw051 were used to amplify and integrate the cassette between codA and cynR genes.

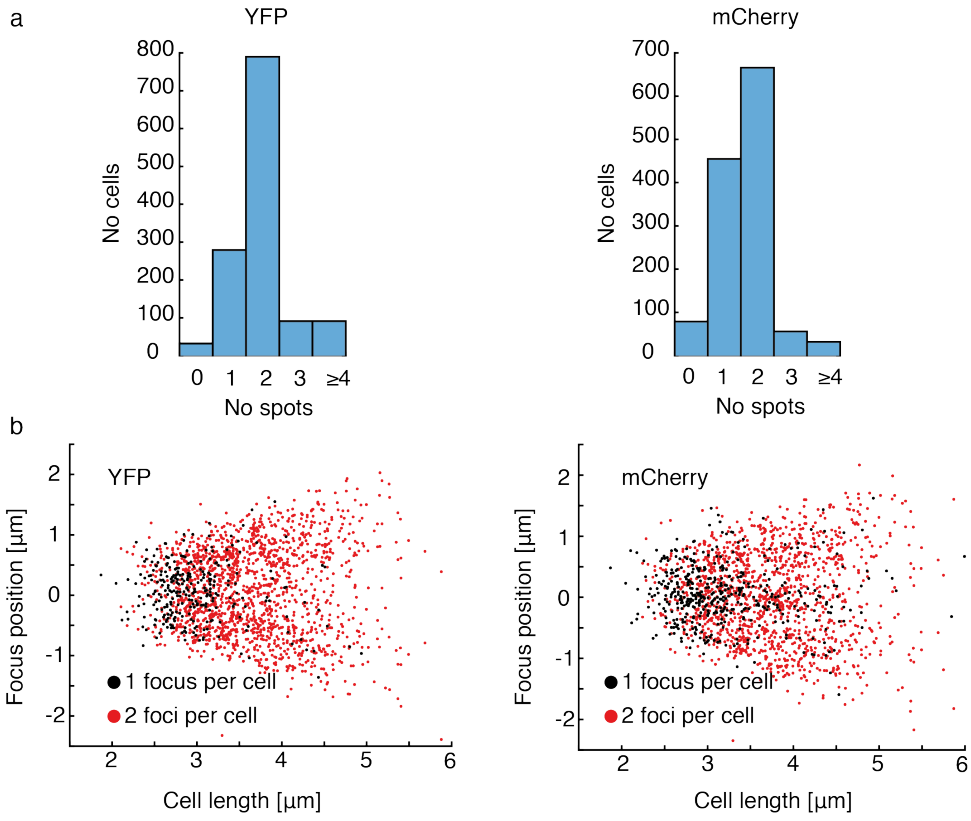


Figure 2.7: **Effects of fluorescent labels on chromosome organization.** (a) Distribution of the number of foci in each channel for pooled data gathered from uninduced cultures from time t = 30 min. Data from all strains combined. (b) Position of the spot along the longitudinal axis of the cell. Positions of foci in cells with 2 foci is shown in red, positions of a focus in cells with only 1 focus is shown in black. Data from all strains combined. In all panels n = 1296 cells.

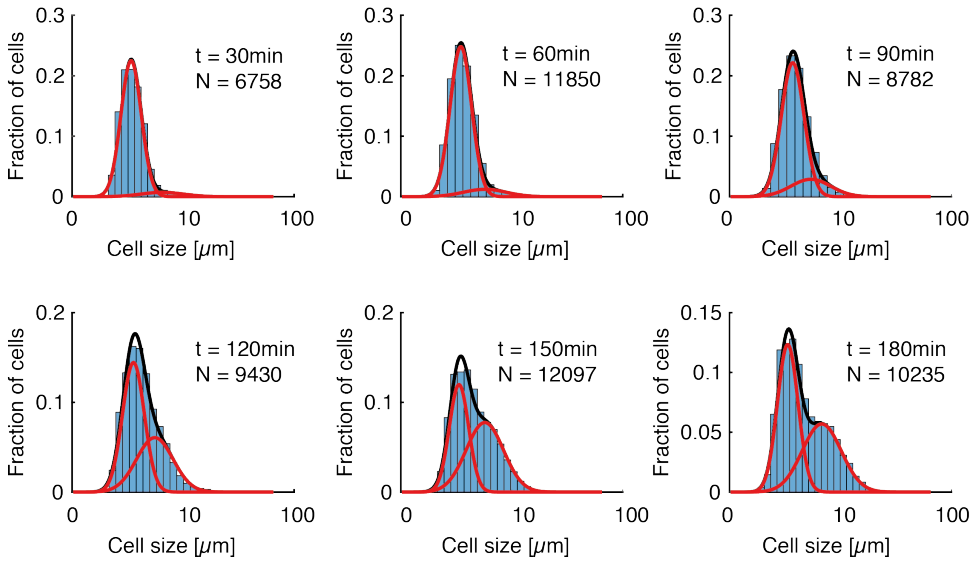


Figure 2.8: **DSBs are induced in a subpopulation of cells.** Fitting of two normal distributions (red) to $\log_1 0$ of pooled data of cell sizes from all tested strains (blue) shows a growing population of elongated cells over time. The sum of the two distributions is shown as a black line.

Table 2.1: List of strains, plasmids, and primers

Strain	Description
TB28	MG1655; Δ*lacIZYA*
Jx097	TB28 parSP1_-30kb
Jx098	TB28 parSP1_-80kb
Jx099	TB28 parSMt1::codA parSP1_+20kb
Jx100	TB28 parSMt1::codA parSP1_-30kb
Jx101	TB28 parSMt1::codA parSP1_-80kb
Jx119	TB28 parSMt1::codA parSP1_+90kb
Jx126	TB28 parSMt1::codA parSP1_+45kb
MG1655	F-, lambda-, rph-1
Jx168	MG1655 frt::recD
Jx169	MG1655 frt::recB
Jx186	TB28 LS6::codA
Jx187	TB28 LS6::codA frt::recD

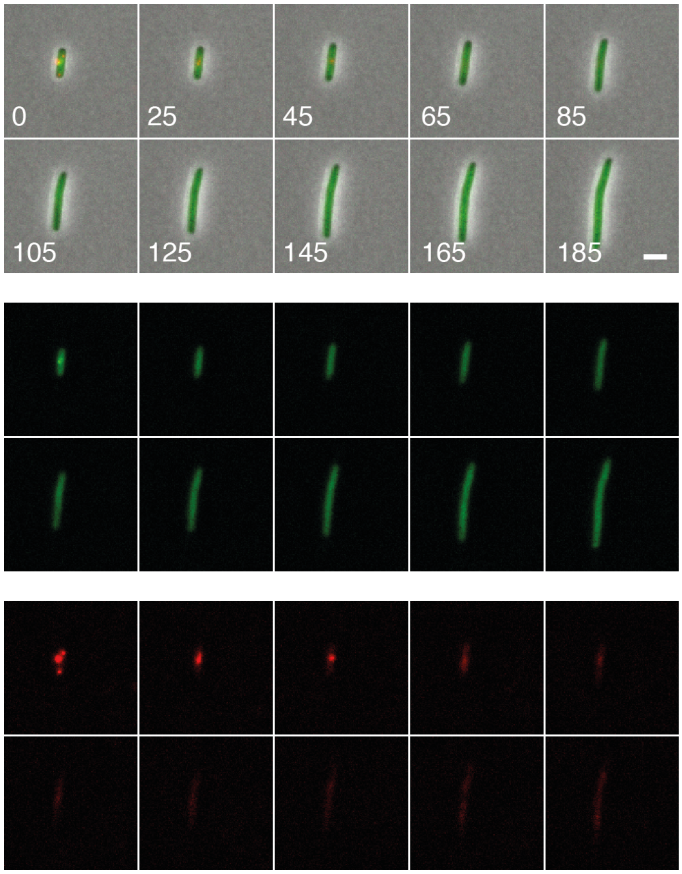


Figure 2.9: Focus lost during the end-resection is not reappearing. Example cell from +30 kb strain showing the irreversible displacement of YFP and mCherry foci. Time in minutes is shown at each frame. Scale bar is 5μm.

Table 2.2: List of plasmids.

Plasmid	Description
pBlueDSBarms-parSMT1-BglII-frtCmR	template for parSMt1 integration
pBlueHTarms-parSMT1-BglII-frtCmR	template for parSP1 intergration
pSC101_LS2χ+	HR reporter plasmid containing χ
pSC101_LS2χ-	HR reporter plasmid lacking χ
pSC101_LS6χ+	template for integration of short repair site
pSC101SceI_deg	expression of SceI fused to LAA degron, SC101 origin
p15aSceI_deg	expression of SceI fused to LAA degron, 15a degron
p2973	expression of parBMt1-YFP and parBP1-mCherry
pLisa5	template for linear integration PCR product

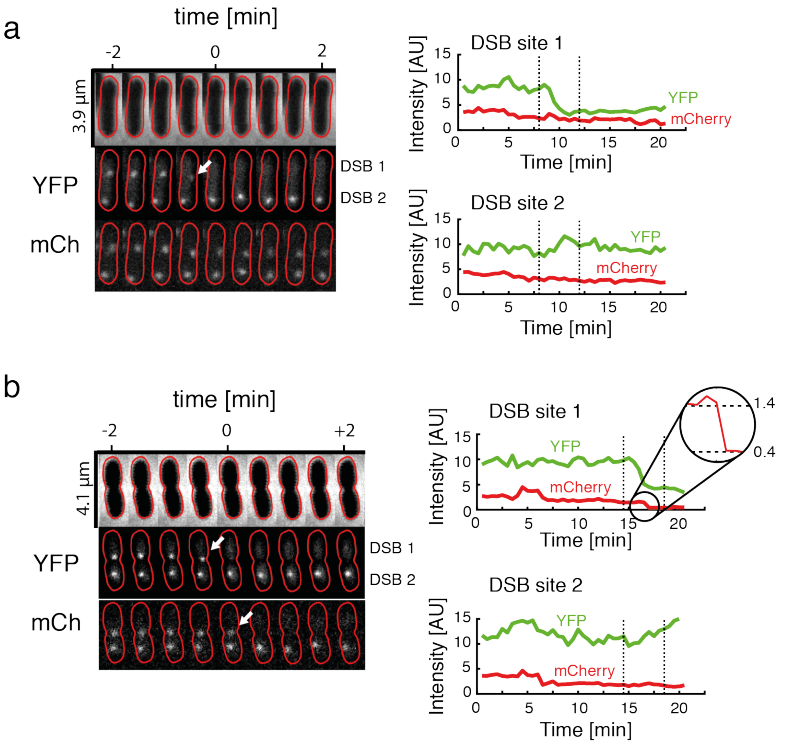


Figure 2.10: **Rapid displacement of foci due to end-resection.** Plots showing intensities of foci over time. Green line shows the intensity of YFP foci, red line shows the intensity of mCherry foci. Dashed line on the line plot shows the time-limits of the microscopy image montage. Intensity was measured by integrating the signal intensity of 3 by 3 pixel square centered at the focus.

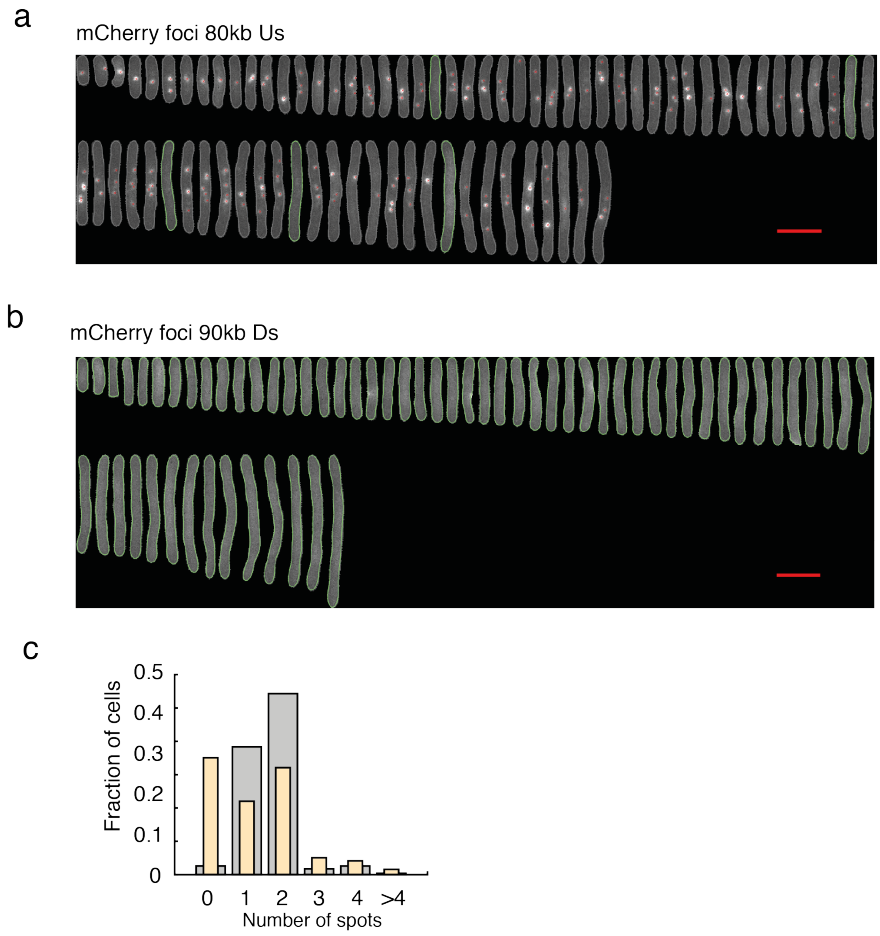


Figure 2.11: **Initiation of replication during DSB repair.** (a) Aligned representative cells from population of DSB-induced cells showing multiple +80 kb. Images were taken 3 hours after induction of the DSB. mCherry channel is shown. Green outlines cells that did not contain mCherry foci, detected using Oufti software. Red scale - 5 μm. (b) Same as in (a) showing +90 kb strain. (c) mCherry foci number distribution in +90 kb strain 3 h after induction of DSBs (yellow; n = 232 cells) and in uninduced culture (grey; n = 2034 cells).

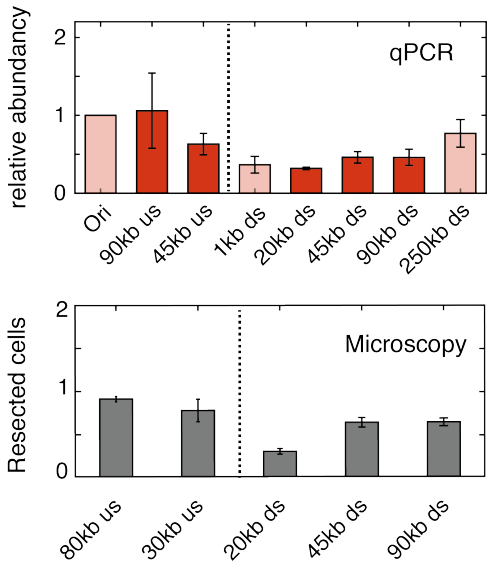


Figure 2.12: **Comparison of end-resection profiles.** Comparison of profiles of end-resection between qPCR and fluorescent microscopy measurements. The microscopy dataset shows results obtained 1 hour after induction of the DSB (adapted from 2.1). In the qPCR dataset markers corresponding to the microscopy fluorescent markers are highlighted. Dashed line indicated position of the DSB (mean ± SD; n = 3).

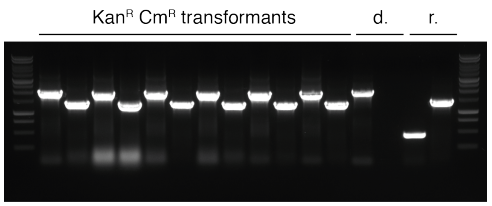


Figure 2.13: **Integration of genomic DNA into the chromosome confirmed by PCR.** PCR screening of colonies growing in presence of chloramphenicol (Fig. 2.4e) confirms the modification of recipient genomes. Two PCR reactions were run for each colony. The first lane for each colony shows the DNA watermark specific to the donor strain, while the second lane shows DNA watermark specific to the recipient strain. 6 colonies were selected. Last 4 lanes show PCR reactions on donor strain and on recipient strain. Promega BenchTop 1 kb DNA ladder was used.

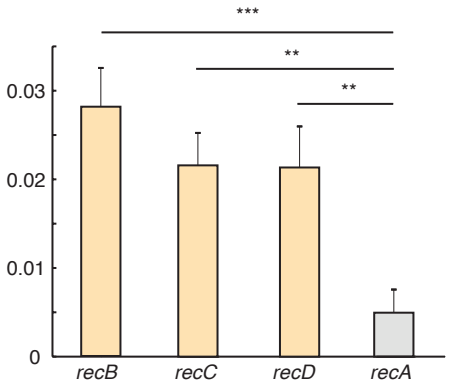


Figure 2.14: **Rate of non-synonymous substitutions in the *recABCD* pathway.** Rates were calculated taking only unique non-synonymous substitutions into consideration. Data represented as mean ± SEM for genes found in all compared genomes. Stars indicate *p*-calues from two-tailed *t*-test (**p ≤ 0.01; ***p ≤ 0.0001).

Table 2.3: List of qPCR and genomic integration primers.

Integration primers	Sequence
+20 kb	AAT TAA TTG CAT TTA AAA AAT ATG TTC TGT GAA CAA GCA TTG TTT GGTCTGCTATGTGGTGCTATC
	GTA TCA TTC TCA ATC ACT CGC ATA TTA CAT TCA CAT AAT GTA TAT GTG TAG GCT GGA GCT GCT TC
+45 kb	AAA AGG TGC CAG AAC CGT AGG CCG GAT AAG GCG TTC ACG CCG CAT GGTCTGCTATGTGGTGCTATC
	AGA CGC GGC AAG CGT CGC ATC AGG CAT CGG AGC ACT TAT TGC CGG GTG TAG GCT GGA GCT GCT TC
+90 kb	CGT TAT AAC AAA AGG GGA GTG CTG AAG GAG TCT GGG CGG GCA ATT GGTCTGCTATGTGGTGCTATC
	ATA ATT ATT TGT TAA ATA ATT GTT TTA TTT CAC ATT GGT TAT ACC GTG TAG GCT GGA GCT GCT TC
-30 kb	CAG ATA TTG CCC GGT CTC CGC GCC CAC GCC CGG CAA CAC GTT AAA GGT CTG CTA TGT GGT GCT ATC
	AAG CTG GCT GAA ATT TAC AGC GAA GCG GGC CTG CCG GAC GGC GTA GTG TAG GCT GGA GCT GCT TC
-80 kb	AGC CAT CTT TTC CCC CTC GCC CCT TTG GGG AGA GGG CCG GGG TGA GGT CTG CTA TGT GGT GCT ATC
	GGA AGA GGG AGT GCG GGA AAT TTA AGC TGG ATC ACA TAT TGC CCC GTG TAG GCT GGA GCT GCT TC
qPCR primers	Sequence
+1 kb f:	CTG CTG GTA ACA ATG TGG AAA G
+1 kb r:	TTG ACC ACT ACT GCG AGA AAG
+20 kb f:	CGAAGAAAGGAGGGCTGAAT
+20 kb r:	TCTCTGAGTGCATCGTTGTG
+45 kb f:	TAC CGT GGA TGG ACG AAG A
+45 kb r:	GAA TCC CGT GCC CGA AAT AA
+90 kb f:	GAC CTG ATG AGC CGT AAC TTC
+90 kb r:	AGT CTT CAC CCA GCT TGT ATT C
-45 kb f:	CCT CGA TAG CCA CGT CAA ATA G
-45 kb r:	CAC CAG TAA GCT GGT TGT GA
-90 kb f:	CCC ATT AAC CGG AGG CAA TA
-90 kb r:	TCG AGG TGG CGA TGT AGT A
+250 kb f:	GCA CGT TAT CCC ACT GAT AGA G
+250 kb r:	CAC CTG GTT CCG TAA CGA TTA T
ori f:	GCGAGGATCGAAGAAGTCATAC
ori r:	GATGCAAATCGTCCGCTCTA
p15a f:	CCGGAAATCGTCGTGGTATT
p15a r:	GTG AGC TGG TGA TAT GGG ATA G
HR product f:	CATGGCTGATGCAATGCGGCG
HR product r:	GAT GCG ATG TTT CGC TTG G

Table 2.4: List of primers used for cloning.

cloning primers	
Jw201	TAC CTG GTT TTT TTT GAT GTT TTT CAT AGA TCC TTT CTC CTC TTT AGA TCT TTT G
Jw202	CTA TCT CCT CCG AAA CTT TCC TGA AA GCTGCTAACGAC-GAAAACTACGC
Jw203	ACG AAA ACT ACG CTC TGG CTG CTT AAG GAT CCA AAC TCG AGT AAG GAT C
Jw204	TAC CTG GTT TTT TTT GAT GTT TTT CAT TAC CAC TCC AGA TTT TGC TGT TC
Jw205	ATG AAA AAC ATC AAA AAA AAC CAG GTA ATG
Jw206	TTT CAG GAA AGT TTC GGA GGA GAT AG
Jw207	TTA AGC AGC CAG AGC GTA GTT TTC GTC GTT AGC AGC TTT CAG GAA AGT TTC GGA GGA GAT AG
Jw208	CTT GCG CCC TGA GTG CTT GCC CCC AGC TGG CAA TTC CGA C
Jw209	TCG CCT TCT TGA CGA GTT CTT CTA AGC GGG ACT CTG GGG TTC GAG AG
Jw210	CTT GCG CCC TGA GTG CTT GCA CGT CTC ATT TTC GCC AGA TAT C
Jw211	TCG CCT TCT TGA CGA GTT CTT CTA ATT TGA TAT CGA GCT CGC TTG GAC
Jw212	TTA GAA GAA CTC GTC AAG AAG GCG
Jw213	GCA AGC ACT CAG GGC GCA AG
Jw245	TTC ATG AAA TCT ATA AAT TAA AGA TTT GTC ACT TAT TGG ATT TAG TAT GCG CTT TCG CTA AGG ATG ATT TCT GG
Jw246	TAA ATG ACA TTA GAC TTA ATA ATA ACA ATA ATT GTC CGC CAT TTA AAT CCG TGT AGG CTG GAG CTG CTT C
Jw358	CAT GGC TGA TGC AAT GCG GCG
Jw359	CGC CGC ATT GCA TCA GCC ATG

REFERENCES

[1] C. Wyman and R. Kanaar, *DNA Double-Strand Break Repair: All's Well that Ends Well,* Annual Review of Genetics **40**, 363 (2006).

[2] S. Shuman and M. S. Glickman, *Bacterial DNA repair by non-homologous end joining,* Nat Rev Micro **5**, 852 (2007).

[3] P. Sung and H. Klein, *Mechanism of homologous recombination: mediators and helicases take on regulatory functions.* Nature reviews. Molecular cell biology **7**, 739 (2006).

[4] J. Y. Lee, T. Terakawa, Z. Qi, J. B. Steinfeld, S. Redding, Y. Kwon, W. A. Gaines, W. Zhao, P. Sung, and E. C. Greene, *Base triplet stepping by the Rad51/RecA family of recombinases,* Science **349**, 977 (2015).

[5] S. T. Lam, M. M. Stahl, K. D. McMilin, and F. W. Stahl, *Rec mediated recombinational hot spot activity in bacteriophage lambda. II. A mutation which causes hot spot activity,* Genetics **77**, 425 (1974).

[6] M. R. Singleton, M. S. Dillingham, M. Gaudier, S. C. Kowalczykowski, and D. B. Wigley, *Crystal structure of RecBCD enzyme reveals a machine for processing DNA breaks,* Nature **432**, 187 (2004).

[7] D. A. Dixon and S. C. Kowalczykowski, *The recombination hotspot χ is a regulatory sequence that acts by attenuating the nuclease activity of the E. coli RecBCD enzyme,* Cell **73**, 87 (1993).

[8] A. S. Ponticelli, D. W. Schultz, A. F. Taylor, and G. R. Smith, *Chi-dependent DNA strand cleavage by RecBC enzyme.* Cell **41**, 145 (1985).

[9] G. R. Smith, *How RecBCD Enzyme and Chi Promote DNA Break Repair and Recombination: a Molecular Biologist's View,* Microbiology and Molecular Biology Reviews **76**, 217 (2012).

[10] A. L. Forget and S. C. Kowalczykowski, *Single-molecule imaging of DNA pairing by RecA reveals a three-dimensional homology search,* Nature **482**, 423 (2012), arXiv:NIHMS150003 .

[11] M. M. Cox, M. F. Goodman, K. N. Kreuzer, D. J. Sherratt, S. J. Sandler, and K. J. Marians, *The importance of repairing stalled replication forks,* Nature **404**, 37 (2000).

[12] A. Kuzminov, *Collapse and repair of replication forks in Escherichia coli.* Molecular microbiology **16**, 373 (1995).

[13] C. Lesterlin, G. Ball, L. Schermelleh, and D. J. Sherratt, *RecA bundles mediate homology pairing between distant sisters during DNA break repair.* Nature **506**, 249 (2014).

[14] Y. Bellaiche, V. Mogila, and N. Perrimon, *I-SceI endonuclease, a new tool for studying DNA double-strand break repair mechanisms in Drosophila,* Genetics **152**, 1037 (1999).

[15] H. J. Nielsen, J. R. Ottesen, B. Youngren, S. J. Austin, and F. G. Hansen, *The Escherichia coli chromosome is organized with the left and right chromosome arms in separate cell halves,* Molecular Microbiology **62**, 331 (2006).

[16] A. Tal, R. Arbel-Goren, N. Costantino, D. L. Court, and J. Stavans, *Location of the unique integration site on an Escherichia coli chromosome by bacteriophage lambda DNA in vivo.* Proceedings of the National Academy of Sciences of the United States of America **111**, 7308 (2014).

[17] M. S. Dillingham and S. C. Kowalczykowski, *RecBCD Enzyme and the Repair of Double-Stranded DNA Breaks,* Microbiology and Molecular Biology Reviews **72**, 642 (2008).

[18] C. Monteilhet, a. Perrin, a. Thierry, L. Colleaux, and B. Dujon, *Purification and characterization of the in vitro activity of I-Sce I, a novel and highly specific endonuclease encoded by a group I intron.* Nucleic acids research **18**, 1407 (1990).

[19] K. C. Keiler, P. R. H. Waller, and R. T. Sauer, *Role of a Peptide Tagging System in Degradation of Proteins Synthesized from Damaged Messenger RNA,* Science **16**, 990 (1996).

[20] X. Wang, X. Liu, C. Possoz, and D. Sherratt J., *The two Escherichia coli chromosome arms locate to separate cell halves,* Genes & Development **20**, 1727 (2006).

[21] B. Liu, R. J. Baskin, and S. C. Kowalczykowski, *DNA unwinding heterogeneity by RecBCD results from static molecules able to equilibrate,* Nature **500**, 482 (2013).

[22] P. R. Bianco, L. R. Brewer, M. Corzett, R. Balhorn, Y. Yeh, S. C. Kowalczykowski, and R. J. Baskin, *Processive translocation and DNA unwinding by individual RecBCD enzyme molecules.* Nature **409**, 374 (2001).

[23] J. M. Schoemaker, R. C. Gayda, and A. Markovitz, *Regulation of cell division in Escherichia coli: SOS induction and cellular location of the SulA protein, a key to lon-associated filamentation and death,* Journal of Bacteriology **158**, 551 (1984).

[24] J. Wiktor, C. Lesterlin, D. J. Sherratt, and C. Dekker, *CRISPR-mediated control of the bacterial initiation of replication,* Nucleic Acids Research **44**, 3801 (2016), arXiv:arXiv:1011.1669v3 .

[25] S. Uphoff, R. Reyes-Lamothe, F. G. de Leon, D. J. Sherratt, and A. N. Kapanidis, *Single-molecule DNA repair in live bacteria,* Proceedings of the National Academy of Sciences **110**, 8063 (2013).

[26] M. Marbouty, A. Le Gall, D. I. Cattoni, A. Cournac, A. Koh, J.-B. Fiche, J. Mozziconacci, H. Murray, R. Koszul, and M. Nollmann, *Condensin- and Replication-Mediated Bacterial Chromosome Folding and Origin Condensation Revealed by Hi-C and Super-resolution Imaging.* Molecular cell **59**, 588 (2015).

[27] M. Spies, P. R. Bianco, M. S. Dillingham, N. Handa, R. J. Baskin, and S. C. Kowalczykowski, *A molecular throttle: The recombination hotspot chi controls DNA translocation by the RecBCD helicase,* Cell **114**, 647 (2003).

[28] I. J. Finkelstein, M.-L. Visnapuu, and E. C. Greene, *Single-molecule imaging reveals mechanisms of protein disruption by a DNA translocase.* Nature **468**, 983 (2010), arXiv:NIHMS150003 .

[29] C. Zierhut and J. F. X. Diffley, *Break dosage, cell cycle stage and DNA replication influence DNA double strand break response,* The EMBO Journal **27**, 1875 (2008).

[30] C. B. Russell, D. S. Thaler, and F. W. Dahlquist, *Chromosomal transformation of Escherichia coli recD strains with linearized plasmids,* Journal of Bacteriology **171**, 2609 (1989).

[31] J. J. Churchill, D. G. Anderson, and S. C. Kowalczykowski, *The RecBC enzyme loads recA protein onto ssDNA asymmetrically and independently of χ, resulting in constitutive recombination activation,* Genes and Development **13**, 901 (1999).

[32] S. T. Lovett, *The DNA exonucleases of Escherichia coli,* EcoSal Plus **4**, 10.1128/ecosalplus.4.4.7 (2011).

[33] R. Anand, A. Beach, K. Li, and J. Haber, *Rad51-mediated double-strand break repair and mismatch correction of divergent substrates,* Nature **544**, 377 (2017).

[34] H. Ochman, J. G. Lawrence, and E. A. Groisman, *Lateral gene transfer and the nature of bacterial innovation,* Nature **405**, 299 (2000).

[35] A. Babić, A. B. Lindner, M. Vulić, E. J. Stewart, and M. Radman, *Direct Visualization of Horizontal Gene Transfer,* Science **319**, 1533 (2008).

[36] Z. Zhu, W. H. Chung, E. Y. Shim, S. E. Lee, and G. Ira, *Sgs1 Helicase and Two Nucleases Dna2 and Exo1 Resect DNA Double-Strand Break Ends,* Cell **134**, 981 (2008).

[37] E. P. Mimitou, S. Yamada, and S. Keeney, *A global view of meiotic double-strand break end resection,* Science **355**, 40 LP (2017).

[38] M. Spies, I. Amitani, R. J. Baskin, and S. C. Kowalczykowski, *RecBCD Enzyme Switches Lead Motor Subunits in Response to χ Recognition,* Cell **131**, 694 (2007).

[39] S. C. Kowalczykowski, *Initiation of genetic recombination and recombination-dependent replication.* Trends in biochemical sciences **25**, 156 (2000).

[40] W. H. Chung, Z. Zhu, A. Papusha, A. Malkova, and G. Ira, *Defective resection at DNA double-strand breaks leads to de Novo telomere formation and enhances gene targeting,* PLoS Genetics **6**, 24 (2010).

[41] S. Sinha and R. J. Redfield, *Natural DNA Uptake by Escherichia coli,* PLOS ONE **7**, e35620 (2012).

[42] T. G. Bernhardt and P. A. De Boer, *The Escherichia coli amidase AmiC is a periplasmic septal ring component exported via the twin-arginine transport pathway,* Molecular microbiology **48**, 1171 (2003).

[43] K. A. Datsenko and B. L. Wanner, *One-step inactivation of chromosomal genes in Escherichia coli K-12 using PCR products,* Proceedings of the National Academy of Sciences **97**, 6640 (2000).

[44] O. Espeli, R. Mercier, and F. Boccard, *DNA dynamics vary according to macrodomain topography in the E. coli chromosome,* Molecular Microbiology **68**, 1418 (2008).

[45] M. A. White, J. K. Eykelenboom, M. A. Lopez-Vernaza, E. Wilson, and D. R. F. Leach, *Non-random segregation of sister chromosomes in Escherichia coli,* Nature **455**, 1248 (2008).

[46] A. Paintdakhi, B. Parry, M. Campos, I. Irnov, J. Elf, I. Surovtsev, and C. Jacobs-Wagner, *Oufti: An integrated software package for high-accuracy, high-throughput quantitative microscopy analysis,* Molecular Microbiology **99**, 767 (2016), arXiv:arXiv:1011.1669v3 .

[47] J. Schindelin, I. Arganda-Carreras, E. Frise, V. Kaynig, M. Longair, T. Pietzsch, S. Preibisch, C. Rueden, S. Saalfeld, B. Schmid, J.-Y. Tinevez, D. J. White, V. Hartenstein, K. Eliceiri, P. Tomancak, and A. Cardona, *Fiji: an open-source platform for biological-image analysis,* Nature Methods **9**, 676 (2012).

[48] B. Azeroglu, J. S. P. Mawer, C. A. Cockram, M. A. White, A. M. M. Hasan, M. Filatenkova, and D. R. F. Leach, *RecG Directs DNA Synthesis during Double-Strand Break Repair,* PLoS Genetics **12**, 1 (2016).

3

SISTER CHROMOSOME CONDENSATION AT THE MIDCELL FACILITATES THE HOMOLOGY SEARCH DURING DOUBLE STRANDED BREAK REPAIR

Homologous recombination (HR) is central for maintaining the genomic integrity, that is often threatened by the formation of double stranded breaks. HR involves a process where physical interactions between the broken strand of DNA and the repair template restore the information lost during the formation of breaks. Copying information from undamaged strands ensures an accurate recovery of information located on damaged strands. How those two homologies find each other in the complex physical and genetic space of a cell remains one of the frontier questions in cell biology. Here, we develop a fluorescence and genetics based approach to confront a hypothesis of a global random search, commonly depicted as the 'needle in a haystack' problem. We find no evidence for this model, and we propose instead that homology search is assisted by relocation of sister loci that brings homologies to the close proximity and highly reduces the search complexity. This notion is supported by a recombination reporter system, which shows that chromosomal double stranded breaks can be repaired with HR sites on freely diffusing plasmids, but not with an ectopic one, constrained by genomic organisation. Together, these results argue against the current homology search model and call for a new mechanism for homology search.

3.1. INTRODUCTION

Homologous recombination (HR) is a wide-spread molecular mechanism found in all branches of life [1]. The two most important functions of HR are: (i) DNA repair of double stranded breaks (DSB) [2], and (ii) rescue of stalled replication forks [3]. Furthermore, in some organisms, HR is involved in the crossing-over reaction during sexual reproduction, contributing to the increase in genetic diversity within species [4]. The versatility of the HR reaction arises from its fundamental principle – it facilitates a physical pairing and exchange of genetic information between two identical homologous sequences [2, 5, 6]. This pairing can be used to recover the information lost during DNA damage, when HR is repairing the DNA, or can be used to exchange parts of homologous chromosomes during crossing-over, using a specific resolution of Holliday junction [1, 2, 5, 6]. Independently of which pathway HR is involved in, the reaction implies an efficient localisation of these homologous sequences within the volume of the cell. Since early discoveries of the genetic recombination it was suggested that HR involves a process of search for homology [7, 8] but experimental evidence elucidating the mechanistic principles of search is still lacking. Our understanding of HR is limited to descriptions of its early steps – end-resection and loading of recombination proteins (RecA/Rad51/RadA), and late steps – resolution of joint molecules [9], but the search process still remains elusive.

The problem of homology search was studied in many outstanding *in vitro* experimental designs [7, 10–13]. In the most straightforward, biochemical experiment the product of homologous pairing of DNA molecules is observed on the agarose gel in the presence of recombination protein, RecA from *Escherichia coli* bacterium [14], or Rad51, originating from yeast *Saccharomyces cerevisiae* [15]. These experiments showed that homologous DNA together with only RecA protein (or its homolog) is enough to perform an efficient search and pairing. Homologous pairing was also directly visualised using single-molecule techniques, such as magnetic-tweezers [11, 16], optical-tweezer [7] or DNA curtains [10] again confirmed that the presence of homology and RecA was sufficient to establish pairing between molecules. Translation of such *in vitro* findings into the context of a cell suggests that the homologous search is controlled by an unidentified mechanism where ssDNA coated with RecA scans the entire cell volume in search for a suitable homologous sequence [7, 10, 13].

Experiments focused on homologous recombination within cells yielded less convincing conclusions regarding the search process. Early discoveries showed that homologous DNA transformed into cells can be processed and integrated into chromosome by the recombination machinery both in eukaryotic [17] and prokaryotic cells [18–20]. Further progress in molecular biology techniques allowed scientists to gain more direct observations of the process of recombination, using fluorescent microscopy [12, 21, 22], PCR [23] and ChIP-seq [24]. Probably the most prominent examples of direct observations of the homology pairing come from studies of fluorescently labelled chromosomal positions of double stranded breaks in yeast [21] and in *E. coli* [12]. In both of those studies, endonucleolitic induction of site-specific DSBs promoted contacts between homologous sites, but the molecular reactions leading to those colocalisation remained unidentified and specific to the tested organism. In *E. coli*, the DSB site moved to the position of the sister chromatid with a directed trajectory [12]. In *S. cerevisiae* cells, such

a directionality was not observed, but another interesting mechanism was involved in the DSB response. DNA damage induced degradation of histones [25], and as a consequence, chromatin liberated from constrains of the genome organisation was more mobile and scanned a larger volume of the nucleus [22]. Such an increased mobility could support a random search, but the broken foci were able to scan only a limited area and not the entire nucleus. ChIP-seq experiments in *S. cerevisiae* showed that Rad51 (RecA homologue) spreads in the proximity of DSB site on the same chromosome and the intensity of the ChIP-seq signal is highest in at the sites where chromosomes interact with each other [24]. Another study, which monitored the amount of recombination product, revealed that the efficiency of the search process drops as the physical distance between homologies increases [23]. Bacterial chromosome condenses in the response to genotoxic stress [26], what may constitute of a mechanism reducing the distance between homologies, thus facilitating an easier search. However, it remains unknown if the condensation was general, or specific to damaged sites. Based on the current results it is particular difficult to conclude to what extent the homology search is driven by random diffusion, but the lack of clear evidence may suggest that another process leads to finding of homology.

Here we provide experimental evidence to test the well-established random-search hypothesis. First, we show that the induction of a DSB does not transmit to changes in cellular positioning and pairing of distant *loci* into which we insert a synthetic homologous repair site. Next, by monitoring foci in the close proximity to a DSB site, we show that, as a response to DNA damage, broken sites localise into the middle of the cell, decreasing the physical distance between homologous sites. By investigating the levels of homologous-recombination products we find that pairings between plasmids and chromosome are much more frequent than inter-chromosomal pairing. We propose that the HR is restricted by the architecture of bacterial genome, and therefore contacts between *loci* placed on one chromosome are too rare to facilitate an efficient whole-chromosome search for homology.

3.2. RESULTS

ECTOPIC RECOMBINATION ON THE BACTERIAL CHROMOSOME

In an attempt to capture the pairing between homologies in a live cell, we developed an *E. coli* strain with a fluorescently labelled inducible DSB site and a fluorescently labelled ectopic homology site. We first placed an L-arabinose inducible I-SceI [27] cut site (CS), flanked by regions of homology and a *parS*/ParB-YFP marker [28] into the *codA* locus of the *E. coli* chromosome (Fig. 3.1a). Next, we made combinations of the CS strain with two strains carrying the I-SceI resistant repair site, flanked by homologies related to the CS and marked by the *parS*/ParB-CFP focus at two distinct genetic positions (RS1 - *ydeO* gene and RS2 – *yeiU*, Fig. 3.1a). Analysis of the foci frequency and positioning showed that the native chromosome organisation is maintained in strains carrying the genetic modifications. Integrated foci occupy defined cellular volumes, consistent with previously published data describing the bacterial chromosome [29]. On a circular chromosome the frequency for occurrence of a genomic marker is related to its genetic distance from the origin of replication. The cut site, which was integrated approximately 1 Mb

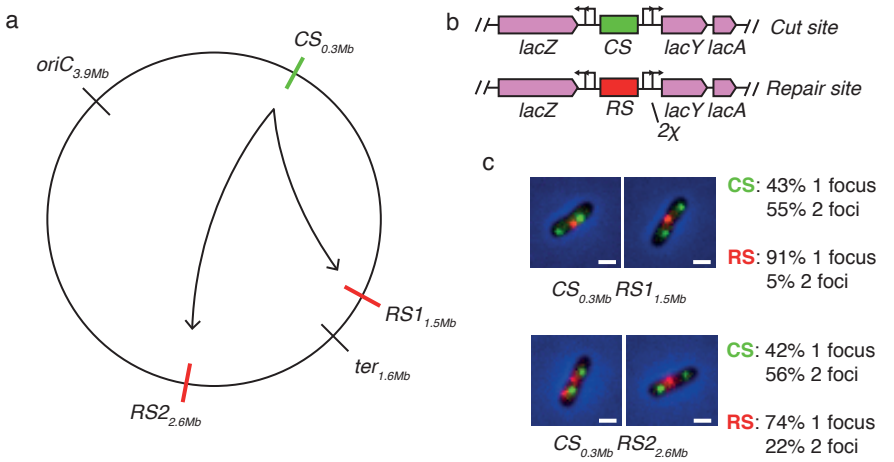


Figure 3.1: **Visualisation of the homology search.** (a) Cartoon representing the *E. coli* chromosome with the positions of origin of replication (*oriC*) and termination of replication (ter) indicated. An inducible cut site (CS) is represented by the green line, and the positions of two repair sites (RS1 and RS2) are shown with a red line. Basepair position on the circular chromosome are also shown. The cut site can be repaired either using the uncut sister or the synthetic repair-sites RS1, or RS2 (arrows). (b) Schematic showing the homologies shared between the CS and the RS, constructs differ only in a short middle fragment. CS and RS are flanked by 2 χ sites at each end. The total length of the homologous regions is 3 kb (pink). (c) Representative cells of two strains that show the positioning of CS (green) and RS (red) ParB foci in the cell. The number of foci in the cell population scales inversely with the distance from *oriC*: the percentages show fraction of cells containing 1 or 2 of each foci. Scale bar is 1 μm.

from the origin was replicated in about half of the cells in growing population (Fig. 3.1c). Repair site 1, 2.2 Mb away from the origin and repair-site 2, 1.6 Mb from the origin, however, were present predominantly as a single copy in the cell.

We hypothesised that after induction of a DSB, cut-site could be repaired either with an intact sister chromosome, or with the synthetic repair site. Repair with the RS would be manifested through a co-localisation of YFP and CFP foci in the cell, while lack of such a co-localisation would suggest that the repair with an ectopic homology did not occur. We induced the formation of DSBs for 1 hour and then analysed cells using fluorescent microscopy. From phase-contrast and fluorescent images, we extracted positions of cells and foci, and calculated the distances between the closest YFP and mCherry markers in each cell. Surprisingly, induction of double stranded breaks did not lead to drastic changes in the distribution of foci distances (Fig. 3.2a). We found that the distribution of distances in the RS2 strain was unchanged, while the RS1 strain was shifted towards shorter lengths which could suggest co-localisation (Fig. 3.2a). Induced culture of RS1 cells was characterised by an increased fraction of cells with 2 CFP foci (23 % induced vs 5.2 % uninduced). Replicated foci in *E. coli* are segregated to opposite cell halves, bringing CFP and YFP spots to a closer proximity, independently of the repair mechanism. Hence, replication of CFP foci is the simplest explanation of shift of the distribution for

that strain.

The general organisation of the chromosome was also insensitive to the DSB induction. Positioning of chromosome-related foci in a population of growing cells is a good approximation of the state of chromosome. We found that the positions of CS foci were the same before and after induction of the DSB (Fig. 3.2c and 3.8). Overlay of the mean positions of YFP spots shows almost perfect overlay for both strains (Fig. 3.2c). The lack of clear pairing of homologous sequences placed on the same chromosome conflicts with the picture of a global search process during the DSB repair. The lack of DSB-induced co-localisations suggests that the DNA breaks in *E. coli* are not efficiently repaired with an ectopic homology. Note that the length of the homology used in our experiments was sufficient to mediate the recombination. The minimal length of homologies estimated in experiments with recombination of PCR fragments was 1 kb while here we provided 1.5 kb of homologies on CS and on RS. Another factor which can mask the co-localisation is the process of end resection. We showed before that end resection can degrade fluorescent foci very efficiently after the formation of DSBs (**Chapter 2**), and in the case that end-resection machinery degraded ParB signal, we could not observe co-localisations under the microscope.

Next we asked whether the presence of ectopic homology site had an influence on cell survival after formation of DSBs. We induced DSBs in cells growing in minimal media supplemented with glycerol. In those conditions, addition of L-arabinose fully induces the arabinose promoter. As result, I-SceI will cut all copies of CS, and the repair can be finished only using the I-SceI-resistant RS. After 4h induction of DSBs, we compared the number of viable cells in cultures in which we induced DSBs, as well as in an un-induced control. By comparing a strain carrying only the cut-site to strains that additionally contained ectopic homologies we found that, surprisingly, the ectopic homology does not increase the DSB resistance of cells. We conclude that ectopic recombination did not happen in the strains. All cultures showed about a 3-fold reduction in the number of viable cells after addition of L-arabinose, regardless of presence of presence of additional homology (Fig. 3.2b).

DSBS CONDENSE AT THE MIDCELL POSITION IN RESPONSE TO DSB

In further attempts to visualise the process of homologous pairing during DSB repair, we integrated *parSP1* sites into distant, upstream (e.g. in the direction of *oriC*) regions of the DSB site. Using the mCherry-ParB-P1 fusion protein, we monitored these markers in cells where DSBs were induced. In previous work we showed that a fraction of 30 kb and almost all 90 kb markers upstream of the *codA* DSB are resilient to RecBCD end resection (**Chapter 2**). The stability of the upstream marker makes them an attractive tool to study the spatiotemporal dynamics of the chromosome during the repair. For these experiments, we used cells which, besides 30 and 80 kb markers, also carried aTc inducible I- SceI-coding plasmid and a YFP marker in the proximity of DSB site, at the *codA* locus (**Chapter 2**). Time-resolved analysis of cells with a DSBs (indicated by a loss of the YFP foci) showed an unexpected condensation of upstream foci as multiple segregated mCherry spots localised into middle position of the cell (Fig. 3.3). This midcell positioning was maintained until eventual disappearance of foci due to end resection (Fig. 3.3a). It has been reported that mCherry-tag can induce spontaneous aggregation of proteins

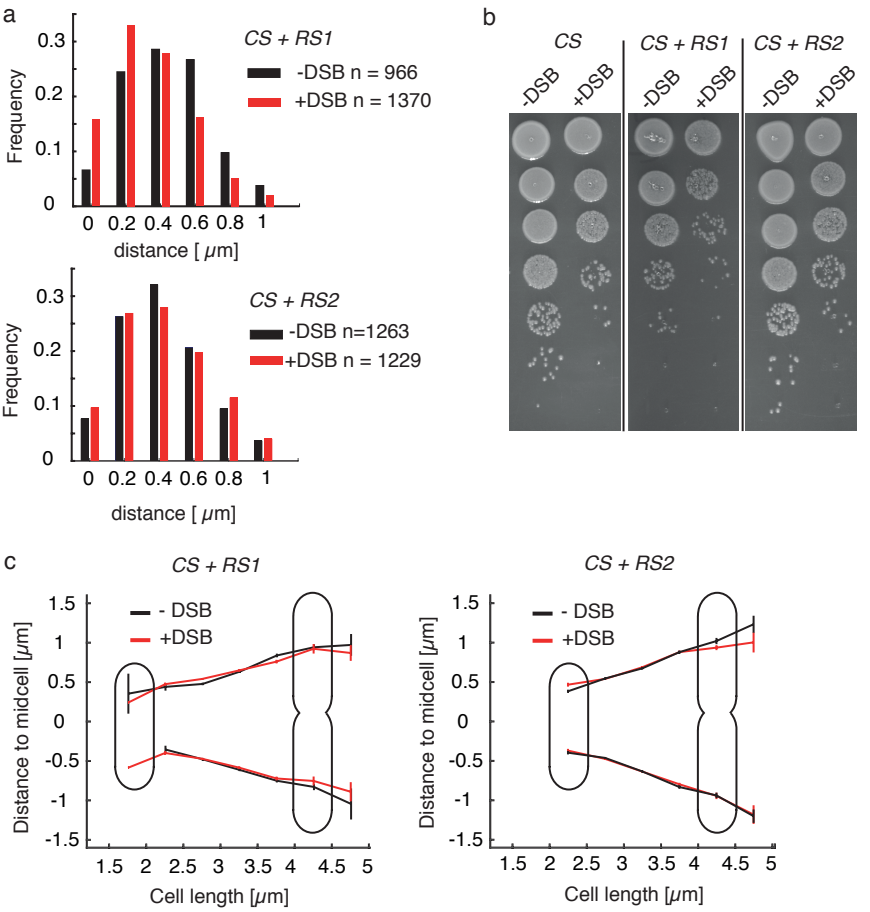


Figure 3.2: **Limited recombination with the ectopic homology.** (a) Distribution of the shortest distances between the CD and the RS foci in the cell, as calculated from microscope images of un-induced (-DSB, black) and induced (+DSB, red) cultures. The number of analysed cells is shown next to the distribution. (b) Survival assay shows no survival advantage after DSB formation on strains carrying the ectopic homology. 10-fold serial dilutions were plated from induced (+DSB) and non-induced (-DSB) cultures (c) Average position of the CS foci in damaged (red) and undamaged (black) cells are shown as a function of cell lengths. The complete dataset is shown in Fig. 3.8. Error bars indicate SEM. The same number of cells was analysed as indicated in (b).

when expressed in the cell [30], creating microscopic artifacts and suggesting false conclusions about sub-cellular organisation of proteins. To exclude that the condensation observed in our experiment was caused by such an aggregation mechanism, we swapped the fluorophores fused to ParB proteins and created a strain in which the proximal DSB site was marked by mCherry and distant, upstream markers were labelled with YFP. Independently of the used fluorescent fusions, upstream markers condensed into the midcell position in response to DSB formation, suggesting the existence of a mechanism fusing the broken chromosomes to the middle of the cell. Pairing of sister foci was reported previously for I-SceI-induced DSBs in *E. coli* [12] and in *Caulobacter cresentus* [31], and was explained as a consequence of homology search and strand invasion assisted by the RecA protein. However, we favour the mechanism which translocates DSB regions into the middle of the cells may be distinct from RecA homology pairing. Based on *in vitro* estimates [10], homology pairing and strand invasion of short homologous molecules can result in minutes-long stable binding. In the case of the mCherry foci pairing observed in our experiments, condensation/segregation cycles happened on faster timescales in contrast with the idea that the pairing is the result of strand invasion (Fig. 3.3b).

CONSTRUCTION OF FLUORESCENT REPORTER FOR HOMOLOGOUS RECOMBINATION

An alternative approach to probe the mechanism and efficiency of homologous recombination in live cells is a reporter system which responds in a countable phenotype as a result of homologous recombination [5, 32]. We designed a new system consisting of the *mCherry* gene flanked by arrays of 3 χ sites on both sides, 3 kb regions of homologies, and a *lac* promoter, driving the expression of the whole synthetic operon (Fig. 3.4a). Next, we interrupted the *mCherry* gene with a short region carrying the I-SceI recognition sequence, and a TAA stop codon, rendering the construct mCherry-negative. We hypothesized that providing a functional ectopic copy of the *mCherry* gene, flanked by the same regions of homology, but lacking a promoter, would result in a mCherry-positive phenotype only after the repair of I-SceI induced DSBs with that synthetic homology (Fig. 3.4b). We integrated the cut-site construct in the position of *codA* locus (1 Mb from *oriC*) of the *E. coli* chromosome, and the repair-site onto a high-copy ColE1-origin plasmid or into the *sseU* locus (0.3 Mb from *codA* locus) of the chromosome. The I-SceI expression system was integrated onto low copy SC101-origin plasmids (Fig. 3.4a). We thus designed a set of experiments in which measurements of the mCherry signal in the population after induction of DSBs provided a quantitative assessment of homologous recombination efficiency.

REPAIR OF DSBS WITH A PLASMID TEMPLATE

To obtain insights into the HR efficiency and to test the performance of our reporter system, we first asked the question whether chromosomal DSB in *E. coli* can be repaired with a homology located on a high-copy number plasmid. We constructed strains carrying a complete reporter system: (i) a strain with a chromosomal cut-site, a plasmid homology, and an aTc inducible I-SceI expression system, (ii) a control strain lacking plasmid homology, and (iii) a control strain lacking the I-SceI expression plasmid. Induction of the DSB in the cells resulted in a phenotype characteristic to the activation of DNA-damage response – the SOS pathway [33]. Activation of the SOS response triggers expression of genes involved in the survival of genotoxic stress. One of the products these

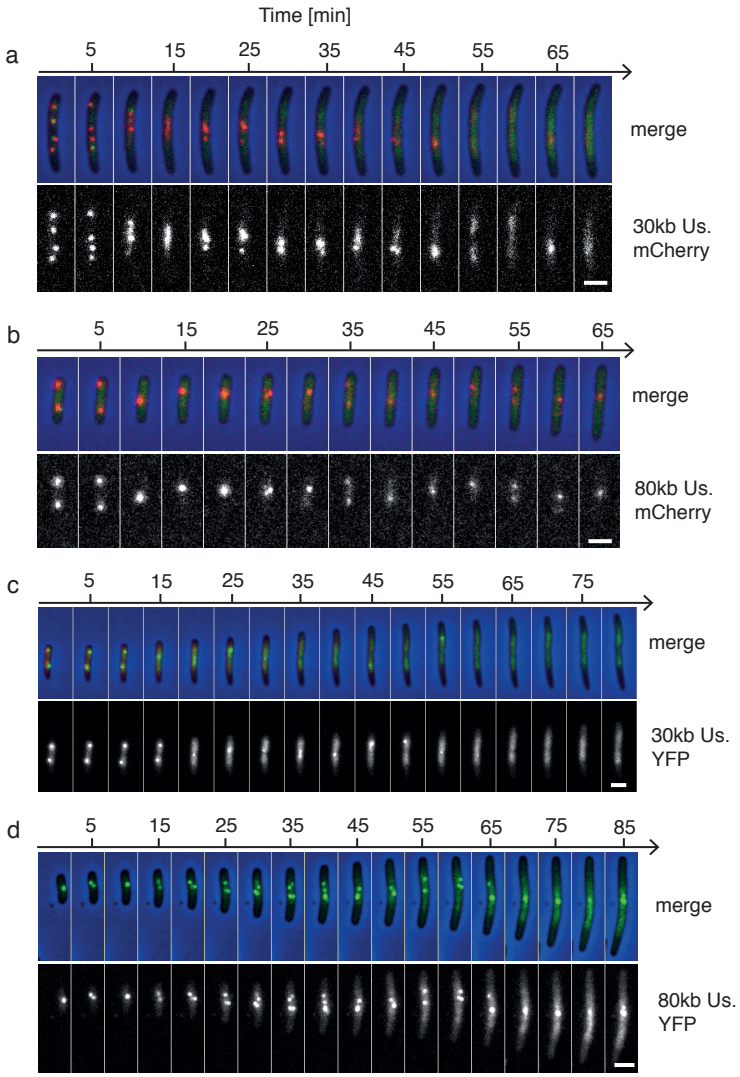


Figure 3.3: **DSB-induced condensation of the DNA regions upstream to the DSB.** Replicated and segregated mCherry foci are undergoing rapid translocation to the midcell position in response to the formation of DSB and induction of the repair. Representative cells undergoing sister foci condensation in response to DSB induction and degradation of the DSB proximal marker (loss of green foci in (a) and (b)). Merged multi-colour images are displayed on top, together with the selected fluorescent channel showing behavior of foci upstream to the DSB region. Scale bar is 2 μm. (a) 30 kb upstream parSP1/ParBP1-mCherry, (b) 80 kb upstream parSP1/ParBP1-mCherry, (c) 30 kb upstream parSP1/ParBP1-YFP, (d) 80 kb upstream parSP1/ParBP1-YFP.

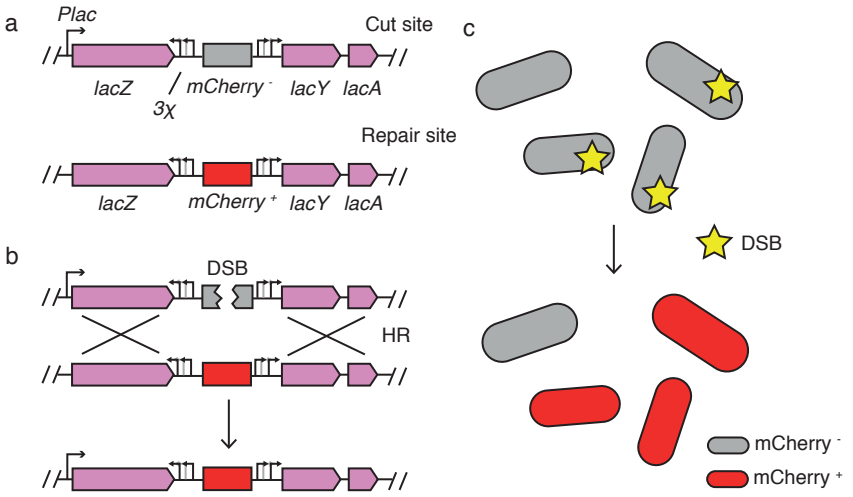


Figure 3.4: **Fluorescent reporter of homologous recombination.** (a) Cartoon showing cut site and repair site. The region coding for the mCherry gene in the cut site was interrupted by the I-SceI recognition site (grey). The lac promoter drives the expression of the entire cut site construct. An array of 3 χ sites flanks the I-SceI recognition site. Total size of homologous regions is 6 kb (purple) (b) Formation of a DSB induces homologous recombination and restores the functional mCherry sequence (red) in the expressed cassette. (c) Formation and repair of DSB (yellow star) with the reporter sequence results in the mCherry signal of individual cells in the population.

genes, the SulA protein, acts as a cell-division inhibitor [34], which postpones division until the DNA damage is removed and effectively occurrence of elongation of cells. We found that induction of DSBs results in presence of elongated bacteria (Fig. 3.5a, bottom-left and bottom-right), which are mostly absent in non-induced culture (Fig. 3.5a, upper row). Moreover, violin plots show that the median of the cell length in the culture increased only when I-SceI enzyme is produced in cells which carried the chromosomal I-SceI recognition sequence (Fig. 3.5c, black line).

Having confirmed that the induction of the I-SceI production creates DSBs on chromosome, we tested the difference in fractions of mCherry-positive cells before and after formation of DSBs. Consistent with our expectations the strain transformed with a plasmid homology rescued the chromosomal *mCherry* gene and the fraction of fluorescent cells was highly increased after induction of DSBs (Fig. 3.5a and b). Interestingly, mCherry signal was not correlated with the cell size (Fig. 3.5 and 3.9). Biochemical experiments show that the number of homologous pairings increases with time of the reaction, and elongated cells induced SOS-response earlier in the experiment than the short ones. Cells of the control strain missing the I-SceI expression system contained some fluorescent cells, but because their fraction was insensitive to the addition of aTc we attributed this fluorescence to leaky expression of mCherry from the high-copy number plasmid. The strain not supplemented with a plasmid homology, but with the I-SceI

recognition site and the I-SceI expression plasmid responded to the formation of DSBs with induction of SOS pathway, but not with changes in fluorescence.

To further confirm that the increase in fluorescence in the reporter system is caused by homologous recombination between the broken site and the repair site, we performed genotypic PCR on induced cells. We selected a primer pair that was sensitive to the product of recombination, that is, which hybridised with the *codA* locus and with the region specific to the repaired *mCherry* gene. Induced culture carrying the complete reporter system contained both primer binding sites and resulted in the formation of a band of the predicted length, demonstrating efficient repair (Fig. 3.10). These data show an application and the robustness of the recombination reporter system. Our results demonstrate that chromosomal DSB can be repaired with a template on a plasmid and that such recombination alters the sequence of the broken site.

CHROMOSOMAL DSBS ARE NOT REPAIRED WITH ECTOPIC HOMOLOGY

Having shown the activity of recombination reporter system on the plasmid, we set out to test whether such a system located on the chromosome can also be used to study the DSB repair. We have integrated the repair site into the *sseU* position (0.3 Mb from *codA*) and combined it with the *codA* cut-site used in previous experiment. The current view of the search process suggests that the ssDNA-RecA complex formed in response to a DSB does probe the volume of a cell to locate the homology [7, 10]. This implies that if we will be able to detect mCherry fluorescence in a number of cells, after the induction of DSBs. However, the result of the experiment was opposite from this expectations: induction of I-SceI expression in cells with an ectopic homology had no influence whatsoever on the intensity of the mCherry fluorescence using the same conditions as in previous experiment (Fig. 3.6a and b). Increasing concentration of aTc yielded a larger fraction of elongated cells, but yet again no increase in mCherry fluorescence was found. (Fig. 3.6b). The lack of recombination between the chromosomal ectopic homology and the chromosomal DSB can be explained by a specific mechanism facilitating sister contacts during the DSB repair. The broken site are likely driven to the proximity of the sister chromosome, similarly to the observed ParB foci condensation (Fig. 3.3). Plasmids, on the other hand, diffuse throughout the whole volume of the cell and accordingly, the broken site can easily contact the diffusing plasmid, and engage in homology recognition, strand invasion and repair.

3.3. DISCUSSION

WHOLE-GENOME SEARCH IS INHIBITED BY THE GENOME ORGANISATION

Chromosomal *loci* are characterised with a remarkably low diffusion coefficient, of only about $10^{-5}\mu m^2 s^{-1}$ [35]. Such low diffusivity of genomic regions is a consequence of genomic spatial organisation and the DNA supercoiling (Fig. 3.7a). A plethora of DNA-binding proteins, such as HU, SMCs, or H-NS or Fis are regulating the spatial structure of the chromosome by specific binding, transcription and translation, transcribed genes non-specifically anchor to macromolecular complexes, or to the membrane [6]. A consequence of all the physical constrains is a very low diffusion coefficient of chromosomal loci, and a very limited radius of gyration [36]. The time required for a diffusing particle to travel a distance is computed by the equation: $t = \frac{x^2}{2D}$, where t defines time, x is the

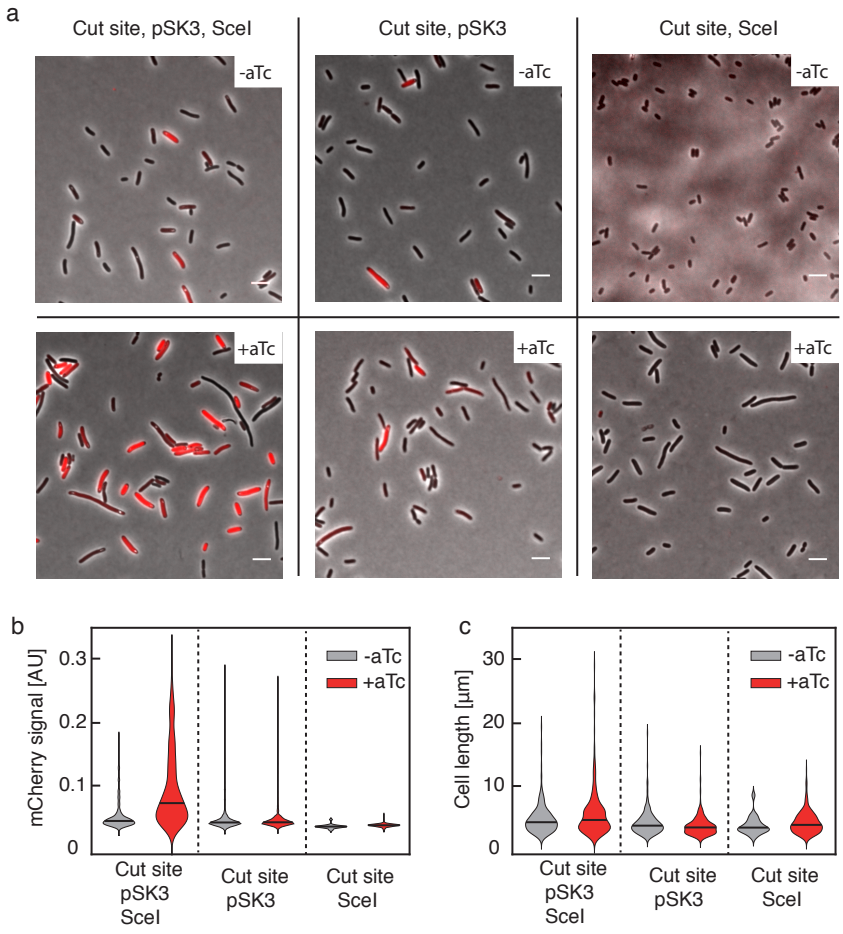


Figure 3.5: **Repair of chromosomal DSBs with a plasmid homology.** (a) Representative images of cells transformed with a complete fluorescent reporter system consisting of the chromosomal cut site, the plasmid carrying the repair template pSK3, and the plasmid coding I-SceI gene (or its incomplete variants, lacking on of two plasmids that consist for the full system), without DSB induction (top row) and after 4h of DSB induction (bottom). Grey – phase contrast channel, red – mCherry channel. Scale bar is 5 μm. (b) Violin plots showing distributions of the mCherry fluorescence signal (datasets correspond to cells shown in (a)). Black line represents median of the plotted values. (c) As in (b) but violin plots represent the distributions of the cell lengths.

distance and D is a diffusion coefficient. Considering that in the *E. coli* replicated and segregated foci can be separated by 1.5 μm, a process driven by random diffusion with $D = 10^{-5}\mu m^2 s^{-1}$ would take 112500 seconds, or 31 hours. In our experiments, ParB foci condensation after the break took less than 10 minutes, which is nowhere close to the value computed based on loci diffusion. This discrepancy of predicted and measured timescale is evidence for the existence of an alternative mechanisms assisting pairing of sister loci during DSB repair. One suggestion, originating from the yeast model, is that the induction of the DNA damage triggers the degradation of histones, what increases the mobility of chromosomal loci [22, 25]. Although such a mechanism can increase the number of contacts between homologies in a fixed unit of time, the mobility of *E. coli* loci should increase 100 times (to reduce t to just under 20 minutes) to satisfy reasonable DNA repair requirements. However, the diffusion coefficients of yeast loci after DSB induction increased only by a factor of 1.523. Another possible mechanism involves the re-cohesion of segregated regions. The SMC protein cohesin plays a crucial role in homologous recombination during DSB repair in eukaryotic cells, where sister chromosomes are kept in a close proximity by cohesin complexes [37, 38]. The presence of proximal, perfect homology drives double stranded breaks towards the HR repair pathway. On the other hand, right after cell division, when the sister homology is not present, cells repair DSBs mainly by means of non-homologous end joining [39]. Such a switch in pathway preference suggests that HR is most efficient when the homology is relatively close and the broken site does not need to scan the entire nucleoid for repair template. Conversely, when the sister chromosome is absent, and the presence of the homology is uncertain, cells engage into the error-prone non homologous end-joining pathway.

HOMOLOGOUS RECOMBINATION WITH ECTOPIC HOMOLOGIES

Recombination between a DNA site on a chromosome and a homology site present on a plasmid, an external linear DNA molecule, or a heterologous chromosome has been frequently used to study the biology of the DNA repair [23, 32] and genome modification [17, 19]. Here we developed a reporter system which upon the repair of a DSB by an ectopic homology signals a mCherry fluorescence in bacterial cells. Surprisingly, we found that plasmid homology can serve as a good template for chromosomal DSB repair, but the same homology integrated onto the chromosome was largely ignored by ssDNA-RecA during homology search. In haploid yeast cells, repair between homologies inserted into different chromosomes can repair efficiently if the repair template is positioned in a close proximity of the DSB, what argues against the global-search model [23]. Moreover, the well-studied *MAT* switching in *Saccharomyces cerevisiae* relies on homologous recombination between genetic elements, but involve chromosome looping, that brings the *MAT* homologies to a close proximity, making global search redundant [40].

We believe that the striking difference in recombination efficiency of plasmid-borne and ectopic chromosomal sites arises from a difference in the mobility of these different DNA structures and from sister-driven homology search. Although ColE1-type plasmids, which we used here to provide a repair template in cells, preferably localise at the cells poles but are not restrained to that position and can diffuse throughout the whole cell volume. Highly mobile molecules can thus interact with a DSB leading to the repair even if the mechanism predominantly favours sister chromosomes contacts. A single genomic DSB can be repaired by homologous recombination with a delivered tem-

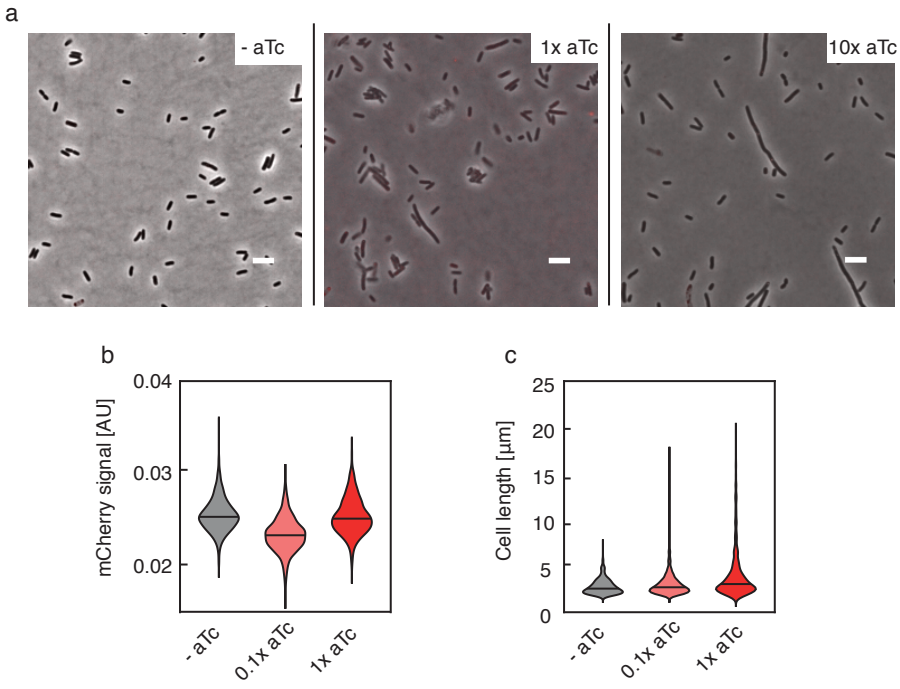


Figure 3.6: **Reporter system integrated into an ectopic position is not a source of repair homology.** (a) Representative images of populations harbouring reporter system, non-induced and induced with two concentrations of aTc. Grey – phase contrast channel, red – mCherry channel (same as in Fig. 3.5a, but here mCherry signal is almost undetected). Scale bar is 5 μm. (b) Violin plots showing distributions of mCherry fluorescence signal (in same populations as in (a)): -atc n = 751 cells, 1x aTc n = 1189 cells, 10x aTc n = 628 cells. Black line represents median of the plotted values. (c) Same as in (b) but violin plots show distributions of cell lengths.

plate in human cells [32], suggesting that the mechanism is conserved between different branches of life. We suggest that the ectopic chromosomal template fails to recombine with a chromosomal DSB site because of (i) a low mobility of the ectopic region, as it is constrained by genome organisation, and (ii) sister chromosome condensation which strongly favouring repair with a replicated template.

DSB REPAIR IN A BIOLOGICAL CONTEXT

To fully understand the nature of DSB repair and homologous search we have to consider the main source of those breaks. It is, in fact, unlikely that the common factors giving rise to DSBs are endonucleases, or gamma radiation, that are usually mentioned in the literature. Instead the vast majority of DSBs will be caused by the most central reaction for life, the replication of DNA. In growing cells, replisome stalling [41], conflicts with translational machinery [42], single-stranded gaps or crosslinks [41] can cause formation of the DSB at the replication fork. Experiments focusing on the spontaneous formation of DSBs in *E. coli* also suggested that active replication is the main source of DNA breaks in growing bacteria [43]. The recently replicated sister chromosomes would be present in a close proximity when a DSB was created at the replication fork. Such breaks would not necessarily scan the large volume of the cell, but would limit the search to a small confined territory. Indeed, low diffusivity and genome organisation are less important when the search happens in a small volume. Another genetic clue which implies repair in the proximity of the replication fork is the orientation of χ sites, which are 8bp, asymmetric DNA motifs (discussed in the **Chapter 2**) regulating the activity of *E. coli* end resection. On the chromosome, χ sites are biased towards an orientation that is stimulating the repair of origin-of-replication side of the DSB (**Chapter 2**). Repair of collapsed replication fork could easily provide an evolutionary pressure to distribute χ sites in such an orientation protecting the origin and neglecting the terminus site.

CONCLUSION

In this chapter we presented a collection of results, that mainly turned out to be negative. Starting with the idea that the homologous search is driven by diffusion and involves scanning of the entire cell volume we expected to capture that process with at least one of tested approaches, involving foci tracking, fluorescent reporters, and survival assays. Our results showed that the popular model of global homology search is inconsistent with the experimental data. Based on our results and the others [6, 12, 31, 44] we suggest a model in which DSBs are directed towards the sister chromosome, which is perfectly homologous to the broken site (Fig. 3.6b). The molecular mechanism of sister condensation is yet to be resolved, with evidence points at involvement of RecN protein [45]. RecN structurally resembles SMC family [46], that is well known for genome compaction and sister chromosomes interactions. RecN is also upregulated during the SOS-response, and genetically is necessary for DSB resistance [26, 46]. The idea of a specific sister driven homology search opens an exciting route towards understanding of the one of the most mysterious steps in the biology of recombination. The results presented in this work can serve as a continuation to such a line of research.

3.4. MATERIALS AND METHODS

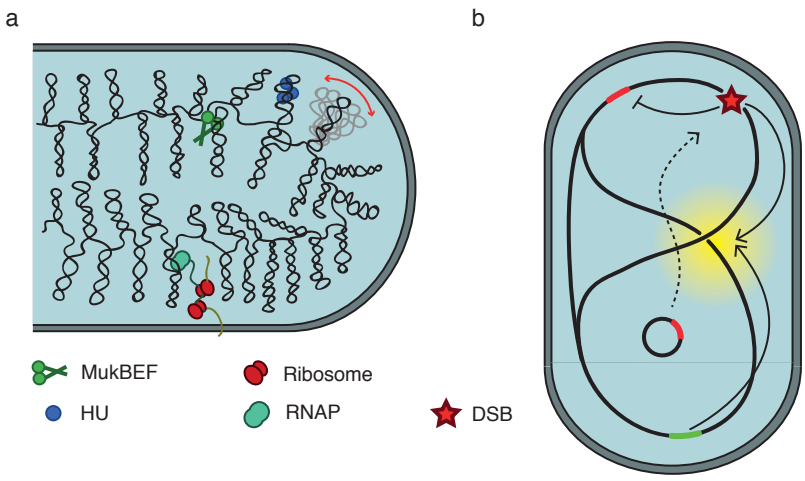


Figure 3.7: **DSB repair in the context of genome organisation.** (a) The bacterial chromosome is organised by multiple factors. Nucleoid-associated proteins (NAPs) such as H-NS, or HU bind and stabilise topological structures in the chromosome. Structural maintenance of chromosomes (SMC, MukBEF complex in E. coli) proteins are believed to form ring structures and induces condensation of distant loci. Ongoing transcription links genes to mRNA molecules, which subsequently are linked to ribosomes and translated peptides. Synthesis of transmembrane proteins may anchor the entire transcription body into the cell membrane, reducing the mobility of the chromosomal loci. Panel adapted from [6, 31] (b) Model of sister-chromosome-condensation assisted homology search. Bidirectional replication of the *E. coli* chromosome results in placement of sister loci in opposite cell halves. After formation of a DSB (red star), nuclear organisation inhibits interactions with the chromosomal ectopic homology (red). Condensation of sister loci into the midcell, however, reduces volume in which homology is present (green) and reduces spatial complexity of the search. Alternatively, a plasmid is a mobile DNA element, that is free from the constrains of the genome organisation and can engage in recombination with chromosomal DNA.

STRAIN CREATION

Strains and plasmids used in this study are listed in Table 3.1. Plasmids p2973 [47] was a gift from Joel Stavans laboratory, p2973switch was created by swapping regions coding fluorescent proteins using Gibson assembly [48]. Plasmids pFHC2973 [28] and pOX6 [12] were gifts from Christian Lesterlin and David Sherratt laboratory laboratory, pSK3 and pSC101SceI were created using Gibson assembly. Genomic integration were prepared using Lambda Red method [18], or P1 transduction to move genetic elements between strains.

IMAGING AND MICROSCOPE DATA ACQUISITION

Wide-field fluorescence microscopy was carried out using Nikon Ti-e inverted microscope equipped with 100x, 1.4NA PlanApo objective and Photometrics Cool-Snap HQ CCD and an environmental chamber (for foci co-localisation and condensation experiments) or CFI Plan Apochromat lambda DM 100x objective, Andor Zyla 4.2 CMOS camera and Lumencor Spectra X LED light source and an environmental chamber (for fluorescent reporter experiments).

For co-localisation experiment cells were grown over night in M9 minimal medium supplemented with glycerol and relevant antibiotics in 30°C and refreshed in fresh minimal medium in 1 to 50 ratio on the following day. When the culture reached early exponential stage 1ml was spun down and cells were re-suspended in minimal media supplemented with L-arabinose (0.2% final concentration), which induce expression of I-SceI from arabinose promoter. After 1h 5 μl of culture was transferred on 1% agarose-M9 pad and several random positions were imaged.

For foci condensation experiment cells were grown over night in M9 minimal media supplemented with glucose and antibiotics in 37°C. In the morning cultures were refreshed in 1 to 50 ratio in fresh media and grown until reached early exponential phase. Anhydrotetracycline (aTc) was added to the culture (1.6 μg/ml final concentration) to induce expression from tetracycline promoter. After 15 minutes of expression cells were transferred onto 1% agarose-M9 pad and imaged in 5 minutes time intervals. Fluorescent recombination reporter cells were grown in Lysogeny Broth (LB) supplemented with antibiotics overnight in 37°C, refreshed in the morning in 1 to 100 ratio in fresh LB media. When cultures reached early exponential stage aTc was added to the culture (4 μg/ml, or 40 μg/ml final concentration) for 4 hours prior imaging. Cultures were transferred onto 1% agarose-M9 pads several random positions were imaged.

IMAGE PROCESSING

Cell outlines and mCherry signal intensities were obtained using microbeTracker [49], or Oufti [50] software from phase contrast images. Centroids of fluorescent foci were localized using custom MATLAB function based on Crocker and Grier routine. Co-localizations were scored for the shortest distance in cells which had at least 1 cut-site foci and only 1 repair-site focus. Average positions of foci were obtained using moving window average. Violin plots were generated using *violin.m* function (Mathworks exchange File ID: #45134).

DSB SURVIVAL ASSAY

For serial dilution survival analysis cells were grown in M9 minimum media supplemented with glycerol and antibiotics in 30°C and DSB were induced by induction of L-arabinose to final concentration of 0.2 % for 4 hours prior to plating on LB media with agarose.

3.5. SUPPLEMENTARY INFORMATION

Table 3.1: List of strains and plasmids used in this study.

Strains	
TB28	MG1655 ΔlacZYA [51]
Jx015	TB28 CS::codA, RS1::ydeO
Jx017	TB28 CS::codA
Jx027	TB28 CS::codA, RS2::yeiU
Jx100	TB28 CS::codA, parSP1::30kb upstream codA
Jx101	TB28 CS::codA, parSP1::80kb upstream codA
Jx117	TB28 $CS_{mCherry}-$::codA
Jx127	TB28 $CS_{mCherry}-$::codA $RS_{mCherry}+$::sseU
Plasmids	
pOX6	I-SceI [12]
pFHC2973	parBMT1-YFP, parBP1-CFP [28]
p2973	parBMT1-YFP, parBP1-mCherry [47]
p2973switch	parBMT1-mCherry parBP1-YFP
pSK3	$RS_{mCherry}+$
pSC101SceI	SceI

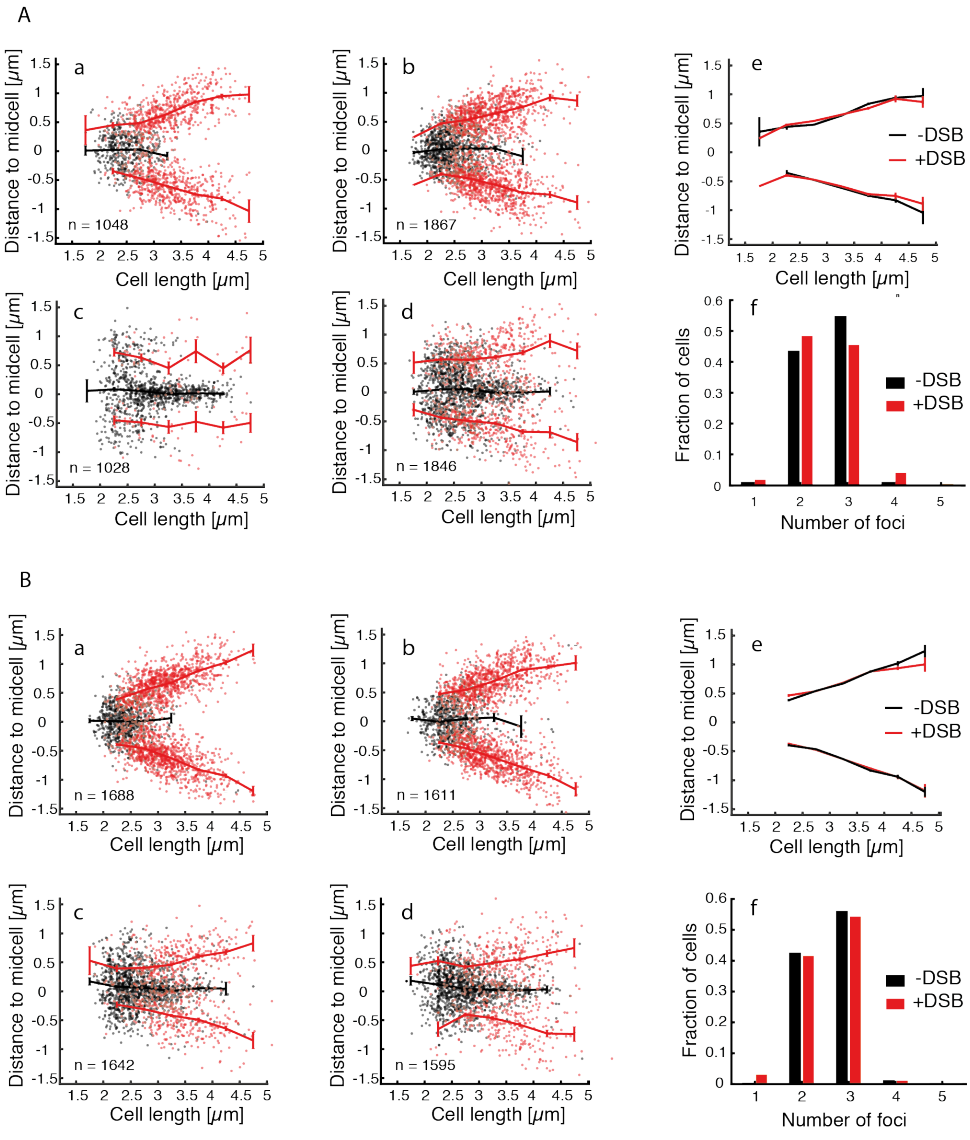


Figure 3.8: **Cut-site and repair-site localization in the cell.** (A) Snapshot analysis of positioning of cut-site foci in the CS RS1 strain before (a) and after (b) 1 h induction. Positions are divided into categories: black dots represent position of a single focus in the cell, red dots represent position of foci in a cell containing 2 CS foci. Solid lines represent averaged position, for cells with 2 foci averages were computed for separate cell halves. Number of analyzed cells containing at least 1 focus is displayed on each graph. Error bars represent SEM Corresponding positioning of RS foci is shown on (c) and (d). (e) Extracted average position of CS foci. (f) Distribution of foci number in cells before (-DSB) and after (+DSB) induction of breaks. (B) Shows corresponding analysis of CS RS2 strain.

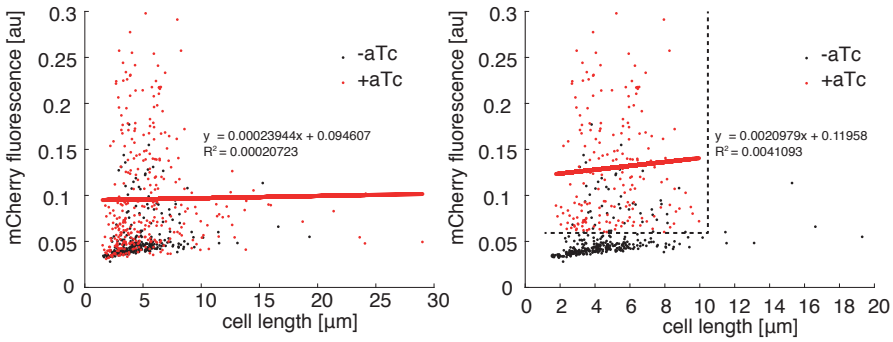


Figure 3.9: Relation between cell length and mCherry signal after induction of DSBs. Each data point is an individual cell, from populations analyzed in Fig. 3.4. Black dots represent cells from non-induced culture, red dots represent cells from culture induced for 4 h. Red line show fitted regression and values are shown as $y = ax + b$. R^2 values for each fit are shown. Right plot shows filtered distribution of induced cells, to refine fitting. Small values of R^2 indicate that the computed regressions poorly explain the variability of signal to cell size ratio.

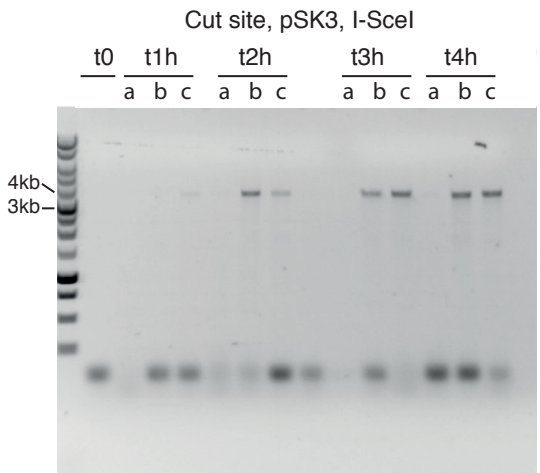


Figure 3.10: Repair template from the plasmid is copied onto the chromosome upon DSB repair. Liquid LB culture of cells was grown until early exponential phase and aliquoted into 3 sub-cultures induced with different concentration of aTc: a) non-induced, b) 4 μm/ml, c) 40 μm/ml. Each hour product of recombination between cut site and repair site was analyzed with PCR. Product of the recombination removes I-SceI recognition site, provides repair-specific primer binding site and results in formation of a 4 kb band.

REFERENCES

[1] C. Wyman and R. Kanaar, *DNA Double-Strand Break Repair: All's Well that Ends Well,* Annual Review of Genetics **40**, 363 (2006).

[2] M. S. Dillingham and S. C. Kowalczykowski, *RecBCD Enzyme and the Repair of Double-Stranded DNA Breaks,* Microbiology and Molecular Biology Reviews **72**, 642 (2008).

[3] S. C. Kowalczykowski, *Initiation of genetic recombination and recombination-dependent replication.* Trends in biochemical sciences **25**, 156 (2000).

[4] C. I. White and J. E. Haber, *Intermediates of recombination during mating type switching in Saccharomyces cerevisiae.* The EMBO Journal **9**, 663 (1990).

[5] J. Renkawitz, C. A. Lademann, and S. Jentsch, *Mechanisms and principles of homology search during recombination,* Nature Reviews. Molecular Cell Biology **15**, 369 (2014).

[6] A. Weiner, N. Zauberman, and A. Minsky, *Recombinational DNA repair in a cellular context: a search for the homology search,* Nature Reviews Microbiology **7**, 748 (2009).

[7] A. L. Forget and S. C. Kowalczykowski, *Single-molecule imaging of DNA pairing by RecA reveals a three-dimensional homology search,* Nature **482**, 423 (2012), arXiv:NIHMS150003 .

[8] J. W. Szostak, T. L. Orr-Weaver, R. J. Rothstein, and F. W. Stahl, *The double-strand-break repair model for recombination,* Cell **33**, 25 (1983).

[9] B. Jedynak and D. Karakos, *Finding a needle in a haystack,* in *Forty-first Annual Conference on Information Sciences and Systems, CISS 2007 - Proceedings* (2007) p. 318.

[10] Z. Qi, S. Redding, J. Y. Lee, B. Gibb, Y. Kwon, H. Niu, W. A. Gaines, P. Sung, and E. C. Greene, *DNA sequence alignment by microhomology sampling during homologous recombination,* Cell **160**, 856 (2015), arXiv:NIHMS150003 .

[11] I. De Vlaminck, M. T. J. van Loenhout, L. Zweifel, J. den Blanken, K. Hooning, S. Hage, J. Kerssemakers, and C. Dekker, *Mechanism of Homology Recognition in DNA Recombination from Dual-Molecule Experiments,* Molecular Cell **46**, 616 (2012).

[12] C. Lesterlin, G. Ball, L. Schermelleh, and D. J. Sherratt, *RecA bundles mediate homology pairing between distant sisters during DNA break repair,* Nature **506**, 249 (2014).

[13] C. Joo, S. A. McKinney, M. Nakamura, I. Rasnik, S. Myong, and T. Ha, *Real-Time Observation of RecA Filament Dynamics with Single Monomer Resolution,* Cell **126**, 515 (2006).

[14] J. J. Churchill, D. G. Anderson, and S. C. Kowalczykowski, *The RecBC enzyme loads recA protein onto ssDNA asymmetrically and independently of Chi, resulting in constitutive recombination activation,* Genes and Development **13**, 901 (1999).

[15] J. H. New, T. Sugiyama, E. Zaitseva, and S. C. Kowalczykowski, *Rad52 protein stimulates DNA strand exchange by Rad51 and replication protein A.* Nature **391**, 407 (1998).

[16] T. van der Heijden, M. Modesti, S. Hage, R. Kanaar, C. Wyman, and C. Dekker, *Homologous Recombination in Real Time: DNA Strand Exchange by RecA,* Molecular Cell **30**, 530 (2008).

[17] T. L. Orr-Weaver, J. W. Szostak, and R. J. Rothstein, *Yeast transformation: a model system for the study of recombination.* Proceedings of the National Academy of Sciences of the United States of America **78**, 6354 (1981).

[18] K. A. Datsenko and B. L. Wanner, *One-step inactivation of chromosomal genes in Escherichia coli K-12 using PCR products,* Proceedings of the National Academy of Sciences **97**, 6640 (2000).

[19] C. B. Russell, D. S. Thaler, and F. W. Dahlquist, *Chromosomal transformation of Escherichia coli recD strains with linearized plasmids,* Journal of Bacteriology **171**, 2609 (1989).

[20] J. M. Skerker, M. S. Prasol, B. S. Perchuk, E. G. Biondi, and M. T. Laub, *Two-component signal transduction pathways regulating growth and cell cycle progression in a bacterium: A system-level analysis,* PLoS Biology **3** (2005), 10.1371/journal.pbio.0030334.

[21] J. Mine-Hattab and R. Rothstein, *Increased chromosome mobility facilitates homology search during recombination,* Nat Cell Biol **14**, 510 (2012).

[22] V. Dion, V. Kalck, C. Horigome, B. D. Towbin, and S. M. Gasser, *Increased mobility of double-strand breaks requires Mec1, Rad9 and the homologous recombination machinery.* Nature cell biology **14**, 502 (2012).

[23] N. Agmon, B. Liefshitz, C. Zimmer, E. Fabre, and M. Kupiec, *Effect of nuclear architecture on the efficiency of double-strand break repair.* Nature cell biology **15**, 694 (2013).

[24] J. Renkawitz, C. A. Lademann, M. Kalocsay, and S. Jentsch, *Monitoring homology search during DNA double-strand break repair in vivo,* Molecular Cell **50**, 261 (2013).

[25] M. H. Hauer, A. Seeber, V. Singh, R. Thierry, R. Sack, A. Amitai, M. Kryzhanovska, J. Eglinger, D. Holcman, T. Owen-Hughes, and S. M. Gasser, *Histone degradation in response to DNA damage enhances chromatin dynamics and recombination rates,* Nature Structural & Molecular Biology (2017), 10.1038/nsmb.3347.

[26] I. Odsbu and K. Skarstad, *DNA compaction in the early part of the SOS response is dependent on RecN and RecA,* Microbiology (Reading, England) **160**, 872 (2014).

[27] Y. Bellaiche, V. Mogila, and N. Perrimon, *I-SceI endonuclease, a new tool for studying DNA double-strand break repair mechanisms in Drosophila,* Genetics **152**, 1037 (1999).

[28] H. J. Nielsen, J. R. Ottesen, B. Youngren, S. J. Austin, and F. G. Hansen, *The Escherichia coli chromosome is organized with the left and right chromosome arms in separate cell halves,* Molecular Microbiology **62**, 331 (2006).

[29] X. Wang, X. Liu, C. Possoz, and D. Sherratt J., *The two Escherichia coli chromosome arms locate to separate cell halves,* Genes & Development **20**, 1727 (2006).

[30] D. Landgraf, B. Okumus, P. Chien, T. a. Baker, and J. Paulsson, *Segregation of molecules at cell division reveals native protein localization,* Nature methods **9**, 480 (2012), arXiv:15334406 .

[31] A. Badrinarayanan, T. B. K. Le, and M. T. Laub, *Rapid pairing and resegregation of distant homologous loci enables double-strand break repair in bacteria,* The Journal of Cell Biology **210**, 385 (2015).

[32] M. T. Certo, B. Y. Ryu, J. E. Annis, M. Garibov, J. Jarjour, D. J. Rawlings, and A. M. Scharenberg, *Tracking genome engineering outcome at individual DNA breakpoints,* Nat Meth **8**, 671 (2011).

[33] J. Wiktor, C. Lesterlin, D. J. Sherratt, and C. Dekker, *CRISPR-mediated control of the bacterial initiation of replication,* Nucleic Acids Research **44**, 3801 (2016), arXiv:arXiv:1011.1669v3 .

[34] J. M. Schoemaker, R. C. Gayda, and A. Markovitz, *Regulation of cell division in Escherichia coli: SOS induction and cellular location of the SulA protein, a key to lon-associated filamentation and death,* Journal of Bacteriology **158**, 551 (1984).

[35] S. Elmore, M. Müller, N. Vischer, T. Odijk, and C. L. Woldringh, *Single-particle tracking of oriC-GFP fluorescent spots during chromosome segregation in Escherichia coli,* Journal of Structural Biology **151**, 275 (2005).

[36] H. C. Lim, I. V. Surovtsev, B. G. Beltran, F. Huang, J. Bewersdorf, and C. Jacobs-Wagner, *Evidence for a DNA-relay mechanism in ParABS-mediated chromosome segregation,* Elife **3**, e02758 (2014).

[37] L. Ström, H. B. Lindroos, K. Shirahige, and C. Sjögren, *Postreplicative recruitment of cohesin to double-strand breaks is required for DNA repair,* Molecular Cell **16**, 1003 (2004).

[38] C. Gelot, J. Guirouilh-Barbat, T. Le Guen, E. Dardillac, C. Chailleux, Y. Canitrot, and B. S. Lopez, *The Cohesin Complex Prevents the End Joining of Distant DNA Double-Strand Ends,* Molecular Cell **61**, 15 (2016).

[39] R. Biehs, M. Steinlage, O. Barton, S. Juhasz, J. Kunzel, J. Spies, A. Shibata, P. Jeggo, and M. Lobrich, *DNA double-strand break resection occurs during non-homologous end-joining in G1 but is distinct to resection during homologous recombination in G2,* Molecular Cell , 671 (2017).

[40] A. Mehta, A. Beach, and J. E. Haber, *Homology Requirements and Competition between Gene Conversion and Break-Induced Replication during Double-Strand Break Repair,* Molecular Cell **65**, 515 (2017).

[41] A. Tubbs and A. Nussenzweig, *Endogenous DNA Damage as a Source of Genomic Instability in Cancer,* Cell **168**, 644 (2017).

[42] D. Dutta, K. Shatalin, V. Epshtein, M. E. Gottesman, and E. Nudler, *Linking RNA polymerase backtracking to genome instability in E. coli,* Cell **146**, 533 (2011), arXiv:NIHMS150003 .

[43] J. M. Pennington and S. M. Rosenberg, *Spontaneous DNA breakage in single living Escherichia coli cells,* Nature Genetics **39**, 797 (2007).

[44] N. Shechter, L. Zaltzman, A. Weiner, V. Brumfeld, E. Shimoni, Y. Fridmann-Sirkis, and A. Minsky, *Stress-induced condensation of bacterial genomes results in repairing of sister chromosomes: Implications for double strand dna break repair,* Journal of Biological Chemistry **288**, 25659 (2013).

[45] E. Vickridge, C. Planchenault, C. Cockram, I. G. Junceda, and O. Espéli, *Management of E. coli sister chromatid cohesion in response to genotoxic stress,* Nature Communications **8**, 14618 (2017).

[46] S. Pellegrino, J. Radzimanowski, D. De Sanctis, E. B. Erba, S. McSweeney, and J. Timmins, *Structural and functional characterization of an SMC-like protein RecN: New insights into double-strand break repair,* Structure **20**, 2076 (2012).

[47] A. Tal, R. Arbel-Goren, N. Costantino, D. L. Court, and J. Stavans, *Location of the unique integration site on an Escherichia coli chromosome by bacteriophage lambda DNA in vivo.* Proceedings of the National Academy of Sciences of the United States of America **111**, 7308 (2014).

[48] D. G. Gibson, L. Young, R.-Y. Chuang, J. C. Venter, C. a. Hutchison, H. O. Smith, C. A. H. Iii, and N. America, *Enzymatic assembly of DNA molecules up to several hundred kilobases.* Nature methods **6**, 343 (2009).

[49] O. Sliusarenko, J. Heinritz, T. Emonet, and C. Jacobs-Wagner, *High-throughput, subpixel precision analysis of bacterial morphogenesis and intracellular spatio-temporal dynamics,* Molecular Microbiology **80**, 612 (2011).

[50] A. Paintdakhi, B. Parry, M. Campos, I. Irnov, J. Elf, I. Surovtsev, and C. Jacobs-Wagner, *Oufti: An integrated software package for high-accuracy, high-throughput quantitative microscopy analysis,* Molecular Microbiology **99**, 767 (2016), arXiv:arXiv:1011.1669v3 .

[51] T. G. Bernhardt and P. A. De Boer, *The Escherichia coli amidase AmiC is a periplasmic septal ring component exported via the twin-arginine transport pathway,* Molecular microbiology **48**, 1171 (2003).

4

INTEGRATION OF LONG LINEAR DNA MOLECULES IN *recD* *Escherichia coli*

Genomic engineering is an asset in the toolkit of laboratory techniques that enables for rational modifications of chromosomal sequences. Here we present a genome-modification method, which relies on Escherichia coli's natural recombination pathway stimulated by the absence of the recD gene. Using this approach, we integrate synthetic PCR-synthetized DNA into E. coli chromosome in a one simple step. This approach shows potential to overcome the current size limitations of current techniques, such as Lambda Red. Indeed, recD-engineering allowed us to integrate very long sequences, such as a 12 kb recombinant cassette, or fragments of purified E. coli genomic DNA. The presented method also can be applied to wild-type cells, as we show that the recD gene can transiently be knocked-down using a CRISPRi system. The simplicity of integration of long fragments presented here can be expected to find many applications, wherever incorporation of large DNA molecules is needed.

4.1. INTRODUCTION

The invention and development of genetic engineering techniques has revolutionised science, medicine, and industry. Production of recombinant human insulin in bacterial cells [1] in the 1970s not only kick-started the biotechnology industry, but also proved that rational modification of genetic code is of enormous value and importance [2]. Since then great technological progress has been made, and genetic engineering is no longer limited to incorporating single genes into small mobile DNA-plasmids, but also allows to directly modify actual genomes. Indeed, genetic engineering has now evolved to the stage that it even provides the for synthesis of, functional chromosomes of both prokaryotic and eukaryotic organisms [3–5].

Currently, a significant amount of chromosomal modifications are done with the use of the natural homologous recombination (HR) pathway [6, 7]. HR is normally involved in DNA-repair and crossing-over reactions, and its efficiency differs among different organisms. Some recombinogenic organisms, such as yeast, can be easily transformed with foreign DNA, whereas other species are not [8, 9]. One of the obstacles in molecular biology was the fact that the main workhorse of molecular biology, *Escherichia coli*, exhibits a very low spontaneous recombination activity when synthetic foreign DNA is introduced (**Chapter 2**) and genomic engineering of *E. coli* wild-type cells proved to be troublesome. Certain mutants of *E. coli* however, are recombination proficient, e.g. for those mutations that were targeting the RecBCD nuclease/helicase complex, either by removing *recBC* genes together with *sbcB* or by removing just the *recD* gene alone [10, 11]. RecBCD is a very potent exonuclease complex capable of destroying double stranded (ds)DNA as it is delivered into the cell, therefore inhibiting any possible recombination events. Mutations inactivating this destructive exonuclease activity greatly increased activity in recombination assays. Recently, we have showed that *recD* knockout cells degrade DNA to a much less extent than wild-type cells (**Chapter 2)** and that this change leads to highly efficient recombination phenotype [10].

The low levels of natural homologous recombination in *E. coli* have also been overcome using phage-derived lambda-red-mediated recombination which can facilitate integrations of synthetic DNA sequence into any position of the bacterial genome [12]. The technique itself was limited to relatively short sequences, no longer than a few kilobases (kb) and integration of long DNA sequences into E. coli chromosome has proved to be particularly difficult. This length limitation was recently highly increased (>100 kb) using bacterial artificial chromosomes (BAC) as a source of dsDNA for lambda-red-recombination [13], but no method allowing for an easy, one step, and efficient integration of long constructs exists. The proposed mechanism of lambda-red-mediated recombination begins with the complete degradation of one of the strands of dsDNA intermediate by the Exo protein. Next, the entire length of single stranded (ss)DNA is coated by Beta proteins, which assists homologous interaction between the lagging strand of the replication fork and the ssDNA substrate [14, 15] (Fig. 4.1). Experimental evidence showing a high efficiency of recombination with single-stranded oligonucleotides complementary to lagging strand strongly supports this ssDNA model [15]. Such a mechanism suggests limitations for the integration of very long fragments, as the formation of ssDNA from a long dsDNA template is more complicated than from the short one, requires more time and exposes ssDNA to the action of common cellular nucleases. More-

over, an increased length of the construct coated by Beta proteins may result in interactions with non-specific regions of the chromosome.

Genomic integration methods which exploit the natural bacterial recombination pathway are an attractive alternative capable of overcoming these limitations. Processing of free dsDNA ends by the bacterial machinery is equal to processing of double stranded breaks [9, 16], which leads to the formation of dsDNA with 3' ssDNA overhangs coated with recombination protein (RecA in *E. coli*) on both ends of the integrated DNA. 3' ssDNA/RecA ends can localise homologous sequences on the chromosome and facilitate integration of the whole construct [17] (Fig. 4.1b). Homologous recombination reaction does not rely on the replication fork and does not require processing of the entire integrated sequence but only the ends of it. Therefore the length of the DNA is not expected to be a limiting factor, once the DNA is delivered inside of the bacterial cell.

Here, we show that Δ*recD* cells can be employed to develop an easy way to disrupt genes and integrate long dsDNA cassettes. Using this technique, we also show that it is possible to move integrated sequences between strains, without the use of P1 phage transduction. And finally we show that transient inhibition of *recD* expression using CRISPRi [18] system can be used to establish an easy and universal platform for *E. coli* genome engineering.

4.2. RESULTS

CHROMOSOMAL TRANSFORMATION IN TEXTITRECD CELLS

We recently reported that the PCR-generated DNA, flanked by regions of at least 1 kb homologies to the chromosome can be integrated into the *lac* operon in Δ*recD* cells (Fig. 4.2a). When the integration was designed to target the regions between *lacZ* and *lacY* genes, we recovered colonies that were resistant to kanamycin, and for which the *lac* operon was modified as expected. Colonies validation using genotyping, targeting *kan* gene and a chromosomal flanking region, resulted in a product only when integrated sequence was present on the genome, further confirming positive integration (Fig. 4.2b). A requisite of Δ*recD* recombination is that the regions of homologies flanking the *kan* gene have to be longer that those used for lambda-red recombination (1-2 kb for Δ*recD* vs 50 bp for lambda-red [12]). A possible explanation for the requirement for longer homologies in case of Δ*recD* integration is the difference in mechanisms of exonucleolitic end-resection and coating by recombination protein. Alternatively, mechanism of DNA insertion by RecA relies on strand invasion and DNA replication and each of those two reactions may require presence of a long intermediate structure, between chromosome and delivered DNA sequence.

ASSEMBLY AND INTEGRATION OF A LONG SEQUENCE INTO LAC OPERON

We hypothesised that in case of RecA-mediated integration, the recombination efficiency in not limited by the length of the DNA molecule, because only homologous ends are involved in the process, excluding the heterologous, middle part from the reaction (Fig. 4.1b). *E. coli* cells are capable of being transformed with very large BAC, therefore delivery of long, linear dsDNA can also be possible. To test the ability of Δ*recD E. coli* to integrate long sequences we designed a 2-step protocol, where first, using Gibson assembly method [19] we combined a 7 kb region containing *tetR, dCas9* and *gRNA* genes

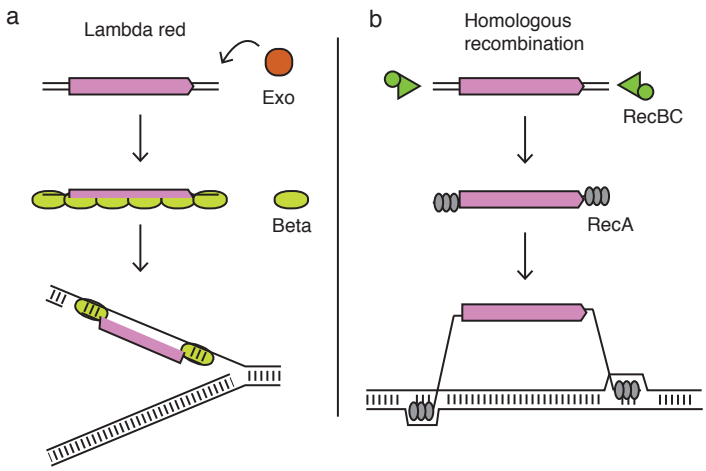


Figure 4.1: **Mechanisms of DNA integration in Lambda Red and RecA recombination.** (a) Lambda Red integration of foregin DNA into the genome. In the first step Exo (red) digests one of two strands of dsDNA, which subsequently is coated by Beta proteins (green). The Beta-ssDNA proteofilament anneals via homologous ends to the lagging strand of the replication fork. (b) Integration of dsDNA via homologous recombination involves an end-resection of dsDNA ends by the RecBC exonuclease (green). Subsequently, ssDNA ends are coated by RecA proteins (blue) and invades homologous sequences on the chromosome, leading to the integration of the heterologous sequence.

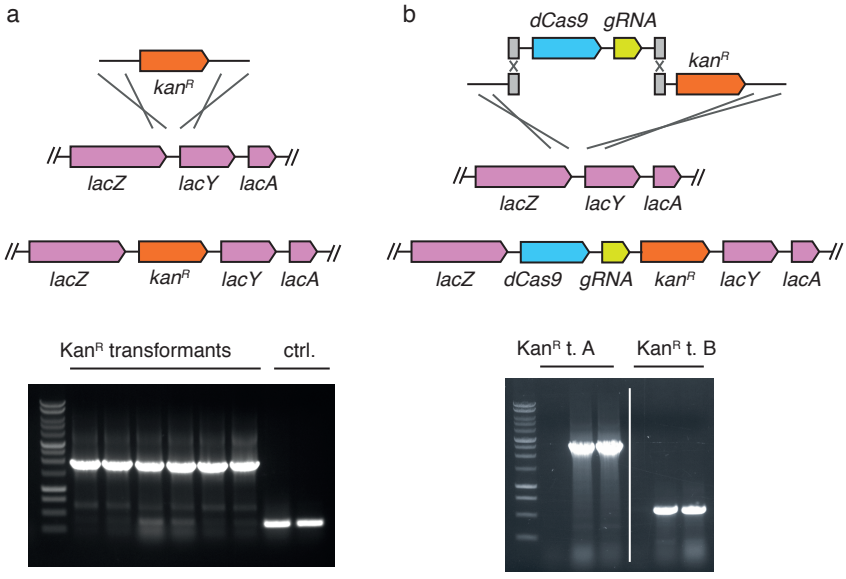


Figure 4.2: **Editing of the *lac* operon with PCR and Gibson assembly products.** (a) in the presence of recombination proteins, integration of dsDNA occurs spontaneously. Homologous arms align with chromosomal regions leading to incorporation of the antibiotic resistance gene. Genotyping PCR confirms the positive integration in all screened colonies (bottom gel, Kan^R transformants). The same PCR on unmodified cells results in small size product. Promega Benchtop 1 kb ladder was used. (b) Long sequences can be prepared using Gibson assembly from multiple shorter fragments prior to integration into *lac* operon. Genotyping PCR confirms the presence of the 5' (Kan^R t. A) and the 3' (Kan^R t. B) end of the construct in the chromosome in the Kanamycin resistant colonies. Promega Benchtop 1 kb ladder was used.

together with two flanking regions, containing genomic homologies and a kanamycin resistance gene. Next we isolated the high-molecular weight product of the assembly reaction from the agarose gel, purified it on a column and used it to electroporate into competent Δ*recD* cells. We detected recombinant colonies, with the designed integration confirmed by a genotyping PCR (Fig. 4.2d). The reaction was only successful in the Δ*recD* background and we found, as expected, no colonies after electroporation of the same construct into the wild-type cells. The final length of the linear DNA used in this experiment was 12 kbp, consisting of 7 kbp of genes to be integrated, flanked by 5 kpb of homologies in total. Due to several steps of purification the final concentration of electraporated DNA was very low - only 5 ng/µl and we used only 1 µl in the reaction. Successful recombination using such small amounts of a long linear DNA proves that genetic engineering in Δ*recD* cells is a very robust and attractive solution for integration of long, or complicated sequences, such as whole operons, or sets of long genes.

HORIZONTAL GENE TRANSFER

Exchange of genetic information between cells is widespread among prokaryotic organisms in a process of horizontal gene transfer (HGT) [20]. Rapid development of resistances to common antibiotic is generally attributed to an ease with which resistance genes can be moved between bacterial cells [21]. Whereas some bacteria can naturally take up and integrate DNA molecules, other including *E. coli* , do not have this ability and the horizontal gene transfer in those species is limited to conjugation, plasmid transformation or phage transduction [22–24]. We hypothesised that the mechanism limiting HGT in wild-type *E. coli* is the same which inhibits integration of linear DNA, and that by removing RecD subunit from cells we can grant cells ability to exchange DNA between cultures. To test that we isolated genomic DNA from a strain carrying chloramphenicol resistance gene on the genome (donor strain) and used that DNA to electroporate into a Δ*recD* strain (recipient strain), which was sensitive to chloramphenicol (Fig. 4.3a). Electroporation with such isolated genomic DNA into Δ*recD* cells indeed led to formation of chloramphenicol-resistant colonies (Fig. 4.3b). PCR genotyping confirmed integration of the foreign DNA into the chromosome, as well as presence of recipient-specific genetic watermark (Fig. 4.3b). Genomic DNA isolation used in this experiment (for details see Materials and methods) produces a collection of long chromosomal fragments corresponding to random chromosomal *loci*. The method of isolation of genomic DNA used in this experiment produces fragments of 30 kb long, therefore on average about 0.65% of all fragments should contain region of interest. As we tested before, transformation with long DNA molecules is possible using electroporation and Δ*recD* cells can easily process and integrate such fragments. Together these results show that HGT among wild-type *E. coli* cells can be strongly enriched by deletion of the *recD* gene.

CRISRPI CONTROL OF RECOMBINATION PHENOTYPE

An ideal genome engineering system has to be easily applicable to cell of versatile genotypes, especially to the wild-type cells. Permanent deletion of the *recD* gene conflicts with to this requirement as modified cells are lacking one of components of a DNA repair machinery. To address this limitation, we have used the CRISPRi system (*c*lustered *r*egularly *i*nterspaced *s*hort *p*alindromic *r*epeats *i*nterference) [18, 25], which can block expression of a gene [26]. Similarly, to the lambda-red recombination system, CRISPRi can be encoded on a single plasmid, containing both parts of the system, the dCas9 protein and the gRNA sequence [26]. The binding specificity of CRISPRi system arises from a 20 nucleotide sequence in the guide RNA (gRNA), and from presence of the Cas9-specyfic PAM (protospacer acquisition motif) sequence on a chromosome, located next to the targeted region (for more information see [27]). Our CRISPRi approach has the potential to be easily applicable to wide variety of *E. coli* cells. We have tested a series of CRISPRi constructs that targeted regions upstream of the *recD* gene. For further experiments, we selected only those constructs that did not show an obvious growth reduction after plating in conditions that induced the CRISPRi system. After selecting the most suitable gRNA, we tested integration efficiency in CRISPR knock-down cells, using the same KanR construct that we used in the previous experiment (Fig. 4.2a). After induction of the CRISPRi targeting *recD* gene and electroporation with dsDNA, we observed colonies growing in the presence of kanamycin PCR genotyping of single colonies confirmed that the linear DNA was integrated into the genome in the desired position.

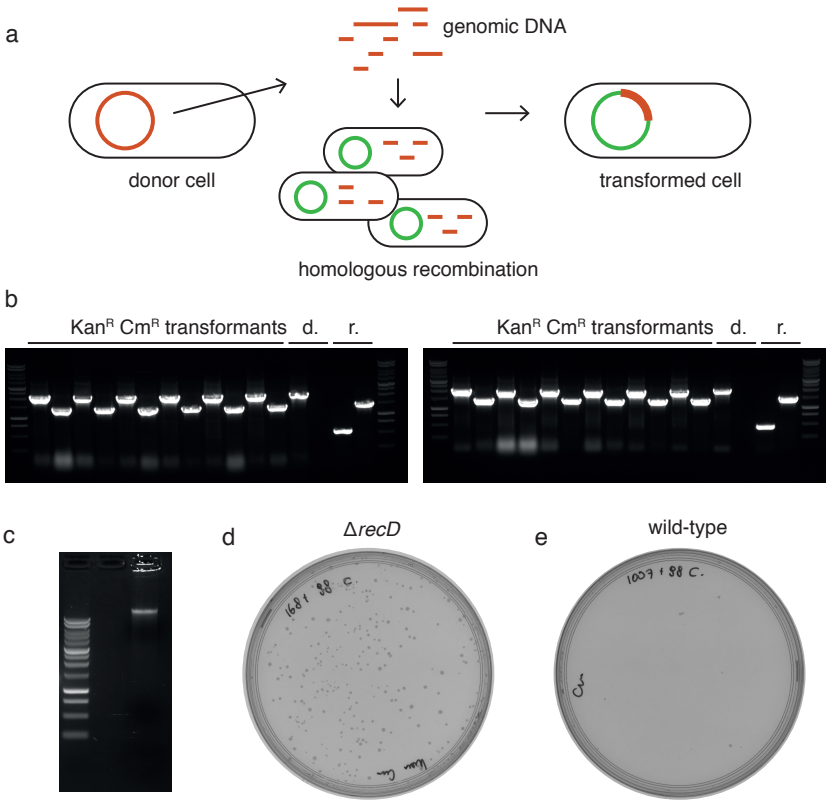


Figure 4.3: **Deletion of *recD* stimulates horizontal gene transfer.** (a) Concept of our horizontal gene transfer assay. Genomic DNA from the donor strain is isolated in a form of library of linear fragments of the chromosome. Next this library is introduced into a *recD* recipient strain and is recombined. Selection markers allows for screening for a desired phenotype. (b) Genotyping PCR analysis confirms the presence of donor and recipient chromosomal DNA watermarks for 2 different HGT assays (left – donor strain Jx097; right – donor strain Jx098) (c) Genomic DNA used in the assay is in the form of high weight linear DNA fragments (white arrow). (d and e) Growth on selection plates is visible after HGT in Δ*recD* background cells (d) but not in wild-type cells (e).

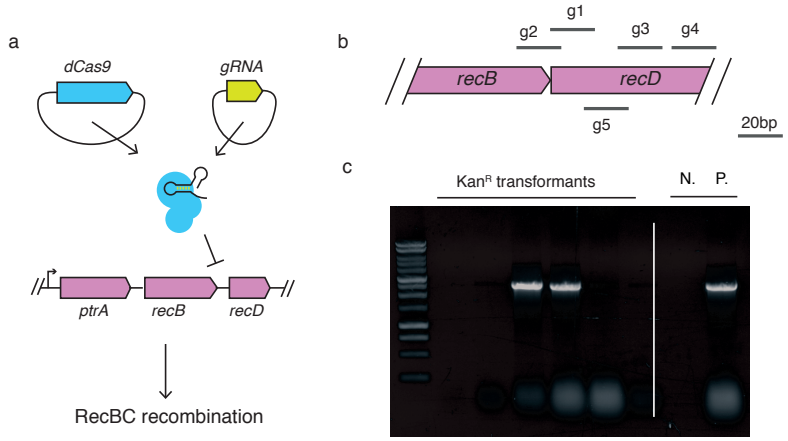


Figure 4.4: **Transient induction of the recombinogenic phenotype.** (a) CRISPRi system consisting of a catalytically dead dCas9 and gRNA, together inhibiting the production of the *recD* gene mRNA by RNA polymerase. As a consequence, cells switch to the RecBC recombinogenic phenotype. (b) Positions targeted by gRNA. Black lines show the position and targeted strand – g1-4 target non-template strand, g5 targets template strand. (c) PCR genotyping of 5 colonies confirms the presence of KanR in the chromosome in 2 colonies that were found to be resistant to Kanamycin. N. – non-transformed strain, P. - Δ*recD* strain carrying integration of Kan^R cassette.

There results show that it is possible to transiently remove the RecD subunit of RecBCD from genetically wild-type cell and thus endow cells with recombination properties on demand.

4.3. DISCUSSION

RecD ENGINEERING AND LAMBDA-RED

Genome engineering has played a critical role in many biological discoveries. Counterintuitively, some more complex organism, such as yeast [9, 19], were easier to modify than the most common and much simpler bacterial model system – *Escherichia coli*. Experimental evidence suggests that the reason for *E. coli* resistance to genomic engineering is the specific mechanism of dsDNA ends degradation by RecBCD complex (**Chapter 2**). Indeed, several methods based on deletions of the RecBCD related genes [10, 11] have been proposed, but none has become a common laboratory tool. On the other hand, the heterologous lambda-red recombination method was positively received by the scientific community and greatly improved *E. coli* genetic studies. Two features made lambda-red so attractive: (i) it can be delivered on a plasmid to any cell and (ii) it required only 36-50 bp of flanking homologies. The fact that such technique was highly welcome in the field of molecular biology is highlighted by the fact that in the time in which this chapter was written, the original paper describing the Lambda Red method was cited more than 9000 times [12].

In this work we introduce a new method, where we combined the recombination potential of Δ*recD E. coli* bacteria with a simplicity brought by a plasmid borne CRISPRi system. Because of the end-specific mechanism in which linear DNA is integrated during Δ*recD* recombination, very long molecules can be integrated using a single electroporation step. According to the current model, RecBC enzyme, in contrast to the Exo protein, prepares only the ends of DNA molecule for the recombination reaction and the middle section of the DNA does not take part in the integration process. Accordingly, the length of the middle segment can be arbitrary long. The integration itself occurs through a RecA-mediated mechanism. RecA recombination is an evolutionary conserved and critical pathway [28], which constitutes for a robust and stable genomic engineering tool. Moreover, cellular concentration of many proteins belonging to DNA repair pathways are not dependent on a cell cycle [17], but rather than on response to DNA damage. Such coupling between DNA damage and protein concentration suggests two important consequences for *recD* recombination approach: the first is that all cells in the test-tube population can be equally suited for recombination, and the second, that possibly bacterial recombinogenic phenotype can be stimulated by induction of DNA damage before delivery of DNA construct into cells. One obvious limitation to *recD* method is that many dedicated cloning strains, such as DH5alpha, or TOP10 cells carry mutations in the *recA* gene, which increase the stability of plasmid sequences by inactivating the homologous recombination pathway.

Here, we performed successful integrations with 12 kb molecules (of which 7 kb was heterologous to the genome) using only 5 ng of purified DNA. In the future we plan to test the length limitations of this technique. Previously discussed fact that RecA recombination happens via ends of the molecule, suggests uncoupling efficiency of integration from the length of the DNA. A possible factor which can potentially lower genomic engineering efficiency is the delivery of DNA into cells, lower for long molecules. However, *Escherichia coli* cells can be routinely transformed with large, 300 kb BAC vectors, showing that also the electroporation of very long, linear molecules can be achieved. Modern PCR polymerases together with DNA assembly techniques, such as Gibson assembly [29], or ligation cycling reaction (LCR) [30] are increasing the lengths of synthetic molecules one can create in the laboratory. LCR was reported to combine 20 DNA parts into 20 kb plasmids and the Gibson protocol in published report [29] was used to combine 166 kb fragment of *Mycobacterium genitalium* with 8 kb of a BAC vector backbone in a process that can easily be adapted to combining fragment of similar size with two handles for homologous recombination.

RELAXED HORIZONTAL GENE TRANSFER

Another interesting feature of *recD* strain is its increased performance in horizontal gene transfer assay. We showed that genomic DNA electroporated into Δ*recD* strain can integrate into the recipient genome in a process of laboratory horizontal gene transfer. Some bacterial species can spontaneously uptake linear DNA from the environment. Wild-type *E. coli* on the other hand are very resistant to environmental transformation with linear DNA and horizontal gene transfer is generally mediated by processes of conjugation or phage transduction. Our observation of chromosomal transformation with genomic DNA in the Δ*recD* background describes a relaxed modus for horizontal gene transfer. One of its straightforward applications is an easy shuffling of genetic elements

between strains, as is commonly done by P1 transduction. Omitting P1 phage in the process can greatly simplify the protocol and reduce the time needed to create combinations of genetically modified strains. Another possible application is a study of the dynamics of horizontal gene transfer in controlled laboratory conditions. Previously such studies were narrowed to specific bacteria, naturally efficient in HGT, or to conjugation [22], but using a Δ*recD* strain, one can design a general laboratory-based experimental work-flow using microfluidic compartments, complex artificial environments, and singe-cell microscopy and genotyping.

CONCLUSIONS

We present a simple and straight-forward method of modifying *E. coli* genomes relying on hijacking its recombination machinery, which in normal conditions is responsible for double stranded DNA break repair. We show that this method can facilitate introduction of long pieces of linear DNA into bacterial chromosome, and that it allows for laboratory horizontal gene transfer. Moreover, we show that a CRISPRi system can be used to induce the Δ*recD* recombinogenic phenotype in almost any genetic background. We estimate that our recD methodology has a big potential in molecular biology laboratories as well as in industrial settings.

4.4. MATERIAL AND METHODS

SEQUENCES, STRAINS AND PLASMIDS

Sequences of primers, plasmids and genotypes of bacteria used in this study are provided in Table 4.1 and 4.2.

CONSTRUCTION OF STRAINS AND PLASMIDS AND INTEGRATION CASSETTES

Escherichia coli K-12 MG1655 was used as the initial strain. Deletion of *recD* in the MG1655 background was done using lambda-red protocol [12]. The deletion was designed to remove only the *recD* gene, while leaving the *ptrA* and the *recB* genes of the same operon intact. The fragment containing *KanR* used for lambda red was simplified using primers Jw261 and Jw262 from pSK3 plasmid (for details see 4.1 and 4.2), resulting in a *KanR* fragment flanked by *frt* sequences and 50 bp of homology to the *recD* genomic region. *KanR* gene was then removed using pCP20 treatment.

The cassette containing *dCas9, gRNA* and *KanR* flanked by regions of genomic homology was prepared as follows: from pdCas9degRna3 plasmid [26], a fragment containing dCas9deg and gRNA was amplified with Jw397 and Jw398 primers. From plasmid pLisa5 (**Chapter 2**) two fragments were amplified with primer pairs: (i) Jw284 and Jw399 and (ii) Jw381 and Jw400. Purified (Wizard® SV Gel and PCR Clean-Up System, Promega) fragments were used for a Gibson assembly reaction and the product of the reaction was separated on agarose gel, 12 kb band corresponding to assembled cassette was cut out and purified (Wizard® SV Gel and PCR Clean-Up System, Promega). 5ng of assembled construct was used in the electroporation.

The *KanR* gene, flanked by genomic homologies was amplified using pLisa5 plasmid as the template and Jw191 and Jw302 primers pair, and purified on a column (Promega) before electroporation.

PURIFICATION OF GENOMIC DNA

Genomic DNA of a strain carrying the CmR gene was prepared using Promega Wizard® Genomic DNA Kit following manufacturers protocol. In summary, 1 ml of overnight culture of Jx097 and Jx098 cells was used to isolate the DNA. After the isolation, the DNA sample was further purified using ethanol precipitation. 100 µl of genomic DNA was mixed with 300 µl of 100% ethanol and 30 µl of 3M sodium acetate and incubated in -20°C for 1h. Then it was centrifuged at 4°C for 45 minutes at maximum speed (Eppendorf 5418R). Supernatant was discarded and the DNA pellet was suspended in 300 µl of 70% ethanol solution and the centrifugation step was repeated. Supernatant was discarded, and the DNA pellet was dried and re-suspended in 100 µl of milli-Q water.

ELECTROCOMPETENT CELLS PREPARATION AND ELECTROPORATION CONDITIONS

Elecrocompetent cells were prepared as follows: overnight culture grown in Lysogeny Broth (LB) was refreshed in fresh medium in a 1 to 100 ratio and grown until reaching OD600 = 0.4–0.6. Then 1 ml of cell culture was chilled on ice, centrifuged and washed in 1 ml of ice-cold milli-Q water twice. Then cells were centrifuged and the cell pellet was re-suspended in 50 µl of ice-cold milli-Q water. 1 µl of 100 ng/µl DNA was used for electroporation of the *KanR* construct, 1µl of 5ng/µl was used for electroporation of the dCas9degRNA3 cassette. After electroporation (Eppendorf Epporator, 0.1cm cuvette, 1250 V), cells were recovered in 1 ml LB media at 37° C degrees and plated on LB-agarose selection plates.

CRISPRI

Transient knock-down of *recD* gene was done in cells co-transformed with pdCas9 and pgRNA plasmids. For plating experiments, cultures were plated on LB-agar plates supplemented with Ampicillin (100 ng/ml) and Chloramphenicol (34 ng/ml) and anhydrotetracycline (100 ng/ml) to induce *dCas9* expression. For the *KanR* cassette integration assay, LB supplemented with antibiotics over-night culture was refreshed 1:100 in LB with antibiotics and 100 ng/ml of anhydrotetracycline. Electrocompetent cells were prepared from an early exponential culture and electroporation was carried out as described above.

4.5. SUPPLEMENTARY INFORMATION

Table 4.1: List of strains and plasmids.

Strain	Description
K12 MG1655	
Jx097	TB28 [31] parSP1_30kb_us_CmR
Jx098	TB28 [31] parSP1_80kb_us_CmR
Jx172	MG1655 *recD::frt*

Plasmids	Description
pLisa5	This work, KanR template flanked by lacZYA arms
pdCas9deg-RNA3	CRISPR [26]
pKD46	Lambda Red [12]
pCP20	Lambda Red [12]
pgRNA	CRISPRi [18]
pdCas9	CRISPRi [18]
pSK3	This work, KanR template

Table 4.2: List of primers used in this study.

Primer	Sequence	Description
Jw261	GAG ATG TTT GCC GGT ATG ACC CTG GAG GAG GCG TAA TGA AAT TGC AAA AG GGATCCGTCGACCT-GCAGTT	KanR knockout
Jw262	CGT CGG ATG CGA CAT GCG TAA CAC TCG TAC GTC GCA TCC GGC AAT TAC GT GTG TAG GCT GGA GCT GCT TC	KanR knockout
Jw397	CGT CTT AAG ACC CAC TTT CAC	
Jw398	GTA TCT TCC TGG CAT CTT CCA G	
Jw284	CAACGTCGTGACTGGGAAAAC	
Jw399	GTG AAA GTG GGT CTT AAG ACG TAGCGACCG-GCGCTCAGCTGG	
Jw381	CAT GCC GGA TGC GGC TAA TG	
Jw400	ggaagatgccaggaagatac GTCCGCAGAAACGGTGCT-GACC	
Jw191	GGATATGTGGCGGATGAGCG	
Jw302	CATGCCGGATGCGGCTAATG	
JW386	CCACTAGT AAT TTC ATT ACG CCT CCT CC GTTTTA-GAGCTAGAAATAGCAAG	gRNA recD1
JW387	CCACTAGT CGC CTC CTC CAG GGT CAT AC GTTTTA-GAGCTAGAAATAGCAAG	gRNA recD2
JW388	CCACTAGT AGG GCA AAT TGC ACA TCC AG GTTTTA-GAGCTAGAAATAGCAAG	gRNA recD3
JW389	CCACTAGT AGG ATG TTC ATC TCC CGC CA GTTTTA-GAGCTAGAAATAGCAAG	gRNA recD4
JW390	CCACTAGT GAAATTGCAAAAGCAATTAC GTTTTA-GAGCTAGAAATAGCAAG	gRNA recD5
Jw096	GG ACT AGT ATT ATA CCT AGG ACT GAG	gRNA cloning

REFERENCES

[1] H. Keen, J. Pickup, R. Bilous, A. Glynne, G. Viberti, R. Jarrett, and R. Marsden, *Human Insulin Produced By Recombinant Dna Technology: Safety and Hypoglycæmic Potency in Healthy Men,* The Lancet **316**, 398 (1980).

[2] D. E. Cameron, C. J. Bashor, and J. J. Collins, *A brief history of synthetic biology,* Nature Reviews Microbiology **12**, 381 (2014).

[3] N. Annaluru, H. Muller, L. A. Mitchell, S. Ramalingam, G. Stracquadanio, S. M. Richardson, J. S. Dymond, Z. Kuang, L. Z. Scheifele, E. M. Cooper, Y. Cai, K. Zeller, N. Agmon, and J. S. Han, *Designer Eukaryotic Chromosome,* Science (New York, N.Y.) **344**, 55 (2014), arXiv:24674868 .

[4] D. G. Gibson, J. I. Glass, C. Lartigue, V. N. Noskov, R.-Y. Chuang, M. A. Algire, G. A. Benders, M. G. Montague, L. Ma, M. M. Moodie, C. Merryman, S. Vashee, R. Krishnakumar, N. Assad-Garcia, C. Andrews-Pfannkoch, E. A. Denisova, L. Young, Z.-Q. Qi, T. H. Segall-Shapiro, C. H. Calvey, P. P. Parmar, C. A. Hutchison, H. O. Smith, and J. C. Venter, *Creation of a Bacterial Cell Controlled by a Chemically Synthesized Genome,* Science **329**, 52 (2010), arXiv:z0024 .

[5] C. A. Hutchison, R.-Y. R.-Y. Chuang, V. N. Noskov, N. Assad-Garcia, T. J. Deerinck, M. H. Ellisman, J. Gill, K. Kannan, B. J. Karas, L. Ma, J. F. Pelletier, Z.-Q. Z.-Q. Qi, R. A. Richter, E. A. Strychalski, L. Sun, Y. Y. Suzuki, B. Tsvetanova, KimS.Wise, H. O. Smith, J. I. Glass, C. Merryman, D. G. Gibson, J. C. Venter, K. S. Wise, H. O. Smith, J. I. Glass, C. Merryman, D. G. Gibson, and J. C. Venter, *Design and synthesis of a minimal bacterial genome,* Science **351**, aad6253 (2016).

[6] P. Mali, L. Yang, K. M. Esvelt, J. Aach, M. Guell, J. E. DiCarlo, J. E. Norville, and G. M. Church, *RNA-guided human genome engineering via Cas9,* Science **339**, 823 (2013), arXiv:arXiv:1011.1669v3 .

[7] J. M. Skerker, M. S. Prasol, B. S. Perchuk, E. G. Biondi, and M. T. Laub, *Two-component signal transduction pathways regulating growth and cell cycle progression in a bacterium: A system-level analysis,* PLoS Biology **3** (2005), 10.1371/journal.pbio.0030334.

[8] W. Jiang, D. Bikard, D. Cox, F. Zhang, and L. A. Marraffini, *RNA-guided editing of bacterial genomes using CRISPR-Cas systems,* Nature Biotechnology **31**, 233 (2013).

[9] T. L. Orr-Weaver, J. W. Szostak, and R. J. Rothstein, *Yeast transformation: a model system for the study of recombination.* Proceedings of the National Academy of Sciences of the United States of America **78**, 6354 (1981).

[10] C. B. Russell, D. S. Thaler, and F. W. Dahlquist, *Chromosomal transformation of Escherichia coli recD strains with linearized plasmids,* Journal of Bacteriology **171**, 2609 (1989).

[11] S. C. Winans, S. J. Elledge, J. H. Krueger, and G. C. Walker, *Site-directed insertion and deletion mutagenesis with cloned fragments in Escherichia coli,* Journal of Bacteriology **161**, 1219 (1985).

[12] K. A. Datsenko and B. L. Wanner, *One-step inactivation of chromosomal genes in Escherichia coli K-12 using PCR products,* Proceedings of the National Academy of Sciences **97**, 6640 (2000).

[13] K. Wang, J. Fredens, S. F. Brunner, S. H. Kim, T. Chia, and J. W. Chin, *Defining synonymous codon compression schemes by genome recoding,* Nature **539**, 59 (2016).

[14] M. Maresca, A. Erler, J. Fu, A. Friedrich, Y. Zhang, and A. F. Stewart, *Single-stranded heteroduplex intermediates in lambda Red homologous recombination.* BMC molecular biology **11**, 54 (2010).

[15] J. A. Mosberg, M. J. Lajoie, and G. M. Church, *Lambda red recombineering in Escherichia coli occurs through a fully single-stranded intermediate,* Genetics **186**, 791 (2010).

[16] J. W. Szostak, T. L. Orr-Weaver, R. J. Rothstein, and F. W. Stahl, *The double-strand-break repair model for recombination,* Cell **33**, 25 (1983).

[17] M. S. Dillingham and S. C. Kowalczykowski, *RecBCD Enzyme and the Repair of Double-Stranded DNA Breaks,* Microbiology and Molecular Biology Reviews **72**, 642 (2008).

[18] L. Qi, M. Larson, L. Gilbert, J. Doudna, J. Weissman, A. Arkin, and W. Lim, *Repurposing CRISPR as an RNA-Guided Platform for Sequence-Specific Control of Gene Expression,* Cell **152**, 1173 (2013).

[19] D. G. Gibson, *Synthesis of DNA fragments in yeast by one-step assembly of overlapping oligonucleotides,* Nucleic Acids Research **37**, 6984 (2009).

[20] S. M. Soucy, J. Huang, and J. P. Gogarten, *Horizontal gene transfer: building the web of life,* Nature Reviews Genetics **16**, 472 (2015).

[21] J. L. Martinez, *Antibiotics and Antibiotic Resistance Genes in Natural Environments,* Science **321**, 365 (2008).

[22] A. Babić, A. B. Lindner, M. Vulić, E. J. Stewart, and M. Radman, *Direct Visualization of Horizontal Gene Transfer,* Science **319**, 1533 (2008).

[23] D. H. Duckworth, *"Who Discovered Bacteriophage?".* Bacteriological reviews **40**, 739 (1976).

[24] J. Lederberg, *Cell genetics and hereditary symbiosis.* Physiological reviews **32**, 403 (1952).

[25] D. Bikard, W. Jiang, P. Samai, A. Hochschild, F. Zhang, and L. A. Marraffini, *Programmable repression and activation of bacterial gene expression using an engineered CRISPR-Cas system,* Nucleic Acids Research **41**, 7429 (2013).

[26] J. Wiktor, C. Lesterlin, D. J. Sherratt, and C. Dekker, *CRISPR-mediated control of the bacterial initiation of replication,* Nucleic Acids Research **44**, 3801 (2016), arXiv:arXiv:1011.1669v3 .

[27] M. H. Larson, L. A. Gilbert, X. Wang, W. A. Lim, J. S. Weissman, and L. S. Qi, *CRISPR interference (CRISPRi) for sequence-specific control of gene expression,* Nature Protocols **8**, 2180 (2013).

[28] Z. Chen, H. Yang, and N. P. Pavletich, *Mechanism of homologous recombination from the RecA-ssDNA/dsDNA structures.* Nature **453**, 489 (2008), arXiv:NIHMS150003 .

[29] D. G. Gibson, L. Young, R.-Y. Chuang, J. C. Venter, C. a. Hutchison, H. O. Smith, C. A. H. Iii, and N. America, *Enzymatic assembly of DNA molecules up to several hundred kilobases.* Nature methods **6**, 343 (2009).

[30] S. de Kok, L. H. Stanton, T. Slaby, M. Durot, V. F. Holmes, K. G. Patel, D. Platt, E. B. Shapland, Z. Serber, J. Dean, J. D. Newman, and S. S. Chandran, *Rapid and Reliable DNA Assembly via Ligase Cycling Reaction,* ACS Synthetic Biology **3**, 97 (2014).

[31] T. G. Bernhardt and P. A. De Boer, *The Escherichia coli amidase AmiC is a periplasmic septal ring component exported via the twin-arginine transport pathway,* Molecular microbiology **48**, 1171 (2003).

5

CRISPR-MEDIATED CONTROL OF THE BACTERIAL INITIATION OF REPLICATION

Programmable control of the cell cycle has been shown to be a powerful tool in cell-biology studies. Here, we develop a novel system for controlling the bacterial cell cycle, based on binding of CRISPR/dCas9 to the origin-of-replication locus. Initiation of replication of bacterial chromosomes is accurately regulated by the DnaA protein, which promotes the unwinding of DNA at oriC. We demonstrate that the binding of CRISPR/dCas9 to any position within oriC blocks the initiation of replication. Serial-dilution plating, single-cell fluorescence microscopy, and flow-cytometry experiments show that ongoing rounds of chromosome replication are finished upon CRISPR/dCas9 binding, but no new rounds are initiated. Upon arrest, cells stay metabolically active and accumulate cell mass. We find that elevating the temperature from 37 to 42 °C releases the CRISR/dCas9 replication inhibition, and we use this feature to recover cells from the arrest. Our simple and robust method of controlling the bacterial cell cycle is a useful asset for synthetic biology and DNA-replication studies in particular. The inactivation of CRISPR/dCas9 binding at elevated temperatures may furthermore be of wide interest for CRISPR/Cas9 applications in genomic engineering.

5.1. INTRODUCTION

The initiation of the replication of a chromosome is an evolutionary conserved process, and the mechanism of initiation is very similar among different organisms [1, 2]. In bacteria, replication is initiated when DnaA proteins recognize and bind to specific sequences - DnaA boxes - within the origin of replication locus (*oriC*). Upon polymerization, they cause adjacent double-stranded DNA to melt, providing a single stranded substrate onto which the replication machinery is loaded and replication begins [3–5]. The process of initiation is tightly regulated in time, and several mechanisms for preventing premature initiation are known in different species. Origin sequestration by SeqA protein, distant DNA tethering to the bacterial membranes, or specific protein regulators interacting with replication initiation machinery are known to act as negative regulators of initiation of the replication [6–8].

Programmable control of the cell cycle has proven to be a powerful tool in cell biology, elucidating many processes involved in replication and metabolism [9, 10]. The dynamic growth of the field of synthetic biology also strives for robust methods for controlling the rate of bacterial replication which can be integrated into genetic circuits [11]. A number of methods are currently available for bacterial cell-cycle control: bacterial cells can be arrested in a pre-replication state using thermosensitive variants of DnaC and DnaA replication-initiation proteins [12], using columns which release newborn cells [13], or using restricted growth conditions [14]. None of these methods, however, provide an easy and chemically inducible control over the bacterial cell cycle.

The CRISPR (*c*lustered *r*egularly *i*nterspaced *s*hort *p*alindromic *r*epeats) system, naturally responsible for bacterial immunity against viruses, has recently been widely engineered and repurposed for many biological applications, from genome engineering and *in vitro* RNA digestion to fluorescent labeling of genomic positions [15–17]. A simple type II CRISPR system, originating from *Streptococcus pyogenes*, consists of only 2 components, a Cas9 protein and small guide RNA (sgRNA). CRISPR/Cas9 system is very efficient in recognizing DNA sequences that are complementary to the specific 20 nucleotides of the sgRNA sequence [18]. A modification of Cas9 protein, which abolishes its nuclease activity, allows CRISPR/dCas9 (dCas9 – nuclease-deficient Cas9) to form such a stable complex with the complementary DNA that, if targeted to the promoter region of the gene, it excludes RNA polymerase from binding to the promoter sequence. Such an approach was used to regulate gene expression in a wide range of organism [19, 20].

Here, we repurpose a CRISPR/dCas9 system for synchronization of *E. coli* cells into the pre-replication state. We do so by inactivating the initiation of replication of the bacterial chromosome by targeting CRISPR/dCas9 to the *E. coli* origin of replication. The replication arrest is found to be highly specific to the origin-of-replication locus and is not observed when dCas9 is targeted to proximal regions of *oriC*. Flow cytometry chromosome-number counting and single-cell fluorescence microscopy show that initiation of replication is blocked very efficiently after the expression of our system in bacterial cells. Furthermore, we show that CRISPR/dCas9 is not active at elevated temperatures, and we exploit that property to recover bacterial cells from the arrested state of replication.

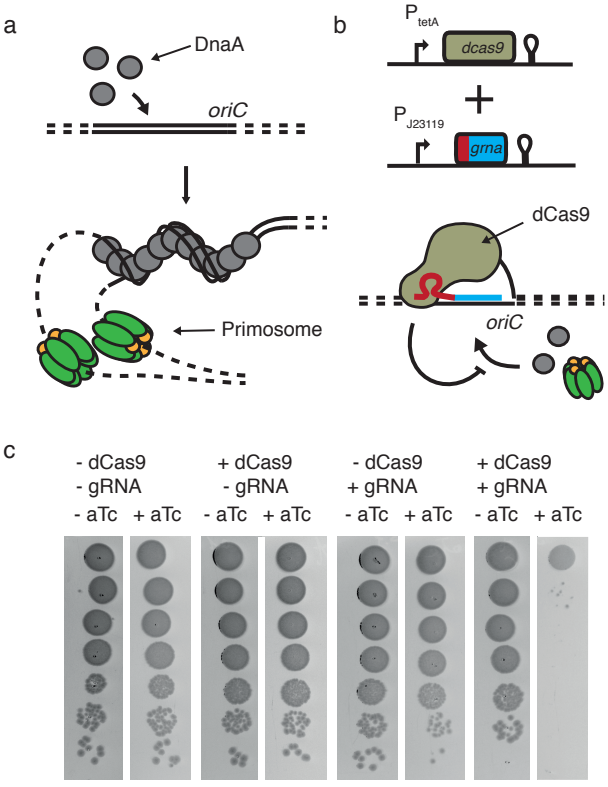


Figure 5.1: **CRISPR/dCas9 system blocks initiation of replication at the oriC locus.** (a) Schematic of initiation of replication by DnaA. Cooperate binding of DnaA proteins to oriC induces unwinding of an adjacent AT-rich region, thus providing single-stranded DNA substrate that is recognized by primosome complexes. (b) The CRISPR/dCas9 system consists of two plasmids, one coding *dCas9* under the control of an aTc-inducible promoter and the other coding *sgRNA* under control of a constitutive promoter. When CRISPR/dCas9 binds to the *oriC* region, DnaA cannot bind and unwind the DNA, and initiation of replication is blocked. (c) Simultaneous expression of dCas9 and *sgRNA* has a lethal effect on cells. Serial ten-fold dilutions of liquid bacterial cultures were plated either on the media supplemented (+aTc) or not (-aTc) with 200 ng/ml of aTc. Only in presence of both CRISPR/dCas9 components, cells are not viable.

5.2. MATERIALS AND METHODS

STRAINS AND CULTURE CONDITIONS

All experiments were done with derivatives of *E. coli* K12 TB28 (MG1655; *lacIZYA* [21]) with the exception of the serial-dilution plating experiment shown in Fig. 5.1, which was done with *E. coli* K12 AB1157 strain with chromosomal loci marked with *lacO* [22]. For serial-dilution plating experiments and genetic manipulations, cells were grown in a Lysogenic broth (LB) at 37°C, except for the experiment testing the thermo-sensitive properties of CRISPR/dCas9, where cells were grown in LB media at 30°C, 37°C or 42°C, as specified. Ampicillin (100 μg/ml) and chloramphenicol (34 μg/ml) were added when required. Expressions of dCas9 and dCas9deg were induced with the addition of anhydrotetracycline (aTc, 200 ng/ml). For microscopy and flow-cytometry experiments, cells were grown in M9 media supplemented with 0.2% glucose at 37°C, or at 42°C for the recovery experiments.

PLASMID AND STRAIN CONSTRUCTION

Supplementary Table 5.1 lists the plasmids and sgRNA targets used in this study. Top10 cells (Thermo Fisher) and the Mix &Go *E. coli* transformation kit (Zymo Research) were used to transform all cloning reactions. Plasmids pdCas9-bacteria and pgRNA-bacteria were obtained from Addgene [23]. Plasmid pdCas9deg was created by restriction digestion and ligation of a PCR fragment obtained by amplification of pdCas9 plasmid backbone with primers Jw098 and Jw099 containing a LAA degradation tag sequence [24] and XhoI restriction sites.

Modifications of sgRNA 20nt sequences were done by PCR amplification of a pgRNA backbone with primers carrying a SpeI restriction site and 20bp of sgRNA sequence. The PCR fragment was digested and ligated into a circular plasmid. A list of primers used to create pgRNA plasmids can be found in supplementary Table 5.2. J23119 constitutive promoter drove the expression of sgRNA.

Plasmid pdCas9deg-gRna3 was created by a CPEC reaction [25], by combining pdCas9deg with the sgRNA region of plasmid pgRNA3 with primers Jw121, Jw122, Jw124 and Jw125. In this construct, pdCas9deg was under the control of an aTc-inducible promoter and gRNA3 was placed under the control of a constitutive J23119 promoter. The strain containing the origin-proximal FROS system and LacI-tagGFP was constructed by P1 phage transduction (as described in [26]) from strain BN1352, carrying an origin proximal *lacO* array [22], and a strain BN1442 carrying a LacI-tagGFP fusion under the control of lactose promoter, into the TB28 strain. Resistances were removed, when possible, using a Flp recombinase expressed from pCP20 [27].

SERIAL-DILUTION EXPERIMENTS

For serial-dilution experiments, cells were grown at 37°C in LB (Lysogeny broth) with addition of ampicilin (100 μg/ml) and chloramphenicol (34 μg/ml) as needed. Tenfold serial dilution were made by diluting 20 μl of cell suspension in 180 μl of LB media in every dilution step. Cells were plated on LB agar plates supplemented with antibiotics, as needed, and incubated at an appropriate temperature for 18 hours before imaging.

FLUORESCENCE MICROSCOPY

Fluorescence microscopy experiments were carried using a Nikon Ti-E microscope with CFI Plan Apochromat λDM 100X objective, Lumencor Spectra X LED light source, Andor Zyla 4.2 CMOS camera, and a Lumencor SpectraX filter set. Images were collected with Nikon NIS software and analyzed using FIJI software [28], microbeTracker suite [29], and custom Matlab scripts and functions.

FLOW CYTOMETRY

Cells were grown in M9 glucose at 37°C (or at 42°C for recovery experiments). Induction with 200 ng/ml of aTc was started in the early exponential phase (OD600nm ~ 0.1). Cell samples were prepared as described in [30], and Syto16 (Life Technologies) was used to stain the DNA. Data were collected with the use of a FACSacn flow cytometer (BD Biosciences) and analyzed with custom Matlab scripts.

5.3. RESULTS

CRISPR/DCAS9 BLOCKS THE INITIATION OF REPLICATION AT THE *oriC* LOCUS

The initiation of replication in *E. coli* is restrained to a relatively short (245 bp) and well-defined DNA sequence called *oriC* (Fig. 5.1A, top) [31]. Because the earliest step of initiation relies on DnaA binding to specific regions of *oriC,* called DnaA boxes, it is possible to design a simple CRISPR/dCas9 system that inhibits the cellular replication-initiation machinery by hindering DnaA – *oriC* interactions (Fig. 5.1b, bottom). To engineer such a system, we used a previously described set of two plasmids, one coding for the dCas9 protein, under control of an aTc inducible tetracycline promoter (pdCas9), and the second plasmid coding for the small guide RNA (sgRNA) under the control of a constitutive promoter (Fig. 5.1b) [23]. Previous results showed that binding of dCas9 to the promoter region of a gene can efficiently block its transcription [20, 23], and targeting it at any protein-recognized DNA sequence may hinder protein-DNA interactions specific to that region. To test if CRISPR/dCas9 binding to the origin of replication can impair the initiation of replication, we constructed and tested a CRISPR/dCas9 that binds to one of the DnaA boxes in the *oriC* (R1 DnaA box) [32].

Serial-dilution plating shows that cells were not viable if dCas9 and sgRNA were expressed simultaneously: bacteria plated on aTc-containing media showed a drastic reduction in survival (Fig. 5.1c, rightmost row). Expression of guide RNA or dCas9 alone had no significant effect on cell survival in the serial-dilution plating experiments (Fig. 5.1c, left). We suggest that the observed effect results from a competition between DnaA and dCas9 binding to the same DnaA box. Tight binding of CRISPR/dCas9 can prevent DnaA from sequence recognition, thus preventing DnaA-filament formation and the subsequent associated replication-bubble opening that starts replication.

CRISPR/DCAS9 INHIBITION OF REPLICATION INITIATION IS SPECIFIC TO THE *oriC* REGION

To test if the observed effect is specifically caused by inhibition of replication initiation, we probed the effect of CRISPR/dCas9 binding to many targets throughout the entire oriC region, as well as adjacent proximal regions, using a library sgRNA constructs. The location of all used sgRNA targets is shown on the Fig. 5.2a, overlaid on a genetic map of

the *oriC* region and its surrounding. Targets displayed in red are binding within the *oriC* whereas green targets are binding outside of the *oriC* origin of replication (Fig. 5.2a).

We find that growth inhibition is only observed when CRISPR/dCas9 is binding to targets within the origin or replication. Binding to regions upstream or downstream of the origin site does not affect cell growth after the induction of CRISPR/dCas9 system with aTc (as observed in the green-labeled serial-plate results in Fig. 5.2c). This indeed indicates a high specificity of the CRISPR/dCas9 system that only blocks the initiation of replication when it binds at any spot within the *oriC* locus.

A potential alternative explanation for the loss of viability of cells is stalling of the replisome during the collision with CRISPR/dCas9 bound to DNA. Indeed, replication-machinery stalling has been reported when the replication fork encountered a repeated sequence of operators that were bound by repressors [33, 34]. However, we find that only CRISPR/dCas9 with guide RNA that are complementary to sequences within the *oriC* lead to cell death, whereas guide RNA targeting sites outside of *oriC*, or other genomic positions tested in other studies, are not lethal. These observations lead to the conclusion that the system is indeed specifically blocking the initiation of the replisome formation, and not its progression.

5

SINGLE-PLASMID CODING SHOWS A SUPERIOR PERFORMANCE OVER A TWO-PLASMID SYSTEM

To minimize non-specific effects of high levels of CRISPR/dCas9 expression, we fused dCas9 protein to a strong LAA degradation signal tag [35], creating a dCas9deg protein. Next, we combined dCas9deg (dCas9 fused to LAA degradation tag) protein and sgRNA3 (Fig. 5.2) on a single low-copy-number plasmid, called pdCas9deg3 (Supplementary Fig. 5.6a). The LAA tag is promoting an active degradation of a dCas9deg by the ClpXP protease, preventing an accumulation of surplus CRISPR/dCas9 in the cell. Such accumulation could potentially trigger a cellular stress by accumulation of inclusion bodies. The dCas9deg is still active in the serial-dilution plating assay, but the LAA tag reduces the aTc sensitivity of the CRISPR/dCas9deg about 10 fold, compared to a original dCas9 protein (Supplementary Fig. 5.6b).

Combination of dCas9deg and pgRNA3 on a single plasmid further improved the performance of the system. The chromosome counting experiment performed with a two-plasmid system (pdCas9 and pgRNA3) shows a replication-inhibited population of bacteria, but also a population of small-sized particles, presumably the remains of dead cells (Fig. 5.7). An identical experiment done with a single-plasmid system shows a well-synchronized culture (Fig. 5.3). Both modifications of CRISPR/dCas9 system, introducing the LAA tag and expressing both components from a single plasmid, proved to be necessary to create a viable cell-cycle control system.

CHROMOSOME COUNTING VERIFIES THAT CRISPR/DCAS9DEG3 SYSTEM INHIBITS THE INITIATION OF REPLICATION

To test the effect of CRISPR/dCas9deg3 binding to the *oriC* region at the level of individual MG1655 *E. coli* bacterial cells, we examined the DNA content of cells with flow cytometry. DNA was stained with Syto16, which provides a quantitative measure for the amount of chromosome equivalents as the fluorescent signal scales linearly with DNA content of a cell. At time zero, early exponential cell culture growing in M9 medium

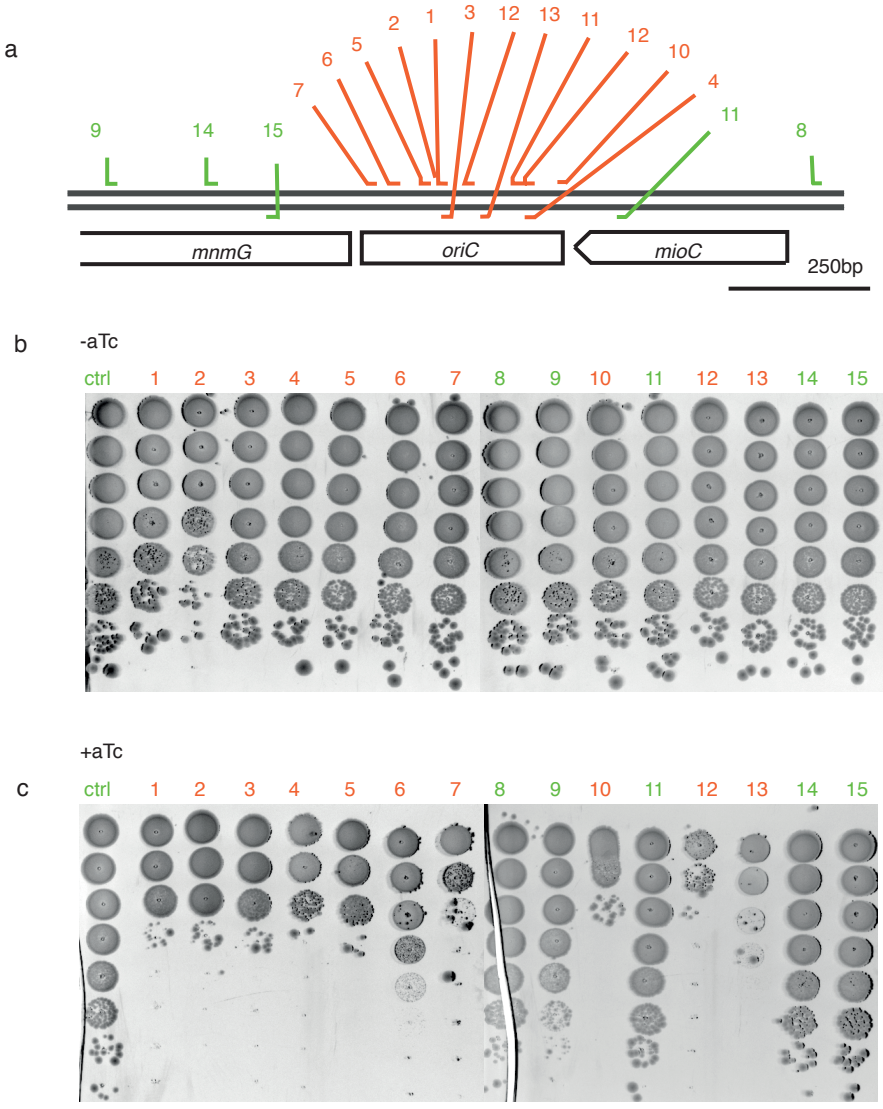


Figure 5.2: **CRISPR/dCas9 inhibition of replication initiation is highly specific to the oriC region.** (a) Genetic map of the *oriC* region. The 245 bp oriC sequence is marked as well as two adjacent gene-coding regions. Each number corresponds to a different gRNA 20nt target. Targets complementary to *oriC* are shown in red, targets binding outside of *oriC* are marked in green. Double line represents the dsDNA; sgRNA targets binding the leading strand are represented to bind the upper strand; gRNA targets complementary to the lagging strand are represented to bind the bottom strand. (b and c) CRPSR/dCas9 inhibits cell growth only when targeted to oriC. Serial ten-fold dilutions of bacterial cultures were plated on media without the addition (b) or with addition (c) of 200 ng/ml of aTc. Numbers corresponds to gRNA targets shown on (a).

supplemented with glucose was split into two smaller cultures, and one of the two was induced with 200ng/ml of aTc.

Upon induction of the CRISPR/dCas9deg3 system, we could clearly distinguish the development of a population of cells that carry only one copy of the chromosome (Fig. 5.3a, shown as a '1c' population on the bottom-left histogram). While cells in the controls contain on average 2 chromosomes, the induced arrest of the initiation of replication leads to a reduction of chromosome content to mostly 1 copy per cell. The shift from 2 to 1 chromosomes is visible already ~60 minutes after the induction and stabilizes after ~180 minutes, and it is clearly visible both on fluorescent intensity histograms and in the appearance of a clearly separated bottom population in the contour plots (Fig.'5.3a). A control experiment with a bacterial culture without the aTc induction shows a typical DNA content distribution of a logarithmic culture in M9 media supplemented with glucose (Fig. 5.3a, right column, mostly 2 chromosomes per cell). Moreover, DNA content profiles for un-induced culture do resemble the cultures not transformed with pdCas9deg3 plasmid, with, or without addition of aTc (Fig. 5.3b). The data indicate that upon blocking the initiation of replication by CRISPR/dCas9deg3 binding, the ongoing replication is finished and cell division leads to cells with 1 chromosome, which cannot be further replicated. In other words, replication forks that were established before the inhibition of initiation are not interrupted by the CRISPR/dCas9deg3 system, and the observed effect is specific to the initial replisome formation at the *oriC*.

These results suggest that the CRISPR/dCas9deg3 system specifically and efficiently blocks the initiation of chromosome replication. Importantly, similar experiments exploiting thermosensitive properties of DnaC2 protein yielded a very comparable result for the chromosome content upon inhibition of initiation of replication [12].

SINGLE-CELL FLUORESCENCE MICROSCOPY OF CRISPR/DCAS9DEG3 ARRESTED CELLS

Next, we investigated the effects of CRISPR/dCas9deg3 binding on cell morphology and chromosome copy number using fluorescence microscopy of MG1655 *E. coli* cells to which we integrated a fluorescent probe (FROS – fluorescent repressor-operator system) in the proximity of *oriC* location [22]. Because origin regions are readily segregated after replication in *E. coli,* we can thus estimate the number of replicated origins by simply counting the number of origin-proximal foci in each cell. As expected, we find that the bacterial cells in which initiation of replication is inhibited contain mainly a single fluorescent focus (Fig. 5.4a, top), meaning that the region labeled with FROS is not replicated, in contrast to un-induced cells where each round of replication doubles the number of foci (Fig. 5.4a, bottom). Analysis of the origin-proximal foci number thus again indicates that the CRISPR/dCas9deg3 efficiently blocks initiation of replication. At 240 minutes after the induction, 73±8% cells contain single origin focus, a fraction about twice as large as at the beginning of the experiment where it was 36±4% (Fig. 5.4c, red line) or for a population growing without the presence of aTc which yielded 36±4% (Fig. 5.4 c, inset).

CRISPR/DCAS9DEG3 BINDING TO *oriC* DOES NOT INHIBIT CELL GROWTH

We next tested whether the inhibition of initiation of replication had an effect on cell growth. Cells from previous experiment were found to elongate after the exposure to aTc: 240 minutes after the induction, the mean cell length had increased to 4.5 ± 1.5 μm,

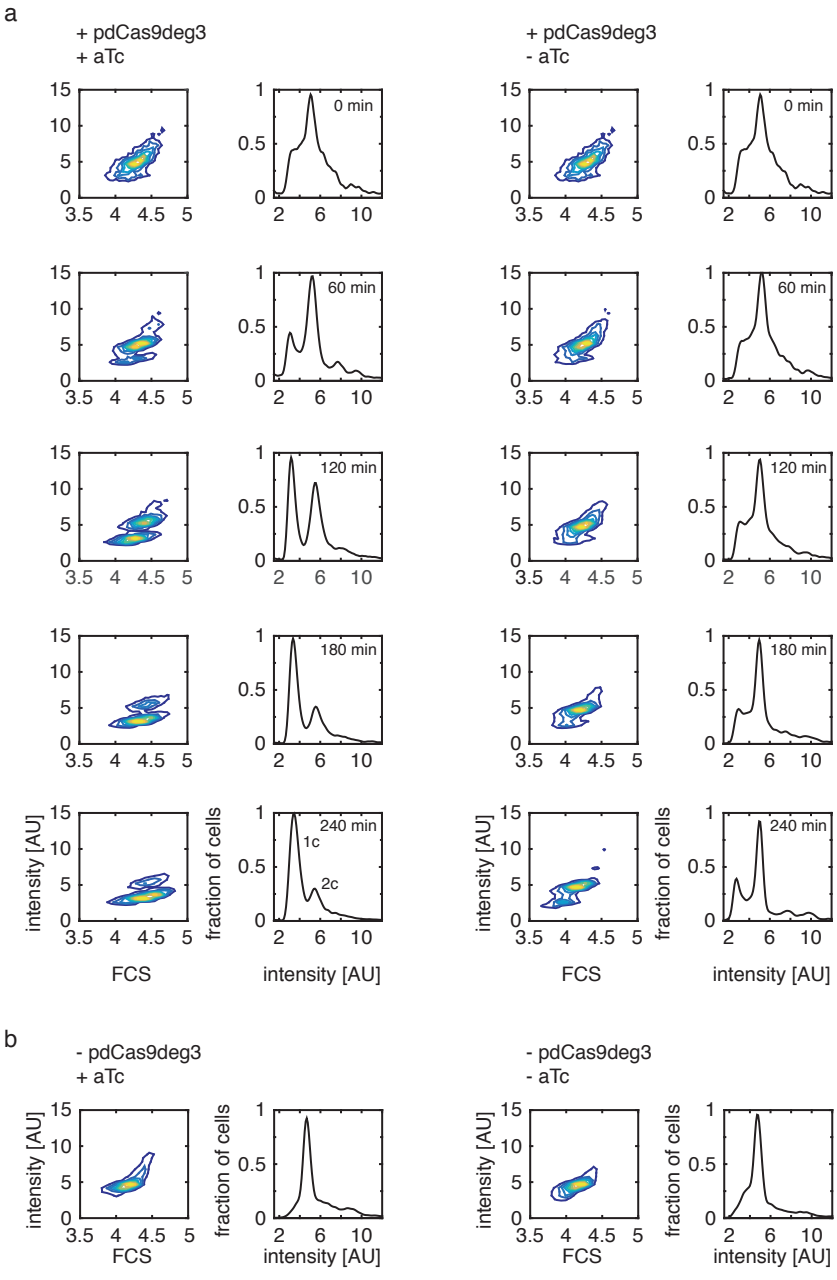


Figure 5.3: **CRISPR/dCas9deg3 binding to oriC inhibits the initiation of replication.** Contour plots of cytograms (left) and fluorescence-intensity histograms (right). (a) Early exponential cell culture was divided into two subcultures, where one (left) was induced by addition of 200 ng/ml aTc and the other one (right) was not. Each hour, cells were fixed, DNA was stained with Syto16, and the fluorescence signal was measured. Induced culture showed an arrested chromosome distribution, which saturated at 180 minutes after induction. Non-induced culture maintained the physiological chromosome content during the entire time of the experiment. The signal equivalent to one or two chromosomes is indicated by '1c' and '2c', respectively, on the 240 minutes-induced histogram. (b) Addition of 200 ng/ml of aTc had no observed effect on the chromosome content of bacteria lacking CRISPR/dCas9deg3 system.

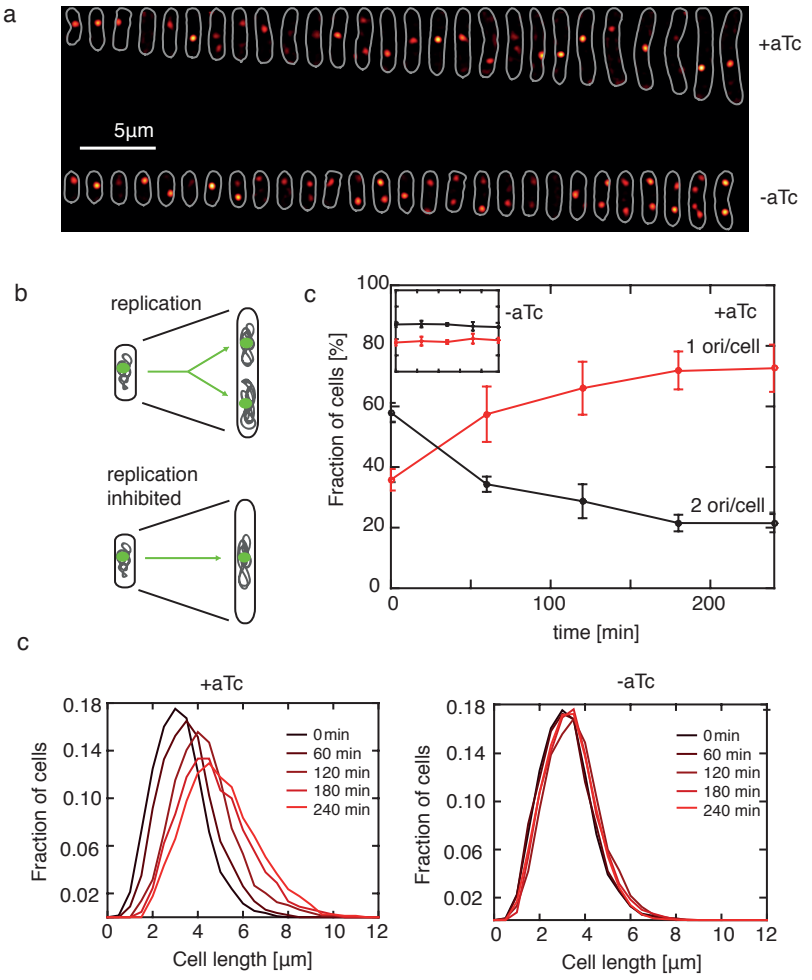


Figure 5.4: CRISPR/dCas9deg3 binding to the origin of replication inhibits the initiation of replication but not the cell growth. (a) Representative images of individual cells grown in M9 glucose at 37°C with (top stripe) or without 200ng/ml of aTc (bottom stripe). All cells images were collected at 240 minutes after the beginning of induction. White color denotes the cell outlines obtained with microbeTracker software. Red spots denote the *oriC* loci signaled by LacI-tagGFP labels. Cells were lined up from the shortest to the longest and are a representative of the bacterial population. Background has been subtracted. (b) Inhibition of replication will lead to an arrested cell phenotype. Arrested cells will not replicate the DNA (hence will have 1 chromosome), but stay metabolically active and increase in size. Green – *oriC*; gray - DNA (c) CRISPR/dCas93 system arrests the cell population, yielding cells with 1 chromosome. After the addition of 200 ng/ml of aTc the population of cells with only 1 *oriC* focus is increasing, and saturates to a fraction larger than 70% after 180 minutes. Population cultured without the CRISPR/dCas9deg3 induction maintains stable level of *oriC* content. Cells were grown in M9 glucose in 37°*C*. Error bars indicate SD for 3 independent experiments for each dataset for each time point, with the exception of t = 240 min, +aTc, where the error bar denotes the mean standard deviation of the dataset. Colors in the inset correspond to those in the main panel. (d) Cell lengths increase after the induction of CRISPR/dCas9deg3-induced arrest. The distribution of lengths of cells in the population is shifting towards longer cells after the induction of arrest. The distribution stabilizes after 180 minutes of induction. Cells are grown as in panel (a) and (c).

which is significantly larger than the initial length of 3.1 ± 0.8 μm (see Fig. 5.4a and d). Cells not induced with aTc maintained a stable distribution of cell length throughout the entire length of the experiment (Fig. 5.4d, right). The increased length of the cells suggests that cell division is inhibited by nucleoid occlusion mechanism, which blocks the formation of the division machinery when *E. coli* chromosomes are not correctly replicated and segregated [36] (Fig. 5.4d, left). Furthermore, the larger cell sizes indicate that the CRISPR/dCas9deg3 system used in our study is not stopping the cell metabolism, as cells are still able to grow and elongate after the induction. These results demonstrate that our system specifically and efficiently interferes with the initiation of replication, while cells are still well viable and able to increase their mass.

REPLICATION CAN BE RESTORED WITH THE THERMO-SENSITIVE PROPERTIES OF CRISPR/DCAS9DEG

In doing these experiments, we unexpectedly discovered that the CRISPR/dCas9 (or for that matter CRISPR/dCas9deg) inhibition of replication is not functional at 42°C. Cells plated on a LB agarose media supplemented with aTc are, surprisingly, able to form colonies at 42°C, whereas cells cultured at 37°C or 30°C do not proliferate in presence of aTc (Fig. 5.5a and b). This is, to our knowledge, the first report of thermosensitive properties of the CRISPR/dCas9 system.

This new property of CRISPR/dCas9 allowed us to reinitiate the previously inhibited replication of chromosomes. Cells transformed with the pdCas9deg3 plasmid were first arrested by the addition of aTc to M9 media supplemented with 0.2% glucose for 2.5 h prior to the start of the temperature-mediated recovery. After arrest, cultures were diluted 10 times in fresh M9 glucose media to reduce the concentration of the inducer, and cultured further at 37°C or at 42°C. The cells growing at 37°C did not show any signs or recovery during the course of an experiment (Fig. 5.5c). At 42°C, however, first replicating cells were visible already after 60 minutes of incubation. 180 minutes after the shift to elevated temperature, a large fraction of replicating cells is clearly visible on the contour plot. (Fig. 5.5d, indicated by the arrow). We can thus exploit this new thermosensitive property of the CRISPR/dCas9 system to reverse effects of inhibition of initiation of replication.

5.4. DISCUSSION

This paper reports a novel system for controlling the bacterial cell-cycle stage using programmable inhibition and re-activation of the initiation of replication of the *E. coli* chromosome. CRISPR/dCas9deg3 efficiently blocks replication at the initiation stage and thus synchronizes a bacterial culture at the pre-replicating state. Arrested cells are still metabolically active, and can restart the replication and proliferation after a switch to 42°C. Our approach is reminiscent of mechanisms based on blocking DnaA from binding to the origin of replication [7, 8] by specific proteins (Spo0A in *B. subtilis*, or CtrA in *C. cresentus*) and our CRISPR/dCas9 based system mimics such naturally occurring mechanisms.

Interestingly, the DNA-binding footprint of CRISPR/dCas9 (~20 nucleotides based on the crystallographic studies [37]) is much shorter than the *oriC* sequence (245 bp [31]), and a single CRISPR/dCas9 molecule thus covers only a small fraction of

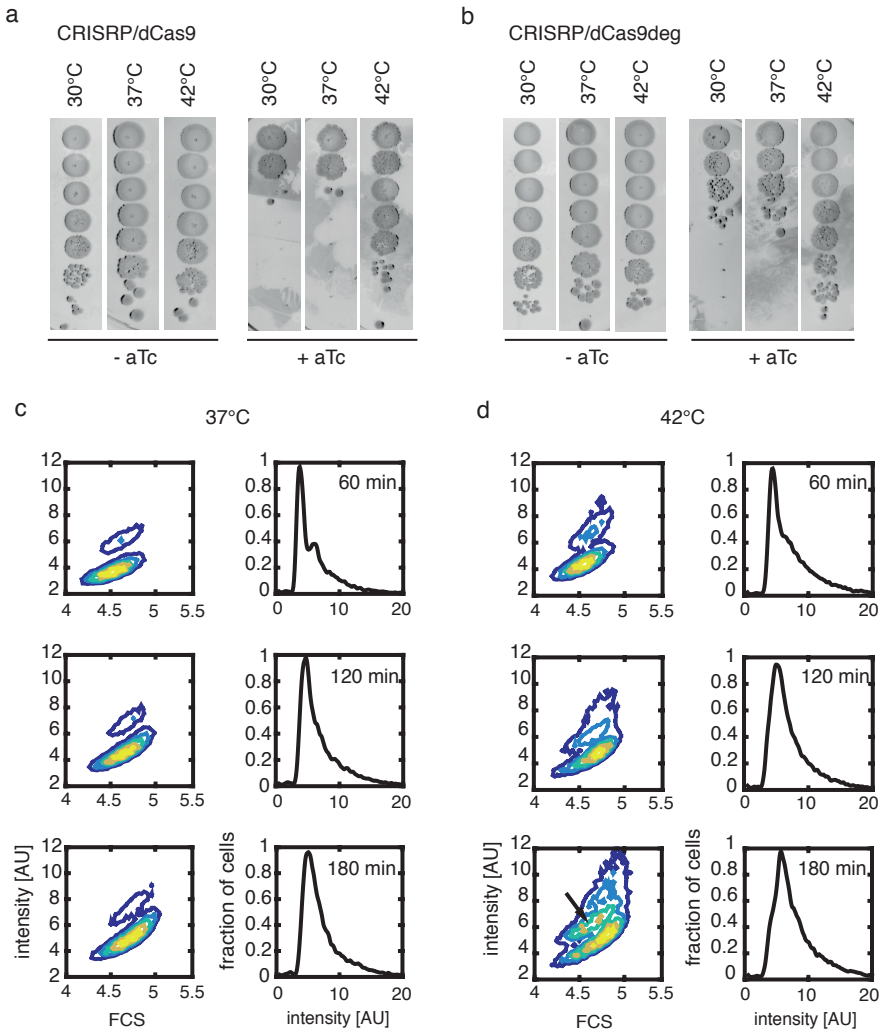


Figure 5.5: **CRISPR/dCas9 and CRISPR/dCas9deg are thermosensitive and the arrest can be reversed.** (a) Effects of temperature on dCas9 and dCas9deg supplemented with pgRNA3. Cells were cultured in LB media, ten-fold serial dilutions were plated with or without 200ng/ml of aTc and cultured at different temperatures. In both cases, the system was supplemented with pgRNA3 guide. For both CRISPR/dCas9 and CRISPR/dCas9deg, the system is not active at elevated temperature and cells are able to form colonies. (b) Cells transformed with pdCas9deg3 were arrested for 150 minutes in M9 glucose at 37°C, the aTc concentration was reduced by a ten-fold dilution of cells in fresh media, and cultures were grown at either 37°C or 42°C. A population of replicating cells can be observed already after 60 minutes at 42°C, and after 180 minutes replicating cells are forming a substantial fraction of the whole population. Contour plots show the DNA content (Syto16 intensity) as a function of cell mass (FCS), while the histograms show the distribution of the DNA content (Syto16 intensity) in the population. An arrow indicates the replicating cells population. The peak on the distribution of DNA content for the same sample is more pronounced and is shifted to the right when compared to the 37°C population. Cells cultured at 37°C are not showing recovery from the arrest.

the entire origin. Nevertheless, that single binding process is highly efficient in preventing the initiation of replication, as shown by our local binding data at many (~10) targets within the *oriC*. This indicates that the entire *oriC* region is important for DnaA-induced replication-bubble formation. CRISPR/dCas9 binding to regions in close proximity to *oriC* showed, however, no effect on the bacterial viability, also verifying that replication initiation is strictly restrained to the *oriC* region, and that the adjacent DNA does not play a vital role in the process of initiation. The system is also active at temperatures of 30°C, at which thermosensitive versions of proteins DnaC and DnaA are able to initiate the replication. Therefore, it can also be used to study processes which require low-temperature growth conditions [12].

In our work, we reported, for the first time to our knowledge, that the system based on CRISPR/dCas9 exhibits thermo-sensitive properties. We used this phenomenon to restart the cell cycle from the arrested phase, but this particular feature of dCas9 systems may find much more general applications. For instance, in Cas9-mediated genome editing, rapid temperature switching may provide control over the dCas9 and likely the wtCas9 activity.

The CRISPR/dCas9deg system controlling the replication can be integrated into any existing genetic circuits. Rational design of genetic circuits lies at the foundation of the rapidly developing field of synthetic biology [38, 39]. In practical application, our CRISPR/dCas9deg can arrest cell growth after the engineered bacterium served its role [11].

Our results show that CRISPR/dCas9 can be used to efficiently control an important biological process, viz., initiation of DNA replication. It provides a new tool to control the cell cycle, which can be used, for example, in studies of the bacterial metabolism or the bacterial genome. We have also shown that CRISPR/Cas9 systems has thermo-sensitive properties, which adds another level of control to this powerful gene-editing system. In summary, we provide a novel, efficient, and simple method to control a stage of bacterial replication using an engineered CRISPR/dCas9deg system.

ACKNOWLEDGEMENTS

We thank Fabai Wu for strain BN1352 and discussions, Hugo Snippert for strain BN1442, Amy Upton for strain AB1157, Erwin Van Rijn for assistance with FACS experiments, and Katarzyna Ginda for discussions. This work was partly supported by the Netherlands Organisation for Scientific Research (NWO/OCW) as part of the Frontiers of Nanoscience program, as well as by European Research Council NanoforBio No. 247072 (CD) and SynDiv 16 669598 (CD). Research in the Sherratt laboratory was supported by the Wellcome Trust (SIA099204/Z/12Z) and the Leverhulme Trust (RP2013-K-017)

5.5. SUPPLEMENTARY INFORMATION

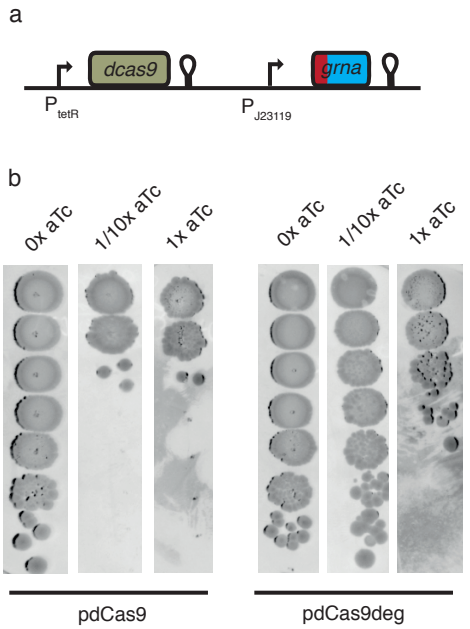


Figure 5.6: **Creation of the pdCas9deg3 plasmid.** (a) A pdCas9deg3 plasmid consist of dCas9deg-coding gene under the control of an aTc inducible promoter and a sgRNA-coding gene under the control of the constitutive promoter with the p15a origin of replication. (b) dCas9deg protein is observed to be less effective in inhibiting the replication than dCas9 that is lacking the degradation tag. Serial ten-fold dilutions of liquid bacterial cultures were plated either on the media supplemented with 20 ng/ml or 200 ng/ml of aTc or lacking aTc. Cells were also co-transformed with pgRNA3.

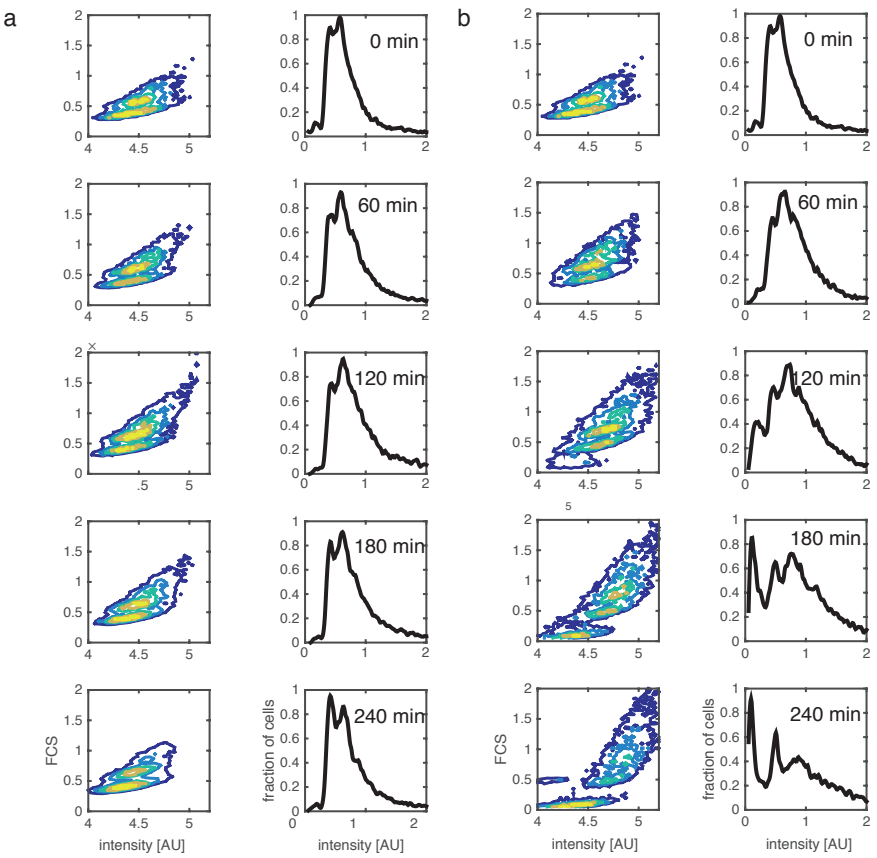


Figure 5.7: **The two-plasmid system does not preserve the cell physiology well.** CRISPR/dCas9 system was expressed from two plasmids, pdCas9deg and pgRNA. Cells were grown in M9 glucose at 37°C. Early exponential culture was induced with 200 ng/ml of aTc and analyzed with flow cytometer every 60 minutes. DNA was stained with Syto16. (a) Uninduced culture maintains a constant DNA/cell concentration throughout the experiment. (b) The induced culture shows an anomalous distribution of cell sizes. The increasing population of small, but highly fluorescent cells suggests that overexpression of CRISPR/dCas9deg system binding to the oriC has negative effects on the bacterial cells.

Table 5.1: List of plasmids.

Name	Genotype	Reference
pdCas9	p15a origin, PtetA-dCas9	Addgene, (23)
pgRNA	ColE1 origin, PJ23119-gRNA	Addgene, (23)
pdCas9deg	p15a origin, PtetA-dCas9deg (AANDENYALAA)	This study
pdCa9deg-gRNA3	p15a origin, PtetA-dCas9deg, PJ23119-gRNA	This study
pgRNA1	CCCTGTGGATAACAAGGATC	This study
pgRNA2	GCACTGCCCTGTGGATAACA	This study
pgRNA3	GCCGGATCCTTGTTATCCACA	This study
pgRNA4	CTACTGTGGATAACTCTGTC	This study
pgRNA5	TTCTATTGTGATCTCTTATT	This study
pgRNA6	TTGAGAAAGACCTGGGATCC	This study
pgRNA7	ATGGATTGAAGCCCGGGCCG	This study
pgRNA8	ATTGTACGCTGTGAACGCGT	This study
pgRNA9	CACGGAACTTCAGTCCCATT	This study
pgRNA10	GTCAGGAAGCTTGGATCAAC	This study
pgRNA11	GAGGCAGAACTCAAAAATTC	This study
pgRNA12	GATCATTAACTGTGAATGAT	This study
pgRNA13	AGCTTATACGGTCCAGGATC	This study
pgRNA14	TTGCGTTTAGTATCCTAAAC	This study
pgRNA15	ATGAGCTGCAACCCGGCGAT	This study
pgRNActrl	CCGCCGTGTCACTTTCGCTTTGG	This study

Table 5.2: List of oligonucleotides.

Name	Sequence (5' - 3')		Purpose
JW096	GGACTAGTATTATACCTAGGACTGAG		pgRNA cloning reverse
JW090	CCACTAGTCCGCCGTGTCACTTTCGCTT GAGCTAGAAATAGCAAG	GTTTTA-	pgRNA ctrl integration
JW092	CCACTAGTCCCTGTGGATAACAAGGATC GAGCTAGAAATAGCAAG	GTTTTA-	pgRNA 1 integration
JW093	CCACTAGTGCACTGCCCTGTGGATAACA GAGCTAGAAATAGCAAG	GTTTTA-	pgRNA 2 integration
JW094	CCACTAGTGCCGGATCCTTGTTATCCACA GAGCTAGAAATAGCAAG	GTTTTA-	pgRNA 3 integration
JW095	CCACTAGTCTACTGTGGATAACTCTGTC GAGCTAGAAATAGCAAG	GTTTTA-	pgRNA 4 integration
JW098	TAACTCGAGTAAGGATCTCCAG		pdCas9 for
JW099	CCCTCGAGTTAAGCAGCCAGAGCGTAGT TTTCGTCGTTAGCAGCGTCACCTCCTAGCTGACTCA		pdCas9 LAA
JW121	TACCTAGGGATATATTCCGCTTCCCTAA AGATCTTTGACAGCTAGCTC		pgRNA for CPEC
JW122	GACCGAGCGTAGCGAGTCGGATCCAGTT CAAAC	CACCGA-	pgRNA rev for CPEC
JW124	GACTCGCTACGCTCGGTC		pdCas9 for CPEC
JW125	GGAAGCGGAATATATCCCTAG		pdCas9 rev CPEC
JW148	CCACTAGTTTCTATTGTGATCTCTTATT GAGCTAGAAATAGCAAG	GTTTTA-	pgRNA 5 integration
JW149	CCACTAGTTTGAGAAAGACCTGGGATCC GAGCTAGAAATAGCAAG	GTTTTA-	pgRNA 6 integration
JW150	CCACTAGTATGGATTGAAGCCCGGGCCG GAGCTAGAAATAGCAAG	GTTTTA-	pgRNA 7 integration
JW151	CCACTAGTATTGTACGCTGTGAACGCGT GAGCTAGAAATAGCAAG	GTTTTA-	pgRNA 8 integration
JW154	CCACTAGTCACGGAACTTCAGTCCCATT GAGCTAGAAATAGCAAG	GTTTTA-	pgRNA 9 integration
Jw158	CCACTAGTGTCAGGAAGCTTGGATCAAC GAGCTAGAAATAGCAAG	GTTTTA-	pgRNA 10 integration
Jw159	CCACTAGTGAGGCAGAACTCAAAAATTC GAGCTAGAAATAGCAAG	GTTTTA-	pgRNA 11 integration
Jw160	CCACTAGTGATCATTAACTGTGAATGAT GAGCTAGAAATAGCAAG	GTTTTA-	pgRNA 12 integration
Jw161	CCACTAGTAGCTTATACGGTCCAGGATC GAGCTAGAAATAGCAAG	GTTTTA-	pgRNA 13 integration
Jw162	CCACTAGTTTGCGTTTAGTATCCTAAAC GAGCTAGAAATAGCAAG	GTTTTA-	pgRNA 14 integration
Jw163	CCACTAGTATGAGCTGCAACCCGGCGAT GAGCTAGAAATAGCAAG	GTTTTA-	pgRNA 15 integration

REFERENCES

[1] A. C. Leonard and M. Mechali, *DNA Replication Origins,* Cold Spring Harbor Perspectives in Biology **5**, a010116 (2013).

[2] N. V. Sernova and M. S. Gelfand, *Identification of replication origins in prokaryotic genomes,* Briefings in Bioinformatics **9**, 376 (2008).

[3] A. Kornberg, *DNA replication,* Biochimica et Biophysica Acta (BBA)-Gene Structure and Expression **951**, 235 (1988).

[4] B.-M. Yung and A. Kornberg, *The dnaA initiator protein binds separate domains in the replication origin of Escherichia coli.* Journal of Biological Chemistry **264**, 6146 (1989).

[5] A. Robinson and A. M. van Oijen, *Bacterial replication, transcription and translation: mechanistic insights from single-molecule biochemical studies,* Nature Reviews Microbiology **11**, 303 (2013).

[6] D. Magnan, M. Joshi, A. Barker, B. Visser, and D. Bates, *DNA Replication Initiation Is Blocked by a Distant Chromosome–Membrane Attachment,* Current Biology **25**, 2143 (2015).

[7] T. Katayama, S. Ozaki, K. Keyamura, and K. Fujimitsu, *Regulation of the replication cycle: conserved and diverse regulatory systems for DnaA and oriC,* Nature Reviews Microbiology **8**, 163 (2010).

[8] R. Siam and G. T. Marczynski, *Cell cycle regulator phosphorylation stimulates two distinct modes of binding at a chromosome replication origin,* The EMBO journal **19**, 1138 (2000).

[9] R. Reyes-Lamothe, C. Possoz, O. Danilova, and D. J. Sherratt, *Independent Positioning and Action of Escherichia coli Replisomes in Live Cells,* Cell **133**, 90 (2008).

[10] J. M. Pennington and S. M. Rosenberg, *Spontaneous DNA breakage in single living Escherichia coli cells,* Nature Genetics **39**, 797 (2007).

[11] J. A. N. Brophy and C. A. Voigt, *Principles of genetic circuit design,* Nature Methods **11**, 508 (2014).

[12] H. L. Withers and R. Bernander, *Characterization of dnaC2 and dnaC28Mutants by Flow Cytometry,* Journal of bacteriology **180**, 1624 (1998).

[13] D. Bates, J. Epstein, E. Boye, K. Fahrner, H. Berg, and N. Kleckner, *The Escherichia coli baby cell column: A novel cell synchronization method provides new insight into the bacterial cell cycle,* Molecular Microbiology **57**, 380 (2005).

[14] D. J. Ferullo, D. L. Cooper, H. R. Moore, and S. T. Lovett, *Cell cycle synchronization of Escherichia coli using the stringent response, with fluorescence labeling assays for DNA content and replication,* Methods **48**, 8 (2009).

[15] B. Chen, L. A. Gilbert, B. A. Cimini, J. Schnitzbauer, W. Zhang, G.-W. Li, J. Park, E. H. Blackburn, J. S. Weissman, L. S. Qi, and B. Huang, *Dynamic imaging of genomic loci in living human cells by an optimized CRISPR/Cas system,* Cell **155**, 1479 (2013).

[16] M. R. O'Connell, B. L. Oakes, S. H. Sternberg, A. East-Seletsky, M. Kaplan, and J. A. Doudna, *Programmable RNA recognition and cleavage by CRISPR/Cas9,* Nature **516**, 263 (2014).

[17] M. Jinek, A. East, A. Cheng, S. Lin, E. Ma, and J. Doudna, *RNA-programmed genome editing in human cells,* Elife **2**, e00471 (2013).

[18] E. Deltcheva, K. Chylinski, C. M. Sharma, K. Gonzales, Y. Chao, Z. A. Pirzada, M. R. Eckert, J. Vogel, and E. Charpentier, *CRISPR RNA maturation by trans-encoded small RNA and host factor RNase III,* Nature **471**, 602 (2011).

[19] M. H. Larson, L. A. Gilbert, X. Wang, W. A. Lim, J. S. Weissman, and L. S. Qi, *CRISPR interference (CRISPRi) for sequence-specific control of gene expression,* Nature Protocols **8**, 2180 (2013).

[20] D. Bikard, W. Jiang, P. Samai, A. Hochschild, F. Zhang, and L. A. Marraffini, *Programmable repression and activation of bacterial gene expression using an engineered CRISPR-Cas system,* Nucleic Acids Research **41**, 7429 (2013).

[21] T. G. Bernhardt and P. A. De Boer, *The Escherichia coli amidase AmiC is a periplasmic septal ring component exported via the twin-arginine transport pathway,* Molecular microbiology **48**, 1171 (2003).

[22] I. F. Lau, S. R. Filipe, B. Søballe, O.-A. Økstad, F.-X. Barre, and D. J. Sherratt, *Spatial and temporal organization of replicating Escherichia coli chromosomes: Escherichia coli chromosome dynamics,* Molecular Microbiology **49**, 731 (2004).

[23] L. Qi, M. Larson, L. Gilbert, J. Doudna, J. Weissman, A. Arkin, and W. Lim, *Repurposing CRISPR as an RNA-Guided Platform for Sequence-Specific Control of Gene Expression,* Cell **152**, 1173 (2013).

[24] K. C. Keiler, P. R. H. Waller, and R. T. Sauer, *Role of a Peptide Tagging System in Degradation of Proteins Synthesized from Damaged Messenger RNA,* Science **16**, 990 (1996).

[25] J. Quan and J. Tian, *Circular polymerase extension cloning of complex gene libraries and pathways,* PloS one **4**, e6441 (2009).

[26] F. Wu, B. G. C. van Schie, J. E. Keymer, and C. Dekker, *Symmetry and scale orient Min protein patterns in shaped bacterial sculptures,* Nature Nanotechnology **10**, 719 (2015).

[27] K. A. Datsenko and B. L. Wanner, *One-step inactivation of chromosomal genes in Escherichia coli K-12 using PCR products,* Proceedings of the National Academy of Sciences **97**, 6640 (2000).

[28] J. Schindelin, I. Arganda-Carreras, E. Frise, V. Kaynig, M. Longair, T. Pietzsch, S. Preibisch, C. Rueden, S. Saalfeld, B. Schmid, J.-Y. Tinevez, D. J. White, V. Hartenstein, K. Eliceiri, P. Tomancak, and A. Cardona, *Fiji: an open-source platform for biological-image analysis,* Nature Methods **9**, 676 (2012).

[29] O. Sliusarenko, J. Heinritz, T. Emonet, and C. Jacobs-Wagner, *High-throughput, subpixel precision analysis of bacterial morphogenesis and intracellular spatio-temporal dynamics,* Molecular Microbiology **80**, 612 (2011).

[30] C. Lesterlin, C. Pages, N. Dubarry, S. Dasgupta, and F. Cornet, *Asymmetry of Chromosome Replichores Renders the DNA Translocase Activity of FtsK Essential for Cell Division and Cell Shape Maintenance in Escherichia coli,* PLoS Genetics **4**, e1000288 (2008).

[31] S. Tabata, A. Oka, M. Takanami, S. Yasuda, and Y. Hirota, *The 245 base-pair oriC sequence of the E. coli chromosome directs bidirectional replication at an adjacent region.* Nucleic Acids Res **11**, 2617 (1983).

[32] C. Weigel, A. Schmidt, B. Rückert, R. Lurz, and W. Messer, *DnaA protein binding to individual DnaA boxes in the Escherichia coli replication origin, oriC,* The EMBO Journal **16**, 6574 (1997).

[33] J. M. Dewar, M. Budzowska, and J. C. Walter, *The mechanism of DNA replication termination in vertebrates,* Nature **525**, 345 (2015).

[34] X. Wang and D. J. Sherratt, *Independent Segregation of the Two Arms of the Escherichia coli ori Region Requires neither RNA Synthesis nor MreB Dynamics,* Journal of Bacteriology **192**, 6143 (2010).

[35] A. Prindle, J. Selimkhanov, H. Li, I. Razinkov, L. S. Tsimring, and J. Hasty, *Rapid and tunable post-translational coupling of genetic circuits,* Nature **508**, 387 (2014).

[36] L. J. Wu and J. Errington, *Nucleoid occlusion and bacterial cell division,* Nature Reviews Microbiology **10**, 8 (2012).

[37] H. Nishimasu, F. Ran, P. Hsu, S. Konermann, S. Shehata, N. Dohmae, R. Ishitani, F. Zhang, and O. Nureki, *Crystal Structure of Cas9 in Complex with Guide RNA and Target DNA,* Cell **156**, 935 (2014).

[38] M. B. Elowitz and S. Leibler, *A synthetic oscillatory network of transcriptional regulators,* J. Biol. Chem **274**, 6074 (1999).

[39] D. E. Cameron, C. J. Bashor, and J. J. Collins, *A brief history of synthetic biology,* Nature Reviews Microbiology **12**, 381 (2014).

SUMMARY

The life of a cell is a never ending effort to maintain, copy, and, eventually, spread the genetic information contained on the strands of the DNA double helix. The genome of the most common prokaryotic laboratory organism, *Escherichia coli*, is contained on a circular chromosome, of approximately 4.6 million basepairs. Before each cell division the genome has to be accurately copied and redistributed between daughter cells. Errors made by the replication machinery may lead to incorporation of wrong nucleotides, sometimes having drastic consequences for the entire organism. Interrupted replication can be a source of highly toxic double stranded DNA breaks (DSB), where both strands of the DNA duplex are broken at the same time. In *E. coli* those breaks are repaired by homologous recombination (HR) pathway. More detailed description of the DNA replication, recombination, and repair processes, together with an overview of the literature can be found in **Chapter 1**.

The first part of this thesis describes experimental work focused on capturing initial steps of DSB repair in live cells. We develop and use a range of diverse techniques to track the separate steps of the HR: the end-resection, and the search for homologous template. End-resection is an early and an essential reaction within homologous recombination where the 5'-terminated strand of the DNA duplex is digested by the RecBCD complex to prepare 3' single-stranded DNA overhangs, to which recombination protein RecA binds. This well-known reaction was extensively studied in diverse in vitro assays, but was not yet directly visualized inside of the live cell.

In **Chapter 2**, using fluorescent microscopy, we observe expansion of end-resection at individual DSB sites. Surprisingly, the end-resection in the cell progresses with the speed of 1000 bps^{-1} similar to what was found in the test tube where RecBCD was processing bare DNA. Then, we used a qPCR assay to verify our observations, and we found that during the end-resection substantial amounts of DNA are being degraded. RecBCD activity in vitro depends on the ratio between concentrations of ATP to Mg^{2+}, and only when Mg^{2+} is in excess over ATP RecBCD acts as an exonuclease, otherwise RecBCD acts only as a helicase without exonuclease activity. In our experiments we observed that the degradation is linked to the end-resection, therefore we propose that in cells the RecBCD reaction is similar to the one in presence of high concentration of magnesium. Deletion of *recD* gene greatly reduced the length of resection and allowed for recombination with short, ectopic homologies, as well as increased the chances of horizontal gene transfer with short, linear sequences. Together, we captured the end-resection in live cells and probed its role in the homologous recombination.

Search for repair template during DSB repair is an unparalleled effort. A broken location has to pair with an intact homology, or else the cell will be fated to die. In **Chapter 3** we confront a popular 'needle in the heystack' model of homology search, that dictates that the search process is random and involves global scanning for suitable sequence. We developed experimental systems to study the search and recombination between ec-

topic homologies within the same *E. coli* chromosome, but each attempt to capture the search failed. We concluded the lack of recombination in our experiments is suggesting that the search process is cell-wide, but biased towards replicated sister template, where the homology is most abundant. When we traced the mobility of chromosomal markers in the proximity of the DSB we found that broken sites rapidly localize in the middle of the cell. Such rapid movement to a single location can be caused by a mechanism that facilitates the search by forming 'repair factory' at a single location of the cell. The most common source of DSBs are stops and interruptions of the replication machinery. We propose that the search process for such is simplified due to close proximity of the repair template at the replication fork.

Genetic engineering is a powerful and essential tool in molecular biology. In **Chapter 4** we experiment with homologous recombination in strains that lack the *recD* gene. When the length of end-resection is reduced, short fragments of DNA can be integrated into a designed location on the chromosome. Recombination mediated by RecA involves only the ends of the DNA, and we propose that integration of long constructs is more efficient in *recD* cells than for widely used lambda-red recombineering. Usually, the transfer of genes between *E. coli* strains is done using P1-phage transduction. We show that shuffling markers between strains can be simple in *recD* cells, using isolated genomic DNA from a donor strain, that is electroporated into *recD* deficient cells. Finally, we use the CRISPRi system to transiently inactivate the expression of *recD*, to create a simple laboratory technique for genome modification.

In the last part of this thesis we move away from recombination and focus on replication. The CRISPR-Cas9 system is a part of bacterial defense system against viral infections. Because of its remarkable ability to recognize programmed DNA sequences, CRISPR-Cas9 became a revolutionary laboratory tool and allowed for an easy modification of genomes. In **Chapter 5**, we describe a CRISPR-dCas9 system designed to control bacterial cell cycle. The replication of the circular chromosome is initiated at the specific location called *oriC*, by the action of DnaA protein filament. When nuclease-deficient dCas9 protein is targeted to bind to the *oriC* before DnaA had a chance to initiate the replication, bacteria are arrested in pre-replication state of cell cycle. Replication-arrested cells halt the division, maintain active metabolism, and grow longer that the usual cells. Surprisingly, the effect of CRISPR-dCas9 can be reversed by a shift to 42°C, where cells reinitiate the replication. We supply this system in a form of ready-to-use, single plasmid.

In this thesis we looked into old biological problems in new ways and from different perspectives. We believe that our direct observations of the steps of the homologous recombination can in the future be extended beyond end-resection, and will help to understand the fundamental principle of the DNA repair. The impact of genomic engineering is not only limited to fundamental research, but also will find its way to clinical applications in the near feature. Understanding the role of each step involved in recombination is necessary to efficiently use knowledge of biochemistry and molecular biology to develop treatments. Only by directly looking at cellular reactions we can fully see the whole and detailed picture of biological processes.

SAMENVATTING

Het leven van een cel is een eindeloze ijver naar het onderhouden, kopiëren en, uiteindelijk, verspreiden van de genetische informatie, die is geherbergd in de strengen van het DNA. Het genoom van het meest onderzochte prokaryotisch organisme in het laboratorium, *Escherichia coli*, is opgeborgen in een circulair genoom van ongeveer 4.6 miljoen baseparen. Voordat een cel zich deelt, moet het genoom accuraat worden gekopieerd en verdeeld over de dochtercellen. Vergissingen gemaakt door de replicatie machinerie kunnen leiden tot incorporatie van verkeerde nucleotiden, met soms drastische consequenties voor het organisme. Onderbroken replicatie kan een bron zijn van zeer toxische, dubbel strengs DNA breuken (DSB), waar beide strengen van de DNA-duplex tegelijk worden doorbroken. In *E. coli* worden deze breuken gerepareerd door middel van homologe recombinatie (HR). Een gedetailleerde beschrijving van de DNA-replicatie, recombinatie en reparatie processen, tezamen met een overzicht van de literatuur is te vinden in **hoofdstuk 1**.

Het eerste deel van dit proefschrift beschrijft experimenteel werk naar het blootleggen van de initiële stappen van DSB-reparatie in levende cellen. We ontwikkelen en gebruiken een divers spectrum aan technieken om de verschillende stappen van de HR te kunnen volgen: de eind-resectie en de zoektocht voor de homologe blauwdruk. Eind-resectie is een vroege en essentiële reactie binnen de homologe recombinatie. Hierbij wordt de 5'-getermineerde streng van de DNA-duplex verteerd door het RecBCD complex om zo een 3' enkelstrengs DNA overhang te maken, waaraan het recombinatie eiwit RecA zich kan binden. Deze reactie is uitgebreid onderzocht in verscheidene in-vitro assays, maar nooit eerder direct afgebeeld binnen in een levende cel.

In **hoofdstuk 2** onderzoeken we de lengte van de eind-resectie op individuele DSB plekken door middel van fluorescentie microscopie. We vinden dat de eind-resectie in de cel zich voort schrijdt met een verrassend hoge snelheid van 1000bps^{-1}, vergelijkbaar met eerdere waarnemingen van de snelheid van dit proces in de reageerbuis, waar RecBCD kaal DNA verwerkt. We hebben een qPCR assay gebruikt om onze waarnemingen te bevestigen en zagen dat tijdens de eind-resectie substantiële hoeveelheden DNA worden verteerd. RecBCD activiteit in-vitro is afhankelijk van de verhouding tussen ATP en Mg^{2+}, en slechts indien Mg^{2+} in overmaat aanwezig is werkt RecBCD als een exonuclease, anders werkt RecBCD als een helicase zonder exonuclease activiteit. In onze experimenten observeren we dat de degradatie is verbonden met de eind-resectie, waarop we veronderstellen dat de RecBCD reactie in cellen vergelijkbaar is met de reactie in aanwezigheid van een hoge concentratie magnesium. Verwijdering van het *recD* gen reduceerde de restrictie lengte sterk, maakte de recombinatie met korte, ectopische homologieën mogelijk en verhoogt de kansen op horizontale gen overdacht met korte, lineaire sequenties. Al met al hebben we eind-resectie in levende cellen vastgelegd en de rol ervan in homologe recombinatie onderzocht.

De zoektocht voor de repareer-blauwdruk tijdens DSB reparatie is een ongeëve-

naarde inspanning. Een gebroken plek moet worden samengebracht met een intacte homologie, anders is de cel gedoemd te sterven. In **hoofdstuk 3** dagen we het populaire 'speld in de hooiberg' model uit, waarin de homologe zoektocht wordt voorgesteld als willekeurig en plaatsvindt aan de hand van een globale scan voor een bruikbare sequentie. We hebben experimentele systemen ontwikkeld om de zoektocht en recombinatie tussen ectopische homologieën binnen hetzelfde E.coli chromosoom te onderzoeken, maar iedere poging om de zoektocht bloot te leggen strandde. We concluderen dat de afwezigheid van recombinatie in onze experimenten suggereert dat het zoekproces niet over de gehele cel plaatsvindt, maar een voorkeur heeft voor de gerepliceerde zusterblauwdruk, waar de homologie zich het meest bevindt. Wanneer we de mobiliteit van chromosomale markers nabij de DSB volgen, vonden we dat gebroken plekken zeer snel lokaliseren in het middel van de cel. Een dergelijk snelle beweging naar één plek kan worden veroorzaakt door een mechanisme dat de zoektocht faciliteert door een 'reparatie fabriek' te vormen op deze locatie in de cel. De meest voorkomende bron van DSBs zijn stops en onderbrekingen van de replicatie machinerie. We veronderstellen dat de zoektocht voor dergelijke breuken wordt vereenvoudigd door de nabijheid van de repareer-blauwdruk bij de replicatie vork.

Genetische modificatie is een krachtig en essentieel gereedschap voor de moleculaire bioloog. In **hoofdstuk 4** experimenteren we met homologe recombinatie in cellijnen waarbij het *recD* ontbreekt. Als de lengte van de eind-resectie is verkleind, kunnen korte fragmenten van DNA worden geïntegreerd in een specifiek ontworpen locatie op het chromosoom. Alleen de uiteinden van het DNA worden gebruikt bij recombinatie die wordt uitgevoerd met behulp van RecA en daarom veronderstellen we dat de integratie van lange constructen efficiënter is in *recD* cellen dan voor de meer gebruikte, lambda-red recombinatietechniek. Het overplaatsen van genen tussen verschillende E. coli cellijnen wordt gebruikelijk gedaan door middel van P1-bacteriophaag transductie. Door gebruik te maken van genomisch DNA van een donor lijn, ingebracht via electroporatie, laten wij zien dat het schuiven van markers tussen cellijnen eenvoudig is in *recD* cellen. Tenslotte gebruiken we het CRISPRi systeem om tijdelijk de expressie van *recD* te inactiveren, waarmee we een simpele laboratoriumtechniek hebben ontwikkeld voor genoommodificatie.

In het laatste deel van dit proefschrift verplaatsen we de focus van recombinatie naar replicatie. Het CRISPR-Cas9 systeem is onderdeel van een bacterieel afweersysteem tegen virale infecties. Door de opmerkelijke eigenschap van dit systeem om voorgeprogrammeerde DNA-sequenties te herkennen, heeft CRISPS-Cas9 zich ontwikkeld tot een revolutionair laboratorium gereedschap voor het eenvoudig modificeren van genomen. In dit laatste **hoofdstuk 5** beschrijven we de ontwikkeling van een CRISPR-dCas9 systeem voor het controleren van de bacteriële celcyclus. De replicatie van het circulaire genoom begint op een specifieke locatie genaamd *oriC*, aan de hand van het DnaA eiwit filament. Wanneer nuclease-deficiënt dCas9 zich doelgericht bindt aan de *oriC* voordat DnaA kans ziet de replicatie te starten, worden bacteriën gestopt in de replicatie fase van de celcyclus. Cellen gestopt in de replicatie, stoppen de celdeling in zijn geheel, maar behouden actief metabolisme en groeien langer dan normale cellen. Verrassend genoeg kan het effect van CRISPR-dCas9 worden omgekeerd door de temperatuur te verhogen tot 42°C, waarbij de cellen hun replicatie herstarten. We bieden dit systeem aan als een

klant-en-klaar, enkel plasmide.

In dit proefschrift hebben we oude biologische problemen bestudeerd op nieuwe manieren en van verschillende perspectieven. We zijn ervan overtuigd dat onze directe observatie van de stappen van de homologe recombinatie in de toekomst de eindrestrictie zullen overstijgen en zullen bijdragen aan het begrip van het basisprincipe van DNA reparatie. Het belang van genoom modificatie zal zich niet slechts beperken tot fundamenteel onderzoek, maar zal in de nabije toekomst ook zijn invloed uitoefenen in medische toepassingen. Begrip van iedere stap van recombinatie is nodig om de kennis vergaard in de biochemie en moleculaire biologie adequaat toe te passen in de ontwikkeling van medicijnen. Slechts door direct de cellulaire reacties bloot te leggen, kunnen we een geheel en gedetailleerd beeld van biologische processen vergaren.

ACKNOWLEDGEMENTS

This thesis would never come to an end without help, support, and simply, presence of many people. First, and foremost I would like to thank my promotor, Cees Dekker, for giving the opportunity to join the lab, and later for giving me the freedom to find my own way during my scientific adventures. I was impressed when I first saw you talk in the IMP in Vienna, I remember thinking that lasers and programming are the future of science. I lucky enough to receive an invitation for an interview in Delft soon after. Now, at the end of my PhD I am grateful for your guidance and for the space you have me to develop my scientific independence to choose my own directions in the lab. You always impressed me with your energy, enthusiasm, and very accurate scientific remarks. Thank you for being my supervisor.

During my PhD I was lucky to spend a rather extended stay in Oxford, in the laboratory of professor David Sherratt. A short visit, initially planned for two days, extended into almost-a-year stay. It was a great time, for work, as well as, after work. It wouldn't be possible without kind and generous hospitality of David. Your knowledge and drive to design killer experiments have always been inspiring and I wish I could have the same scientific instinct as you have. Thank you for letting me stay in the lab and guiding me in times when I was very lost in the experiments and results.

I've got also tremendous amount of help in Oxford from Christian Lesterlin. Christian, thank you for all the help you have given me, even though you didn't have to. You helped me to make my first steps towards the field of microbiology and microscopy. I don't think I could finish any experiment without your help, or at least finish them well. I will always remember the time we spent outside the lab, having a beer, or playing the guitar in the garden. I'm sure your group in Lyon will be very successful and most importantly, very happy. And finally, Christian, thank you for joining my PhD defense as a member of my committee. Good luck with Nelly, Olivia and Virgile.

Special thanks goes to the members of my committee, who agreed to dedicate their precious time and effort to read, and correct my thesis, and come to Delft on the day of my defense: Sander Tans, Jan Hoeijmakers, Stan Brouns, Olivier Espeli, and Nynke Dekker.

In Delft, I had pleasure to work with great students. Thank you Félicia, for being brave enough to be my first student and showing enthusiasm even when the project turned out to be rather difficult. Good luck for the future! Marit, you cannot even imagine how happy I was to work with you. It was only a pleasure to supervise such a motivated, self driven, and hard working student as you. You weren't afraid to try new experiments, even the ones that were most insane and difficult. Thank you for staying after finishing your project. Without your help, the details of end-resection in live cells would still be a mystery. Lisa, thank you for your help on recombination project. Plasmids you cloned made were amongst the biggest and best ever made. I am sure that the student association is blooming under your presidency. Amelie, thank you for coming all the way from

Germany to work on recombination project. I was impressed that even when the project wasn't proceeding in a right way you didn't loose the focus. Good luck with your own scientific adventures.

Days wouldn't be as good as they were, without all the friends from the CD-lab. I had a great honor to share the office with the finest characters. Adi, we started almost at the same time, and I will always think of you when I will be thinking of my PhD. I enjoyed our discussions, listening to your crazy stories, taking lengthy and frequent coffee breaks, and having drinks after work. If you say that cricket is the best sport there is, it must be true. Good luck with your Himalayan adventures. Rutger, thank you for being the voice of a reason, and all your support and wisdom, good luck in Utrecht. When I thought that our office couldn't be any better, we got Daniel! Daniel, I don't know how can you do it, but I have never seen more happy and cheerful young man. Even when it's raining, you act as it was sunny, because the sun always shines around you. Good luck with everything. We should have beers on the sunny side of the Tango more often. Anthony, I am very happy we could be office mates. You showed me the joy of the innocent office pranks, healthy snacks, and the finest whiskey. Wayne, it was most unfortunate that you couldn't stay with us longer, and had to change the office. I enjoyed our conversations about life, science, and psychology in fast food restaurants. I hope you enjoyed working out with me, as much as I enjoyed working out with you.

All the other members that unfortunately could not fit in the same office with me also made my time in Delft more pleasant. Federico, we shared some stories, usually full of emotions, and made some great music. The actual Beatles would be surely proud of us. Thanks for being yourself. Jorine, thanks for all the support, discussion about science, and life, tea breaks, and cheesecakes. Thank you also for critically looking at my figures and correcting the manuscript, especially all the missing 'the'. My greetings also go to Stephanie (please remember about me when your biotech company will be hiring) , Michel, Parul, Anna, Sabina, Amanda (thank you for helping me move into the Netherlands and all the support), Yaron (please, keep being yourself), Siddharth (who would think that making life is so hard), Mahipal, Aleksandre, Greg, Je-Kyung, Alessio (thanks for scoring all those ridiculous goals), Yoones, Laura, Daniel Shi, Allard, Jacob (thank you for our 5-minute philosophical discussions and help with programming), Jaco (your support and knowledge in the lab were priceless), Dimitri, Jelle, Sung Hyun, Fabai, Rifka, Tim and Meghan, Daniel B., Magnus, Gautam, Marijn, Calin, Fabai, Felix, Erwin, and Hugo.

The Bionanoscience department is a collection of extremely interesting minds and characters and I was lucky enough to share my coffee and lunch breaks many of them: Pauline, Nicole, Kim, Alicia (thank you for the delicious brownie!), Helena (thank you for showing me that Mexican food is so tasty), Viktorija, Mathia, Rebecca, Ilja, Natalia, Essengül, David, Andrew, Maarten, Sriram, Sumit (good luck with your replication project!), Hyun (thank you for reading and correcting the RecBCD manuscript), Jonas, Dominik, Luis, Sam, Mohamed, Stanley, Luuk, Sam, Theo (your scientific suggestions were always accurate), Marek, Mehran, Patrick, Carsten, Misha. I could always rely on excelled support and help from Amanda, Susanne, Emmylou, Dijana, and Jolijn.

Special thanks goes also to people I've met during my stay in Oxford. First, I'd like to thank Kasia for being the best flatmate and friend. Never I had a better and nicer break-

fasts than on the legendary Coolidge Close house. Weekend pizza lunches with Jose, Florence, Rogelio, Robbie, Jackie, Piotr and Alex were also amazing. At the biochemistry department I enjoyed the company of many great researchers, thank you for being there Amber, Lidia, Rachel, Amy, Kasia and Paweł, Jacob, Stephan, Gianluca, Christopher, Eric, Sophie and Alan, and Zack.

I would never arrive to Delft if I didn't meet Cees in Vienna and the only reason I was there was the help from Tim Clausen. Tim offered me to join a project in his group at the IMP, where every day was an adventure. My gratitude also goes to Sonja, whose guidance was unparalleled, Doris, Marcin, Ricardo and Abhiyan.

I was very lucky when during the last year of my master degree course I was offered a project in the laboratory of professor Jarosław Marszałek in Gdańsk, where I was working under excellent supervision of Rafał Dutkiewicz. The first project is one of the most important scientific experiences and I'm very glad that mine was under the guidance of Rafał.

Finally, I would never be where I am now without the help, inspiration, and love from my family. My parents were always believing in my, and supporting me, even in the moments when I didn't really believe in myself. Dziękuję za to że zawsze mieliście dla mnie czas. I could always count on the help from my sister, especially when programming and on-line shopping were concerned. I wish I could be as smart as you. My brother was always a role model, I've always tried to imitate you, and here I am, defending my PhD half a year after you. Special thanks goes also to Michał for reading the CRISPR/dCas9 manuscript.

Jakub Wiktor, May 2017

CURRICULUM VITÆ

Jakub Marek WIKTOR

26-02-1987 Born in Gdańsk, Poland.

EDUCATION

2006–2009 B.Sc Biology
University of Gdańsk, Gdańsk, Poland

2009–2011 M.Sc of Biotechnology
University of Gdańsk, Gdańsk, Poland

2012–2017 Ph.D. Bionanoscience
Technische Universiteit Delft, Delft, The Netherlands
Thesis: Live cell studies of bacterial DNA replication, recombination, and degradation.
Promoter: Prof. dr. C. Dekker

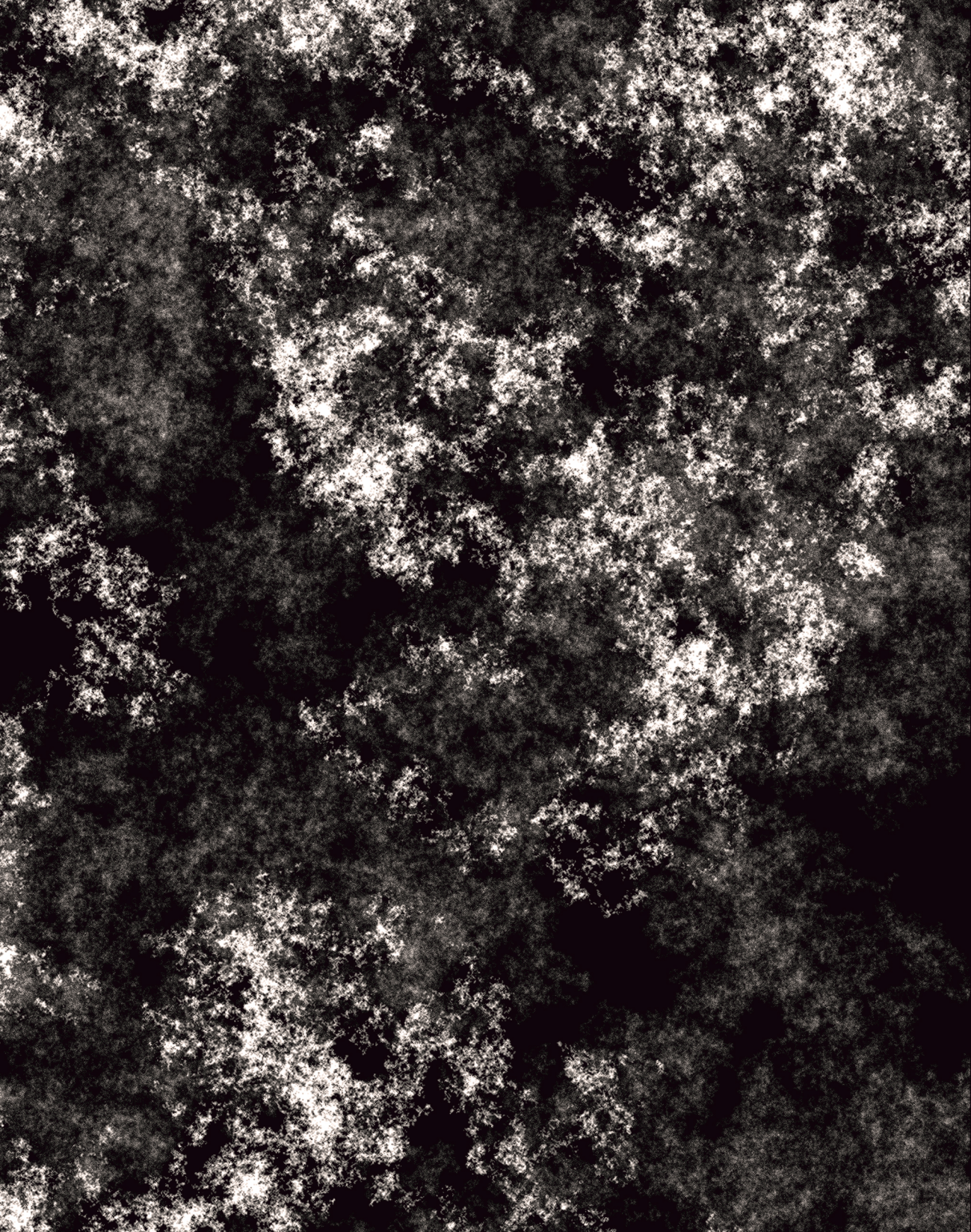
List of Publications

4. **J. Wiktor**, J. van der Torre, C. Dekker, *Integration of long constructs in ΔrecD Escherichia coli*, *manuscript in preparation.*

3. **J. Wiktor**, M. van der Does,L. Buller, D. J. Sherratt, C. Dekker, *Extensive and asymmetric end-resection by RecBCD in vivo during double stranded break repair, manuscript in preparation.*

2. **J. Wiktor**, C. Lesterlin, D. J. Sherratt, C. Dekker, *CRISPR-mediated control of the bacterial initiation of replication*, Nucleic Acids Res **44**, 3801 (2016).

1. S. Nolivos, A. L. Upton, A. Badrinarayanan, J. Müller, K. Zawadzka, **J. Wiktor**, A. Gill, L. Arciszewska, E. Nicolas, D. Sherratt, *MatP regulates the coordinated action of topoisomerase IV and MukBEF in chromosome segregation*, Nature Communications **7**, 10466 (2016).



Casimir PhD series 2017-22
ISBN: 978-90-8593-306-9