Skolkovo Institute of Science and Technology

THE ROLE OF GENOME MAINTENANCE PROTEINS IN PRIMED CRISPR ADAPTATION BY THE TYPE I-E CRISPR-CAS SYSTEM

Doctoral Thesis

by

ELENA KURILOVICH

DOCTORAL PROGRAM IN LIFE SCIENCES

Supervisor
Professor Konstantin Severinov

Moscow - 2023

I hereby declare that the work presented in this thesis was carried out by myself at Skolkovo Institute of Science and Technology, Moscow, except where due acknowledgement is made, and has not been submitted for any other degree.

Candidate (Elena Kurilovich)

Supervisor (Prof. Konstantin Severinov)

Abstract

Bacteria and archaea use CRISPR-Cas adaptive immunity systems to interfere with viruses, plasmids, and other mobile genetic elements. In *Escherichia coli*, immunity is acquired upon integration of 33-bp spacers into CRISPR arrays. DNA targets complementary to spacers get degraded and serve as a source of new spacers during a process called primed adaptation. Precursors of such spacers, prespacers, are ~33-bp double-stranded DNA fragments with a ~4-nt 3' overhang. The detailed mechanism of conversion of targeted DNA to spacers in CRISPR array during primed CRISPR adaptation remains to be identified. It is reasonable to assume that cellular machinery engaged in genome maintenance and DNA metabolism also contributes to primed spacer acquisition. In this work, we investigated the role of proteins involved in genome stability maintenance in spacer acquisition by the *Escherichia coli* type I-E CRISPR-Cas system. Using FragSeq and the biochemical approach, we show that RecJ is the main exonuclease trimming 5' ends of prespacer precursors, though its activity can be partially substituted by ExoVII. The RecBCD complex participates in 5' end trimming by allowing single-strand specific RecJ to process double-stranded regions flanking prespacers. We also demonstrate that *recJ* deletion greatly decreases CRISPR adaptation efficiency by affecting prespacer generation and influences the accuracy of spacer incorporation.

Primed adaptation efficiency is also dramatically inhibited in double mutants lacking *recB* and *sbcD* but not in single mutants, suggesting independent involvement and redundancy of the RecBCD and SbcCD pathways in spacer acquisition. Moreover, the absence of RecBCD and SbcCD affects the pattern of acquired spacers. The results reveal intricate functional interactions of genome maintenance proteins with CRISPR-Cas machinery.

A project focusing on primed CRISPR adaptation in *Pseudomonas aeruginosa* was undertaken in parallel with the main work. The results show that bacteriostatic antibiotics promote the evolution of CRISPR-Cas immunity by delaying the production of mature phage particles. These findings can be generalized to other conditions that slow down the

speed of phage development. The work shows that, in addition to defective phages and nucleases that cleave phage genomes, the time of phage development inside the infected cell is a key determinant of the acquisition of CRISPR-Cas immunity against the phage. The results further support the evidence that the longer the DNA substrates are present in the cell, the higher is the chance for a CRISPR-Cas system to acquire new spacers from them.

Publications

1. **Kurilovich, E.**, Shiriaeva, A., Metlitskaya, A., Morozova, N., Ivančić-Baće, I., Severinov, K., & Savitskaya, E. (2019). Genome maintenance proteins modulate autoimmunity mediated primed adaptation by the *Escherichia coli* Type I-E CRISPR-Cas system. *Genes*, *10(11)*, 872.

2. Shiriaeva, A., Kuznedelov, K., Fedorov, I., Musharova, O., Khvostikov, T., Tsoy, Y., **Kurilovich, E.**, Smith, G., Semenova, E., Severinov, K. (2022). Host nucleases generate prespacers for primed adaptation in the *E. coli* type I-E CRISPR-Cas system. *Sci. Adv.*, *8(47)*, 8650.

3. Dimitriu, T., **Kurilovich, E.**, Lapinska, U., Severinov, K., Pagliara, S., Szczelkun, M., Westra, E. (2021) Bacteriostatic antibiotics promote CRISPR-Cas adaptive immunity by enabling increased spacer acquisition. *Cell Host Microbe*, *30(1)*, 31-40.e5.

Conferences

1. **Kurilovich, E.**, Shiriaeva, A., Metlitskaya, A., Morozova, N., Ivančić-Baće, I., Severinov, K., Savitskaya, E. The impact of cellular genome maintenance proteins on primed CRISPR adaptation by the *Escherichia coli* Type I-E CRISPR-Cas system. 3rd International Conference on CRISPR Technologies, September 16-18, 2019, Würzburg, Germany: poster presentation

2. **Kurilovich, E.**, Shiriaeva, A., Kuznedelov, K., Fedorov, I., Musharova, O., Khvostikov, T., Tsoi, Y., Semenova, E., Severinov, K. The role of the host nucleases in prespaser generation during primed CRISPR adaptation in the Type I-E CRISPR-Cas system. FEMS Conference on Microbiology, June 30 - July 2, 2022, Belgrade, Serbia: poster presentation

Acknowledgements

I would like to express deep gratitude to my supervisor Prof. Konstantin Severinov for his great support, guidance, and invaluable contribution to my development as both a researcher and person. I am also immeasurably grateful to Ekaterina Savitskaya, who sadly is no longer with us. Ekaterina introduced me to the field in 2014, provided guidance and kind support along the whole way, and initiated the project that resulted in the current Thesis.

I would like to thank Skoltech for all the great research and education opportunities it provided. I am grateful to Prof. Westra and Tatiana Dimitriu, for kindly hosting me in Exeter University and teaching a variety of new techniques.

I would also like to thank colleagues for their high impact on the current work, namely: Anna Shiriaeva, Anastasia Metlitskaya, Natalia Morozova, Kirill Datsenko, Olga Musharova, Daniel Vyhovskyi, Maria Logacheva, Konstantin Kuznedelov.

I am grateful to my Individual Doctor Committee and Jury members, Maria Sokolova, Prof. Gelfand, Prof. Mojica, Prof. Fineran, Prof. Bolt, Prof. Kotelevtsev, and Prof. Khrameeva, for their valuable comments and suggestions, which helped to greatly improve the quality of the work.

I would also like to thank my family and my husband Aleksandr Kurilovich for their great support and help.

Table of Contents

List of Symbols, Abbreviations

A – adenine
Abi – abortive infection
Acr – anti-CRISPR
A-EJ – alternative end-joining
AMR – antimicrobial resistance
ATP – adenosine triphosphate
bp – base pair
bp/s – base pairs per second
BREX – bacteriophage exclusion
C – cytosine
CARF - CRISPR-associated Rossmann fold
Cas - CRISPR-associated sequence
Cascade – CRISPR-associated complex for antiviral defense
cGAMP –cyclic guanosine monophosphate–adenosine monophosphate
CBASS – cyclic oligonucleotide-based anti-phage signaling system
CFU – colony forming unit
CRISPR – Clustered Regularly Interspaced Short Palindromic Repeats
CRISPRi – CRISPR intefrerence
crRNA – CRISPR RNA
DNA – deoxyribonucleic acid
dNTPs – deoxynucleotides
d.p.i. – day post-infection
DSB – double-strand break
dsDNA – double-stranded DNA
eAgo – eukaryotic Argonaute
G – guanine
GTP – guanosine-5'-triphosphate
HD – histidine–aspartate domain
HEPN – higher eukaryote and prokaryote nucleotide-binding
HIP – half-integrated prespacers
H-NS – histone-like nucleoid-structuring protein
IHF – integration host factor
IPTG – isopropyl β-D-1-thiogalactopyranoside
kb – kilobase
Kd – dissociation constant
kDa – kilodalton
LB – Luria-Bertani broth
Mb – megabase
MBC – minimum bactericidal concentration
MGEs – mobile genetic elements

MIC – minimum inhibitory concentration
MMEJ – microhomology-mediated end joining
MT – methyltransferase
NHEJ – non-homologous end joining
nM – nanomolar
nt – nucleotide
NT-strand – nontarget strand
OD – optical density
PAGE – polyacrylamide gel electrophoresis
pAgo – prokaryotic Argonaute
PAM – protospacer adjacent motif
PBS – Phosphate-buffered saline
PCR – polymerase chain reaction
PPS – priming protospacer
pre-crRNA – precursor CRISPR RNA
RE – restriction endonuclease
RM – restriction-modification
RNA – ribonucleic acid
RPM – reads per million mapped reads
rRNA – ribosomal RNA
SAM – S-adenosyl methionine
scaRNA – small CRISPR-Cas-associated RNA
SEM – standard error of mean
SF1 – Superfamily 1
SF2 – Superfamily 2
SM – surface modification
SSB – ssDNA-binding protein
ssDNA – single-stranded DNA
T – thymine
TA – toxin-antitoxin
TE buffer – Tris-EDTA buffer
TCEP – tris(2-carboxyethyl)phosphine
TIR – terminal inverted repeat
tracrRNA – *trans*-activating RNA
T-strand – target strand
TSD – target site duplication
UV – ultraviolet
μm – micrometer
μM – micromolar

List of Figures

Chapter 1. Literature review.

1.1. Bacterial defense systems

Billions of years of evolution and constant arm races between prokaryotes and invading mobile genetic elements led to the appearance of various defense strategies on both sides. Major groups the bacterial and archaeal defense systems may be divided into are: 1) innate immunity (restriction-modification, Argonaute, BREX), 2) adaptive immunity (CRISPR-Cas systems) and 3) dormancy or programmed cell death (toxin-antitoxin, abortive infection).

1.1.1.Innate immunity

1.1.1.1. Restriction-Modification systems

The first bacterial immunity system discovered was restriction-modification (RM), which uses DNA modification and restriction enzymes to mark the host genome and degrade unmodified MGE (mobile genetic elements), correspondingly. RM systems mainly protect bacteria and archaea from phages and other invading DNA elements, which are not modified. With a few exceptions, the genes encoding R and M parts are typically located in one operon (Ershova et al., 2012; G. G. Wilson, 1991). There are several types of RM systems known, which differ by the content of enzymes and requirement of ATP/GTP.

Type II RM systems are the simplest and typically comprise separate restriction endonuclease (RE) and methyltransferase (MT) polypeptides, though some subtypes encode restriction and modification domains within the same protein (Ershova et al., 2012). MT modifies bacterial DNA, adding a methyl group from S-adenosyl methionine (SAM) to cytosines (at N^4 or C^5), and adenines (at N^6) in the recognition sequence, which is usually palindromic (Sistla & Rao, 2004). RE recognizes unmodified recognition sequence in invading genetic elements and introduces cuts inside or close to this sequence, thus protecting the host. Due to their simplicity, the ability to specifically recognize particular sequences and introduce DNA breaks in predictable sites, and their great variety, Type II restriction enzymes are widely used in molecular biology (Pingoud et al., 2014).

Type I RM systems are more complex and contain three genes coding for restriction (R), modification (M), and specificity (S) subunits of the RSM complex. The R subunit also contains a helicase superfamily II ATPase domain. Upon recognition of a specific sequence, the RSM complex translocates along DNA using the energy of ATP hydrolysis to ultimately introduce DNA breaks in non-specific sites (Dryden, 2004; Janscak et al., 1999; Powell et al., 1993; Seidel et al., 2004). DNA cleavage is proposed to be activated by any barrier blocking the translocation (Janscak et al., 1999). The RSM complex acts as a RE on unmethylated DNA, or as a MT on semi-methylated DNA, methylating the N^6 position of an adenine in the recognition site using SAM a methyl group donor.

Type III RM systems have some features of both Types I and II: it has two subunits, R and M, but R also has an ATPase domain. The recognition sites are methylated only on one strand, and two unmodified sites in inverse orientations are required for restriction (Meisel et al., 1992). Two RM complexes translocate DNA after recognition of specific sites, remaining bound to them (Dryden et al., 2001; Reich et al., 2004). Stalled translocation activates restriction, and each complex cuts one strand.

Type IV RM systems likely appeared as a result of a long-term coevolution of RM systems and MGEs and specifically target genetic elements protected from the action of Type I-III RM systems by DNA modifications (Stewart et al., 2000).

Defense systems resembling classical RM systems but utilizing phosphorothioation as a mark for self/nonself discrimination are called DND (Ho et al., 2015; D. You et al., 2007). The *dndABCDE* and *dndFGH* genes are required for restriction and modification, respectively, and their functions are now only partially established.

Bacteriophages have evolved several mechanisms to protect from RM defense, with avoidance of the recognition sites being the simplest one (Tock & Dryden, 2005). Some phages encode RM-inhibiting proteins, such as Ocr of phage T7 (Bandyopadhyay et al., 1985), or have their own modification enzymes to avoid the recognition by restriction enzymes (Warren, 1980). Another anti-RM strategy is the use of unusual bases, such as 5-hydroxymethyluracil in some *B. subtilis* phages or hydroxymethylcytosine in T-even phages (Krüger & Bickle, 1983; Warren, 1980). The P1 phage encodes proteins that are

injected together with phage DNA into infected cells and block restriction sites in phage DNA, hiding them from bacterial restriction enzymes (Iida et al., 1987). Another known strategy is to stimulate the bacterial M protein to modify the phage DNA (Zabeau et al., 1980), or to degrade cofactors required for the restriction activity of Type I and III restriction endonuclease enzymes (Studier & Rao Movva, 1976).

1.1.1.2. Argonaute proteins

Argonaute proteins were initially discovered in eukaryotes (eAgos). They perform RNA interference, cleaving RNAs that pair with small interfering RNAs (siRNA) bound to eAgo (Lingel & Izaurralde, 2004; J. Liu et al., 2004). Inactive variants of eAgos in complex with microRNAs were demonstrated to block the translation of complementary RNAs in a reversible way (Hutvagner & Simard, 2008). Prokaryotic Agos (pAgos) were found to be the ancestors of eAgos (E. V. Koonin, 2017). They defend bacteria and archaea from invading nucleic acids. In contrast to eAgos, which are present in ~65% of all sequenced eukaryotic genomes, pAgos are present in only ~30% of archaeal and ~10% of bacterial genomes (Swarts, Makarova, et al., 2014). Both eukaryotic and prokaryotic Agos contain a nuclease PIWI domain and guide-binding MID and PAZ domains, though PAZ is present only in some prokaryotic Agos (Makarova et al., 2009). pAgos with a full set of domains are called “long”, while “short” pAgos have only PIWI and MID. Different types of pAgos may use short RNA or DNA molecules as guides to recognize complementary targets, which is followed by the introduction of single-strand breaks. The majority of studied pAgos use ssDNA as a guide and prefer dsDNA as targets (Swarts, Jore, et al., 2014; Willkomm et al., 2017; Yuan et al., 2005). Interestingly, catalytically inactive variants of pAgos are often genetically associated with other nucleases, suggesting that these pAgos perform target recognition followed by target cleavage by other, recruited nucleases (Swarts, Makarova, et al., 2014). The guide molecules are proposed to originate from degradation products of foreign mRNAs, though the precise mechanism is still to be studied (Olovnikov et al., 2013). Some pAgos can perform guide-independent target cleavage, thus potentially generating guides for subsequent directed targeting (Swarts et al., 2017; Zander et al., 2017). Given their simplicity, pAgos are promising tools for

biotechnology. Establishing the conditions to ensure high target and guide binding specificity is one of the main problems for Ago applications (Kuzmenko et al., 2019). Moreover, Agos lack helicase activity and may rely on additional cellular proteins to unwind dsDNA targets (Hunt et al., 2018; Kuzmenko et al., 2019).

1.1.1.3. BREX

Bacteriophage exclusion (BREX) systems are widespread among bacteria and archaea and are classified into six subtypes based on the number and the order of genes in the system (Goldfarb et al., 2015). All BREX systems known to date have two common proteins with the PglZ alkaline phosphatase and PglY ATPase domains (Goldfarb et al., 2015). The PglX site-specific methyltransferase domain is also present in most BREX systems and is suggested to be responsible for self/nonself discrimination by modifying the host DNA (Goldfarb et al., 2015; Gordeeva et al., 2019). Though this makes BREX similar to RM systems, it is functioning differently, apparently blocking the invading DNA replication without degrading it (Goldfarb et al., 2015; Gordeeva et al., 2019). Consistently, removing the *pglX* gene does not lead to the host genome degradation, though makes the cell susceptible to the phage (Goldfarb et al., 2015). A complex of multiple BREX proteins is suggested to scan the DNA for unmethylated recognition sites and block phage replication origins (Shen et al., 2013). Some phages were found to be resistant to BREX (Goldfarb et al., 2015), implying the existence of anti-BREX mechanisms. The T7-encoded Ocr protein was demonstrated to protect the phage not only from RM of Type I, but also from BREX, by associating with the PglX methyltransferase (Isaev et al., 2020).

1.1.2. Adaptive immunity

Adaptive immunity of bacteria and archaea is represented by CRISPR (**c**lustered **r**egularly **i**nterspaced **s**hort **p**alindromic **r**epeats)-Cas (**C**RISPR-associated sequence) systems, which are more comprehensively described in Chapter 2.2. They are present in ~85% of archaea and ~40% of bacteria (Makarova et al., 2020). CRISPR-Cas systems use

pieces of foreign DNA to form a unique memory record in a host genome, such that the next time an invader comes it is specifically recognized and destroyed. CRISPR arrays are composed of short (25-50bp) palindromic direct repeats separated by unique spacers and an upstream leader sequence required for the acquisition of new spacers (Deveau et al., 2010; Wei et al., 2015). There can be several, from 2 to 20, CRISPR arrays in the genome, each of them containing from 1 to several hundred CRISPR repeats. Many bacteria encode multiple CRISPR-Cas systems in their genome. For example, *Serratia sp. ATCC 39006* encodes type I-E, I-F, and III-A systems (Patterson et al., 2016).

Some Cas proteins form an effector complex necessary for CRISPR-Cas-mediated immunity function. Other Cas proteins are required for immunity acquisition. The structure of CRISPR loci, gene composition, and Cas protein sequences define several types and subtypes of CRISPR-Cas systems (Makarova et al., 2020).

There are three major stages of CRISPR-Cas immunity: 1) adaptation (spacer acquisition), 2) expression, and 3) interference. During CRISPR adaptation, the fragments of foreign DNA are recognized, bound by Cas adaptation enzymes, and incorporated into the CRISPR array at the leader-proximal end, becoming new spacers. If this is the first time bacteria encounter an invader, the process is called naïve adaptation (Datsenko et al., 2012; Fineran & Charpentier, 2012). During the expression stage, the array is transcribed into a long precursor CRISPR RNA (pre-crRNA), which is then processed into mature short crRNAs (Brouns et al., 2008; Pougach et al., 2010). Finally, during the interference step Cas-crRNA effector complexes recognize protospacers, invader DNA sequences complementary to the crRNA spacer part. This is followed by the degradation of foreign nucleic acid by the nuclease component of the effector (Brouns et al., 2008; Garneau et al., 2010).

To be recognized, a protospacer should be accompanied by a specific PAM sequence (protospacer-adjacent motif) (Mojica et al., 2009). PAM sequence is excised from pre-spacers before integration, such that spacers in CRISPR arrays are not targeted (Westra et al., 2013). Moreover, PAM unequivocally determines the way a new spacer is

inserted into the CRISPR array, assuring that the corresponding future crRNA will be protective (Datsenko et al., 2012; Sashital et al., 2012; Swarts et al., 2012).

Genetic elements may acquire point escape mutations in PAM or protospacer sequences to avoid recognition by bacterial CRISPR-Cas systems (Semenova et al., 2011). In turn, bacteria "learned" to fight this through primed CRISPR adaptation, or priming (Datsenko et al., 2012). This type of CRISPR adaptation is coupled with CRISPR interference. Priming was demonstrated for the Type I and Type II CRISPR-Cas systems (Datsenko et al., 2012; Nussenzweig et al., 2019; Swarts et al., 2012). CRISPR adaptation machinery is attracted to target-bound effector-crRNA complex, leading to a robust CRISPR adaptation process, incorporating new spacers from DNA located *in cis* to the targeted protospacer (Datsenko et al., 2012; Savitskaya et al., 2013; Swarts et al., 2012). The primed CRISPR adaptation yield is strongly increased by mutations in PAM or mismatches of the protospacer sequence with crRNA (Datsenko et al., 2012; Fineran et al., 2014). Priming mechanism allows the bacteria to update the memory record, increase the immune diversity in the population, and prevent the propagation of mutated genetic elements, which could otherwise overcome CRISPR immunity. CRISPR diversity appears important at both individual and population levels. The existence of several different spacers against the invader in one cell makes it harder for the foreign DNA element to avoid CRISPR immunity, as it would need to acquire more escape mutations (Levin et al., 2013). In population, escape mutants that managed to avoid the CRISPR-Cas system in one cell, may not survive in other cells possessing different spacers, due to the limited ability to acquire escape mutations (van Houte et al., 2016).

Some phages encode specific anti-CRISPR proteins (Acrs) to protect themselves from the host CRISPR-Cas immunity. Acrs affect different stages of CRISPR-Cas immunity, preventing crRNA or target binding (D. Dong et al., 2017; Meeske et al., 2020; Thavalingam et al., 2019; Zhu et al., 2019), or blocking the nuclease activity of the effector (Harrington et al., 2017). Acrs for almost every CRISPR-Cas system type were characterized. The diversity of phage anti-CRISPR proteins is suggested to be one of the main drivers of the diversification of the CRISPR-Cas types and subtypes (Y. Li & Bondy-

Denomy, 2021). Acrs are promising tools for CRISPR-Cas applications since controlling the nuclease activity of Cas proteins allows to increase the editing accuracy. Recently a CRISPR adaptation-specific inhibition by the truncated Acr was also demonstrated (Philippe et al., 2022), though its mechanisms are still to be determined.

1.1.3. Abortive infection

This type of defense includes mechanisms that prevent the infection of the entire population by sacrificing individual infected cells while preventing the generation of viral progeny. These mechanisms usually act in the later stages of infection when other defense systems fail (Lopatina et al., 2020). The infected cell either becomes dormant or dies, with both outcomes preventing the formation of progeny phage (Lopatina et al., 2020). Abortive infection (Abi) systems are diverse and include TA (toxin-antitoxin), Rex, PifA, CBASS (cyclic oligonucleotide-based antiphage signaling), and certainly many others.

TA systems are widely encoded on both plasmids and chromosomes (Jensen & Gerdes, 1995; Pandey & Gerdes, 2005). They encode a toxin, which is typically a stable protein able to affect some essential cellular systems such as replication, translation, or cell wall synthesis, and a much less stable antitoxin, which may be a protein or a non-coding RNA (Gerdes et al., 1986; Greenfield et al., 2000; Jensen & Gerdes, 1995; Ogura & Hiraga, 1983). Antitoxin should be constantly expressed in sufficient levels to inhibit the toxin function. When a cell is undergoing stress or infection, the expression levels of TA system components are altered. When the antitoxin concentration decreases below a certain threshold, the toxin causes growth arrest or cell death. TA modules are categorized into eight classes, according to their toxin-antitoxin interaction mechanisms (Gerdes et al., 2005). In types I-VII, toxin is a protein which is neutralized by a small RNA antitoxin (types I and III) or in other types, by a protein. In type VIII, toxin is a small RNA inhibited by an anti-sense antitoxic RNA.

The Rex system of *E. coli* consists of two genes, RexA and RexB, acting together to stop cell growth and abort the infection by some lambdoid phages as well as some T4,

T5, and T7 strains (Toothman & Herskowitz, 1980). RexA presumably senses phage-produced DNA-protein complexes and activates the transmembrane RexB protein (Snyder, 1995). Activated RexB forms an ion channel and causes loss of membrane potential, ATP level drop, and growth arrest. There is evidence showing that infected Rex+ cells enter the stationary phase rather than undergo suicide (Slavcev & Hayes, 2003). The T4 phage has evolved a mechanism to overcome the Rex defense. T4 RIIA and RIIB proteins inhibit Rex by a yet unknown pathway, though it was shown to be concentration-dependent (Shinedling et al., 1987).

Another Abi system PifA is encoded by the *pifA* gene on F plasmid and protects *E. coli* from the T7 phage. Membrane-associated PifA senses capsid protein Gp10 and dGTPase inhibitor Gp1.2 (Schmitt & Molineux, 1991). Upon activation, PifA leads to ATP leakage and cell death due to the membrane damage (Schmitt et al., 1991). The mechanism of sensing is still unknown. Mutations in both *1.2* and *10* genes allow T7 to avoid recognition and to overcome the PifA defense (Molineux et al., 1989).

CBASS (cyclic oligonucleotide–based antiphage signaling system) systems were discovered in bacteria relatively recently. They are present in ~10% of known bacteria and archaea and are likely to be the ancestors of eukaryotic cGAS-STING systems (D. Cohen et al., 2019; Kranzusch et al., 2015). They include several mechanisms utilizing an oligonucleotide cyclase producing cyclic molecules made of two or three nucleotides in response to phage infection (Whiteley et al., 2019). Cyclic oligonucleotides serve as secondary messengers of infection activating a cell-killing effector (D. Cohen et al., 2019), which can be a phospholipase or another protein, depending on the type of system. One of the types encoded by *Vibrio cholerae* El Tor and *E. coli* TW11681 includes the DncV protein producing cGAMP and a phospholipase CapV. Phage infection activates DncV, which in turn induces CapV (Severin et al., 2018), leading to membrane degradation and cell lysis (D. Cohen et al., 2019). CBASS system was shown to be activated before the mature phage particles are formed inside the infected cell, so that cell lysis does not lead to more infections (D. Cohen et al., 2019).

1.2. CRISPR-Cas function and classification

There are plenty of types and subtypes of CRISPR-Cas systems known, differing from each other by the composition and structure of effector complexes and target cleavage mechanisms. All systems can be divided into two classes: Class 1 systems are much more abundant and possess multi-subunit effector complexes. Systems with single-subunit effector enzymes belong to Class 2. Class 1 and 2 CRISPR-Cas systems likely originated independently (E. V. Koonin & Makarova, 2019). In recent years, the need for more compact and efficient genome editing tools and the implementation of novel bioinformatics approaches (Shmakov et al., 2015) greatly stimulated the discovery of novel CRISPR-Cas systems, particularly of Class 2 (Burstein et al., 2017; Murugan et al., 2017; Stella et al., 2017). Each type of CRISPR-Cas systems has a signature protein distinguishing it from other types. According to the current classification, Type I, III and IV belong to Class 1 and include 16 subtypes. Class 2 contains Type II, V and VI with 17 subtypes (Makarova et al., 2020). Besides differences in effector complex composition, Class 1 and 2 utilize different mechanisms of pre-crRNA processing: Class 1 uses a multisubunit effector complex, sometimes with the help of additional Cas proteins, to cut pre-crRNA. Processing in Class 2 systems is performed by RNAse III with the help of *trans*-activating RNA (tracrRNA) encoded in CRISPR loci (Deltcheva et al., 2011) or by the effector protein itself (East-Seletsky et al., 2016; Fonfara et al., 2016).

Class 2 systems attract particular interest in biotechnology due to their compact size. The effector comprises one single protein: Cas9 for Type II, Cas12 for Type V, and Cas13 for Type VI. Types and subtypes within the class mainly differ by their gene composition and pre-crRNA processing mechanisms. In Type II and some Type V subtypes, pre-crRNA processing is performed by RNAse III with the help of tracrRNA (Deltcheva et al., 2011). TracrRNA has a 5’ region partially complementary to pre-crRNA repeat (which is not palindromic). They form duplexes which are recognized and cleaved by RNAse III, leading to the formation of mature crRNAs (Deltcheva et al., 2011). TracrRNA molecule remains bound to crRNA and the effector complex after processing.

The 3’ end of tracrRNA contains several structural elements interacting with the effector and is required for target recognition and cleavage (Briner et al., 2014).

Types II and V differ in their effectors' domain organization: Cas9 has RuvC-like and HNH nuclease domains, each cutting one of the target DNA strands (Jinek et al., 2012; Makarova et al., 2006). Cas12 has only one RuvC-like domain responsible for cleaving both strands (Swarts & Jinek, 2019). Cas13 has two HEPN domains (Abudayyeh et al., 2016). After being activated by recognition of a complementary target, which is RNA, it demonstrates non-specific *in trans* RNAse activity and thus might promote cell dormancy or programmed cell death and slow down phage infection, which allows to consider Type VI systems as Abi mechanisms (Abudayyeh et al., 2016; Meeske et al., 2019).

The most popular gene editing system utilizes Cas9 of *Streptococcus pyogenes* (SpyCas9) and a fused single guide RNAs (sgRNAs), instead of a complex of individual crRNA:tracrRNA (Cong et al., 2013; Jinek et al., 2012). To control gene expression without editing, the catalytically inactivated dCas9 (**d**ead Cas9) variant is used. The system is called CRISPRi (CRISPR interference) (Qi et al., 2013), though, in contrast to RNAi, which prevents translation, it mainly affects gene transcription. dCas9 can be used not only for gene repression but also for gene activation if fused to the transcription activation domains (C. Dong et al., 2018; Gasiunas et al., 2012). Programmable RNA-targeting can be performed by Cas13. Given its smallest size among all known CRISPR effectors, Cas13 is a promising tool for the RNA interference applications in medicine, biotechnology, and research (Jinek et al., 2012). However, the “collateral” activity of Cas13 makes specific RNA targeting problematic, though many successful application cases are already described. Besides Cas13, Cas9 variants able to target RNA are also known (Strutt et al., 2018).

Interestingly, several Type II CRISPR-Cas systems were found to naturally perform gene regulation functions important for bacterial virulence. In *Francisella novicida* strain U112 possessing Type II-B CRISPR-Cas system, the expression of bacterial lipoprotein (BLP) is downregulated by the complex of Cas9, tracrRNA, and a small CRISPR–Cas-associated RNA (scaRNA), which is transcribed from the locus near the CRISPR array

(Sampson et al., 2013). At the same time, CRISPR array itself and other Cas proteins are not required for this function. BLP was shown to play a crucial role in bacterial virulence, as it is recognized by the innate immune system of the host. Thus, the repression of BLP expression by CRISPR-Cas system promotes bacterial infection, though the details of the mechanism remain unidentified. Similarly, *blp* expression was shown to be upregulated in Δ*cas9* mutants of *Streptococcus agalactiae* and *Riemerella anatipestifer*, possessing Type II-A and II-C CRISPR-Cas systems, respectively (Ma et al., 2018; Y. Wang et al., 2019). In several other strains, Cas9 was also demonstrated to regulate the adhesion, cytotoxicity, and survival during infection in the host (Gao et al., 2019; Saha et al., 2020; Shabbir et al., 2018; Spencer et al., 2019).

Class 1 systems, particularly of Type I, are much more abundant in nature than Class 2, though not so popular for gene editing purposes because of their complexity. Type I systems utilize Cascade (**C**RISPR-**as**sociated **c**omplex for **a**ntiviral **de**fense)-crRNA effector complex to recognize and bind complementary protospacers and recruit the Cas3 translocase/nuclease for target degradation. All Type I systems are now classified into 7 subtypes, from I-E to I-G (Makarova et al., 2020). Type I-E and I-F systems are lacking *cas4* gene, encoding an accessory adaptation protein. In subtype I-G the *cas1* and *cas4* adaptation genes are fused. The key interference protein Cas3 is encoded by a distinct *cas3* gene in all subtypes except for I-F, where it is fused with *cas2*. I-A, I-C, and I-G systems are supposed to be derivatives of I-B, differing in the order and composition of genes (Makarova et al., 2020). In subtype I-A, the Cas3 helicase and nuclease domains are encoded by two genes, *cas3'* and *cas3"* respectively, and *cas8* is split into two genes. Subtype I-D has a couple of special features such as Cas10d instead of Cas8 and *cas3"* encoded as a part of *cas10d* (Makarova, 2011).

Type I CRISPR-Cas systems are promising for long-range genome editing in eukaryotic cells (Dolan et al., 2019). Besides, Type I-E CRISPR-Cas system with deleted *cas3* was applied for transcription repression in bacteria (Rath et al., 2015).

The additional role of the Type I-F CRISPR-Cas system in virulence was shown for *Pseudomonas aeruginosa* PA14, where Cas3 targets the mRNA of LasR protein, which is crucial for recognition by the host immunity (R. Li et al., 2016). Type I-C CRISPR-Cas system of *Myxococcus xanthus* was proposed to be involved in the fruiting body development, as the expression of the *M. xanthus cas* genes is tightly regulated by the inter- and intracellular signals and was demonstrated to be activated in the fruiting body, but not in the peripheral cells (P. Viswanathan et al., 2007). In *Salmonella*, which possess Type I-E CRISPR-Cas system, the deletion of *cas3* reduced the biofilm formation and virulence (Cui et al., 2020), suggesting its role in gene regulation.

Several defective, "minimal" variants of I-B, I-E, and I-F subtypes are also known (Faure et al., 2019; Peters et al., 2017; Shmakov et al., 2018) (Makarova et al., 2020). Such defective systems do not possess components for target degradation or adaptation, but are capable of pre-crRNA processing and target recognition and often have the associated CRISPR arrays (Faure et al., 2019). Defective I-F and I-B variants were first discovered in Tn7-like transposons and suggested to mediate crRNA-guided incorporation of MGE into specific DNA loci (Peters et al., 2017). This function was also demonstrated experimentally (Klompe et al., 2019). Several variants of defective type I-E systems found in *Streptomycetaceae* include those having *tnsBC* genes instead of *cas1* and *cas2* (suggesting that transposon genes replaced the adaptation module in this system (Faure et al., 2019)) and those lacking *cas1*, *cas2,* and *cas3* but associated with STAND superfamily NTPases and a membrane protein, pointing to the possible signaling or regulating role of such systems (Shmakov et al., 2018).

Type III systems are characterized by the presence of *cas10* gene encoding a large multidomain subunit. Type III-A and III-B encode Csm and Cmr effector complexes, respectively. The interference mechanism is quite different from the one of Type I systems. The effector complex recognizes nascent RNA transcripts complementary to crRNA and then, presumably, nicks the DNA from which transcription occurs. Thus, Type III CRISPR-Cas systems target DNA in a co-transcriptional way. Interestingly, the recognition of the target by the Type III system induces the production by Cas10 of a cyclic

oligoadenylate (cOA), which serves as a secondary messenger of infection and activates CRISPR-associated non-specific RNAse Csm6, which possesses nucleotide-binding CARF (CRISPR-Associated Rossmann Fold) and HEPN (Higher Eukaryote and Prokaryote Nucleotide-binding) RNAse domains (Kazlauskiene et al., 2017; Niewoehner et al., 2017). Activated Csm6 targets both phage and bacterial RNA, which may lead to cell dormancy or death. Thereby, Type III CRISPR-Cas systems may act as Abi systems and protect the whole population from the phage in case the first lines of defense are overcome (Kazlauskiene et al., 2017; Niewoehner et al., 2017).

In contrast to effector complexes, CRISPR adaptation enzymes are conserved for both CRISPR-Cas classes and represented by the Cas1 integrase, the Cas2 structural protein, and, for some systems, auxiliary proteins Cas4, Csn2 and a reverse transcriptase (Makarova et al., 2020). Cas1 is the primary enzyme performing spacer integration. Cas2 possesses a nuclease activity *in vitro* (Beloglazova et al., 2008; Nam, Ding, et al., 2012), but it is unnecessary for the spacer integration reaction (Nuñez et al., 2014a; Yosef et al., 2012). Cas1 is the most conserved CRISPR-associated protein throughout all types (Haft et al., 2005; Makarova et al., 2006). Together with CRISPR arrays and Cas4, it is suggested to originate from casposons, a family of transposons encoding a Cas1-homologous transposase (E. V. Koonin & Krupovic, 2015). Casposons are flanked by terminal inverted repeats (TIRs) and target site duplications (TSDs), similarly to the typical DNA transposons (Krupovic et al., 2017). No Cas2 homologs have been found in casposons so far, though it is also proposed to originate from the ancestral casposon (E. V. Koonin & Krupovic, 2015). The integration of a new spacer into the CRISPR array is similar to the reaction of casposon integration. Both proceed via two nucleophilic reactions and result in duplication of the repeat in which the integration occurs (Beguin et al., 2016). Cas1 and Cas1-homologous casposases target the border between the leader and the first CRISPR repeat or between a tRNA gene and target site duplication (TSD), respectively. In both cases, the resulting ssDNA regions are filled in by an unknown polymerase. CRISPR arrays

have likely evolved from casposon TSDs, and the leader sequence presumably originated from the adjacent tRNA gene sequence (Krupovic et al., 2017).

CRISPR-Cas interference and adaptation modules are proposed to have evolved independently and must have frequently recombined to form different types and subtypes. Accordingly, the phylogenetic tree of Cas1 variants only partially corresponds to the effector proteins phylogeny of overall CRISPR-Cas classification. Recombination between modules was demonstrated (Garrett et al., 2011; Hudaiberdiev et al., 2017; Vestergaard et al., 2014).

1.3. CRISPR interference

1.3.1. CRISPR interference in Type I systems

In the Type I systems, CRISPR array is transcribed into a long precursor CRISPR RNA (pre-crRNA), which is recognized and processed by one of the components of the Cascade complex containing an RNA recognition motif (RRM): Cas5 (CasD) for Type I-C system (Nam, Haitjema, et al., 2012) or Cas6 (CasE) for other subtypes (M. Li et al., 2013; Przybilski et al., 2011; Sashital et al., 2011). The pre-crRNAs are cleaved inside each repeat, and the final crRNA consists of a spacer flanked by sequences corresponding to parts of the repeat. After the processing step, crRNA remains bound to Cas5/Cas6 and, therefore, Cascade (Jore et al., 2011; Nam, Haitjema, et al., 2012).

Cascade-crRNA complex initiates the interaction between a spacer in crRNA and a protospacer of the target DNA through the formation of an R-loop (Brouns et al., 2008; Jore et al., 2011). The protospacer strand which is complementary to crRNA is defined as the target strand (T-strand). The recognition of protospacer requires the presence of short PAM directly adjacent to it. It could be recognized by the large Cas8 or, in case of I-D system, the Cas4 subunit (Sashital et al., 2012; Shiimori et al., 2018). The presence of PAM allows the system to distinguish foreign DNA (protospacer, has PAM) from cellular DNA (spacers in CRISPR array, lack PAM) (Deveau et al., 2008). Preferred PAMs vary across different Type I systems.

The unwinding of target DNA and the formation of heteroduplex begins in the 8 bp "seed sequence" (seed) proximal to PAM, and then is extended to the remainder of the protospacer so that the protospacer strand not complementary to crRNA (nontarget, or NT-strand) is completely displaced (Jore et al., 2011; Semenova et al., 2011). Mismatches in seed between the protospacer and crRNA abolish or strongly decrease interference (Datsenko et al., 2012; Semenova et al., 2011). After the full R-loop formation, Cascade changes its conformation, recruiting the Cas3 protein (Mulepati & Bailey, 2013; Sinkunas et al., 2013; Westra et al., 2012). Cas3 has two catalytic domains, an N-terminal HD and a C-terminal SF2 domain, possessing 3'-5' exonuclease and helicase activities, respectively (Westra et al., 2010). In Type I-A and I-B systems, the nuclease and the translocase/helicase parts of Cas3 are encoded as separate proteins, while in Type I-F system *cas3* gene is fused with *cas2* (Makarova, 2011). Cas3 cleaves the NT-strand of the target DNA, and then proceeds with further DNA unwinding and degradation in the presence of ATP (Mulepati & Bailey, 2013; Westra et al., 2012).

The Cas1-Cas2 adaptation complex was shown to associate with Cascade/Cas3 during CRISPR interference (Dillard et al., 2018) and it is suggested that Cas3 degradation products may fuel CRISPR adaptation machinery (Künne et al., 2016).

1.3.2.CRISPR interference in Type I-E system

Escherichia coli K12 possesses a type I-E CRISPR-Cas system, which is well characterized. There are 2 active CRISPR arrays, CRISPR I and CRISPR II, containing 12 and 6 spacers, respectively, and 8 *cas* genes, associated with CRISPR I loci and organized in 2 operons (Figure 1) (Pul et al., 2010). The first one is controlled by P*cas* promoter and contains *casA (ygcL/cse1/cas8e), casB (ygcK/cse2/cas11e), casC (ygcJ/cse4/cas7e), casD (ygcI/cas5/cas5e)* and *casE (ycgH/cse3/cas6e)* genes encoding CasA (Cas8), CasB (Cas11), CasC (Cas7), CasD (Cas5) and CasE (Cas6) subunits, which make up the 11-subunit 405 kDa Cascade effector complex in a ratio 1: 2: 6: 1: 1 (Jore et al., 2011), and the *ygbT* and *ygbF* genes encoding the Cas1 and Cas2 proteins required for adaptation (Nuñez et al., 2014b). The Cas3 translocase/nuclease gene is transcribed separately from

P*cas3* promoter. The **i**nter**g**enic region *ygc**L**-ygc**B*** (*casA-cas3*) is named IGLB and, besides P*cas*, contains divergent promoter anti-P*cas*, located 80bp upstream from P*cas* and generating 150-200 nt RNA complementary to the 3' end of *cas3* transcript (Pul et al., 2010). Whether this RNA has a particular regulatory function is not known.

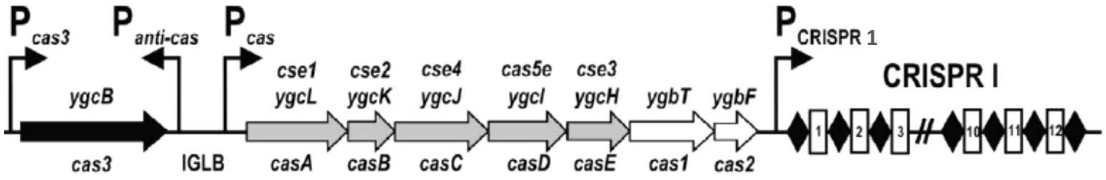


Figure 1. The Type I-E CRISPR-Cas system in *E. coli* K12 (Westra et al., 2010). The location of one of the two CRISPR arrays (CRISPR I) and associated *cas* genes is shown. Arrows indicate promoters (P*cas3*, P*anti-cas*, P*cas*, P*crispr1*). The intergenic region between *ygcL* and *ygcB* (IGLB) is indicated. Genes encoding Cascade are shown in grey, and those encoding the CRISPR adaptation module are shown in white. Black diamonds indicate CRISPR repeats, while white rectangles indicate spacers.

Each of the two CRISPR arrays is transcribed from promoters located in the leader regions (P*crispr1* and P*crispr2*). The resulting pre-crRNAs is cut by CasE, a protein with two ferredoxin-like RNA-recognizing domains. Though CasE remains bound to the Cascade-crRNA effector complex during all steps of CRISPR-Cas immunity, being also an essential structural part required for the interference step, CasE alone can process pre-crRNA (Jore et al., 2011). CasE recognizes specific sequences in stem-loop structures formed by CRISPR repeats and introduces cuts 8 nt upstream of each spacer sequence, forming ~60nt crRNAs, comprising spacers flanked by the repeat fragments (Brouns et al., 2008). The crRNA ends are bound by the opposite ends of Cascade, which has a seahorse-like structure (Jore et al., 2011). It presents the crRNA spacer, divided in 6 intervals, on the surface of a helical backbone made of CasC subunits hexamer. The 3' end of crRNA is bound by CasE at the "head" of the Cascade complex, and the 5' end is placed within the "tail" formed by the CasA, CasC, and CasD subunits (R. N. Jackson et al., 2014; Wiedenheft, Lander, et al., 2011). The CasB dimer connects the head and the tail and

stabilizes the complex, but does not interact with crRNA directly (R. N. Jackson et al., 2014).

Invader DNA targeting starts from PAM recognition by the specific Loop-1 motif of the CasA subunit, which leads to local destabilization of the target DNA (Sashital et al., 2012). PAM recognition facilitates the formation of a duplex between crRNA and the first 8 nt of the complementary protospacer strand (seed) (Semenova et al., 2011). Mutations in the seed were shown to be critical for target recognition efficiency, while mutations introducing single mismatches in other positions have little effect on interference efficiency. The consensus PAM determined by the bioinformatic analysis for the *E. coli* CRISPR-Cas system is AWG (Mojica et al., 2009). Later *in vivo* experiments showed that ~20 out of 64 possible PAM variants support at least some CRISPR interference in *E. coli* (Musharova et al., 2019).

Upon target recognition, NT-strand of the protospacer is displaced and R-loop is formed, extending all the way to the PAM-distal end of the protospacer (Jore et al., 2011; Sashital et al., 2012; Xiao et al., 2017). Full R-loop formation causes structural changes in Cascade: CasB subunits move along the backbone to the tail, causing the rotation of CasA. Cas3 is then recruited through the interaction with CasA (Hochstrasser et al., 2014; Xiao et al., 2017). Such a conformation-dependent mechanism allows to minimize the off-target effects, increasing the specificity of target degradation.

Cas3 is a ~ 100 kDa protein possessing 3'-5 'exonuclease and translocase/helicase activities. It has an N-terminal histidine-aspartate (HD) nuclease domain and a C-terminal ATP-dependent 3'-5' superfamily 2 (SF2) helicase domain. Upon binding the DNA-bound Cascade and recognizing the displaced NT-strand, the Cas3 HD nuclease introduces a cut 7-11 nt from PAM (Hochstrasser et al., 2014; Mulepati & Bailey, 2013). After that, Cas3 unwinds and moves along the NT-strand in the 3’-5’ direction, using the ATP hydrolysis energy and performing unidirectional exonucleolytic DNA degradation (Mulepati & Bailey, 2013; Sinkunas et al., 2013). According to the current model (Dillard et al., 2018), Cas3 remains bound to Cascade-crRNA and translocates the DNA, which leads to the formation of a DNA loop on the T-strand, which is cut when the complex encounters other

DNA-binding proteins. The T-strand is presumably degraded by another Cas3 molecule, which may unwind the R-loop and bind the DNA which was in heteroduplex with crRNA, displacing the effector complex (Gong et al., 2014; Howard et al., 2011; Mulepati & Bailey, 2013). In the absence of ATP, Cas3 performs nicking of the NT protospacer strand without subsequent translocation and degradation of DNA (Mulepati & Bailey, 2013).

Interestingly, the CRISPR-Cas system in *E. coli* is not active in natural strains. Transcription of the *cas* operon is inhibited by the H-NS (histone-like nucleoid-structuring) protein (Pul et al., 2010; Westra et al., 2010). The LeuO transcription factor was demonstrated to be an antagonist of H-NS, preventing it from binding to *cas* promoters, at least when overexpressed (Westra et al., 2010). No physiological conditions favoring CRISPR activation were found in *E. coli* as of the time of this writing. Be as it may, all components of *E. coli* CRISPR-Cas are functional, and genetically engineered strains lacking *hns* genes or having *cas* genes under inducible promoters are interference and adaptation proficient and widely used in research.

1.4. CRISPR adaptation

1.4.1.CRISPR adaptation in Type I systems

There are two scenarios of how immunity may be acquired by the Type I CRISPR-Cas systems when a foreign DNA agent enters the cell. If this is the first time bacteria meet the invader, and there is no record of this invader DNA in the CRISPR array, naïve CRISPR adaptation takes place (Datsenko et al., 2012; Fineran & Charpentier, 2012). This process requires the Cas1 and Cas2 proteins only (Datsenko et al., 2012; Yosef et al., 2012). It is believed that the Cas1-Cas2 proteins bind DNA fragments generated by RecBCD *in vivo* (Levy et al., 2015). Other sources are also likely. If there is a mismatch between a spacer in crRNA and a protospacer in invading DNA, or the PAM sequence of the protospacer is suboptimal, the Cascade-crRNA complex can still recognize the protospacer and induce the process of highly efficient spacer acquisition, called primed CRISPR adaptation, or priming (Datsenko et al., 2012). Fully matching targets induce priming as well, though

spacers are not acquired as efficiently due to the robust degradation of the target, leading to its rapid disappearance (Krivoy et al., 2018; Severinov et al., 2016). Priming requires the interference machinery and likely uses fragments generated by Cas3 nuclease as a source of new spacers (Datsenko et al., 2012; Künne et al., 2016). Numerous other cellular nucleases and helicases also affect both naïve and primed adaptation (Ivančić-Baće et al., 2015; Levy et al., 2015; Radovčić et al., 2018), though the exact mechanisms are still to be investigated.

Cas1 and Cas2 individually form homodimers and both have metal-dependent nuclease activity. Together they constitute a heterohexamer $(Cas2)_2(Cas1)_4$, where symmetric Cas2 dimer (Cas2-Cas2' (J. Wang et al., 2015)) is located between two asymmetric Cas1 dimers (Cas1a-Cas1b and Cas1a'-Cas1b' (J. Wang et al., 2015)), forming a dumbbell-like structure (Figure 2) (Nuñez et al., 2014a; Nuñez, Harrington, et al., 2015; J. Wang et al., 2015). Cas1 has a molecular weight of ~30 kDa and is structurally divided into an N-terminal β-sheet domain and a C-terminal α-helical domain with an active endonuclease site made of several conserved metal-binding residues (Figure 3) (Babu et al., 2011; T.-Y. Kim et al., 2013; Wiedenheft et al., 2009). Cas2 is a ~10kDa protein with a core ferredoxin fold (Beloglazova et al., 2008; Nam, Ding, et al., 2012; Samai et al., 2010). In the Cas1-Cas2 integrase complex, asymmetric Cas1 dimers perform the catalytic functions, while the Cas2 active sites are unnecessary for spacer integration (Nuñez et al., 2014a; Yosef et al., 2012). Overall, there are 4 Cas1 catalytic sites in the complex, though only Cas1a and Cas1a' catalytic sites are active during CRISPR adaptation. Cas1b and Cas1b' are responsible for interactions with Cas2 (Figure 2) (Nuñez et al., 2014a; J. Wang et al., 2015), while there are no interactions between Cas2 and Cas1a or Cas1a'. Cas2 also ensures non-specific interactions with prespacers, which are crucial for CRISPR adaptation (Nuñez et al., 2014a; J. Wang et al., 2015).

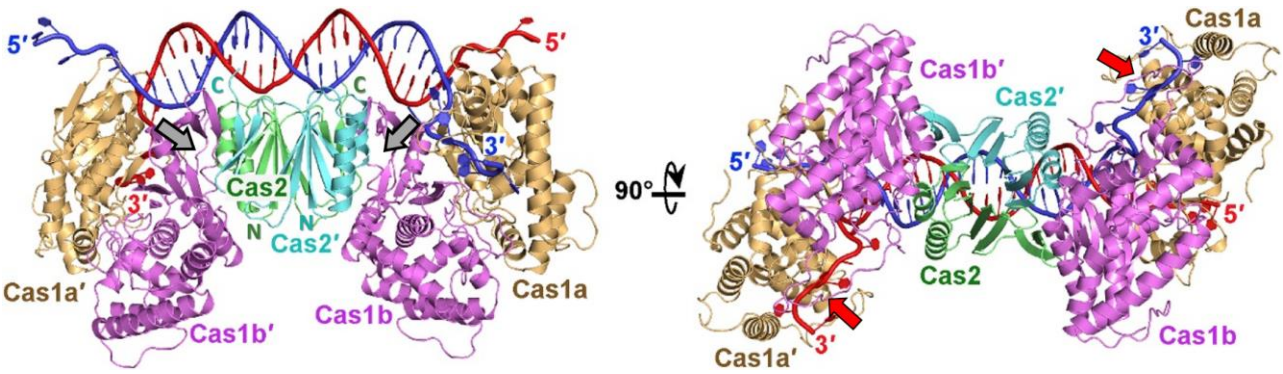


Figure 2. The crystal structure of the *E. coli* Cas1-Cas2 complex bound to dual-forked DNA, comprising 23 bp duplex with 6 nt 5' overhangs and 10 nt 3' overhangs, containing the PAM-complementary 5′-CTT-3′ sequence. Cas1a and Cas1a′ are colored in light orange, Cas1b and Cas1b′ are colored in magenta, and the Cas2 monomers are in green and cyan. The two DNA strands are colored red and blue, respectively. On the left, N-terminal β-sheet domains of Cas1b and Cas1b' interacting with Cas2 dimer are indicated by grey arrows. On the right, Cas1a and Cas1a' active sites in the C-terminal α-helical domains are indicated by red arrows. Adapted from (J. Wang et al., 2015).

The preferred size and structure of DNA substrates bound by the Cas1-Cas2 complex vary across different CRISPR-Cas types and are determined by the distance between Cas1 dimers and their relative orientation.

Besides Cas1 and Cas2, some Type I CRISPR-Cas systems employ an additional Cas4 protein for CRISPR adaptation. Cas4 in a complex with Cas1-Cas2 ensures the proper spacer orientation and PAM specificity. It recognizes and cleaves the PAM-containing 3' end to the length suitable for incorporation into the CRISPR array (Shiimori et al., 2018). In I-U and I-B systems, the Cas4 domain is fused with Cas1 (Almendros et al., 2019).

1.4.2. CRISPR adaptation in Type I-E system

Structural data and the results of the *in vitro* experiments suggest a 23 bp DNA duplex with 5 bp 3' overhangs on both sides to be the preferred substrate for the *E. coli* Cas1-Cas2 complex (Nuñez, Harrington, et al., 2015). Similar constructs with 5' overhangs are not integrated by Cas1-Cas2 (Nuñez, Lee, et al., 2015). The duplex part of the DNA molecule is stretched across the flat surface of the Cas1-Cas2 integrase, fixed via non-

specific DNA-protein interactions. The single-stranded 3’ ends are located in the channels of catalytic Cas1a and Cas1a’ subunits (Figure 2) (Nuñez, Harrington, et al., 2015; J. Wang et al., 2015). Y22 residues of Cas1a and Cas1a’ were shown to unwind the protospacer ends and displace the 5’ end extensions from the active site, stabilizing the structure (Nuñez, Harrington, et al., 2015; J. Wang et al., 2015). The distance between two Y22 residues determines the 23bp length of the bound DNA duplex (J. Wang et al., 2015). The sequence complementary to PAM (3’-TTC-5’) is specifically recognized inside a pocket formed by the catalytic Cas1a subunit (Figure 3) (J. Wang et al., 2015). The C-terminal tail of Cas1b covers the Cas1a catalytic pocket and also interacts with PAM. *In vitro* Cas1 was demonstrated to cleave 3’ ends of the dual-forked DNA substrate inside the PAM-complementary sequence at position 5nt from the 23-bp duplex region (between C28 and T29, Figure 3), resulting in a 33-nt product (J. Wang et al., 2015).

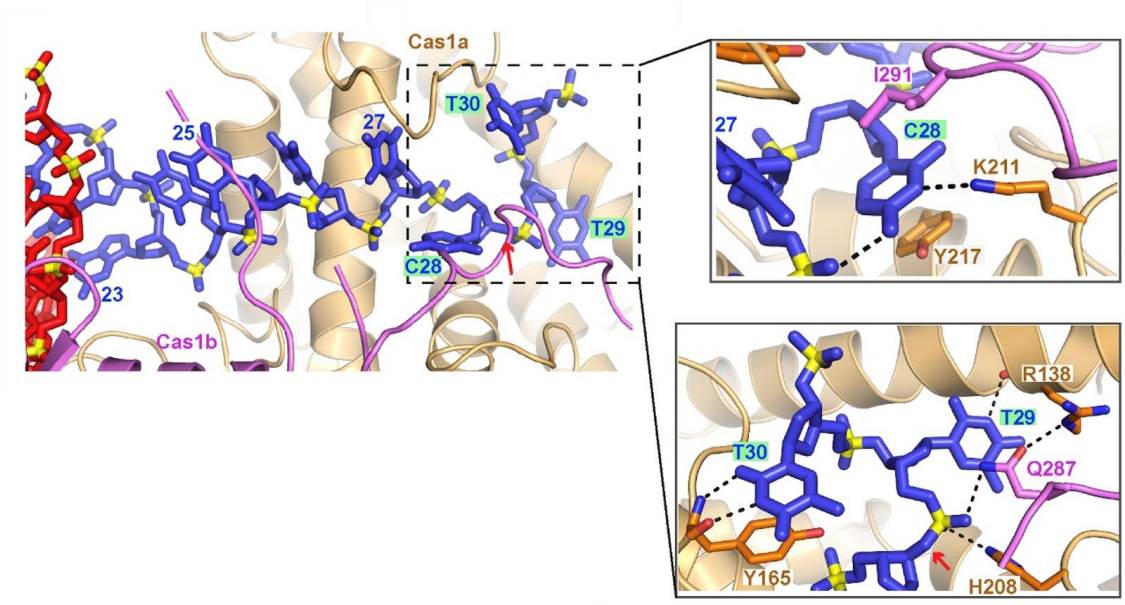


Figure 3. View of the 3′ overhang containing PAM-complementary sequence located in the C-terminal domain of Cas1a and covered by the C-terminal tail of Cas1b. The PAM-complementary nucleotides (C28, T29, T30) are indicated by green background. On the right, the detailed sequence-specific interactions between Cas1 and PAM-complementary nucleotides are zoomed-in. The site of DNA cleavage by Cas1 *in vitro* is indicated by the red arrow. Adapted from (J. Wang et al., 2015).

The integration reaction is similar to retroviral integration and starts with the Cas1-catalyzed nucleophilic attack by the 3'-OH of the prespacer at the phosphodiester bond between the leader and the first repeat in the CRISPR array, leading to the formation of a half-site product (Figure 4) (Nuñez et al., 2016; Rollie et al., 2015). Next, the second 3'-OH attacks another strand, at the first repeat-spacer junction (Nuñez et al., 2016; Rollie et al., 2015). As a result, a product is formed resembling a new dsDNA spacer flanked by the two ssDNA repeat regions. These regions are filled in, presumably, by DNA polymerase I, since deletion of *polA* was shown to prevent CRISPR adaptation (Ivančić-Baće et al., 2015). The remaining nicks are ligated by an unknown ligase.

Interestingly, spacer integration *in vivo* always occurs at the proximal end of the leader sequence, before the first repeat, while *in vitro* new spacers are incorporated after any repeat in a CRISPR array located on a supercoiled plasmid (Nuñez, Lee, et al., 2015). It is assumed that the cellular IHF factor, which binds to and bends the leader sequence and attracts Cas1-Cas2 complex, is responsible for integration specificity observed *in vivo* (Nuñez et al., 2016). The IHF binding site is located inside the leader sequence upstream of the Cas1-Cas2 binding site (Figure 4), at positions -9 to -35 from the first CRISPR repeat (Nuñez et al., 2016). In a supercoiled plasmid, it is suggested that DNA is already in a bent conformation supporting the integration in the absence of IHF. *In vitro*, spacer acquisition is abolished with a linear or relaxed CRISPR array containing DNA substrates (Nuñez, Lee, et al., 2015).

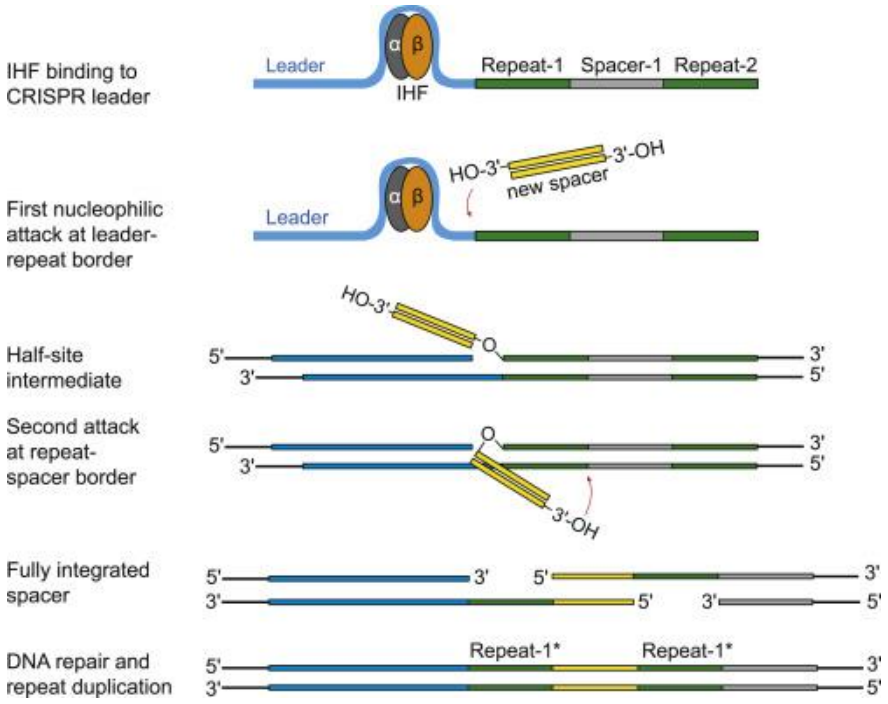


Figure 4. Spacer integration into the CRISPR array in *E. coli* (Nuñez et al., 2016).

The Type I-E CRISPR-Cas systems lack the accessory Cas4 protein, therefore, it should employ other mechanisms to ensure the proper orientation of spacers in CRISPR array to provide immunity. The structure of spacer precursors formed during primed CRISPR adaptation in *E. coli in vivo* was determined. Most detected prespacers contain a 33 bp duplex with a 4 nt 3' overhang on the CTT-containing strand (Musharova et al., 2021; Shiriaeva et al., 2019). The asymmetrical prespacer structure may determine the proper spacer orientation in the array (Musharova et al., 2021; Shiriaeva et al., 2019), while the PAM specificity is provided by the base-specific recognition of PAM-complementary sequence by Cas1.

Both primed and naïve adaptation are suggested to proceed through the same steps, though the differences are expected to arise in prespacer formation mechanisms. Naïve CRISPR adaptation is proposed to use ssDNA products generated by RecBCD and other cellular nucleases during DNA repair, replication termination, and resolution of chromosomal concatemers (Ivančić-Baće et al., 2015; Levy et al., 2015; Radovčić et al.,

2018). Only about half of newly acquired spacers during naïve CRISPR adaptation correspond to protospacers with the canonical AAG PAM (Yosef et al., 2012). In contrast, ~95% of spacers acquired by primed adaptation derive from protospacers with the AAG PAM, and most originate *in cis* with respect to the protospacer from which priming is initiated (Savitskaya et al., 2013). Another important feature of priming is a strong strand bias for the new spacers selection, which is determined by the localization of the priming protospacer (PPS). In *E. coli*, most spacers map on the NT- strand upstream the PPS, or from the T-strand downstream the PPS (Musharova et al., 2021; Shiriaeva et al., 2019). Interestingly, the pattern is reversed in Type I-F system, which could be due to the different locations of the Cas3 cleavage sites in the R-loop (Shiriaeva et al., 2019). The efficiency of spacer acquisition during priming has a gradient distribution, being maximal close to the PPS site and decreasing further away from it (Shiriaeva et al., 2019). In contrast, new spacers are acquired from both strands without any preference during naïve CRISPR adaptation (Musharova et al., 2018).

Primed CRISPR adaptation is thought to proceed via the formation of a primed adaptation complex (PAC), which consists of target-bound Cascade, Cas1-Cas2, and Cas3 (Dillard et al., 2018; Musharova et al., 2021; Redding et al., 2015). In agreement with this model, prespacer fragments were recently shown to be associated with Cas1-Cas2 and Cas3 during priming (Musharova et al., 2021). Cas3 is suggested to move along the DNA in a 3'-5' direction as a part of PAC, unwinding and degrading it. Other cellular nucleases may also take part in this process. The degradation products are then passed directly to Cas1-Cas2, which samples the PAM-containing fragments and directs them for CRISPR adaptation (Künne et al., 2016; Musharova et al., 2021). The free 3' and 5' ends of bound prespacers are trimmed to the proper prespacer length either by the Cas1 nuclease itself (J. Wang et al., 2015) or with the help of other cellular nucleases.

Though branched dsDNA fragments were initially demonstrated to be the preferred substrates for binding and incorporation into the CRISPR array (Nuñez, Harrington, et al., 2015), subsequent single-molecule experiments showed that Cas1-Cas2 can capture various substrates, including ssDNA. Particularly, Cas1-Cas2 binds 3′-TTC-5′ -containing

ssDNA fragments and then anneals them to the complementary strands, leading to the formation of a dsDNA prespacer precursor (S. Kim et al., 2020). Together with PAC formation, such a mechanism allows Cas1-Cas2 to rapidly bind ssDNA fragments directly transferred from Cas3 and ensures high efficiency of primed CRISPR adaptation.

1.5. Genome maintenance systems

Different external (UV-, X-, gamma-rays, chemical agents) and internal (replication defects) factors can damage bacterial DNA. There are several types of DNA damage threatening genome stability in bacteria, among which double-strand breaks (DSBs) are the most critical. In the absence of any external factors, DSBs commonly occur in bacteria when a replicative fork collapses after facing a ssDNA nick on a template strand (Kuzminov, 2001). Such DSBs are called one-ended. Two-ended DSBs are formed when both DNA strands are cleaved simultaneously by external factors. In *E. coli*, both types of DSBs are repaired mostly by homologous recombination (HR), using intact double-stranded DNA copy as a template (Figure 6). The mechanism relies on several conserved proteins and involves the formation of a nucleoprotein filament, which performs homology search and sequence exchange. The details of the HR mechanism are discussed below.

1.5.1. The RecBCD pathway

The majority (99%) of DSBs in *E. coli* are repaired through the RecBCD pathway (Dillingham & Kowalczykowski, 2008; Emmerson, 1968; Howard-Flanders & Theriot, 1966; Willetts & Mount, 1969). The RecBCD complex is a highly processive nuclease-helicase. According to the current model, RecBCD binds blunt or nearly blunt dsDNA ends (Taylor & Smith, 1985) and starts unwinding and degrading both strands, using ATP energy (Hickson et al., 1985). Upon facing a Chi site, RecBCD stops DNA degradation and initializes HR, thus repairing the break.

RecB is a 134 kDa protein containing several SF1 helicase motifs in its N-terminal part and different nuclease motifs in the C-terminal part. Isolated RecB is a weak 3’-5’ helicase and ssDNA-dependent ATPase (Hickson et al., 1985). Its N-terminal part binds RecC. No particular functional motifs were found in RecC, a 129 kDa protein. It is required

for Chi recognition and stimulates the activity of RecB (Masterson et al., 1992). Structural (Cheng et al., 2020; Singleton et al., 2004) and mutational (Amundsen et al., 2016; Handa et al., 2012; Schultz et al., 1983) data suggest that the 3’ DNA end is passed through RecB and recognized in the tunnel of RecC. The smaller 67 kDa RecD subunit also contains SF1 helicase motifs and exhibits 5’-3’ helicase and ssDNA-dependent ATPase activities (H. W. Chen et al., 1997; Dillingham et al., 2003).

RecB and RecD are the motors, translocating on opposite DNA strands. Notably, the helicase activity of the RecBCD complex is much higher than that of RecB or RecD alone and was measured to be above 1000 bp/s (Bianco et al., 2001; Dillingham et al., 2003). One of the motors is moving faster, which causes the formation of a loop (Taylor & Smith, 2003). The nuclease of RecB is responsible for degradation of both strands (J. Wang et al., 2000). The 3’-end is suggested to be degraded more efficiently than the 5’-end, which is presumably directed by RecD to the nuclease domain (J. Wang et al., 2000). Another view suggests that RecBCD does not degrade DNA and only unwinds the two strands during translocation (Taylor et al., 1985; Taylor & Smith, 1995b). To proceed with recombination, RecBCD must recognize a properly oriented specific motif 5'-GCTGGTGG-3' called Chi (crossover hotspot instigator), through the RecC subunit (Lam et al., 1974; Malone et al., 1978). It should be recognized in the upper, 3’-terminated strand. The recognition causes a sequence of transformations: the 3’-5’ activity of the complex is attenuated, while the 5’-3’ activity is upregulated, resulting in the formation of a 3’-tail terminated with Chi, which presumably remains bound to RecC for some time after the recognition (Spies et al., 2007). RecBCD is no longer able to recognize and cut Chi on downstream DNA (Taylor & Smith, 1992). RecB then induces the loading of RecA on the 3’-tail, with each RecA monomer binding to three nucleotides (Z. Chen et al., 2008). As a result, a nucleoprotein filament capable of promoting HR is formed (Persky & Lovett, 2008) (Figure 5). When the complementary strand is recognized, a D-loop is formed followed by subsequent recombination through the formation of the Holliday junction. RecG or RuvABC and PriA are then recruited to the recombination intermediate and restart DNA synthesis (Persky & Lovett, 2008). RecG and RuvABC move and resolve the

junction, while PriA mediates the binding of DnaB helicase, thus restoring DNA synthesis (Marians, 2004).

In vitro, RecBCD possesses a Mg^{2+}- and ATP-dependent nuclease activity. A high concentration of free Mg^{2+} ions is required for the RecBCD nuclease activity (Eggleston & Kowalczykowski, 1993), while ATP chelates Mg^{2+} (J. E. Wilson & Chin, 1991). In conditions of low Mg^{2+}/ATP concentration ratio, RecBCD unwinds both DNA strands without degrading them (Eichler & Lehman, 1977) and makes a site-specific nick at the Chi-containing strand, 4-6 nt to the 3' side from Chi (Ponticelli et al., 1985; Taylor et al., 1985). When Mg^{2+} is in excess, RecBCD nuclease activity is stimulated, leading to degradation of the upper, Chi-containing strand into up to Chi, while the lower strand is cut into longer fragments (Dixon & Kowalczykowski, 1995; Taylor & Smith, 1995b). The products of the Chi-containing strand degradation in these conditions were shown to vary from tens to hundreds nt, while the lower strand is cut into fragments of a kilobase length (Dixon & Kowalczykowski, 1995). As it is hard to estimate the real intracellular Mg^{2+} and ATP concentrations, whether RecBCD degrades both strands before Chi or only unwinds them, is not yet determined.

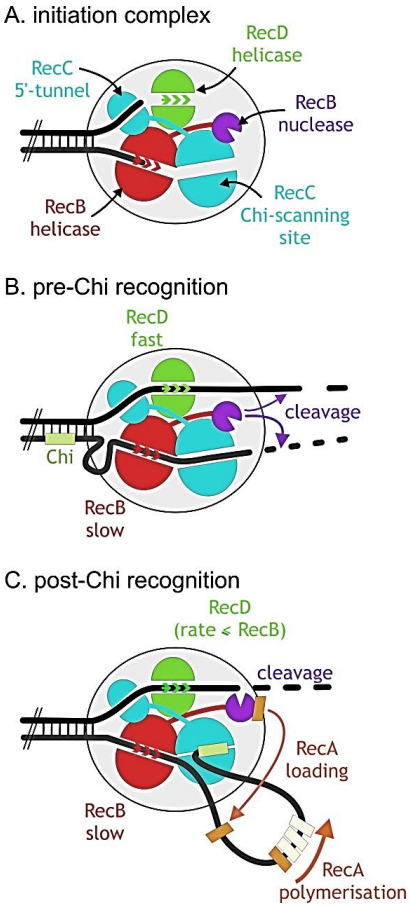


Figure 5. The RecBCD enzyme mechanism. Adapted from (Dillingham & Kowalczykowski, 2008).

In the absence of RecD, the RecBC complex is a helicase lacking nuclease activity (Anderson et al., 1997). This activity may be substituted by other cellular nucleases, such as the 5'-3' exonucleases RecJ and ExoVII (Dermić, 2006). Unlike *recB* or *recC* mutants, *recD* mutants are perfectly viable and recombination-proficient (Amundsen et al., 1986; Biek & Cohen, 1986), and were demonstrated to induce RecA loading without the requirement of Chi sequence (Churchill et al., 1999). In the absence of RecJ and ExoVII, *recD* mutants are poorly viable and deficient in recombination (Dermić, 2006; Dermić et al., 2006). RecJ is the main 5'-3' exonuclease, while ExoVII, which is also a 3'-5' exonuclease, is less efficient and provides the backup (Dermić, 2006). RecA mutant was

shown to be recombination-deficient, as it is strictly required for the final step of homologous recombination (Willetts & Mount, 1969).

RecBCD pathway is responsible for the repair of both one-ended DSBs, formed during the replication fork collapse, and two-ended DSBs (Figure 6). In the former case, when the moving replication fork encounters a nick in a template, it results in the formation of a DSB in one of the fork arms, recognized and processed by the RecBCD complex up to the Chi sequence. Upon Chi binding, RecBCD loads RecA on a free 3' end, which induces the invasion into the intact homologous DNA duplex. The Holliday junction is formed which is resolved mainly by the RuvABC complex. After that, PriA mediates the reassembly of the replisome, and the replication fork is restored. Two-ended DSBs are processed similarly, with the difference that two RecBCD complexes are required to process two DBSs. The RecA-bound ssDNA ends invade the homologous sequence in a sister chromosome. The resulting Holliday junctions are cleaved, and PriA mediates the formation of two converging replication forks, which restore the chromosomes (Figure 6).

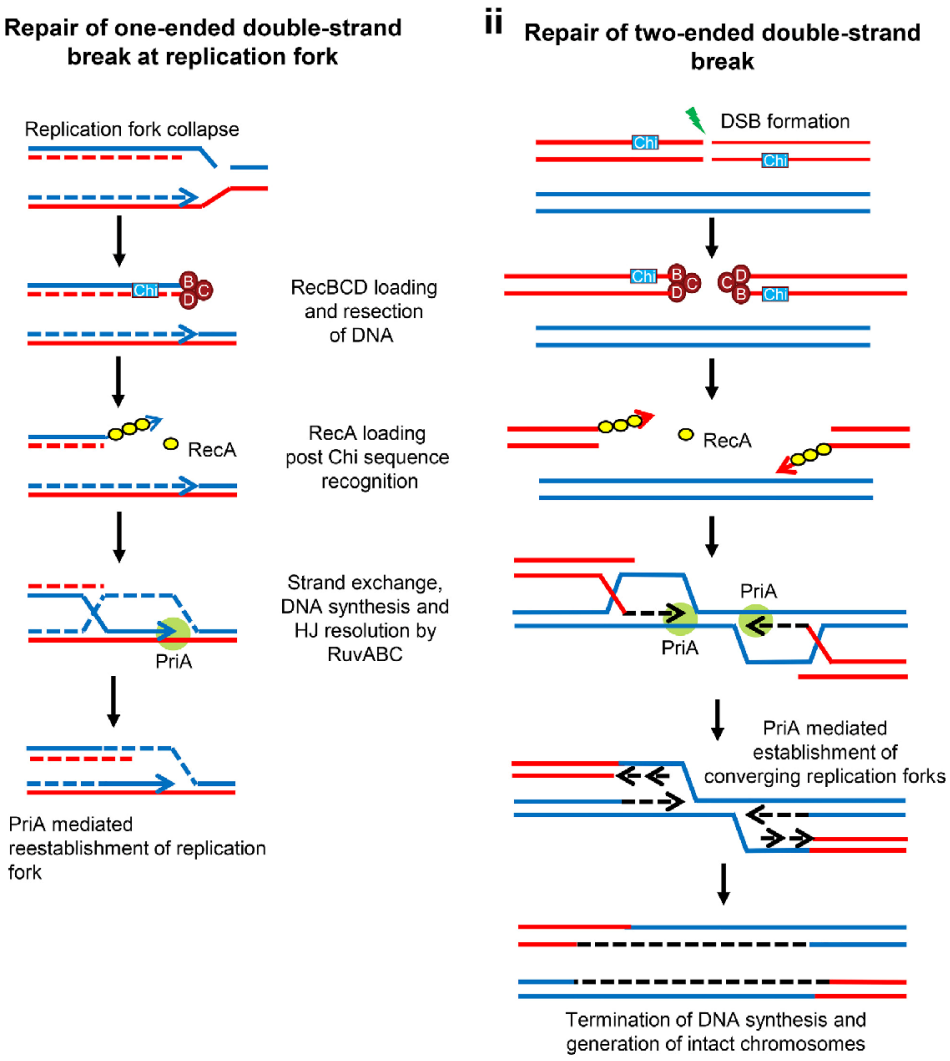


Figure 6. The mechanism of one-ended and two-ended DSBs repair by the RecBCD pathway. Adapted from (Sinha et al., 2020).

The RecBCD ability to degrade DNA is becoming particularly useful in case of infection by dsDNA phages (Behme et al., 1976; Benzinger et al., 1975). Some phages encode RecBCD inhibitors which bind free DNA ends (Appasani et al., 1999; Silverstein & Goldberg, 1976; Williams & Radding, 1981) or mimic dsDNA ends (Murphy, 1991; Silberstein & Cohen, 1987). Chi sites were earlier suggested to function as a mark for distinguishing self from non-self, as their frequency in bacterial DNA is much higher than

random (Dillingham & Kowalczykowski, 2008). However, recently it was demonstrated that Chi sequences appear only ~7 times more frequently than could be expected, and this overrepresentation might be explained by the *E. coli* codon usage and ORF orientation biases (Subramaniam & Smith, 2022). It was demonstrated that the DNA of those phages that lack RecBCD inhibitors are enriched with Chi sequences, thus protecting DNA from excessive degradation (Bobay et al., 2013). Still, the Chi abundance for P1- and T5-like phages is rather explained by the requirement of recombination for the phage packaging rather than for protection from the RecBCD-induced DNA degradation (Bobay et al., 2013; Subramaniam & Smith, 2022).

1.5.2.RecFOR pathway

If the RecBCD pathway is broken, the RecFOR pathway, which is usually responsible for single-strand DNA breaks repair, partially substitutes it. If 5'-3' exonucleases ExoI (SbcB) and SbcCD are inactive, the level of DSB repair is restored to the wild-type level (F. P. Gibson et al., 1992; Lloyd & Buckman, 1985). In this case, the RecQ helicase binds blunt or 3'-overhanging DNA ends and unwinds them, while RecJ binds ssDNA of at least 7 nt long and degrades it in the 5'-3' direction, creating single-stranded 3'-end extensions, which are covered by the single-strand DNA binding protein (SSB) (Han, 2006). The loading of RecA is next induced by the RecFOR complex bound to the junction site of single-stranded and double-stranded DNA. Alternatively, the RecOR complex binds the 3' end of ssDNA extension and facilitates RecA loading. HR then continues via the standard way.

The SbcB protein, also known as Exonuclease I, interacts with SSB (Molineux & Gefter, 1975). By performing highly processive 3' end trimming it participates in gap repair and blunting of DNA ends to allow RecBCD binding (Kushner et al., 1971; Thoms & Wackernagel, 1998). The SbcCD complex recognizes and cleaves hairpin structures formed by palindromic sequences *in vitro* (J. C. Connelly & Leach, 1996) and *in vivo* (Eykelenboom et al., 2008). It was reported to possess dsDNA exonuclease and ssDNA 3'–5' exonuclease activities (J. Connelly, 1999).

1.5.3.A-EJ and MMEJ

As HR may only be helpful in the presence of a complementary DNA template, i.e., during cell division, it cannot operate during stationary phase, dormancy, or sporulation. In eukaryotes, there is a distinct mechanism called NHEJ (non-homologous end joining) which restores DSB without the need for a template. While eukaryotes prefer NHEJ for DSB repair, bacteria mostly rely on HR and possess no or reduced NHEJ mechanisms. The NHEJ pathway in eukaryotes includes several proteins ensuring efficient and accurate repair. Some bacteria, such as *Mycobacterium* or *Bacillus subtilis*, utilize a compact version of NHEJ, consisting of just two enzymes, Ku and ligase D (Shuman & Glickman, 2007). *E. coli* lacks these proteins, though it was discovered to have an alternative end-joining (A-EJ) mechanism, relying on ligase A (Chayot et al., 2010).

Archaeal alternative NHEJ pathway, called microhomology-mediated end joining (MMEJ), utilizes short homology regions in two DSB ends for repair. In *Haloferax volcanii*, Rad50-Mre11 processes the DSB ends, allowing homology regions to anneal (Delmas et al., 2009; Sharma et al., 2015). The flapped DNA is then cleaved by Fen1, and the gaps are filled by the cellular polymerase and ligase activities. MMEJ was shown to be the prevalent DNA repair pathway, though HR may act after MMEJ to restore the lost genetic information (Delmas et al., 2009).

1.6. Interactions of repair enzymes and CRISPR-Cas systems

It is logical to assume that cellular repair enzymes should become activated during CRISPR-Cas system function, as DNA breaks occur during both CRISPR interference and incorporation of new spacers. On the other hand, the observed domain structure similarity of some of Cas proteins with the DNA repair enzymes initially suggested their possible roles in DNA repair (Makarova et al., 2002). Indeed, *cas1* deletion in *E. coli* was demonstrated to affect DNA damage sensitivity and chromosome segregation (Babu et al., 2011). The negative association between Type II CRISPR-Cas systems and NHEJ was observed in bacterial genomes, which can be explained by the inhibition of NHEJ by Csn2

protein (Bernheim et al., 2017). In contrast, the components of the DSB repair system in *Proteobacteria*, namely RecBCD, SbcCD, and SbcB, were found to frequently co-occur with CRISPR-Cas systems (Bernheim et al., 2019).

There is also multiple evidence of the interplay between CRISPR-Cas systems and DNA repair in archaea. For instance, in *Sulfolobus islandicus*, possessing Type I-A CRISPR-Cas system, CRISPR-associated factor Csa3a was demonstrated to simultaneously control the expression of the CRISPR adaptation and DNA repair genes and activate the DNA damage response (DDR) (T. Liu et al., 2015, 2017; Z. Liu et al., 2020). One of the activated DNA repair mechanisms induces the transfer of chromosomal DNA between cells so that it can be used for homologous recombination. Such crosstalk is proposed to help evade self-immunity, as self-derived spacers are acquired frequently (~7%) (T. Liu et al., 2017). In naturally polyploid *Haloferax volcanii*, autoimmunity was shown to be well tolerated and resolved mainly by recombination between two CRISPR arrays flanking *cas* genes (Stachler et al., 2017). Cas1 of the *H.volcanii* Type I-B CRISPR-Cas system was demonstrated to be crucial for cell growth in the oxidative stress conditions (Wörtz et al., 2022). The authors propose Cas1 to act similarly to Fen1 repair protein, which cleaves the flapped DNA intermediates formed during DNA repair and replication. Moreover, in *H.volcanii* the Cas3 translocase/helicase activity was found to participate in DNA repair by restraining HR together with Mre11-Rad50 and promoting MMEJ (Miezner et al., 2023).

For the Type I-E CRISPR-Cas system, it was demonstrated that during naïve CRISPR adaptation spacers are predominantly selected from the regions on chromosomal DNA with replication terminators TerC and TerA, where replication forks are stalled (Neylon et al., 2005), and limited by Chi motifs (Levy et al., 2015). Another hotspot was found to be located near the CRISPR array. ssDNA gaps in both strands are formed during spacer incorporation, which should result in DSB formation when faced by the replication fork. Furthermore, introducing a dsDNA break by I-SceI endonuclease resulted in another CRISPR adaptation hotspot, which was also limited by Chi (Levy et al., 2015). Thus, substrates for naïve CRISPR adaptation were suggested to be derived from replication and

concomitant DNA repair (Figure 7) (Levy et al., 2015). Since RecBCD is the main enzyme restoring DSBs, the products of DNA degradation by RecBCD could serve as a source of spacers for naïve CRISPR adaptation. In agreement with this model, *recBCD* mutations were shown to disrupt naïve CRISPR adaptation, and the expression of the T7 RecBCD inhibitor gp5.9 abolishes naïve CRISPR adaptation (Ivančić-Baće et al., 2015; Levy et al., 2015; Radovčić et al., 2018). Moreover, Cas1 was shown to physically interact with RecBC and RuvB (Babu et al., 2011), (Levy et al., 2015).

In the absence of RecD, RecBC possesses helicase activity but lacks nuclease activity (Palas & Kushner, 1990). Interestingly, naïve CRISPR adaptation is reduced in *recD* mutant, but fully restored in the *recD recA* mutant (Radovčić et al., 2018). RecBC is able to load RecA in the absence of RecD without Chi requirement (Churchill et al., 1999). Therefore, the formation of RecA filament likely inhibits CRISPR adaptation, and the nuclease activity of the RecBCD complex is not required for naïve CRISPR adaptation (Radovčić et al., 2018). Furthermore, more than 80% of new spacers are associated with a canonical AAG PAM in *recD recA* mutant (Mitić et al., 2023), in contrast to only 50% in the wild type. It suggests that when RecBCD and RecA are present, the available ssDNA substrates are present for a short time, which is not enough for Cas1-Cas2 to select and bind proper PAM-containing fragments (Mitić et al., 2023). Interestingly, cells with the mutant RecB^{1080A}CD complex, which lacks nuclease but retains the full helicase activity, were shown to have the wild-type level of CRISPR adaptation (Radovčić et al., 2018). Thus, though the reduced level of naïve CRISPR adaptation in *recD* mutant initially suggested RecBCD to produce substrates for naïve CRISPR adaptation by its nuclease activity, it is rather explained by the decreased processivity of RecBC helicase, compared to the full RecBCD enzyme (Biek & Cohen, 1986; Masterson et al., 1992; Palas & Kushner, 1990; Radovčić et al., 2018).

A higher preference for foreign DNA during protospacer selection during naïve adaptation could be explained by a higher number of replication forks on plasmids or phages and a higher abundance of Chi in chromosomal DNA (Levy et al., 2015). Moreover, free unprotected ends of a linear phage DNA can be recognized and degraded by RecBCD

straightaway upon entering the cell (Behme et al., 1976; Benzinger et al., 1975), thus ensuring the higher preference for CRISPR adaptation machinery for phage DNA. Nevertheless, the high abundance of Chi sequences in bacterial genome and their strong orientation preference in chromosomal DNA are explained by the *E. coli* codon usage and co-directed replication and translation processes, rather than by the recombination-driven selection (Subramaniam & Smith, 2022).

While RecBC activity is crucial for naïve CRISPR adaptation, it was found to be inessential for primed CRISPR adaptation. Thus, the production of CRISPR adaptation precursors during priming seems to mostly rely on fragments produced by Cas3 (and possibly other nucleases) during PAC translocation. RuvC and RecA were demonstrated to be inessential for both primed and naïve CRISPR adaptation, but *recG* and *priA* deletions abolish primed adaptation in a phage lambda infection model (Ivančić-Baće et al., 2015). It is proposed that the replication fork is blocked by the Cascade-crRNA complex bound to the target, followed by the recruitment of RecG and PriA. PriA binds to the free 3' end of the leading strand, while RecG remodels the blocked fork and the Cascade R-loop, allowing Cas1-Cas2 to bind its substrate (Figure 7) (Ivančić-Baće et al., 2015).

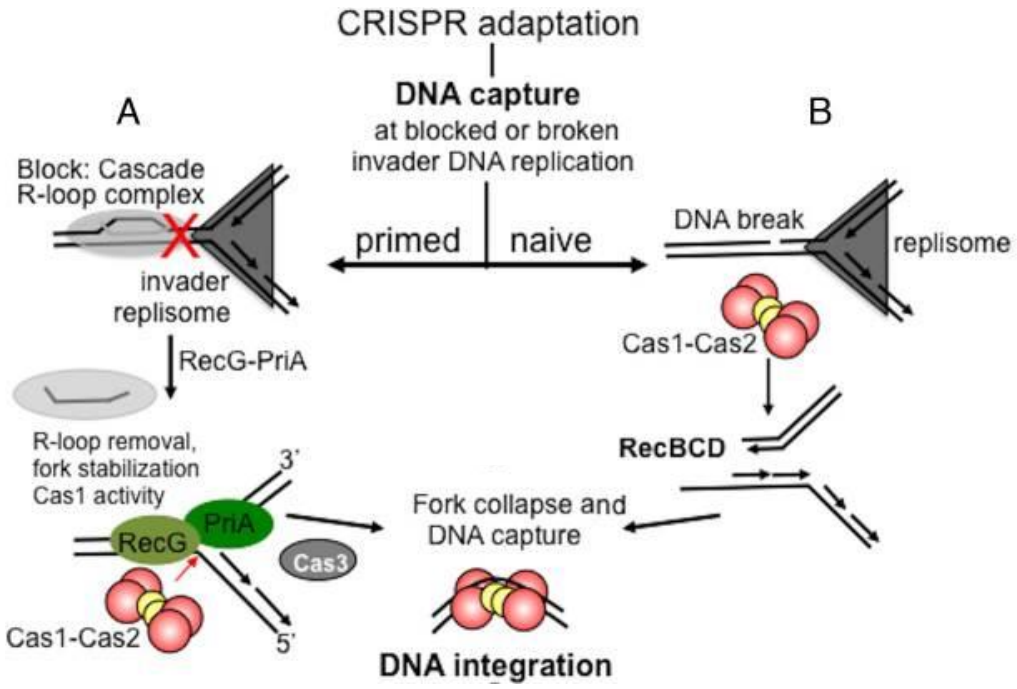


Figure 7. The proposed model illustrating the roles of the cellular DNA repair proteins in primed and naïve CRISPR adaptation. (A) The replication fork is blocked upon facing the R-loop formed by the Cascade-crRNA complex. The blocked fork is recognized by RecG and PriA. RecG helicase activity remodels the fork and removes the R-loop. Further DNA degradation provides Cas1-Cas2 with substrates for CRISPR adaptation. (B) In the absence of priming, DSBs formed during DNA replication are repaired by RecBCD, which, together with other cellular nucleases, produces substrates for capture by Cas1-Cas2. The figure was adapted from (Ivančić-Baće et al., 2015).

In the *Streptococcus thermophilus* CRISPR-Cas system, Cas2 has an additional DnaQ domain, which was shown to participate in the 3' end trimming of prespacers (Drabavicius et al., 2018). *In vitro*, DnaQ-like 3'-5' exonucleases ExoT and PolIII trim adaptation precursors to correct size (S. Kim et al., 2020), though whether they participate in prespacer trimming *in vivo* is not known.

1.7. Type I-F CRISPR-Cas system of *Pseudomonas aeruginosa*

1.7.1. CRISPR-Cas system of Pseudomonas aeruginosa PA14

One of the model organisms for studying the acquisition of CRISPR-Cas immunity is *P. aeruginosa* PA14, which carries a type I-F CRISPR-Cas system. It has two divergently oriented CRISPR arrays, CRISPR1 and CRISPR2, flanking *cas* genes and containing 13 and 21 32nt spacers, respectively (Figure 8). CRISPR2 has a spacer partially matching the phage DMS3vir genome (Cady et al., 2012). *P. aeruginosa* PA14 CRISPR-Cas system is active under laboratory conditions (Cady & O'Toole, 2011) and was demonstrated to efficiently acquire new spacers during DMS3vir infection due to priming (Cady et al., 2012).

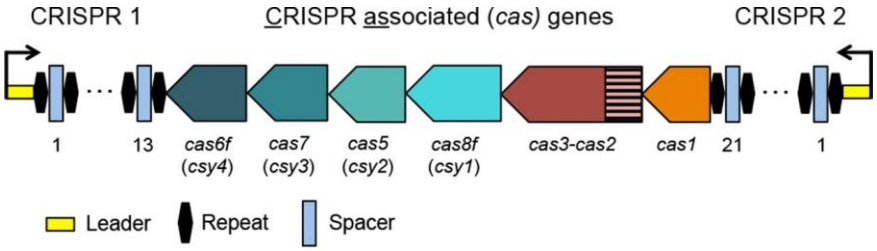


Figure 8. The Type I-F CRISPR-Cas system of *Pseudomonas aeruginosa* PA14.

The Cas3 and Cas2 proteins in the Type I-F system are fused forming a single Cas3-Cas2 protein, which interacts with Cas1 and is required for CRISPR adaptation (Richter et al., 2012). The structure of the *P. aeruginosa* effector Cascade complex is highly similar to that of *E. coli*. The effector is composed of Csy1-4 proteins and has a $Csy1_1$:$Csy2_1$:$Csy3_6$:$Csy4_1$ stoichiometry. It has a molecular weight of 350 kDa, which is 55 kDa less than the *E. coli* Cascade, and lacks a CasA-like tail that is involved in non-specific interactions with DNA (Jore et al., 2011; Wiedenheft, Van Duijn, et al., 2011). Csy4 is responsible for the binding of pre-crRNA and its cleavage into crRNAs (Cheng et al., 2020). crRNA of the type I-F system has a 3' terminal phosphate, unlike crRNA generated by CasE of the *E. coli* system, which has 2',3'-cyclic phosphate (Wiedenheft,

Van Duijn, et al., 2011). The first 8 nt on the crRNA 5' end constitute a seed sequence, that must be complementary to target DNA for efficient recognition (Wiedenheft, Van Duijn, et al., 2011).

The consensus PAM supporting CRISPR interference in the Type I-F system is GG (Mojica et al., 2009; Richter et al., 2014). According to the current model, upon target recognition by the Cascade, Cas3-Cas2 is recruited, supporting priming through interaction with Cas1 (Richter et al., 2012). The complex translocates on an NT-strand in the 3'-5' direction, unwinding the DNA and selecting fragments for CRISPR adaptation.

Under laboratory conditions, *P. aeruginosa* acquires CRISPR-Cas immunity against DMS3vir in a minimal M9 medium but preferentially evolves surface modifications (loss of phage receptor) in a rich LB medium (Westra et al., 2015). The density of parasites is proposed to determine the preferred type of defense, with high phage density favoring receptor mutations and low density favoring resistance due to CRISPR-Cas (Westra et al., 2015).

1.7.2. Clinical relevance of P. aeruginosa

P. aeruginosa is an important opportunistic human pathogen classified as a priority one pathogen by the World Health Organization. Infections are commonly treated with antibiotics, despite the emergence of multidrug-resistant strains that are selected during extended antibiotic treatment, for example, in the lungs of people with cystic fibrosis (Langendonk et al., 2021). Over the past few years, interest in phage-antibiotic combination therapies has increased in the hope that phage and antibiotics can act in concert to control infections (Kortright et al., 2019; Segall et al., 2019; Tagliaferri et al., 2019; Torres-Barceló & Hochberg, 2016). However, around 36% of clinical *P. aeruginosa* isolates were found to possess functional CRISPR-Cas systems, 30% of Type I-F, and 6% of Type I-E (Cady et al., 2011). *P. aeruginosa* clones that acquire CRISPR-Cas immunity can escape these trade-offs and retain virulence (Alseth et al., 2019). Antibiotics can also affect the evolution of phage resistance (Torres-Barceló et al., 2018), for example, because bacterial mutation rates can be enhanced in response to antibiotics exposure (Baharoglu &

Mazel, 2011; Kohanski et al., 2010), thus increasing the probability that they acquire mutations in phage receptor genes. *P. aeruginosa* PA14 rapidly evolves CRISPR-Cas immunity against DM3vir cultured in nutrient-limited media (van Houte et al., 2016; Westra et al., 2015) but mostly evolves phage resistance through surface modification (SM) (loss or mutation of the type IV pilus, which is the DMS3vir receptor) in nutrient-rich broth and in artificial sputum medium that mimics the cystic fibrosis lung environment it commonly colonizes (Alseth et al., 2019; Westra et al., 2015). Unlike CRISPR-Cas immunity, mutation of the phage receptor leads to attenuated virulence (Alseth et al., 2019), and it is therefore key to predict and manipulate which mechanism bacteria use to become phage resistant under clinically relevant conditions.

Chapter 2. Project objectives

The detailed mechanism of conversion of DNA in a locus targeted by Cascade-crRNA and Cas3 to spacers in CRISPR array during primed CRISPR adaptation by Type I CRISPR-Cas systems remains to be identified. It is assumed that fragments formed during Cas3-mediated DNA degradation are feeding primed CRISPR adaptation machinery. Other cellular proteins must also participate in this process as different genome maintenance systems components were earlier shown to modulate CRISPR adaptation (Ivančić-Baće et al., 2015; Levy et al., 2015). The proposed model of primed CRISPR adaptation in *E. coli* is presented in Figure 9.

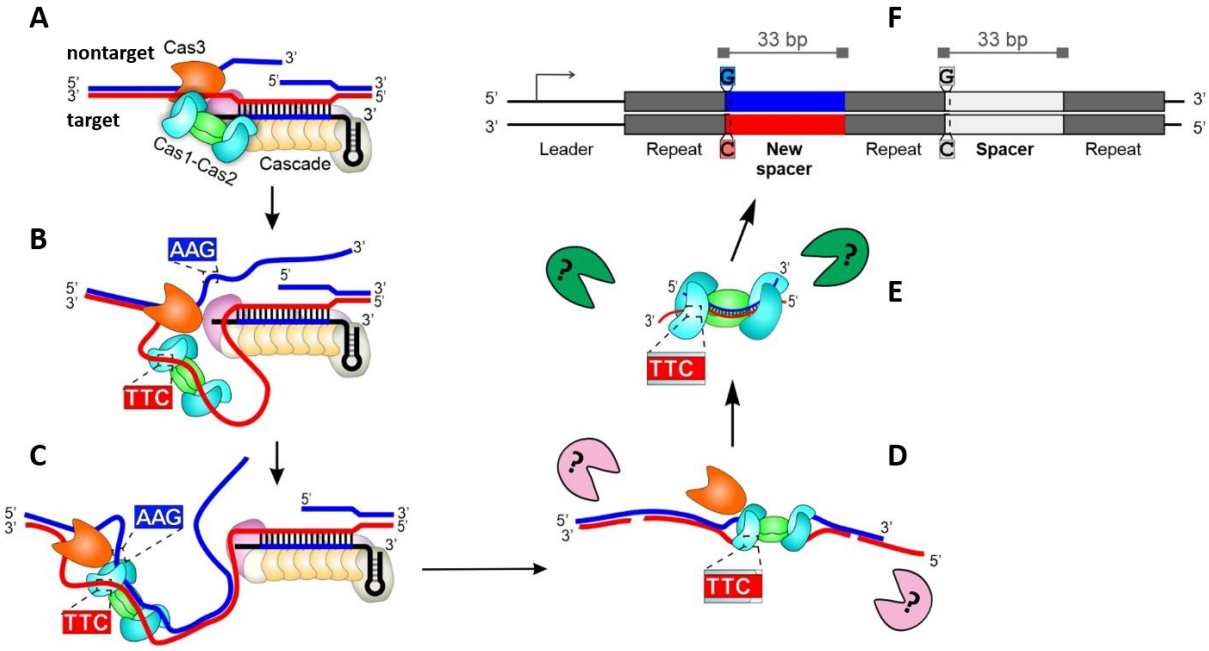


Figure 9. The model of primed CRISPR adaptation in *E. coli*. (A) PPS is recognized in the target strand by the Cascade-crRNA complex, primed acquisition complex (PAC) is assembled, and the nontarget strand is cleaved by Cas3. (B) PAC translocates, separating the DNA strands. Cas1 recognizes the PAM-complementary sequence, and Cas1-Cas2 binds the adjacent single-stranded protospacer. (C) Cas1-Cas2 anneals single-stranded TTC-associated protospacer to the complementary AAG-associated strand. (D) The prespacer precursors formed as a result of DNA degradation by Cas3 and other cellular nucleases are bound

by Cas1-Cas2. (E) Prespacer precursors are trimmed by the unknown nucleases to form prespacers. (F) Prespacers are incorporated into the CRISPR array and become new spacers.

This project aimed to study the role of different cellular genome maintenance proteins in primed CRISPR adaptation by the Type I-E CRISPR-Cas system *in vivo*.

The following objectives were set in this work:

1. Check the requirements of different recombination/repair pathways for CRISPR adaptation *in vivo*
2. Study the impact of genome maintenance proteins on adaptation efficiency
3. Reveal the effects of different genome maintenance proteins on spacer choice
4. Investigate the impact of the cellular DNA repair proteins on the generation of spacer precursors

In parallel to the main work, an additional project was done, focusing on primed CRISPR adaptation by the Type I-F CRISPR-Cas system in *Pseudomonas aeruginosa* PA14. *P. aeruginosa* is a human pathogen of high clinical relevance, which causes severe infections, commonly treated with antibiotics. The problem of multidrug resistance raised the interest in phage-antibiotic combination therapies. However, around 36% of clinical *P. aeruginosa* isolates were found to possess functional CRISPR-Cas systems. The project aim was to investigate the effects of antibiotics on the CRISPR-Cas immunity development during phage infection. The objectives were:

1. Conduct evolution experiments with *P. aeruginosa* PA14 infected by the DMS3vir phage in both rich and minimal media in the presence of different antibiotics
2. Study the impact of antibiotics on the CRISPR-Cas immunity development
3. Check the effects of antibiotics on the fitness of the resistant clones and phage production

Chapter 3. Materials and methods

3.1. Bacterial strains

The KD403 strain was described in (Shiriaeva et al., 2019). P1 transduction was used to replace *recB, recC, recD, sbcB, sbcD, recJ, xseA,* and *recQ* genes in the strain KD403 with the cassette encoding the gene of kanamycin resistance as described previously (Moore, 2011). BL21-AI strains with deletions Δ*recB* Δ*recJ* and Δ*recJ* Δ*xseA* were constructed by the same method. Keio collection strains (Baba et al., 2006) were used as donor strains. To obtain double mutants, the kanamycin cassette was flipped out from single knockouts using FLP recombinase expressed from pCP20 plasmid as described (Datsenko & Wanner, 2000), and then the second round of P1 transduction was applied to introduce the second deletion. The pCas1-2 plasmid for the expression of *cas1* and *cas2* genes in prespacer efficiency assay was described earlier (Yosef et al., 2012). The deletions in obtained colonies were confirmed by PCR with primers annealing to the gene of kanamycin resistance and those annealing to the regions flanking the mutation site. The genotypes of mutant KD403 derivatives, except those containing Δ*xseA* and Δ*recQ* mutations, were additionally verified by sequencing of genomic DNA with Illumina MiniSeq and analyzing the mutations by the breseq pipeline (Deatherage & Barrick, 2014).

3.2. The analysis of acquired spacers during primed CRISPR adaptation

3.2.1.Primed adaptation assay

Cells were grown overnight at 37 °C in Luria-Bertani (LB) broth. Aliquots of the cultures were diluted 100-fold into fresh 20 mL of LB broth and grown at 30 °C. After reaching OD600 0.3, the cultures were divided into two aliquots (7 mL each), and IPTG (isopropyl β-D-1 thiogalactopyranoside) and L-arabinose were added to one of two aliquots at the final concentration 1 mM each. The cultures were grown at 30 °C for 5 h, washed with 10 mL of 1x PBS buffer, resuspended in 700 μL of 1× PBS buffer, and stored at −20 °C. Genomic DNA was purified from cultures by phenol-chloroform extraction as described in (Shiriaeva et al., 2019). CRISPR expansion was monitored by PCR with

primers annealing to Spy^{ihN} spacer (auto-Spy^{ihN}-R, 5′- aatagcgaacaacaaggtcggttg-3′) and the leader of the CRISPR array (LDR-F2, 5′- atgctttaagaacaaatgtatacttttag-3′). ImageJ (Rasband, 2012) was used for quantitative analysis of adaptation efficiency. To measure the relative intensities of the bands on agarose gel corresponding to extended CRISPR arrays, the intensities in rectangles that cover corresponding areas were divided by the sum of intensities of the same size areas of upper (extended) and lower (unextended) bands and normalized to the values calculated for samples without induction. Three independent experiments were done for each strain. Pairwise t-test with Holm correction for multiple comparisons was used for statistical analysis of differences in adaptation efficiency between different strains.

For extraction of CRISPR arrays expanded during CRISPR adaptation three independent 100 μL amplification reactions containing 20–50 ng genomic DNA were pooled, PCR products corresponding to expanded CRISPR array were gel purified using GeneJET Gel Extraction Kit (Thermo Fisher Scientific, Vilnius, Lithuania) and sequenced in pair-end mode with MiniSeq Illumina System at Skoltech Genomics Core facility. Two independent experiments were analyzed by Illumina sequencing for each strain.

3.2.2.Data processing

Raw sequencing data were processed using ShortRead and BioStrings R packages (Morgan et al., 2009; Pagès et al., 2008) using custom scripts. Illumina reads were trimmed with a quality score cutoff = 20. Paired reads were merged with the following parameters: gap opening = −10, gap extension = −4. Reads containing two or more repeats (with up to two mismatches) were selected. The 33-bp segments were considered spacers and were mapped to the KD403 genome sequence with no mismatches and only unique mapping allowed. A minor fraction of non-uniquely mapped spacers (i.e., those originating from rRNA operons) was excluded from the analysis. Spacers that have been inaccurately incorporated, i.e., those shifted by a few nucleotides with respect to the consensus PAM or inserted in an opposite orientation (Shmakov et al., 2014) were also discarded from the analysis except when inaccurately incorporated spacers were specifically analyzed (see

below). Custom R scripts used for spacer analysis are available upon request. Spacers were counted in bins of 10 kb separately for target and nontarget strands and normalized to the total spacer counts. For plotting spacer distribution, we averaged results across two independent experiments, also, to directly compare the curves, maximal values for mutants and wild-type were adjusted to make them the same. A portion of spacers associated with AAG PAM as well as a portion of spacers with internal AAG trinucleotide was calculated for 400-kb region spanning the PPS area in bins of 10 kb separately for nontarget and target strands. Spacers that mapped next to protospacers associated with AAG PAM but shifted up to 2 bp up- or downstream, and/or inserted in an opposite orientation, were considered as incorrectly incorporated. Their numbers were also calculated in 10-kb bins in the 400-kb PPS-centered area. A pairwise proportional test that accounts for multiple testing at p-value of 0.001 was applied to compare the values obtained for mutant strains with corresponding values for the wild-type strain.

3.3. Microscopy

Cells of cultures grown with or without induction of *cas* gene expression were analyzed using a LIVE/DEAD bacterial viability kit (Thermo Scientific) at 5 hr after induction in cell chambers made as described in (Shiriaeva et al., 2019). Fluorescence microscopy was performed using Nikon Eclipse Ti-E inverted microscope. Fluorescence signals in green (living cells) and red (dead cells) fluorescent channels were detected using Semrock filter sets YFP-2427B and TxRed-4040C, respectively, and the numbers of corresponding cells were calculated. The image analysis was performed using ImageJ (Fiji) with ObjectJ plugin used for measurements of cell length.

3.4. The analysis of CRISPR adaptation efficiency by high-throughput sequencing

3.4.1.Sample preparation

Cell lysates were prepared by concentrating cells 20× in Milli-Q water and heating at 95°C for 5 min. Cell debris was removed from lysates by centrifugation at 16g for 1 min. Amplified mixtures of extended and nonextended CRISPR arrays were purified using the

GeneJET PCR Purification Kit. Sequencing libraries were prepared using NEBNext Ultra II Library Prep Kit (New England Biolabs) and sequenced on the MiniSeq, HiSeq4000, or NovaSeq 6000 Illumina platforms in 2 × 150 bp paired-end read mode. The sequencing coverage was planned based on a visual assessment of spacer adaptation efficiency using agarose gel electrophoresis so that at least a thousand reads corresponding to expanded CRISPR arrays were expected in the sequencing data.

3.4.2.Data analysis

Bioinformatic analysis was performed in RStudio using ShortRead and Biostrings packages (Morgan et al., 2009; Pagès et al., 2008). Reads containing more than 10% of positions with Phred quality score <20 were discarded, and adapter sequences were trimmed from read 3′ ends. Plots were generated using ggplot2 (Hadley Wickham, 2016). Sequence logos were generated using the ggseqlogo package (Wagih, 2017). Forward and reverse reads overlapping by at least 20 nt were merged. The CRISPR repeat sequence was searched with two mismatches allowed. For estimation of primed adaptation efficiency, the number of merged reads with 1, 2, 3, or 4 CRISPR repeats corresponding to arrays with 0, 1, 2, or 3 newly acquired spacers was calculated. Primed adaptation efficiency was calculated as the total number of newly acquired spacers divided by the total number of CRISPR arrays.

3.5. Prespacer efficiency analysis

3.5.1.Prespacer efficiency assay

The assay was performed as described earlier (Shipman et al., 2016; Shiriaeva et al., 2019) except for the polymerase chain reaction (PCR) stage that was performed with primers containing 5′-terminal extensions corresponding to NEBNext Illumina adapter sequences. The libraries were purified, indexed, and sequenced on the MiniSeq or HiSeq4000 Illumina platforms in 2 × 150 bp paired-end read mode as described above. Oligonucleotides used for electroporation are listed in Table 1.

Table 1: Oligonucleotides used for prespacer efficiency assay.

#	Transforming oligo names	Transforming oligo sequences
1.	33/37F	5’P-GCCCAATTTACCGCGCGATCGGGTGTTTGGTGA-3’
	33/37R	3’-GGTTCGGGTTAAATGGCGCGCTAGCCCACAAACCACT-P5’
2.	32/36F	5’P-CCCAATTTACCGCGCGATCGGGTGTTTGGTGA-3’
	32/36R	3’-GGTTCGGGTTAAATGGCGCGCTAGCCCACAAACCAC-P5’
3.	34/38F	5’P-AGCCCAATTTACCGCGCGATCGGGTGTTTGGTGA-3’
	34/38R	3’-GGTTCGGGTTAAATGGCGCGCTAGCCCACAAACCACTG-P5’

The sequencing coverage was planned based on a visual assessment of spacer adaptation efficiency using agarose gel electrophoresis so that at least a thousand reads corresponding to expanded CRISPR arrays were expected in the sequencing data.

3.5.2.Data analysis

Forward and reverse reads overlapping by at least 20 nt were merged, and CRISPR repeat sequences were searched with two mismatches allowed. Newly acquired spacer sequences corresponding to the top spacer strand were extracted from CRISPR arrays containing at least two repeats. Newly acquired spacers were mapped with three mismatches allowed to the Bl21-AI genome, the pCas1-2 plasmid, and the sequences of oligonucleotides used for electroporation. The total number of CRISPR arrays, the total number of oligo-derived spacers, and the number of properly processed oligo-derived spacers were calculated. We define properly processed oligo-derived spacers as 33-bp oligo-derived spacers that start with the PAM-derived G. Oligo acquisition efficiency was calculated as the number of properly processed oligo-derived spacers per CRISPR array (%) further adjusted to remove the batch effect caused by replicating experiments on different days. In all cases, the batch-group design was balanced, i.e., the study groups were equally represented in all batches. Therefore, we removed the batch effect by subtracting the mean in each batch from all measurements in that batch resulting in zero-centered values (Nygaard et al., 2016).

3.6. FragSeq analysis

3.6.1.Sample preparation for FragSeq

Total genomic DNA was isolated by phenol-chloroform extraction, and fragments smaller than ~700 nt were purified using the Select-a-Size DNA Clean & Concentrator kit (Zymo Research) as described previously (Shiriaeva et al., 2019). FragSeq libraries were prepared using the Accel-NGS 1S Plus DNA Library Kit (Swift Biosciences) with modifications to the standard protocol recommended by the manufacturer to retain small fragments. The libraries were sequenced on NextSeq 500, HiSeq4000, or MiniSeq Illumina platforms in 2×75 or 2×150 bp paired-end modes. The sequencing coverage needed was determined based on the results of library size evaluation such that at least a few thousand 31- to 40-nt reads were expected. Library size evaluation was performed using the Bioanalyzer 2100 system.

3.6.2.FragSeq data analysis

Reads were mapped to the KD403 reference genome using a custom-written R script. During the first stage of library preparation following the protocol of the Accel-NGS 1S Plus DNA Library Kit (Swift Biosciences), a low complexity tail with an average length of 8 bases (and up to ~12 bases) mostly composed of C and T nucleotides is ligated to 3′ ends of fragments. To account for these tails, the last 15 nt were removed from 3′ ends of forward reads followed by forward read mapping with several mismatches allowed (up to 5% of read length). The 5′ ends of mapped forward reads correspond to the fragments' 5' ends. The first 20 nt of reverse reads, which include the tails and the 3′-end nucleotides of fragments, were mapped without mismatches allowed. If a read was not aligned or was not properly oriented relatively to the corresponding forward read, 1 nt of a presumptive tail was removed from the 5′ end of the reverse read and the first 20 nt of the trimmed reverse read were mapped again. The cycles of reverse read alignment and trimming by 1 nt were repeated until the appropriate hit was found but not more than 15 times. The positions of 5′ ends of mapped reverse reads were regarded as positions of the fragments' 3' ends.

3.7. *in vitro* methods

3.7.1.Protein purification

Cas1 and Cas2 were purified as described previously (Ramachandran et al., 2020). RecBCD and RecBC were purified as described previously (Taylor & Smith, 2003).

3.7.2.Exonuclease footprinting

For exonuclease footprinting, unlabeled and 3′-fluorescein-labeled DNA oligonucleotides purchased from IDT were used to assemble prespacer precursor substrates. Exonuclease RecJ and ExoVII reactions were carried out on double-forked DNA substrates containing a central 23-bp duplex flanked with 5′ and 3′ overhangs (Figure 26). For ExoV (RecBCD) nuclease footprinting, a double-forked DNA substrate was supplemented with a complementary (to 5′ overhang) strand to provide a blunt end configuration of the substrate (Figure 27). Exonuclease reactions were performed with 0.2 μM DNA substrate bound by the 1.6 μM Cas1-Cas2 complex, in 10 μl of binding buffer [40 mM tris-HCl (pH 7.5), 50 mM NaCl, 10 mM $MgCl_2$, 0.5 mM tris(2-Carboxyethyl)phosphine (TCEP), and bovine serum albumin (BSA; 0.1 mg/ml)] supplemented with 1 mM adenosine triphosphate (ATP) when RecBCD was used. Before these reactions, the Cas1-Cas2 complex was reconstituted in the binding buffer (without magnesium) by incubating Cas1 and Cas2 subunits at the molar ratio of 2:1 on ice for at least 20 min. This mixture was diluted 10-fold into the binding reaction with DNA substrates and incubated at room temperature for 20 min. Exonuclease reactions were initiated by the addition of 30 U of RecJ or 10 U of ExoVII (both from NEB), or RecBC or RecBCD (final concentration of 50 nM) to 10 μl of binding reaction and incubated at 37°C for 30 min. The reactions were terminated by the addition of 15 μl of formamide loading buffer containing heparin (0.4 mg/ml) and heated in a boiling water bath for 1 min. The digested DNA fragments were resolved by urea-denaturing 8% polyacrylamide gel electrophoresis (PAGE). The gel was screened and imaged with an Amersham Typhoon scanner.

3.7.3.Electrophoretic mobility shift assay

DNA duplexes were formed by mixing 5′-[^{32}P]–labeled strands and unlabeled strands at a molar ratio of 1:2 in Nuclease-Free Duplex Buffer (IDT), heating for 2 min at 95°C, and slowly cooling to 20°C. The Cas1-Cas2 complex was reconstituted as described in the previous section. The binding reaction was carried out in 10 µl of binding buffer [40 mM tris-HCl (pH 7.5), 50 mM NaCl, 10 mM $MgCl_2$, 0.5 mM TCEP, and BSA (0.1 mg/ml)] containing 25 nM DNA substrate and 2 µM Cas1-Cas2 complex. Binding reactions were incubated for 30 min at 37°C, loaded onto a native precast 4 to 20% gradient polyacrylamide gel (Novex, Invitrogen), and run in tris-glycine buffer at 20 mA for 2 hours. The bound complexes were visualized with Typhoon phosphorImager and calculated by ImageQuant software (GE Healthcare).

3.7.4.Permanganate probing

$KMnO_4$ probing was performed with 10 nM 5′-[^{32}P]–labeled DNA substrate and 800 nM Cas1-Cas2 complex in 10 µl of binding buffer [40 mM tris-HCl (pH 7.5), 50 NaCl, 10 mM $MgCl_2$, 0.5 mM TCEP, and BSA (0.1 mg/ml)]. The binding reaction was incubated at 37°C for 30 min, and the probing reaction was initiated by adding $KMnO_4$ to a final concentration of 2.5 mM. Reactions were incubated for 10 s and quenched by the addition of 10 µl of 1% 2-mercaptoethanol. The reactions were extracted with a phenol-chloroform mixture, followed by ethanol precipitation. DNA pellets were dissolved in 50 µl of freshly prepared 1 M piperidine and heated at 95°C for 10 min. Piperidine was removed by chloroform extraction, and DNA was ethanol-precipitated. Pellets were dissolved in 10 to 15 µl of formamide loading buffer, and products were separated by denaturing 11% PAGE and visualized with Typhoon phosphorImager (GE Healthcare).

3.8. Experiments with *Pseudomonas aeruginosa* PA14

3.8.1.Bacterial strains

Evolution experiments used *P. aeruginosa* UCBPP-PA14 (PA14) (Cady et al., 2012). UCBPP-PA14 *csy3::lacZ* was used for phage stock amplification and phage titre

determination. Competition experiments used a surface mutant (SM) derived from PA14 *csy3::lacZ* and a CRISPR-resistant mutant (BIM-2sp, bacteriophage insensitive mutant with 2 additional acquired spacers against DMS3vir) derived from PA14, both of which have been previously described (Westra et al., 2015). For the microfluidics experiments, we used PA14 *flgK::Tn5B30(Tc^R)* (O'Toole & Kolter, 1998).

All bacterial strains were grown at 37 °C in LB broth or M9 medium (22 mM Na_2HPO_4; 22 mM KH_2PO_4; 8.6 mM NaCl; 20 mM NH_4Cl; 1 mM $MgSO_4$; and 0.1 mM $CaCl_2$). All liquid cultures were grown with 180 rpm shaking.

3.8.2.Phages

Evolution experiments used lytic phage DMS3vir (Cady et al., 2012). DMS3vir and a mutant expressing anti-CRISPR against PA14 I-F system, DMS3vir-AcrIF1, were used for determination of the resistance phenotypes (van Houte et al., 2016). Phage stocks were obtained from lysates prepared on PA14 *csy3::lacZ* and stored at 4 °C.

3.8.3. *Determination of antibiotic activity*

For MIC (minimum inhibitory concentration) determination, overnight cultures (~$5*10^9$ cells/mL) were diluted 10^4-fold in LB medium. 20 µL of the diluted cultures were inoculated into 96-well microplate wells containing 180 µL of LB supplemented with antibiotics using 2-fold serial dilutions of the antibiotic. After 18 h growth at 37 °C, MIC was determined as the lowest antibiotic concentration with no visible growth. To determine the MBC (minimal bactericidal concentration), the content of wells with no visible growth was plated on LB-agar and further incubated overnight. MBC was defined as the lowest antibiotic concentration resulting in 99.9% decrease in initial inoculum cell density (< 5 CFU in 100 µL). MBC/MIC ratio was used to estimate if antibiotic activity was bacteriostatic or bactericidal: a high MBC/MIC ratio indicates that the concentration sufficient to prevent growth is much lower than the concentration required to kill the majority of cells (Pankey & Sabath, 2004). In our assay, antibiotics with average MBC/MIC ratio >1 were the ones that are commonly recognized as being bacteriostatic (Tm, Erm, Chl and Tc).

3.8.4.Evolution experiments

Evolution experiments were performed in glass vials containing 6 mL growth medium and appropriate antibiotics at the concentrations shown in Table 2.

Table 2: Antibiotics used in the study.

abbreviation	antibiotic	target	MIC	Dose	MBC/MIC	static/cidal
Chl	chloramphenicol	translation	30	25	2.67	static
Tc	tetracycline	translation	10	2.5	2.5	static
Erm	erythromycin	translation	100	5	3.67	static
Tm	trimethoprim	THF synthesis	6.25	2.5	8	static
Gen	gentamycin	translation	1.25	0.625	1.17	cidal
Strep	streptomycin	translation	25	12.5	1.33	cidal
Carb	carbenicillin	peptidoglycan	25	2.5	1.17	cidal
Cip	ciprofloxacin	DNA gyrase	0.2	0.1	1.33	cidal

Antibiotic concentrations were chosen which were below the MIC and did not affect cell densities after 24h growth too drastically (more than 10-fold) in the absence of phages. 60 µL from overnight cultures were co-inoculated with 10^4 plaque-forming units (p.f.u.) of phage DMS3vir. 1:100 volume was then transferred every 24 h into fresh medium for 3 days. Each treatment contained 6 biological replicates. Cell densities and phage titers were monitored with serial dilution in M9 salts (after chloroform treatment for phages), and enumeration of colonies on LB-agar and enumeration of plaques on a lawn of PA14 *csy3::lacZ* cells. The identification of phage resistance type (sensitive, CRISPR-Cas or SM) was performed by cross-streaking 24 randomly selected colonies on DMS3vir and DMS3vir-AcrIF1 phages: SM clones are resistant to both phages and have a characteristic smooth colony morphology, whereas clones with CRISPR-Cas immunity are resistant to DMS3vir but sensitive to DMS3vir-AcrIF1 (van Houte et al., 2016).

3.8.5.Determination of bacterial growth rate by optical density

Overnight cultures were diluted 100-fold into fresh growth media. Growth of 200 μL of culture was measured in a 96-well plate by measuring optical density at λ=600nm (OD600) for 14 to 24 h at 37 °C in a BioTek Synergy 2 Plate reader, with 5 s shaking before each measurement. All growth curves were performed in at least 8 replicates. The exponential growth rate in LB was determined in R using the package growthrates (Petzoldt, 2022). The growth rate was then calculated from cultures inoculated without phage, between t_{max} -3 h and t_{max}.

3.8.6.Determination of bacterial doubling time by microfluidics

The mother machine device was fabricated and handled as previously reported (Bamford et al., 2017; Cama et al., 2020). Briefly, overnight cultures in LB were spun down via centrifugation for 5 minutes at 4000 rpm at room temperature (Eppendorf 5810 R). The supernatant was filtered twice (Medical Millex-GS Filter, 0.22 μm, Millipore Corp.) and used to re-suspend the bacteria to an OD600 of 75. 2 μl of the bacterial suspension was injected into the microfluidic mother machine device and incubated at 37 °C until there were 1-2 bacteria in the lateral side channels. Fluorinated ethylene propylene tubing (1/32” × 0.008”) was connected to the inlet and outlet holes and connected to a computerized pressure-based flow control system (MFCS-4C, Fluigent) controlled by MAESFLO software (Fluigent) and outlet reservoir respectively. Spent media was flushed through the device to wash excess bacteria out of the main channel at 300 μL/h for 8 minutes to completely exchange the fluid in the device and tubing. The chip was mounted on an inverted microscope (IX73 Olympus, Tokyo, Japan) and images were acquired in bright-field via a 60×, 1.2 N.A. objective (UPLSAPO60XW, Olympus) and a sCMOS camera (Zyla 4.2, Andor, Belfast, UK) with a 0.03 s exposure. The microfluidic device was moved by two automated stages (M- 545.USC and P-545.3C7, Physik Instrumente, Karlsruhe, Germany, for coarse and fine movements, respectively) to image multiple fields of view in a sequential manner. The imaging setup was controlled by LabView. After acquiring the first set of images, we flowed each of the investigated antibiotics dissolved

in LB at the appropriate concentration at 300 µL/h for 8 minutes before lowering the flow rate to 100 µL/h for 3 hours. The entire assay was carried out at 37 °C in an environmental chamber surrounding the microscope. Bacterial doubling times were extracted from the acquired image sets as previously reported (Łapińska et al., 2019). Briefly, we tracked each bacterium and its progeny throughout each experiment, and doubling times were measured as the lapses of time between successive bacterial divisions that were assessed by eye through the images loaded in ImageJ (Schneider et al., 2012) and considered to have happened when two daughter cells became clearly distinguishable from their respective parental cell.

3.8.7.One-step phage growth assays

Overnight cultures of PA14 were first diluted into 6 mL growth medium ± antibiotic treatment in glass vials (N=4). For experiments with 30 min of pre-exposure to antibiotics, cells were diluted 25-fold into fresh media with antibiotics and grown for 30 min before phage addition. For experiments with 12 h of pre-exposure to antibiotics, cells were diluted 100-fold into fresh media with antibiotics and grown for 12 h before phage addition. Bactericidal treatments were excluded from further analysis because they caused a 4-fold to 570-fold decrease in cell density after 12 h, making it impossible to determine the latent period of phage under those conditions. After growing in the presence of antibiotics, approximately $5*10^7$ p.f.u. of DMS3vir were added in each vial, and vials were vortexed and incubated at 37 °C for 15 min, allowing phage adsorption. Cultures were then diluted 1000-fold into 6 mL growth medium ± antibiotic treatment to limit further adsorption and re-infection, vortexed again, and transferred to 24-well plates for parallel processing. Samples were taken immediately (t=0) and then approximately every 20 minutes. The first samples were diluted in M9 salts and plated on LB-agar to quantify cell densities; all samples were chloroform-treated and plated on PA14 *csy3::lacZ* lawns. Phage densities measured after chloroform treatment correspond to the sum of free phages and mature phage particles inside infected cells.

3.8.8.Determination of antibiotic effects on infection success

Four overnight cultures of PA14 were diluted in parallel 100-fold into LB with or without antibiotics. After 2 h growth at 37 °C, DMS3vir phages were added to a final concentration of 1000 p.f.u./mL (equivalent to 5 p.f.u. in the 200 µL total volume in each well) and the vials were vortexed. After 15 min at 37 °C, vials were vortexed again and 24∗ 200 µL of each individual culture were aliquoted into 24 wells of a 96-well plate. Plates were incubated at 37 °C for 22 h, then 20 µL of each well were spotted on a lawn of PA14 *csy3::lacZ* cells in two replicates. With an average phage inoculum of 5 phages, the distribution of phages across wells is expected to follow a Poisson distribution with 0.7% wells containing 0 phages and 1.3% wells containing more than 10 phages. The control treatment with no antibiotics was consistent with this, as 1 in 96 wells produced no lysis. Lysis indicated that the founding phages reproduced. The number of wells in which phages failed to reproduce was counted for each treatment, and significance was determined by chi-square tests between antibiotic and no-antibiotic treatment.

3.8.9.Measurement of mutation towards SM

To evaluate the frequency of SM cells in the absence of phage selection, cells were grown in LB ± antibiotic treatment for 24 h. After 24 h, cultures were serially diluted in M9 salts, then dilutions were plated both on LB-agar to calculate total cell density, and on LB-agar containing a high concentration of DMS3vir, which was generated by covering the agar surface with a phage stock of 10^8 p.f.u./µL. A pilot experiment confirmed that all colonies growing on top of DMS3vir were phenotypically SM. The density of SM mutants was calculated by counting the number of colonies growing on top of DMS3vir. Three independent experiments were run with 6 experimental replicates each.

3.8.10.Spacer acquisition assay

20 µL of PA14 overnight culture were first diluted 1:50 into 1 mL LB with or without antibiotics in 24-well plates, in 8 replicates per treatment. After 30 min of growth at 37 °C, $2*10^9$ DMS3vir phages were added per well, and cultures were incubated at 37°C

for 3h. The density of phage-sensitive cells was measured by plating 100 µL on LB-agar after 10^4-fold dilution in M9 salts. The density of phage-resistant cells was measured by directly plating 100 µL of cultures on LB-agar without dilution: the phage density on these plates was sufficient to prevent the growth of sensitive colonies. The majority of colonies had a smooth morphology characteristic of SM clones. We confirmed that smooth colonies were resistant to both DMS3vir and DMS3vir-AcrIF1, whereas non-smooth colonies were resistant to DMS3vir but sensitive to DMS3vir-AcrIF1, and were, therefore, CRISPR immune. In each culture, the proportion of CRISPR-Cas immune clones within the total population of resistant clones (CRISPR-Cas and SM) was calculated.

3.8.11.Competition assays

Competition experiments were performed in 6 mL LB supplemented in the presence or absence of antibiotics. They were initiated by inoculating 60 µL of a 1:1 mix of LB overnight cultures of CRISPR-Cas immune (BIM-2sp) and surface mutant (SM) clones. For treatments including phages, $8*10^9$ p.f.u. of DMS3vir were added per vial. Samples were serially diluted at 0 and 24 h and plated on LB agar supplemented with 50 µg/mL X-Gal (5-bromo-4-chloro-3-indolyl-B-D-galactopyranoside), to determine the ratio of the surface mutant that carries the *lacZ* gene and therefore forms blue colonies, and the BIM-2sp, which forms white colonies. The selection rate of the CRISPR-Cas clone was calculated as m_{BIM2}-m_{3A}, with m the Malthusian parameter defined as $\log(\text{density}(t_1)/\text{density}(t_0))$ (Lenski et al., 1991). We used selection rate rather than relative fitness because some treatments led to an absolute decline in the abundance of the CRISPR-Cas clone.

3.8.12.Quantification and statistical analysis

All statistical analyses were done with R version 3.4.1, and package cowplot (Wilke, 2017). Individual Student t-tests were used to compare each treatment to the associated no-antibiotic treatment. For each experiment, statistical parameters are reported in the figure legends or within the results section.

Chapter 4. Results

4.1. Genome maintenance proteins modulate primed adaptation by *E. coli* Type I-E CRISPR-Cas system

The results presented in this section are based on the following publication: **Kurilovich, E.**, Shiriaeva, A., Metlitskaya, A., Morozova, N., Ivančić-Baće, I., Severinov, K., & Savitskaya, E. (2019). Genome maintenance proteins modulate autoimmunity mediated primed adaptation by the *Escherichia coli* Type I-E CRISPR-Cas system. *Genes*, *10(11)*, 872.

The author performed genomic DNA purification and CRISPR array amplification for sequencing, and all sequencing data analysis. Mutant self-targeting strains were constructed by the author and Dr. A. Metlitskaya. The experiment with bacterial cultures was performed by Dr. A. Shiriaeva. Microscopy was done by Dr. N. Morozova. Sequencing was performed at the Skoltech Genomics Core facility.

4.1.1. Primed CRISPR adaptation is impaired in ΔrecJ, ΔrecB ΔrecJ and ΔrecB ΔsbcD mutants

The impact of the cellular repair enzymes on primed CRISPR adaptation was studied in *E. coli* KD403 strain carrying an engineered inducible self-targeting CRISPR-Cas system (Figure 13A) (Shiriaeva et al., 2019). KD403 contains a mini CRISPR array with a spacer targeting a non-essential *yihN* gene in its genome and *cas* genes under the control of inducible promoters. Upon *cas* genes induction, the protospacer in the *yihN* gene is recognized, and extensive DNA degradation at both sides of the protospacer is observed (Shiriaeva et al., 2019). The degradation of host DNA requires the Cas3 nuclease activity. Induction of self-targeting leads to cessation of growth: compared to uninduced control, the number of colony-forming units is decreased by several orders of magnitude. Induced cells become elongated, indicating activation of SOS response. They remain viable for at

least several hours post-induction and acquire new spacers, mostly from host DNA surrounding the *yihN* protospacer (referred below as priming protospacer, PPS) (Shiriaeva et al., 2019).

We used the self-targeting model to assess the effect of mutations in several host genomic DNA maintenance systems on the adaptation process. First, we evaluated the role of the RecBCD complex responsible for homologous recombination induced by double-strand DNA breaks. We used P1 transduction to generate derivatives of KD403 with *recB*, *recC*, or *recD* genes replaced with kanamycin resistance cassette. As in the parental, wild-type KD403 strain, induction of self-targeting in the mutants caused SOS response, as judged by cell elongation (Figure 12), cessation of culture growth (Figure 10), and reduction in colony-forming units (Figure 11), but did not increase the number of dead cells 5 h after the addition of *cas* genes inducers as judged by differential staining of live and dead cells (Figure 12, 13B).

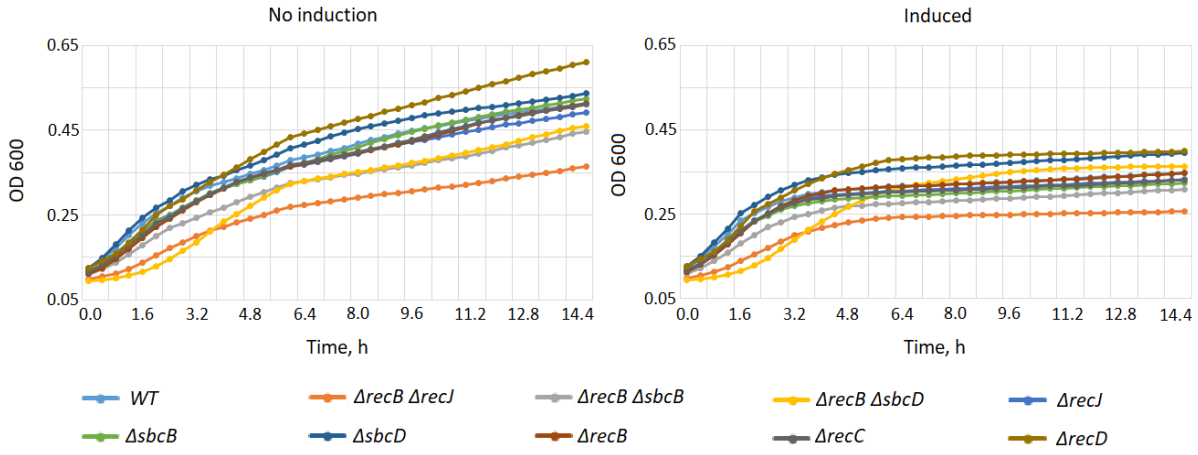


Figure 10. Growth curves showing OD600 of indicated cultures of self-targeting cells in the absence (left) or in the presence (right) of *cas* gene inducers. Aliquots of cultures collected five hours post induction were used for further analyses.

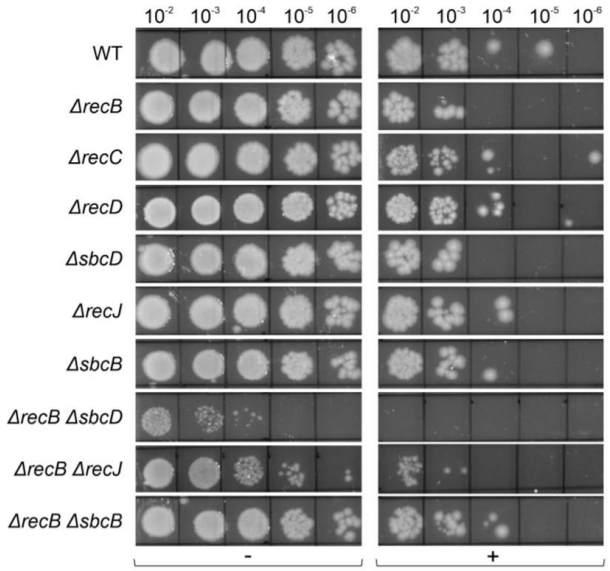


Figure 11. Aliquots of serial dilutions of cultures of indicated self-targeting cells with (right) and without (left) inducers of *cas* gene expression were spotted on LB plates five hours after induction. The results of overnight growth at 37ºC are presented.

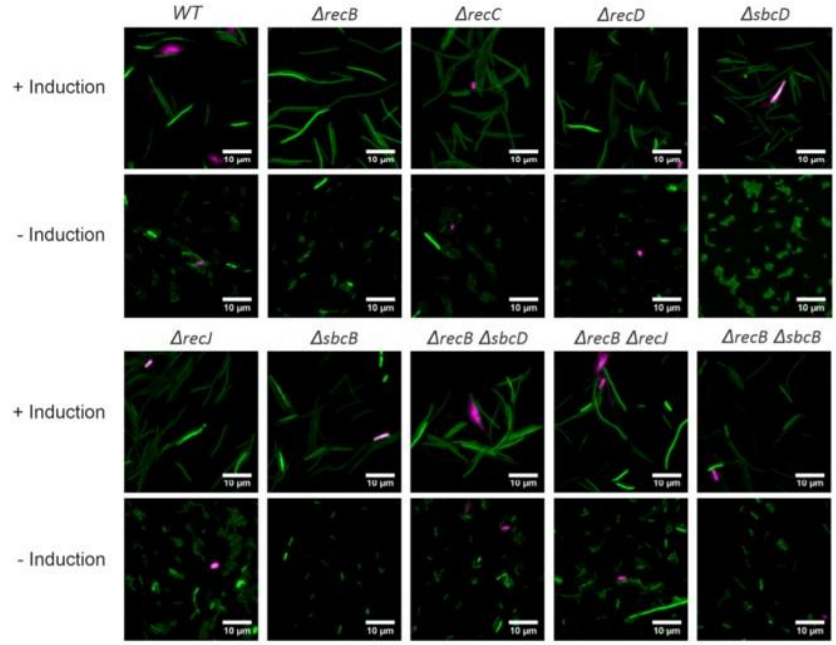


Figure 12. Representative images of cells from cultures with and without induction of self-targeting. Red cells are dead, green cells are alive. Cells become elongated in response to double-stranded DNA breaks (SOS response).

In *recBCD* single mutants, spacer acquisition, detected by the appearance of PCR amplicons corresponding to extended arrays, was observed at levels comparable to that observed in the parental KD403, suggesting that neither RecBC nor RecBCD is required for primed adaptation (Figure 13C).

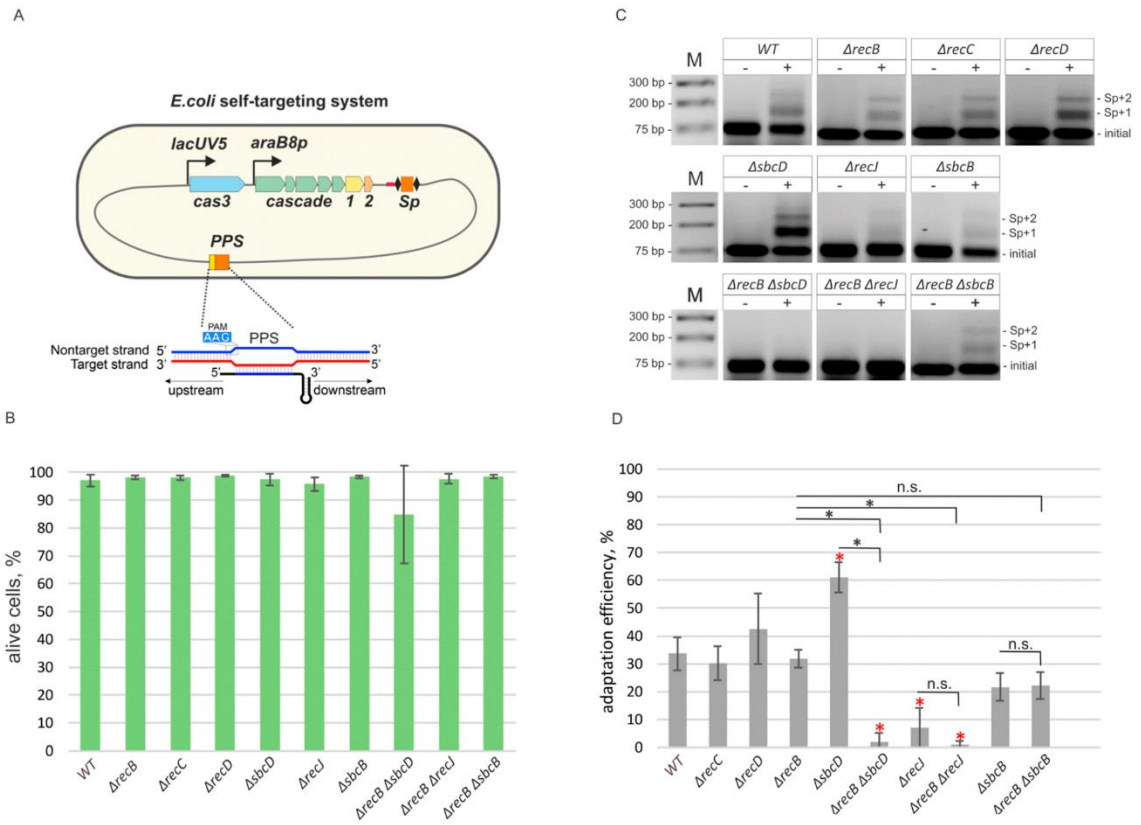


Figure 13. Components of genome maintenance pathways contribute to primed adaptation efficiency. (A) A self-targeting *E. coli* strain KD403 used to study the effects of non-CRISPR genes of primed adaptation is schematically shown. KD403 contains the *cas3* gene under IPTG-inducible lacUV5 promoter and the *cse1, cse2, cas5, cas6, cas7, cas1, cas2* operon under the control of arabinose-inducible promoter araB8p. A minimized CRISPR array contains a single spacer (Sp) matching the priming protospacer (PPS) in the non-essential *yihN* gene. The leader region is shown in red. The structure of the R-loop complex on PPS is shown in the inset. (B) Cultures of KD403 and derivatives with mutations in *recB, recC, recD, sbcD, recJ, sbcB, recB sbcD, recB recJ*, and *recB sbcB* genes were grown in the presence ("+") or in the absence ("−") of inducers and the number of live cells 5 h post-induction was measured by live fluorescence microscopy. (C) Spacer acquisition determined by PCR amplification of the CRISPR array in uninduced and induced cells 5 h post-induction. Amplification products were resolved by agarose gel electrophoresis. Amplicons

corresponding to initial, unexpanded CRISPR array and arrays expanded by one ("Sp+1") and two ("Sp+2") spacers-repeat units are indicated. A representative result of one of three independent experiments is shown. (D) Adaptation efficiency was measured as a percentage of intensities of upper bands on agarose gels shown in panel C (correspond to extended arrays), to the sum of intensities of upper and lower bands and normalized to values obtained for samples grown in the absence of inducers. The mean and standard deviations for three independent experiments are shown. Samples that significantly differ from wild-type KD403 are marked by red asterisks (pairwise t-test, p-value < 0.01). The brackets above the bars indicate pairwise t-test comparisons between double mutants and corresponding single mutants (n.s., non-significant differences at p-value ≥0.01; black asterisks, significant differences at p-value < 0.01).

By constructing appropriate KD403 derivatives, we next assessed the role of 3'-5' single-strand specific exonuclease SbcB, 5'-3' single-strand specific exonuclease RecJ, and SbcD, a component of a single- and double-stranded DNA 3'-5' exonuclease and endonuclease SbcCD. Induction of self-targeting in Δ*sbcB*, Δ*sbcD*, and Δ*recJ* mutant cultures led to an increase in cell length (Figure 12), cessation of culture growth (Figure 10), and reduction in colony forming units (Figure 11) comparable to what was seen in parental KD403, and had either no effect on spacer acquisition (Δ*sbcB* mutant) or led to a ~2-fold increase (Δ*sbcD* mutant) or ~4-fold decrease (Δ*recJ* mutant) in adaptation efficiency (Figure 13D).

Taking into consideration that some genes analyzed may encode enzymes with redundant activities, we assessed primed adaptation in double mutants Δ*recB* Δ*recJ*, Δ*recB* Δ*sbcD*, and Δ*recB* Δ*sbcB*. Induction of self-targeting in double mutants led to an increase in cell length (Figure 12), cessation of culture growth (Figure 10), and reduction in colony forming units (Figure 11) at levels comparable to those seen in parental KD403 cell cultures. As judged by live microscopy, the viability of double mutants was either not or mildly affected (15% dead cells for Δ*recJ* Δ*sbcD* strain 5 h post-induction (Figure 12, Figure 13B), these cells also formed minute colonies in the absence of inducers (Figure 11)), which allowed us to compare the adaptation efficiency. While adaptation was not affected in the Δ*recB* Δ*sbcB* double mutant, it was not detectable in the Δ*recB* Δ*recJ* and

Δ*recB* Δ*sbcD* mutants, suggesting that the presence of functional *cas3* and *cas1-cas2* alone is not sufficient for spacer acquisition at wild-type level in these cells (Figure 13C,D).

4.1.2. Deletions in recB, recC, recD, and sbcD genes affect the choice of spacers acquired during primed adaptation

To get a deeper insight into the role of genes under study in primed adaptation, amplicons corresponding to extended CRISPR arrays that acquired new spacers were analyzed by high-throughput sequencing (see Materials and Methods). Sequences from Illumina reads flanked by CRISPR repeats were treated as spacers, extracted, and mapped to the KD403 genome. It has been shown that in KD403 newly acquired spacers correspond to protospacers located in an area surrounding the PPS with a characteristic gradient of acquisition efficiency falling as the distance from the PPS increases (Shiriaeva et al., 2019). Further, acquired spacers (taken as sequences of non-transcribed strand of the CRISPR array), preferentially map to the nontarget (NT) strand upstream of the PPS and the target (T) strand downstream of it, with more than 96% of spacers originating from sequences with an AAG PAM. According to previously published data, 97% of spacers acquired by the KD403 cultures originate from a 400-kb region upstream and downstream of the PPS. 59% of spacers map on the NT-strand upstream of the PPS while 41% of spacers map on the T-strand downstream of it (Shiriaeva et al., 2019; Strotskaya et al., 2017) (Figure 14). 98% of these "strand-biased" spacers originate from sequences associated with an AAG consensus PAM, a hallmark of primed adaptation (Datsenko et al., 2012) (Figure 15). No spacers acquired due to priming at PPS were detected in DNA amplified from Δ*recB* Δ*recJ* and Δ*recB* Δ*sbcD* double mutants, in agreement with the lack of detectable PCR bands corresponding to expanded arrays in Figure 13C. Therefore, they were excluded from further analysis. Though the overall yield of spacer acquisition in Δ*recB*, Δ*recC*, and Δ*recD* mutants, as judged by PCR analysis of CRISPR array amplicons, was the same as in parental KD403, mapping of acquired spacers revealed a clear alteration in their pattern. Specifically, compared to the "wild-type" strain, the area, from which spacers were acquired in the mutants was extended further away from the PPS in both directions. A

similar, though less pronounced trend was observed for spacers acquired by the Δ*sbcD* mutant (Figure 14). The results were highly consistent between the experiments. In contrast, compared to parental KD403, no changes were observed in the size of the area from which spacer acquisition occurred in Δ*recJ* or Δ*sbcB* single mutants or in the Δ*recB* Δ*sbcB* double mutant compared to the Δ*recB* single mutant.

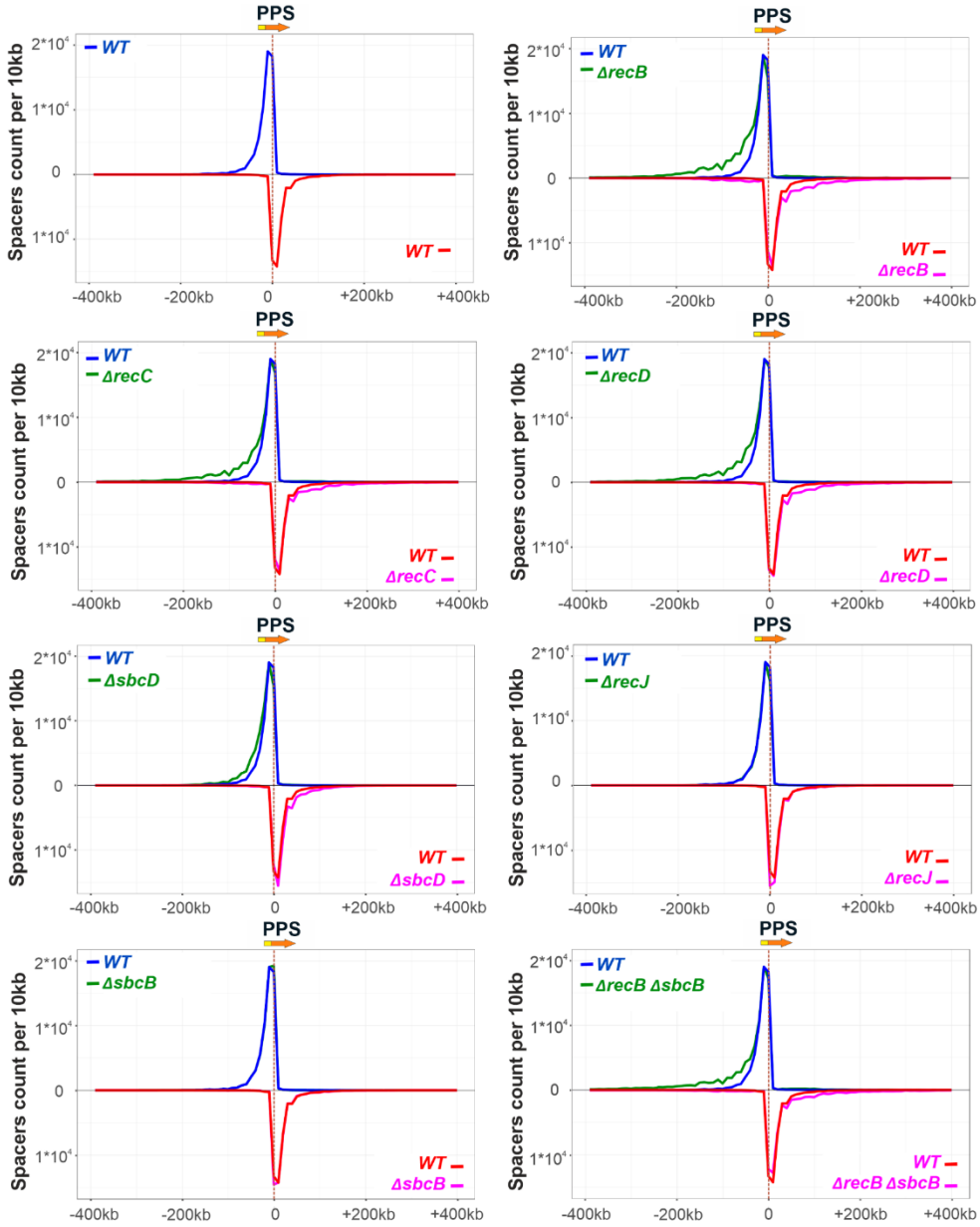


Figure 14. Components of genome maintenance pathways affect the extent of the area around the priming protospacer (PPS) from which spacers are selected. At the top left panel, unique spacers acquired in parental KD403 are mapped to the area of the genome 400 kb upstream and downstream of the PPS. The orientation of PPS is shown by a red arrow at the top, with yellow fragment at the beginning of the arrow denoting the protospacer adjacent motif (PAM) of PPS. The numbers of spacers mapping to 10-kb bins at each side of the PPS are plotted in a strand-specific manner. A curve showing the distribution of spacers mapping to NT-strand is colored blue and plotted above the X-axis; a curve showing the distribution of spacers mapping to the T-strand is colored red and plotted below the X-axis. The rest of the panels show normalized comparisons of spacer acquisition profiles between the parental ("*WT*") KD403 and indicated mutants in which spacer acquisition was detected (see Figure 13). In each panel, the wild-type profile is shown by blue and red colored lines, while the mutant profiles are green (NT strand) and magenta (T strand). Averaged results from two biological replicates are shown. For Δ*recJ* and Δ*sbcB* mutants, the mapping curves are completely superimposed with the *wt* and are hardly seen.

For all strains analyzed, the average association with AAG for spacers mapping to "preferred" strands upstream and downstream of the PPS was at a level of at least 97% (Figure 15). This was also true for PPS-distal "extra" spacers acquired in Δ*recB*, Δ*recC*, Δ*recD*, and Δ*sbcD* mutants from regions from which acquisition for parental KD403 was negligible (Figure 16).

A minor (~3%) fraction of spacers acquired by induced KD403 cultures was mapped to the T-strand upstream of the PPS and to the NT-strand downstream of it. The acquisition of these spacers is clearly driven by the PPS recognition, as they cluster in the area around the PPS and in this sense originate through priming. Interestingly, only 79% and 74% of these spacers originate from sequences associated with the AAG PAM (Figure 15). These values, while clearly higher than those reported for naïve adaptation (40–50%) (Yosef et al., 2013), are considerably below the values typical for primed adaptation, suggesting that a mechanistically distinct process is involved. Yet, the area of acquisition of such spacers, mapping to opposite strands, was extended in Δ*recB*, Δ*recC*, Δ*recD*, and Δ*sbcD* mutants and unaffected by deletions of *recJ* or *sbcB*, similar to the situation observed with spacers acquired during the *bona fide* primed adaptation process. Surprisingly, when analyzing the AAG bias of minor spacers acquired in different KD403

mutant derivatives, we consistently observed significantly more AAG-associated spacers in the Δ*recB* mutant but not in the Δ*recD* mutant (Figure 15). We assume that the RecBCD complex directly participates in the creation of precursors of such spacers. Although the detailed mechanism remains to be established, our observation highlights the distinct roles of helicase and nuclease activities of RecBCD in primed adaptation.

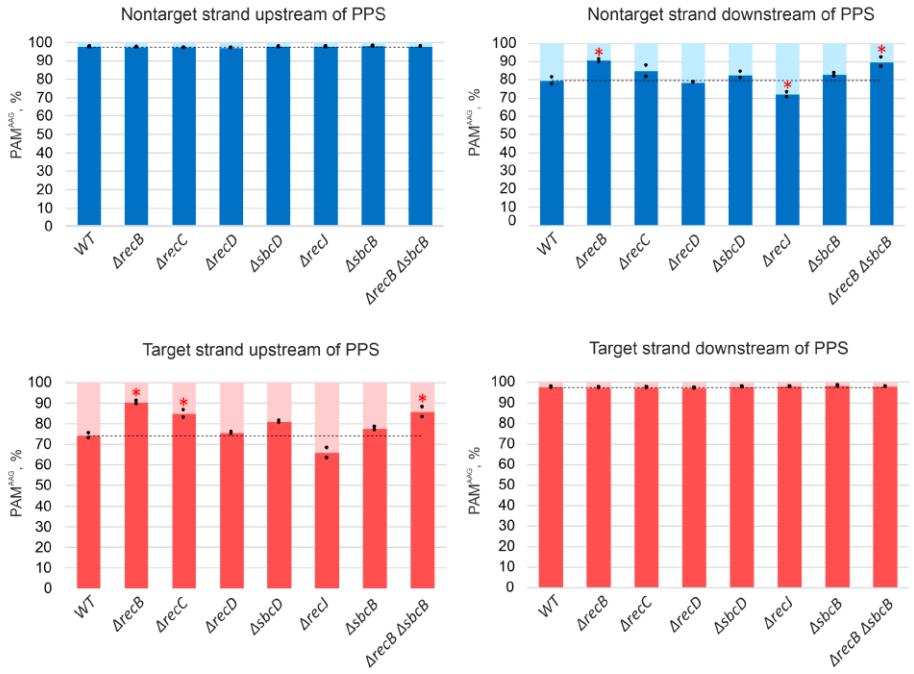


Figure 15. Association with the AAG PAM of spacers acquired by KD403 and its derivatives. Spacers acquired by indicated cells were mapped to an area within 200 kb upstream and downstream of the PPS, and the percentage of mapping sites with an AAG PAM was determined separately for spacers mapping to upstream and downstream T and NT-strands. The heights of the bars show mean values obtained in two independent experiments (black dots show values obtained in each individual experiment). The dotted line shows the average level for the parental KD403 ("*WT*") strain. Red asterisks mark values that significantly differ from the *wt* level according to pairwise proportional test with a p-value of 0.001.

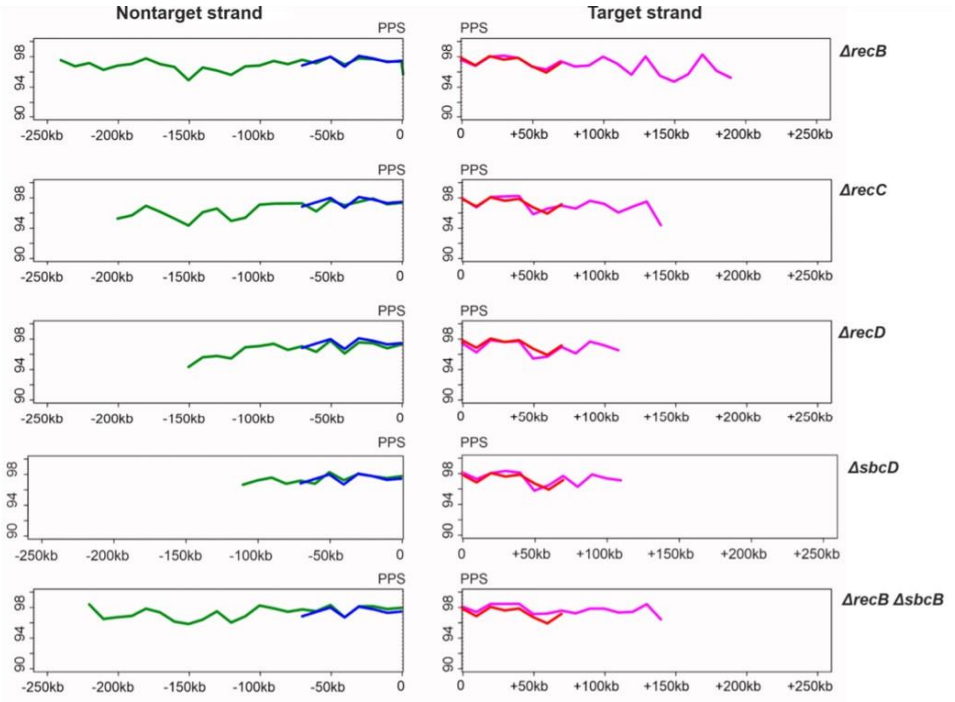


Figure 16. Association with the AAG PAM of spacers acquired by KD403 and its derivatives in an extended area around the PPS. The percentage of spacers mapping to sequences with appropriately positioned AAG PAM is plotted as a function of distance upstream and downstream of the PPS in 10-kb bins. Blue and red lines represent, correspondingly, values for the NT-strand upstream of the PPS and the T-strand downstream of the PPS for parental KD403. Green and purple lines represent the corresponding values for mutants. Only bins with a number of spacers above a threshold of 500 counts were taken into consideration.

4.1.3.Deletion of recJ influences prespacer integration

The deletion of *recJ* decreased primed adaptation efficiency ~4-fold (Figure 13C, D). Simultaneous deletion of *recJ* and *recB* made primed adaptation undetectable (Figure 13C, D). Since, in contrast to Δ*recBCD* mutants, the absence of *recJ* did not shift the area which spacers were selected from (Figure 14), we assumed that the pathways that involve RecBCD and RecJ in primed adaptation are non-redundant. In other words, each pathway could make its own, independent, contribution, and abolishing both pathways prevents primed adaptation. To get further insights into the role of RecJ, we investigated the repertoire of acquired spacers in the Δ*recJ* mutant. Spacers were clearly acquired less

accurately in the absence of RecJ than in all other strains (Figure 17). The percentage of inaccurately incorporated spacers that have been shifted relative to the AAG PAM in protospacers (shifters) or inserted into the CRISPR array in the opposite orientation (flippers) was 2.25% in parental KD403 and Δ*recBCD* and Δ*sbcD* mutants, compared to 3% in the Δ*recJ* mutant (a 30%, highly statistically significant increase). Second, the choice of acquired spacers was affected in the Δ*recJ* mutant in a very specific way. Recently, it was shown that spacers efficiently acquired in the course of primed adaptation are depleted of internal AAG trinucleotides (Musharova et al., 2018). The percentage of spacers containing internal AAG trinucleotide was slightly, yet, significantly increased in the Δ*recJ* mutant (Figure 17B). Consistently, compared to parental KD403, the AAG PAM bias was slightly decreased (though the effect is marginally significant) for acquired spacers mapping to poorly used strands around the PPS in Δ*recJ* cells (Figure 15).

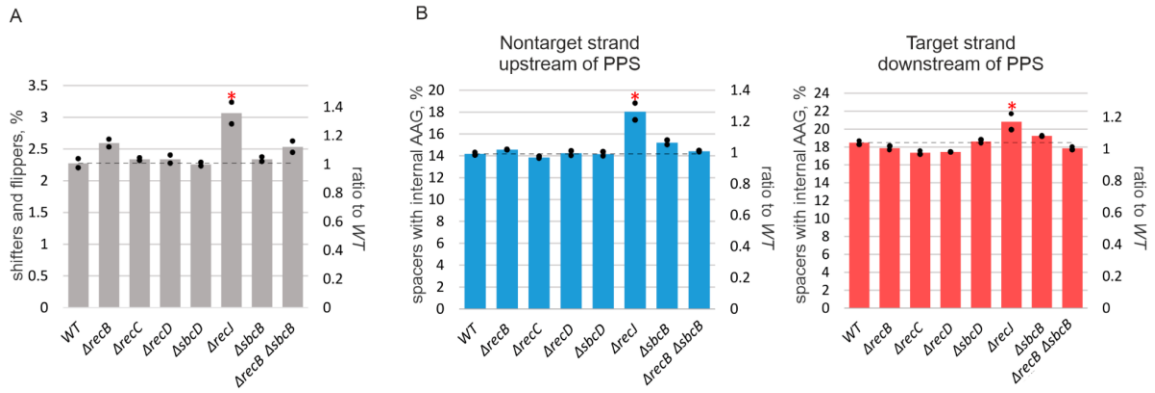


Figure 17. RecJ affects spacer acquisition precision and specificity. Spacers acquired by indicated cells were mapped to an area within 200 kb upstream and downstream of the PPS for KD403 and indicated derivatives. (A) Percentage of inaccurately incorporated spacers (both shifters and flippers) in indicated cultures. The numbers on the right axis show the ratios of inaccurately incorporated spacers to those obtained for *WT* cells. (B) Percentage of spacers with internal AAG trinucleotides. Data for spacers mapping to NT-strand upstream of the PPS and T-strand downstream are shown by blue and red bars, correspondingly. The heights of the bars show mean values obtained in two independent experiments (black dots show values obtained in each experiment). The numbers on the right axis show the ratios of spacers with internal AAG trinucleotides to those obtained for *wt* cells. The dotted line shows the average level for the parental KD403 ("*WT*") strain. Red asterisks mark values that significantly differ from the *wt* level according to pairwise proportional test with a p-value of 0.001.

4.2. Host nucleases generate prespacers for primed adaptation by the *E. coli* type I-E CRISPR-Cas system

The results presented in this section are based on the following publication: Shiriaeva, A., Kuznedelov, K., Fedorov, I., Musharova, O., Khvostikov, T., Tsoy, Y., **Kurilovich, E.**, Smith, G., Semenova, E., Severinov, K. (2022). Host nucleases generate prespacers for primed adaptation in the *E. coli* type I-E CRISPR-Cas system. *Sci. Adv.*, *8(47)*, 8650.

The author contributed to the work by preparing NGS libraries for sequencing and performing the analysis of new spacers in *xseA* mutant (Figure 23). Data curation and the rest of the analysis was done by Dr. A. Shiriaeva. *In vitro* experiments were performed by Dr. K. Kuznedelov.

The results provided in sections 4.3.1 and 4.3.2 were partially presented in the PhD Thesis of Dr. A. Shiriaeva (2020). As they were obtained after the first publication (Kurilovich et al. 2019), and are directly relevant to the topic of the current thesis, they are included below.

4.2.1. E. coli RecBCD and RecJ are jointly required for prespacer generation during primed adaptation in vivo

To investigate the roles of genome maintenance proteins in prespacer formation and trimming, DNA short fragments were purified from *wt* KD403, Δ*recD*, Δ*recB*, Δ*recC*, Δ*recJ*, and Δ*recB* Δ*recJ* derivatives and sequenced according to the FragSeq approach (Shiriaeva et al., 2019). Uninduced *wt* (-Ind) sample was used as a control. The amounts of fragments were adjusted to account for the preferential loss of shorter fragments. As expected, enrichment of 31-40-nt fragments around the PPS in induced *wt* self-targeting cells was observed (Figure 18). The sharpest central part of the enrichment peak is within 50 kb around the PPS. In what follows, if not stated otherwise, we analyze fragments located not farther than 25 kb from the PPS. Similar enrichment peaks were also observed

in coverage plots of all studied mutants except for Δ*recB* Δ*recJ*, suggesting that prespacer generation is abolished when both RecBCD and RecJ are inactivated.

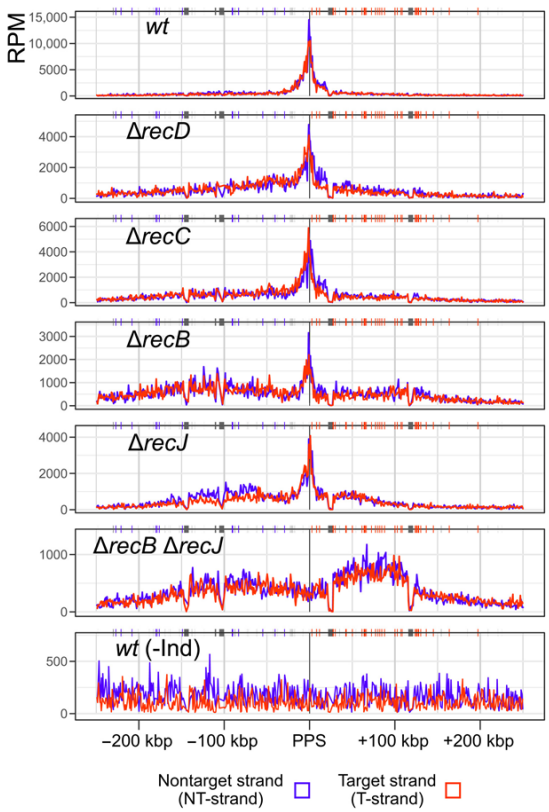


Figure 18. Detection of prespacers in self-targeting *E. coli* strains affected in homologous recombination. Normalized sequence coverage of PPS ± 250 kb (RPM, in 1-kb bins) by 31-40-nt fragments after adjusting fragment abundances using fragment-length specific loss coefficients. Coordinates on the X-axes show the distance from the PPS. Fragments mapped to the NT-strand are shown in blue; fragments mapped to the T-strand are in red. Chi sites oriented to activate RecBCD moving away from the PPS are shown above each coverage plot as blue or red vertical lines. Grey boxes indicate repetitive regions of the genome. Reads matching such regions were excluded from the analysis, leading to an apparent decrease in fragment coverage.

Previously, it was reported that in *wt* cells most fragments originating from the NT-strand upstream of the PPS (i.e., at the PAM-proximal side of the PPS; Figure 19) and from the T-strand downstream of the PPS (i.e., at the PAM-distal side of the PPS) were ~33 nt

in length and their 5’ ends were generated by cleavage of a phosphodiester bond within the 5’-AAG-3’ sequence. We will refer to these fragments as “AAG-associated fragments”. In contrast, most fragments originating from the T-strand upstream of the PPS and from the NT-strand downstream of the PPS were ~37 nt in length and their 3’ ends were generated by cleavage of the phosphodiester bond 2 nt to the left of the 3’-TTC-5’ motif complementary to PAM. We will refer to these fragments as “TTC-associated fragments” (Figure 19). Once annealed, complementary AAG- and TTC-associated fragments should produce prespacers with a double-stranded region of ~33 bp and a short 3’ overhang on the PAM-derived end. The ratios of fragments from the two complementary strands (NT/T) varied in the range of 0.77-1.6, suggesting that most fragments have a complementary pair and form double-stranded prespacers.

To determine the effects of *recB*, *recC*, *recD*, and *recJ* deletions on AAG-associated fragments, for each strain we combined 31-40-nt fragments from the NT-strand upstream and from the T-strand downstream of the PPS (Figure 19). Likewise, we combined fragments from the T-strand upstream and the NT-strand downstream of the PPS to determine the effects of *rec* deletions on TTC-associated fragments (Figure 19).

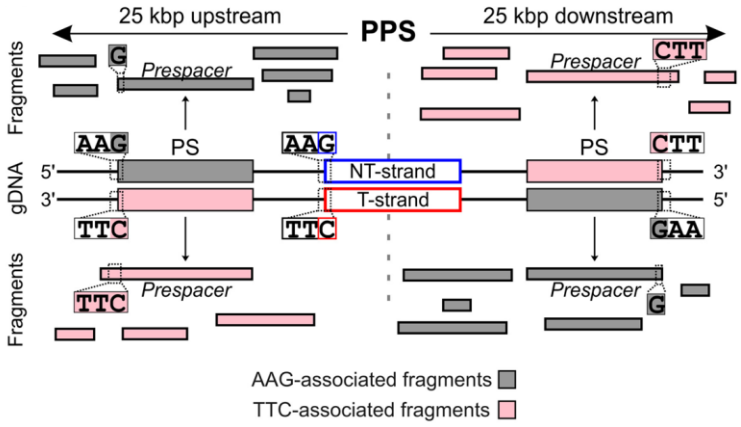


Figure 19. A 50-kb region around the PPS is schematically presented. Oppositely oriented protospacers (PS) that are predominantly selected as spacers during primed adaptation are depicted to the left (upstream) and to the right (downstream) of the PPS. Fragments mapping to the NT-strand upstream of the PPS and to the T-strand downstream of the PPS are shown in gray. In the wild-type, AAG-associated fragments are found

in this group. Fragments mapping to the T-strand upstream of the PPS and to the NT-strand downstream of it are shown in pink. In the wild-type, TTC-associated fragments are found in this group.

While in most mutant strains we detected AAG- or TTC-associated fragments with frequencies and distributions the same as or very similar to that of *wt* , no enrichment with AAG- or TTC-associated fragments was observed in the Δ*recB* Δ*recJ* mutant (Figure 18, Figure 20). Instead, the distribution of fragments in this mutant was similar to that observed in the uninduced *wt* control where no detectable adaptation takes place. We thus conclude that the absence of spacer acquisition in the Δ*recB* Δ*recJ* mutant is due to the absence of prespacers.

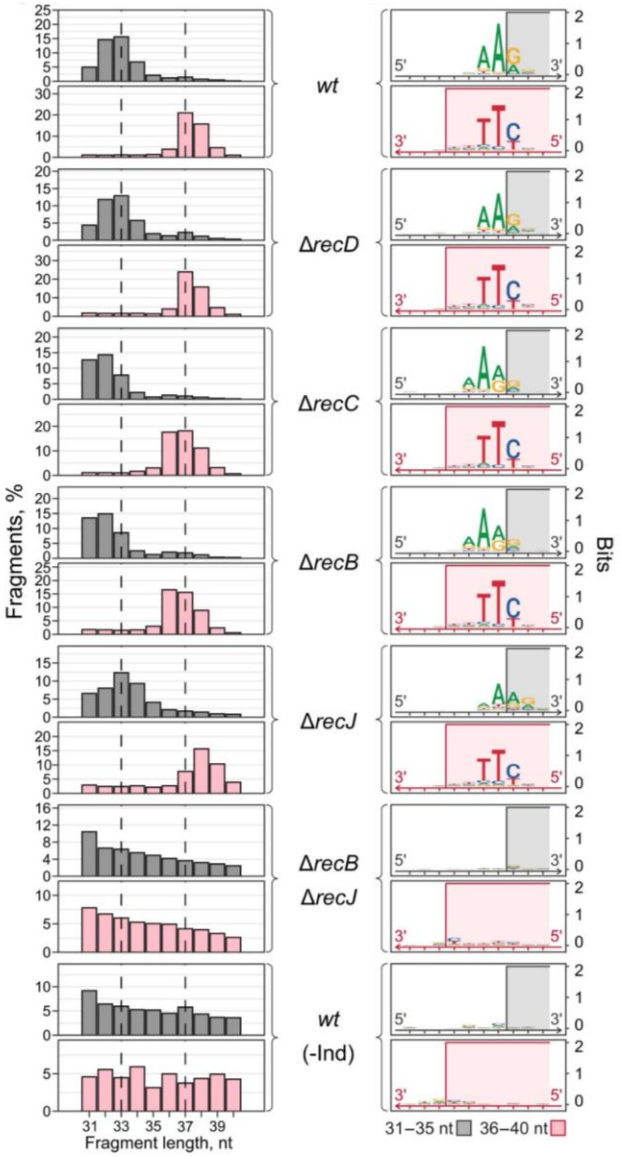


Figure 20. On the left, length distributions of 31-40-nt fragments originating from the 50-kb PPS region. 100% corresponds to all 31-40-nt fragments from both strands in this region. Fragment abundancies were adjusted using fragment-length specific loss coefficients. On the right, sequence alignments of fragments' ends and adjacent genomic regions. Green rectangles represent 5' ends of 31-35-nt AAG-associated fragments. Pink rectangles represent 3' ends of 36-40-nt TTC-associated fragments.

4.2.2.The RecBCD helicase and RecJ nuclease participate in the processing of prespacer 5' ends

Though AAG- and TTC-associated fragments were detected in all studied mutants except for the Δ*recB* Δ*recJ* strain, fragment lengths were slightly different in Δ*recJ*, Δ*recC*, and Δ*recB* mutants compared to *wt* (Figure 20). In particular, while 37-nt fragments constituted the major fraction of TTC-associated fragments in *wt*, 38-nt fragments were predominant in the Δ*recJ* strain. Likewise, a higher fraction of the AAG-associated fragments was one nucleotide longer in this mutant than in the *wt* strain. On the contrary, higher percentages of fragments one nucleotide shorter than those observed in *wt* were detected in Δ*recB* and Δ*recC* strains. The differences in fragment length distributions might be caused by altered processing of the 5', 3', or both ends of prespacer precursors.

To analyze variations in prespacer ends, we first determined chromosomal coordinates of all possible 33-bp protospacers associated with 5'-AAG-3'/3'-TTC-5' PAMs and located not farther than 25 kb from the PPS in the correct orientation (375 protospacers with 5'-AAG-3' in the NT-strand upstream of the PPS and 490 protospacers with 5'-AAG-3' in the T-strand downstream of the PPS). The coordinates of such protospacers should i) coincide with properly processed 33-nt prespacer coordinates (as is the case with 33-nt prespacer strands starting with a G) or ii) lie within not fully processed prespacer coordinates (as is the case with 37-nt prespacer strands having the 3'-NNTTC-5' motif on their 3' ends). Finally, when a prespacer is processed to a length shorter than the protospacer length (33 bp), prespacer coordinates should lie within the protospacer coordinates. We calculated distances from the 5' and 3' ends of all 31-40-nt AAG- and TTC-associated fragments to the boundaries of their corresponding protospacers. Given that in the crystal structure of the *E. coli* Cas1-Cas2 bound to a model substrate the central 23-bp region is in a double-stranded form, while the terminal 5-nt regions on both sides are single-stranded and likely more exposed to degradation (Nuñez, Lee, et al., 2015; J. Wang et al., 2015), we selected for our analysis only those fragments that spanned the central 23-nt protospacer parts. Negative values were assigned to calculated distances if the positions of fragment ends were shifted towards the centers of protospacers; positive values were

assigned if the positions of fragment ends lay beyond the corresponding protospacer boundaries; zero values were assigned if the positions of fragment ends coincided with positions of protospacer boundaries.

The distributions of calculated distances for the *wt* and the mutants are shown in Figure 21 (the Δ*recB* Δ*recJ* strain was not included in this analysis due to low number of PAM-associated fragments and high background of unrelated fragments, presumably generated by random DNA fragmentation). For every strain considered, the expected differences in distance distributions for PAM-derived and PAM-distal 3' ends were observed (Figure 21B). The major fraction of fragments had the 3'-NNTTC-5' (~60%) or 3'-NNNTTC-5' (~20%) motif on their PAM-derived 3' ends, which corresponds to the distance of +4 and +5, respectively. PAM-distal 3' ends either coincided with protospacer boundaries (distance = 0) or were truncated by 1 nucleotide (distance = -1) in ~60-70% of fragments. Overall, none of the nucleases encoded by the affected genes has a considerable effect on the processing of prespacer 3' ends.

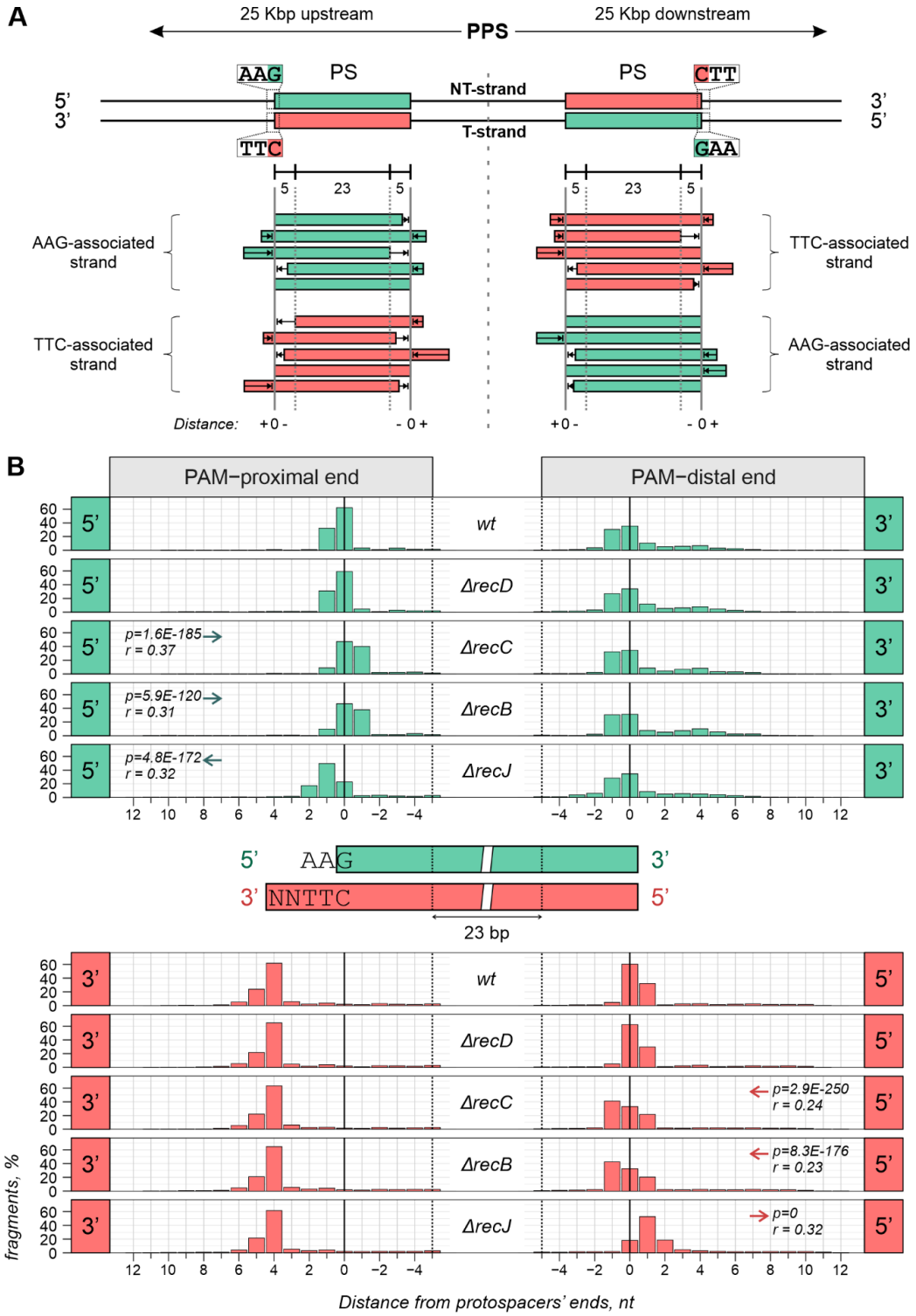


Figure 21. Processing of prespacer ends varies in different *E. coli* strains. (A) A strategy for selection of potential prespacer fragments. Coordinates of all 865 possible 33-bp protospacers with an adjacent 5'-AAG-

3'/3'-TTC-5' PAM were determined in a 50-kb region centered at the PPS. Upstream of the PPS, only protospacers with the 5'-AAG-3' motif in the top (NT) strand were selected. Downstream of the PPS, only the protospacers with the 5'-AAG-3' motif in the bottom (T) strand were selected. For each protospacer, the coordinates of its central 23-bp region were determined. 31-40-nt fragments that contained the central 23-bp protospacer parts were selected for further analysis. All fragments that mapped to the protospacer strand associated with the 5'-AAG-3' motif were pooled (shown in green). Fragments mapped to the protospacer strand associated with the 3'-TTC-5' motif were also pooled (shown in pink). For each fragment, distances from its ends to corresponding protospacer ends were calculated. Zero values were assigned to the distances if the fragments' ends coincided with the protospacers' ends. Positive values were assigned to the distances if fragments' ends lay outside protospacers. Negative values were assigned to the distances if fragments' ends lay within protospacers. (B) Distribution of the distances from fragments' ends to the ends of protospacers in indicated strains. Values within individual plots are p-values of one-sided Mann Whitney U tests computed in comparisons with the *wt* and effect size values r. Only p-values less than 0.05 and r-values greater than 0.1 are shown. The arrows indicate the direction of the shift along the X-axis in samples compared with the *wt*.

Unlike the asymmetric 3’ ends, PAM-derived and PAM-distal 5’ ends in *wt* were processed symmetrically with respect to the protospacer boundaries (Figure 21B): in ~60% of fragments, both 5’ ends coincided with protospacer boundaries (distance = 0); in ~30% of fragments, an additional nucleotide was present (distance = +1). Both 5’ ends were on average longer in the Δ*recJ* mutant than in *wt* (Mann Whitney U test, p≈0, effect size r = 0.32). Only ~20% of fragments had 5’ ends coinciding with protospacer boundaries, while 50% of fragments had an extra nucleotide (distance = +1) and fragments with two additional nucleotides (distance = +2) became prominent (~20%). Significant differences with *wt* were also revealed in Δ*recB* and Δ*recC* mutants (Mann Whitney U test, $p \leq 5.9E-120$, r = 0.23-0.37), but in this case 5’ ends were processed more excessively than in *wt*, such that ends located at the -1 distance became prominent (~40% of fragments). Interestingly, in strains lacking RecB or RecC, the shapes of the distributions calculated for the PAM-distal and PAM-derived 5’ ends were different. Specifically, for PAM-derived 5’ ends formed in these mutants, the distance equal to 0 was most prominent, while for PAM-distal 5’ ends, it was the -1 distance (Figure 21B).

The result suggests that recognition of the PAM influences the way 5' ends are processed but only when no RecBCD or RecBC complex is present in the cell.

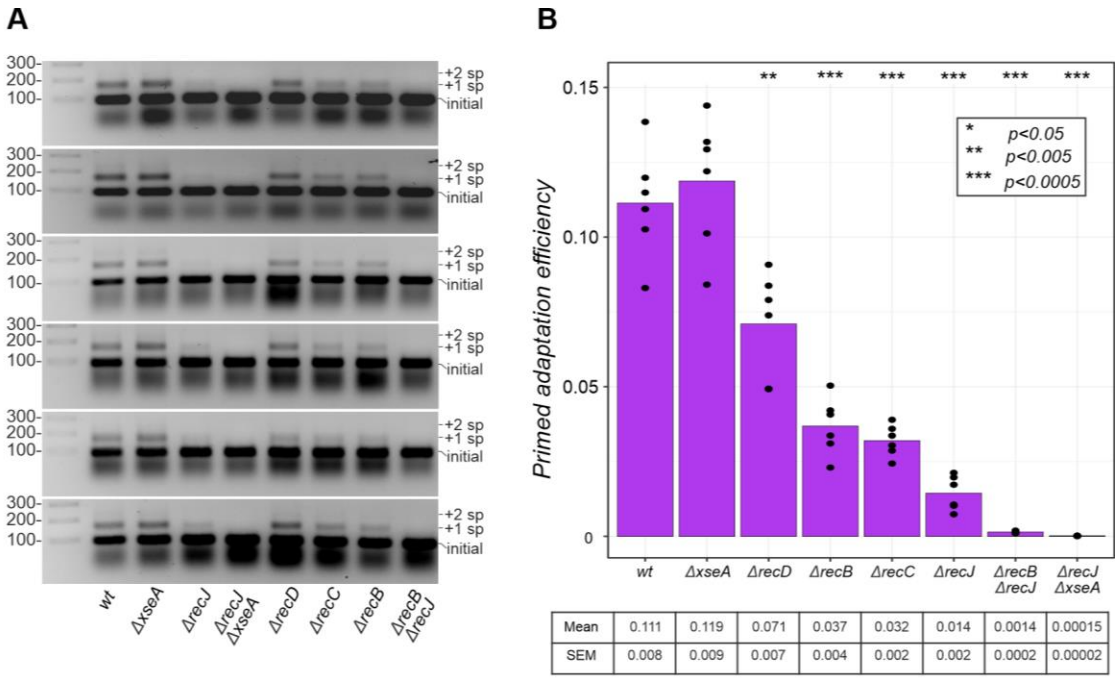
No changes in 5' end processing were found in the Δ*recD* mutant. The RecBC complex has helicase but no nuclease activity. Therefore, the helicase activity of RecBCD is required for the generation of proper prespacer ends. Altogether, we conclude that the RecJ 5'-3' exonuclease activity and the RecBCD helicase activity are involved in the processing of prespacer 5' ends during primed adaptation.

4.2.3.CRISPR adaptation efficiency is decreased in KD403 mutant derivatives

We re-assessed the efficiency of CRISPR adaptation using high-throughput sequencing of newly acquired spacers. In agreement with previous results, compared to the wild-type self-targeting strain, we observed a decrease in primed adaptation efficiency by ~90% in Δ*recJ* and by ~99% in the Δ*recB* Δ*recJ* strains. We also detected a less prominent, ~36% decrease in the Δ*recD* strain, and ~70% decreases in both Δ*recB* and Δ*recC* strains (Figure 22). In the previous work, no differences in primed adaptation efficiency between the *wt* and either Δ*recD,* Δ*recB,* or Δ*recC* strains were observed. This discrepancy is likely explained by a more sensitive and quantitative method used to detect spacer acquisition in the current work (high-throughput sequencing versus quantification of agarose gel bands intensity used previously).

We also inactivated the large ExoVII subunit encoded by the *xseA* gene in *wt* and Δ*recJ* backgrounds. ExoVII exonuclease (a heterodimer of the products of *xseA* and *xseB* genes) has both 3′→5′ and 5′→3′ exonuclease activities on single-stranded DNA (Chase & Richardson, 1974). The results show that the deletion of *xseA* alone did not influence the adaptation efficiency. In contrast, primed adaptation was dramatically (~700-fold) decreased in the Δ*recJ* Δ*xseA* mutant (Figure 22).

Overall, our observations suggest that i) host nucleases RecJ, and RecBCD participate in prespacer generation, ii) ExoVII may be responsible for prespacer generation in the Δ*recJ* strain and iii) the decrease in spacer acquisition in mutant strains is caused by reduced prespacer amounts or modified prespacer structures (or both).



	wt	ΔxseA	ΔrecD	ΔrecB	ΔrecC	ΔrecJ	ΔrecB ΔrecJ	ΔrecJ ΔxseA
Mean	0.111	0.119	0.071	0.037	0.032	0.014	0.0014	0.00015
SEM	0.008	0.009	0.007	0.004	0.002	0.002	0.0002	0.00002

Figure 22. Primed adaptation efficiency is decreased in Δ*recD,* Δ*recB,* Δ*recC,* Δ*recJ,* Δ*recB* Δ*recJ,* and Δ*recJ* Δ*xseA* mutants. On the left panel, amplification products of unexpanded CRISPR arrays ("initial") and CRISPR arrays with one ("+1 sp") or two ("+2 sp") additional spacer-repeat units resolved by electrophoresis in agarose gel are shown. A molecular-weight size marker is shown on the left. Results of 6 independent biological replicates are presented. On the right panel, primed adaptation efficiency is calculated as the ratio of newly acquired spacers to the total number of CRISPR arrays determined by high-throughput sequencing of CRISPR array amplicons. Asterisks indicate statistically significant differences between a mutant and the *wt* ($p<0.05$) in pairwise Welch's t-test. Points represent individual values obtained in 6 biological replicates. For each strain, the mean and standard error of the mean are shown in the table under the bar graph.

Newly acquired spacers in Δ*xseA* strain were mapped on the reference genome. No changes in spacer acquisition pattern were observed, compared to *wt* (Figure 23).

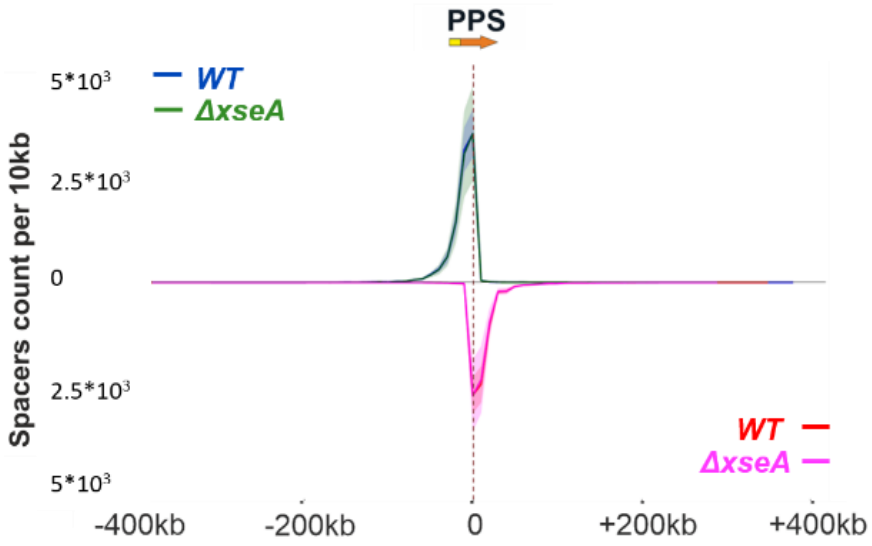


Figure 23. The *xseA* mutation does not affect the extent of the area around the priming protospacer (PPS) from which spacers are selected. The normalized comparison of spacer acquisition profiles between the parental ("*WT*") KD403 and Δ*xseA* mutant is shown. The wild-type profile is shown by blue and red colored lines, while the mutant profiles are green (NT-strand) and magenta (T-strand). Averaged results for six biological replicates are shown. The mutant mapping curves are completely superimposed on the *wt* curves and are hardly seen. The shaded area represents the range of values.

4.2.4.Lowered efficiency of prespacer generation rather than modified structure of prespacer ends causes a decrease in primed adaptation efficiency in ΔrecJ, ΔrecB, and ΔrecC mutants.

To test if lowered primed adaptation efficiencies in mutants are caused by modifications of prespacer ends, we used an oligo electroporation assay. *E. coli* cells containing a single CRISPR array and expressing *cas1* and *cas2* from a plasmid were transformed either with a 33/37-nt canonical prespacer, a 34/38-nt prespacer with 5' ends elongated by 1 nt as observed in Δ*recJ*, or a 32/36-nt prespacer with 5' ends shortened by 1 nt as observed in Δ*recB* and Δ*recC* (Figure 24A, B). An amplified mixture of extended and unextended CRISPR arrays in cells harvested 2 hr after transformation was sequenced and the percentage of arrays containing an oligo-derived spacer was calculated. We found that elongation or shortening of prespacer 5' ends did not significantly decrease the

frequency of oligo-derived spacers or the accuracy of integration (Figure 24C, D). Similarly, changes in oligo lengths had no or a very small effect on their incorporation in the array of the double *recB recJ* mutant (Figure 24E, F). Therefore, the altered structures of prespacer ends in Δ*recJ,* Δ*recB,* and Δ*recC* do not explain low spacer acquisition efficiencies in these mutants. We also tested the efficiency of acquisition of oligos into the CRISPR array of the Δ*recJ* Δ*xseA* strain. No significant changes were observed (Figure 24G, H).

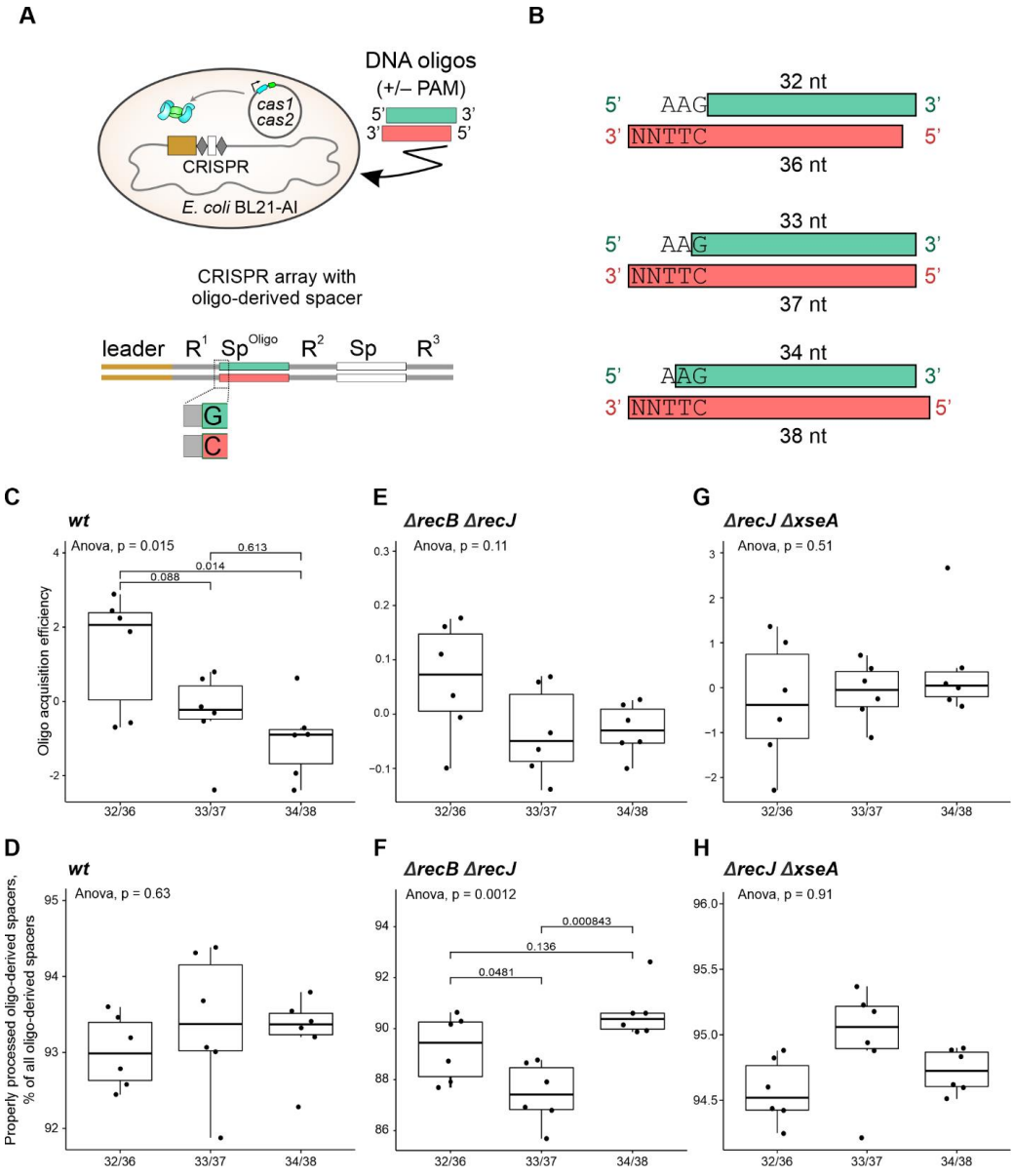


Figure 24. Shortening or extending the prespacer 5' ends by 1 nt does not affect prespacer acquisition efficiency. (A) Prespacer acquisition efficiency assay. Top, electroporation of a synthetic DNA oligo into cells containing a CRISPR array and a plasmid that directs expression of cas1 and cas2. Bottom, oligo-derived spacers integrated into the CRISPR array. (B) Oligonucleotides used in prespacer acquisition efficiency assay. The results of oligo electroporation into the *wt* (C, D), Δ*recB* Δ*recJ* (E, F), and Δ*recJ* Δ*xseA* (G, H) strains are shown. (C, E, G) Zero-centered oligo acquisition efficiencies after removal of batch effects by subtracting the mean of the measurements for each batch. The measurements were taken 2 hours after electroporation. p-values calculated in one-way ANOVA and Tukey HSD tests are shown. (D, F, H) Percent

of properly processed oligo-derived spacers among all oligo-derived spacers. Properly processed oligo-derived spacers are defined as 33-bp spacers starting with the PAM-derived G/C pair. p-values of one-way ANOVA are reported.

To estimate the efficiency of prespacer generation in mutant strains, we calculated the amounts of prespacer-like fragments in 1-Mb region surrounding the PPS. Prespacer-like fragments were defined as 31-40-nt fragments whose ends are located at the following distances from protospacer boundaries: -1, 0, 1, 2 for both 5' ends; -2, -1, 0, 1 for PAM-distal 3' ends, and 2, 3, 4, 5, 6 for PAM-derived 3' ends. We extended the region around the PPS compared to the analysis presented above since in *recBCD* mutants spacers are acquired from a wider region compared to the *wt*. The number of prespacer-like fragments around the PPS was normalized to the number of prespacer-like fragments in a 1-Mb "control" region distant from the PPS (Figure 25A, C). Since this control region is not subject to CRISPR interference, fragments mapping to this region represent a nonspecific background.

When the 1-Mb PPS-containing region was assessed, the Δ*recJ* mutation led to a ~70% decrease in prespacer generation efficiency (Figure 25C). Very few prespacers were detected in the Δ*recB* Δ*recJ* strain (a decrease of ~95% compared to the *wt*). Prespacer generation was also decreased, by ~35%, in Δ*recB* and Δ*recC*, but not in the Δ*recD* strain (Figure 25C). The lowered prespacer generation efficiency in the Δ*recD* mutant may be masked by the widening of the area around the PPS where prespacers are generated. Therefore, we also estimated prespacer generation efficiency in a smaller, 0.1-Mb, area centered at the PPS (Figure 25D). This analysis revealed a decrease in the amount of prespacers formed in the Δ*recD* strain, which was less prominent than in the *recB* and *recC* mutants.

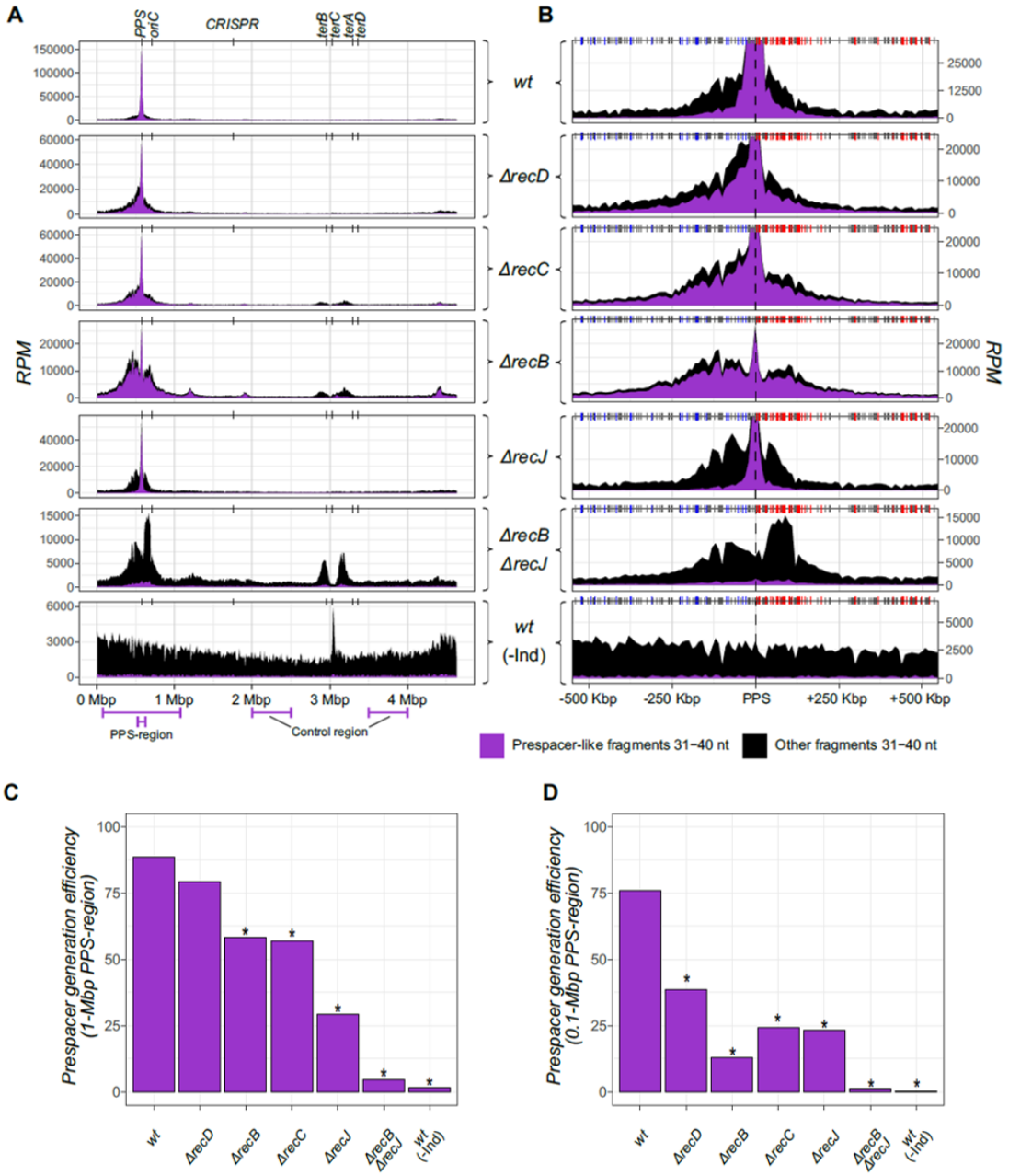


Figure 25. Chromosomal distribution of 31-40-nt fragments. (A) Normalized sequence coverage (RPM, in 10-kb bins) by 31-40-nt fragments. Coordinates on the X-axes represent positions on the chromosome. *oriC* - replication origin; *terA*, *terB*, *terC*, *terD* - sites of replication termination; "CRISPR" – denotes the CRISPR array. Prespacer-like fragments are shown in purple. All other 31-40-nt fragments are shown in black above prespacer-like fragments. (B) A close-up of the sequence coverage from A shown for the 1-Mb PPS region (coordinates on the X-axes represent distance from the PPS). Chi sites oriented to activate RecBCD moving

away from the PPS are shown above each coverage plot as blue or red vertical lines. Grey boxes indicate repetitive regions of the genome. Reads matching such repetitive regions were excluded from the analysis, resulting in an apparent decrease in fragment coverage. (C) Prespacer generation efficiency in the 1-Mb PPS region calculated as the ratio of prespacer-like fragments from the 1-Mb PPS region to prespacer-like fragments from the 1-Mb control region. Coordinates of the 1-Mb PPS and control regions are shown in A. (D) Prespacer generation efficiency in the 0.1-Mb PPS region calculated as the ratio of prespacer-like fragments from the 0.1-Mb PPS region to prespacer-like fragments from the 1-Mb control region. Coordinates of the 0.1-Mb PPS region and 1-Mb control region are shown in A. * indicates $p<0.05$ for a Chi-square test comparing the amounts of prespacer-like fragments in the control and the PPS region between *wt* and other samples.

Overall, these results suggest a key role of RecJ in prespacer generation and demonstrate that RecBCD also participates in the process. Therefore, a decrease in prespacer generation efficiency is the likely reason for the low adaptation level in Δ*recD*, Δ*recB*, Δ*recC*, Δ*recJ*, and Δ*recB* Δ*recJ* strains.

Non-prespacer-like fragments could be generated during the FragSeq procedure or represent specific DNA fragments that exist in the cell. In Figure 25A-B, prespacer-like fragments are indicated with purple color, other fragments are black. While this work is dedicated to prespacer-like fragments, it is evident that non-prespacer-like fragments have specific patterns of accumulation that are affected by *rec* mutations.

4.2.5.RecJ trims single-stranded 5′ ends in the presence of Cas1-Cas2 up to protospacer boundaries in vitro

Our *in vivo* results suggest that 5′ ends of prespacer precursors formed in self-targeting cells are trimmed by the RecJ 5′-3′ exonuclease up to the boundaries of mature 33/37-nt prespacers. To reconstruct this process *in vitro*, we used a model DNA substrate based on a previously characterized dual forked substrate containing a 23-bp central duplex flanked by 5-nt 3′ and 5′ noncomplementary overhangs (Nuñez, Harrington, et al., 2015). To achieve stronger binding by Cas1-Cas2 and make the substrate more similar to prespacers observed *in vivo* (Musharova et al., 2021; Shiriaeva et al., 2019), we introduced

the 3′-TTC-5′ PAM and two additional nucleotides in the PAM-proximal 3′ overhang. To provide the recognition and processing of 5′ overhangs by RecJ (Han, 2006; Lovett & Kolodner, 1989), we extended both single-stranded 5′ ends to 19 nt.

The affinity of the Cas1-Cas2 complex toward the resulting extended DNA substrate was estimated using electrophoretic mobility shift assay. The Cas1-Cas2 concentration needed to bind essentially all the substrate under the conditions of the experiment was 1.6 μM [a dissociation constant (Kd) of ~0.5 μM]. Parallel reactions containing DNA substrates 3′-terminally labeled at either of the two strands were treated with RecJ in the presence or absence of saturating amounts of Cas1-Cas2. In the absence of Cas1-Cas2, RecJ processed ~30% of substrate molecules to very short fragments that were not resolved on the gel (Figure 26, lanes 3 and 10). Approximately 20 to 50% of substrates were digested up to 5 nt within the duplex, reflecting RecJ's limited digestion within dsDNA connected to a single-stranded 5′ end (Han, 2006). In contrast, in the presence of Cas1-Cas2, digestion stopped at protospacer boundaries on both sides of the DNA substrate, yielding a mature 33/37-nt prespacer and a 33/38-nt version (Figure 26, lanes 5 and 12).

The cleavage specificity of RecJ was compared to that of *E. coli* ExoVII exonuclease. As can be seen from Figure 26 (lane 11), in the absence of Cas1-Cas2, ExoVII completely digested the bottom strand of the model substrates to products that were not resolved on the gel. Most of the top strand was processed similarly, but some products located at the prespacer boundary (33 nt), as well as shorter (29 to 30 nt) digestion intermediates, were observed (Figure 26, lane 4). In the presence of Cas1-Cas2, the DNA substrate was stabilized against ExoVII digestion, as expected; however, the digestion products were 1 to 2 nt longer than those generated by RecJ (Figure 26, lanes 6 and 13).

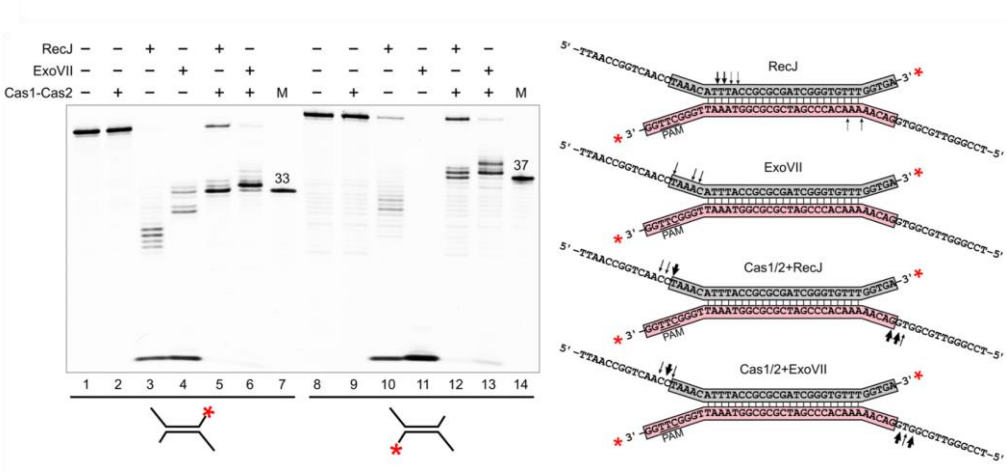


Figure 26. Processing of prespacer 5′ ends by RecJ, ExoVII, and RecBCD nucleases. A double-forked DNA substrate composed of a 23-bp central duplex with single-stranded extensions (shown on the right) was labeled at one of the 3′ ends with fluorescein (shown by asterisks) and treated with RecJ or ExoVII exonucleases in the presence or absence of Cas1-Cas2.

Considering also the results on primed adaptation efficiency in Δ*recJ*, Δ*xseA*, and Δ*recJ* Δ*xseA* mutants, and the detection of prespacers with extended 5′ ends *in vivo* in the Δ*recJ* mutant, the *in vitro* data is thus consistent with the idea that ExoVII performs the prespacer 5’ end processing in Δ*recJ* strain and is responsible for nearly all of the residual adaptation in this mutant. The results of the oligo transformation experiments also suggest that both RecJ and ExoVII are important for prespacer generation but not for their incorporation into CRISPR arrays.

4.2.6.RecBCD-assisted processing of double-stranded prespacer ends by RecJ in vitro

Our *in vivo* data indicate that the RecBCD helicase activity promotes adaptation, but its nuclease activity may be dispensable since nuclease-deficient RecBC promotes prespacer formation. One can envision that when Cas1-Cas2 binds to a prespacer precursor that has terminal double-stranded regions, RecBCD unwinds these regions, enabling access of RecJ to 5′ ends. To test this idea *in vitro*, we used a double-forked DNA substrate with an extended PAM-proximal 5′ end annealed to a complementary oligonucleotide such that

a 15-bp double-stranded segment with a blunt end was formed (Figure 27). In the absence of Cas1-Cas2, RecBCD fully degraded this substrate, while RecJ demonstrated very little activity (lanes 3 and 5), suggesting that a blunt end is a suitable substrate for the entry of RecBCD but not RecJ, as expected (Han, 2006; Taylor & Smith, 1985, 1995a). RecBC had little effect on the substrate in the absence of Cas1-Cas2 (lane 4), also as expected, as it is a helicase devoid of nuclease activity. We attribute the partial degradation of the substrate to low–molecular weight products (Figure 27, lane 4) to the small amount of RecBCD contaminating our RecBC preparation, which was obtained during RecBCD purification as a side fraction that lost the weakly held RecD subunit.

When the substrate was bound by Cas1-Cas2, the addition of RecBC had no considerable effect, while RecBCD removed 3 to 4 nt from the double-stranded end, generating a PAM-proximal 5′ end located 15 to 16 nt upstream of the protospacer boundary (Figure 27, lane 7), which is about the distance from the forward end of RecBCD to the nuclease active site (Singleton et al., 2004). Thus, RecBCD nuclease is unable to generate 5′ ends of mature prespacers. When either RecBCD or RecBC was added to Cas1-Cas2–bound DNA substrate together with RecJ, the substrate was fully converted to a 33-nt fragment whose 5′ end was located precisely at the prespacer boundary (Figure 27, lanes 8 and 9). We conclude that the RecBCD helicase unwinds the dsDNA flanking the protospacer bound to Cas1-Cas2, cuts it 15 to 16 nt from the PAM, but allows further digestion by RecJ to produce a correct prespacer.

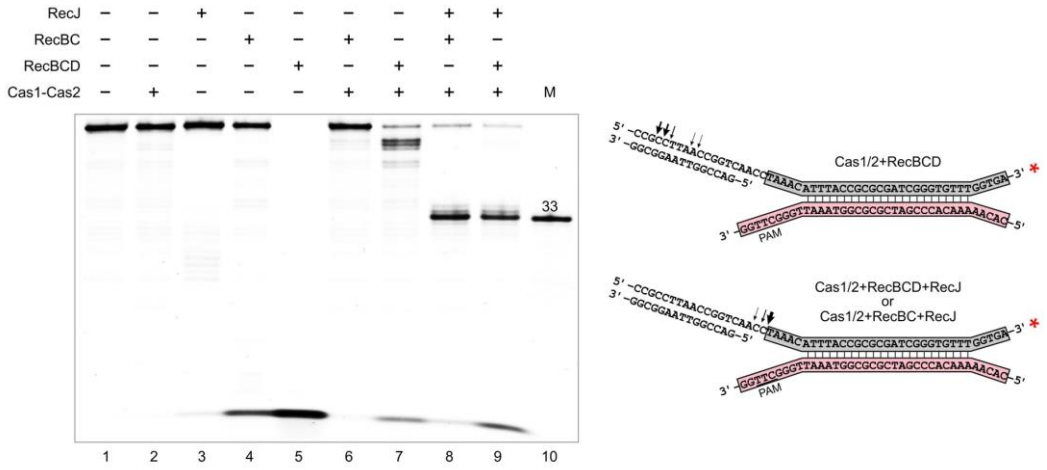


Figure 27. Processing of prespacer 5′ ends by RecJ and RecBCD nucleases. A modified double-forked DNA substrate (shown on the right) was labeled with fluorescein and used to study RecBC- and RecBCD-assisted processing of double-stranded terminal regions by RecJ. Reaction products were resolved by denaturing PAGE. Fluorescein-labeled 33- and 37-nt oligonucleotides (highlighted, respectively, in gray and pink in schemes on the right) were used to map PAM-proximal and PAM-distal boundaries of the DNA substrate protection by Cas1-Cas2. Only the fluorescein-labeled 33-nt oligonucleotide was used as a marker. Arrows show the positions of cleavage sites observed in the presence of different proteins.

Since the RecBCD helicase activity is required for prespacer generation, in the absence of RecBCD, other helicase(s) could be involved. One such helicase is RecQ of the RecFOR pathway of homologous recombination, which also includes RecJ (Umezu et al., 1990). We generated self-targeting Δ*recQ* and Δ*recB* Δ*recQ* mutants and assessed their primed adaptation efficiency. While the *recQ* mutant was not statistically different from *wt*, the double mutant was strongly affected (a sevenfold defect compared to the *recB* single mutant; Figure 28), indicating that RecQ can contribute to primed adaptation, at least when RecBCD is not active.

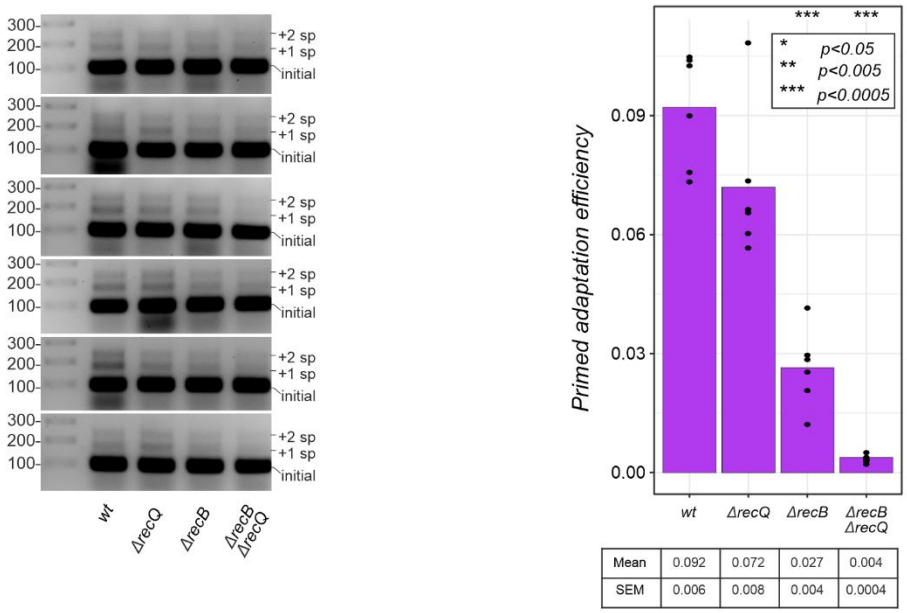


Figure 28. Efficiency of primed adaptation in the *wt*, Δ*recB*, Δ*recQ*, and Δ*recB* Δ*recQ* strains. On the left panel, amplification products of unexpanded CRISPR arrays ("initial") and CRISPR arrays with one ("+1 sp") or two ("+2 sp") additional spacer-repeat units resolved by electrophoresis in agarose gel are shown. A molecular-weight size marker is shown on the left. Results of 6 independent biological replicates are presented. On the right panel, primed adaptation efficiency is calculated as the ratio of newly acquired spacers to the total number of CRISPR arrays determined by high-throughput sequencing of CRISPR array amplicons. Asterisks indicate statistically significant differences between a mutant and the *wt* ($p<0.05$) in pairwise Welch's t-test. Points represent individual values obtained in 6 biological replicates. For each strain, the mean and standard error of the mean are shown in the table under the bar graph.

4.3. Detection of the half-integrated prespacers *in vivo* during primed CRISPR adaptation

The integration of a new spacer into CRISPR array proceeds in several steps. First, 3'-OH of the prespacer attacks the phosphodiester bond between the leader and the first repeat in CRISPR array, leading to the formation of the half-site intermediate. Then, the second 3'-OH attacks another strand, at the first repeat-spacer junction (Nuñez et al., 2016). The product of these reactions is a new dsDNA spacer flanked by the two ssDNA repeats, which are then amplified and ligated.

A new method developed in our lab (Vyhovskyi et al., not published) allows to detect the half-integrated intermediates of CRISPR adaptation (Figure 29). A double-strand adapter with a 5' overhang is ligated to the DNA isolated from cells with induced CRISPR adaptation. The adapter is ligated to all free 5' ends in DNA, including half-integrated prespacers (HIPs). In order to amplify HIPs specifically, PCR reaction is done with a pair of primers, one of which anneals to the adapter and another one anneals to the CRISPR array. Depending on the location of the primer on the CRISPR array, it is possible to amplify HIPs from the top or the bottom strand. The second adapter is introduced during another PCR, followed by the library indexing and sequencing on the Illumina platform.

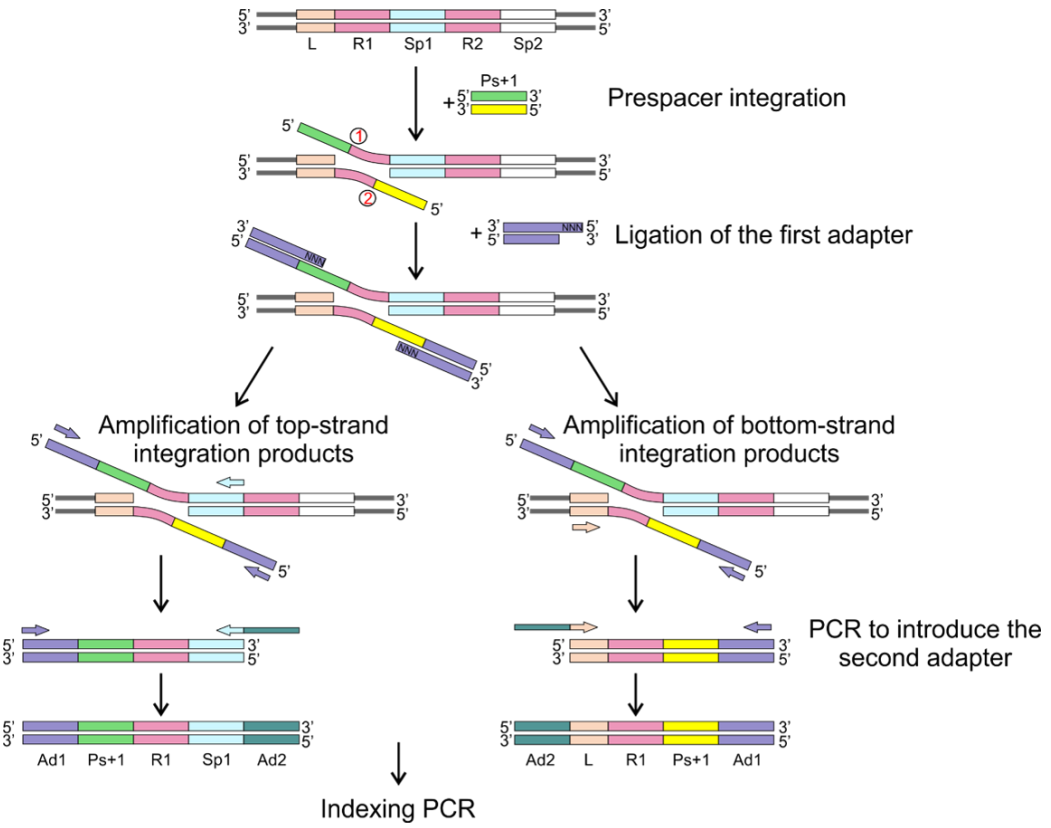


Figure 29. Detection of half-site intermediates *in vivo* (Vyhovskyi et al., not published). The figure was provided by Dr. Anna Shiriaeva.

We used the method of HIPs detection to compare the length and distribution of HIPs in induced self-targeting KD403 and derived mutant strains. The author prepared samples for HIP detection. Sequencing data analysis and pictures were obtained by Dr. A. Shiriaeva.

The identified HIPs were mapped on KD403 genome (Figure 30). For most mutant strains, the distribution looks similar to *wt*, except for the *recJ xseA* mutant, for which the pattern is much more flattened and does not have a pronounced peak at PPS. Consistently, CRISPR adaptation efficiency in the *recJ xseA* mutant was strongly reduced (Figure 22). We also identified HIPs for *recB recJ* mutant, though we could not detect prespacers in this strain (Figure 25). Thus, HIP detection is a more sensitive method than FragSeq and should be used when adaptation efficiency is low.

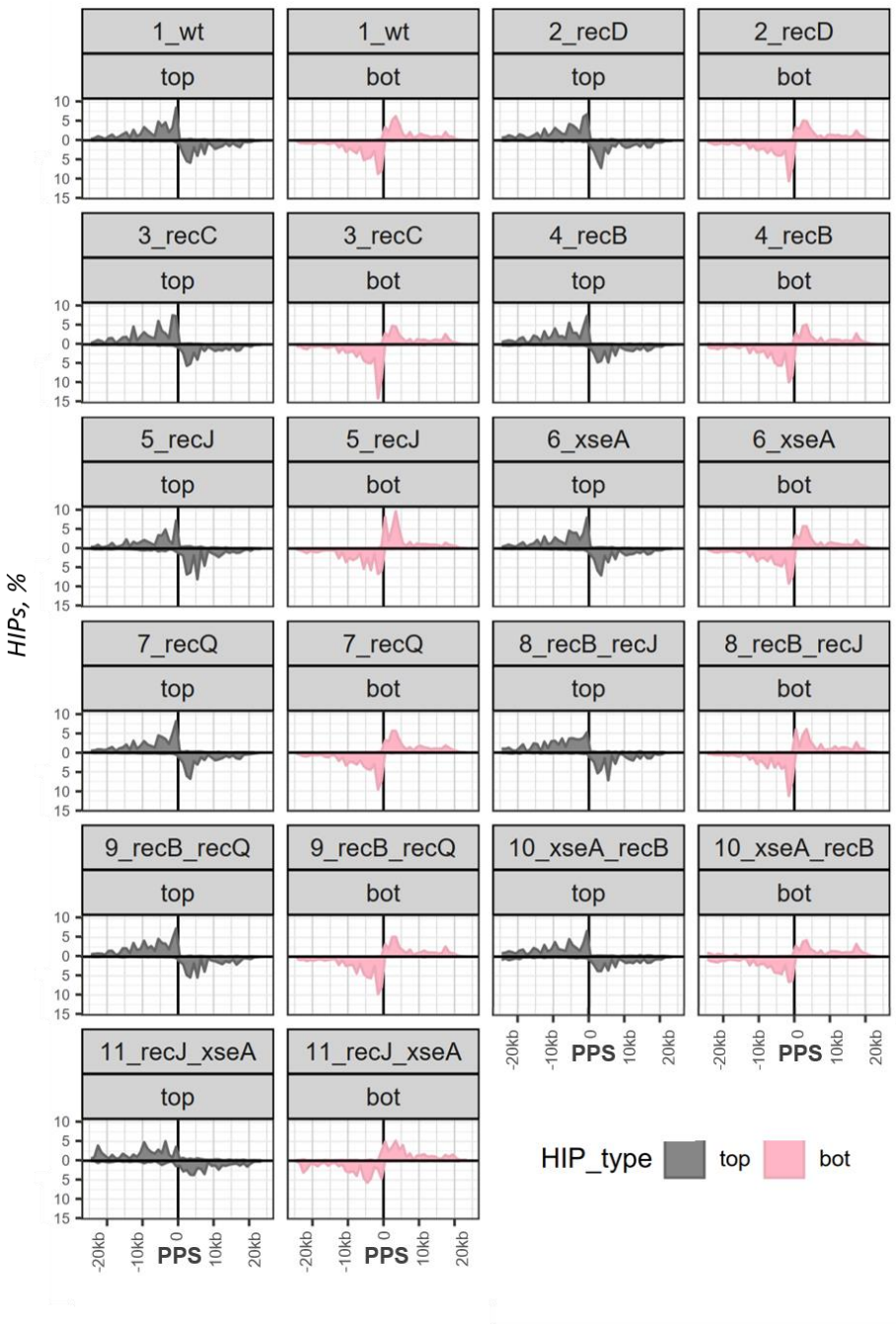


Figure 30. Chromosomal distribution of half-integrated prespacers. The mean is shown (where number of replicates >1).

Next, we assessed the length distribution of identified HIPs (Figure 31). In line with the results on prespacer length distribution (Figure 20), both top and bottom HIPs are generally 1 nt shorter for *recB*, *recB recQ*, and *recC* mutant strains, while for *recJ* mutant most HIPs are 1 nt longer. For *recB recJ* strain, most HIPs are 1 nt longer than in *wt*, while the amount of shorter HIPs is also high, which points to the additive effect of *recB* and *recJ* mutations. One could suggest that there are other proteins compensating for the helicase RecBCD activity in *recB* mutant, or that a part of prespacer precursors formed during primed CRISPR adaptation have single-strand 5' ends which are processed by RecJ or ExoVII without participation of RecBCD. Interestingly, there is no significant effect of double *recB recQ* mutation on HIPs length and distribution compared to *recB* (Figure 31), though it substantially reduced CRISPR adaptation efficiency (Figure 28).

For the *recJ xseA* mutant, the HIP length distribution resembles a non-specific background, consistent with the abolished CRISPR adaptation in this strain.

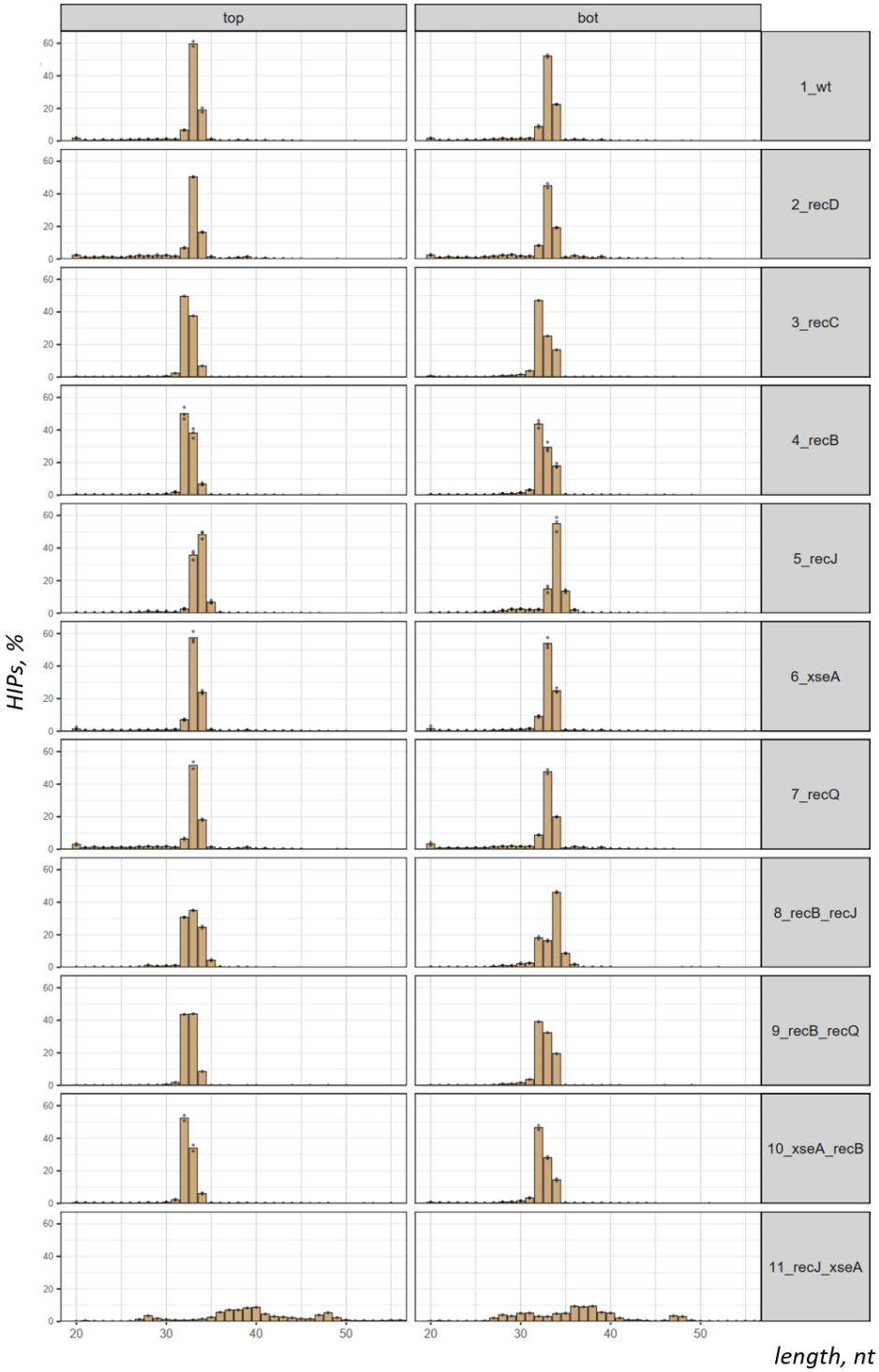


Figure 31. Length distribution of HIPs mapped to the 50-kb PPS region. The mean is shown (where number of replicates >1).

Thus, using the new method, we were able to detect the half-site intermediates of primed CRISPR adaptation in self-targeting strain and its mutant derivatives. Overall, the obtained results are consistent with the data on prespacer analysis. They additionally prove that ExoVII encoded by *xseA* likely replaces the RecJ function in its absence, and double *recJ xseA* mutation abolishes the formation of adaptation intermediates. In contrast, double *recB recJ* mutation, which also dramatically reduces CRISPR adaptation, allows the formation of CRISPR adaptation intermediates of the required length, suggesting the additive roles of RecB and RecJ in prespacer formation. Additionally, the results propose that RecQ has no significant or direct role in primed CRISPR adaptation.

4.4. Bacteriostatic antibiotics promote CRISPR-Cas adaptive immunity by enabling increased spacer acquisition

The results presented in this section were obtained in parallel with the work on the main project. While they are not directly related to the main project objectives, they provide important insight into CRISPR-Cas immunity and are therefore included in the thesis. A part of these data was published in the following article:

Dimitriu, T., **Kurilovich, E.**, Lapinska, U., Severinov, K., Pagliara, S., Szczelkun, M., Westra, E. (2021) Bacteriostatic antibiotics promote CRISPR-Cas adaptive immunity by enabling increased spacer acquisition. *Cell Host Microbe, 30(1),* 31-40.e5.

The author of this Thesis participated in evolution experiments in LB medium and measured MIC and MBC for antibiotics (Figure 32A). The author also performed evolution experiments in M9 medium and analyzed the results (Figure 39, Figure 40). The rest of the data described in this section were obtained by Dr. T. Dimitriu.

4.4.1.Bacteriostatic antibiotics promote CRISPR-Cas immunity

To understand the implications of antibiotic treatment on the evolution of CRISPR-Cas immunity in *P. aeruginosa*, we studied how they influence the evolutionary dynamics of *P. aeruginosa* PA14 in response to phage DMS3vir. We infected PA14 cultures grown in a rich medium supplemented with sub-inhibitory concentrations of 8 different antibiotics with phage DMS3vir. Of these antibiotics, four are bactericidal (ciprofloxacin [Cip], streptomycin [Strep], gentamycin [Gen] and carbenicillin [Carb]) and four are bacteriostatic (chloramphenicol [Chl], tetracycline [Tc], erythromycin [Erm], and trimethoprim [Tm]) against *P. aeruginosa*.

Of the bacteriostatic antibiotics, Chl and Tc had the strongest effect on bacterial exponential growth rate at the concentrations used, whereas Erm had little effect on exponential growth but instead slowed growth at a later stage when bacteria reached high densities (Figure 32).

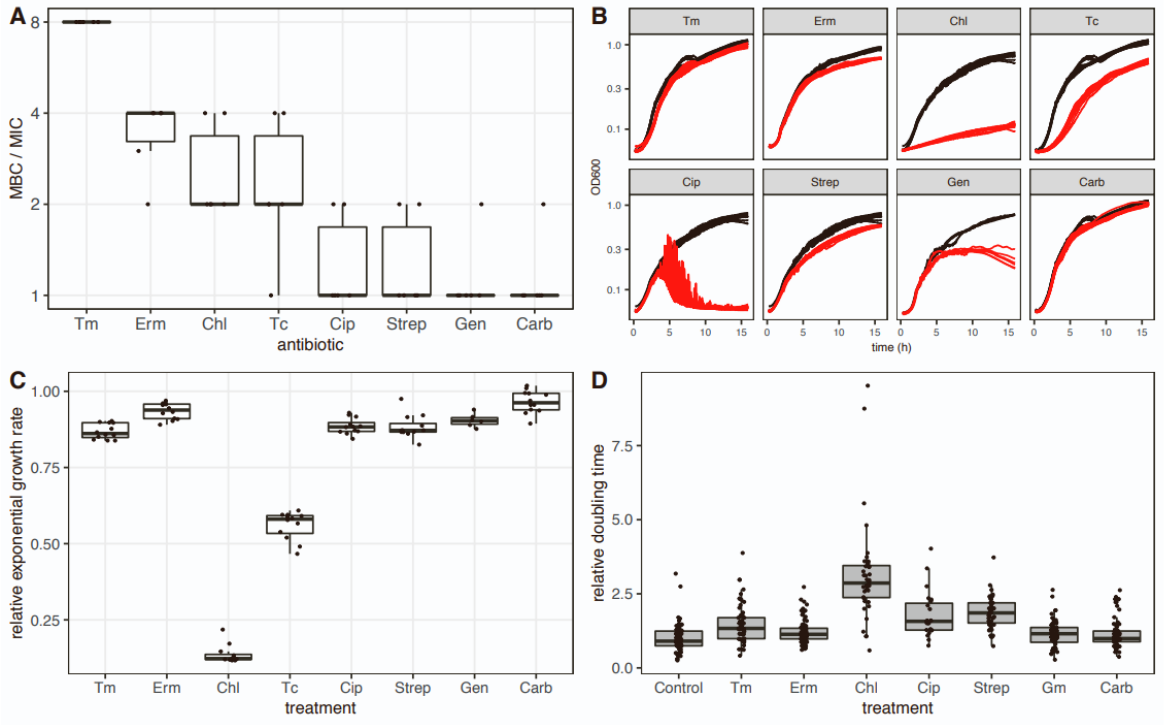


Figure 32. The effects of different antibiotics on *P. aeruginosa* PA14 growth. (A) Determination of bactericidal or bacteriostatic effects of antibiotics. MBC (minimum bactericidal concentration) vs MIC (minimum inhibitory concentration) concentration ratio is shown for each antibiotic. Bacteriostatic antibiotics have on average MBC/MIC ≥ 2, whereas bactericidal antibiotics have MBC/MIC ~1. (B) OD600 growth curves for all tested antibiotics (no antibiotic treatment in black, antibiotic treatment in red), at the antibiotic concentrations used for the evolution experiment. The shape of curves with Cip treatment was due to the formation of cell aggregates. (C) Antibiotic effect on the exponential growth rate, measured from OD600 growth curves. For each antibiotic, the relative exponential growth rate was calculated as the ratio of the exponential growth rate with antibiotic to the average exponential growth rate of no antibiotic control during the same experiment. (D) Antibiotic effect on cell doubling time, measured in microfluidic channels. In A, C, and D, the center value of the boxplots shows the median, boxes show the first and third quartile, and whiskers represent 1.5 times the interquartile range; individual data points are shown as dots.

Most antibiotics delayed phage epidemics and subsequently the acquisition of phage resistance (Figure 33A, B). Nonetheless, at 3 days post-infection (d.p.i.) phage resistance was essentially fixed in all cultures (Figure 33C).

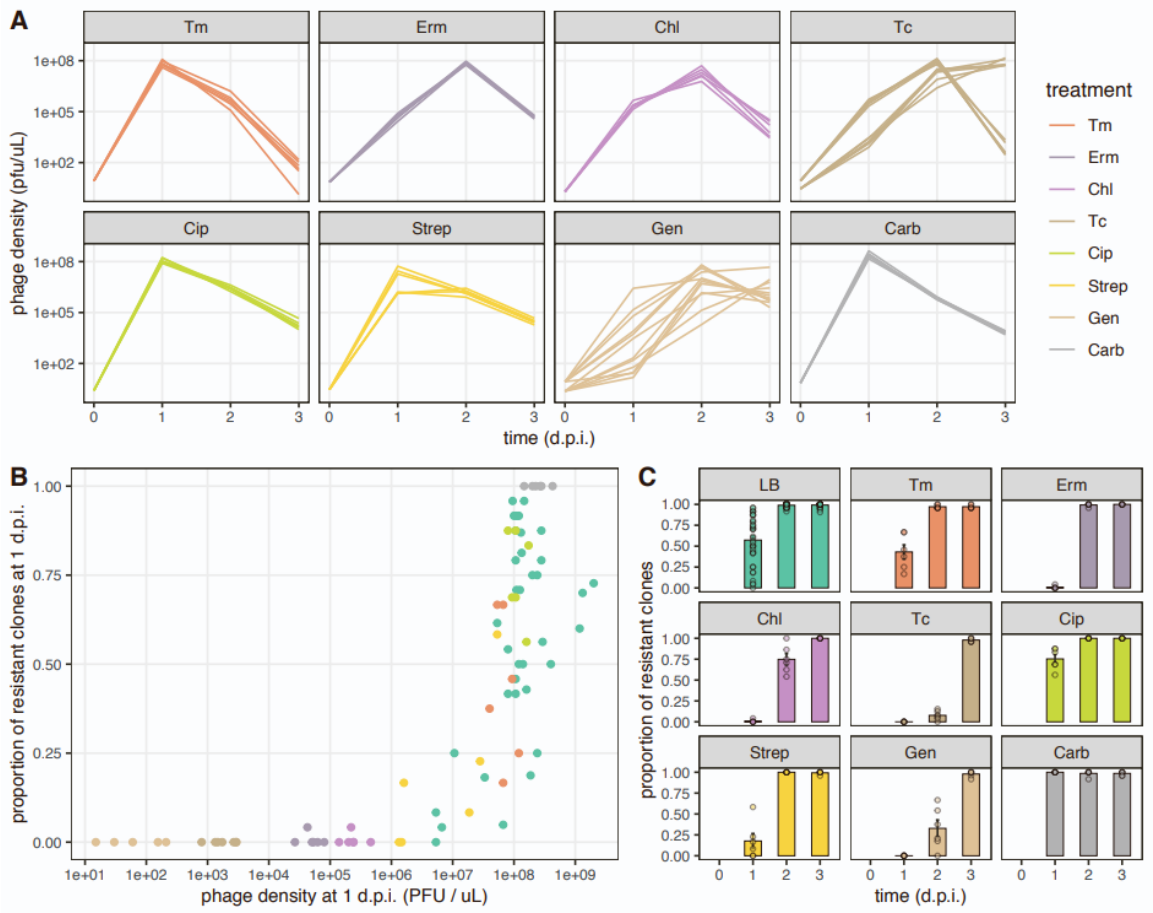


Figure 33. (A) Phage population dynamics. Densities of phages are shown as a function of time of each antibiotic treatment. (B) The proportion of clones resistant to DMS3vir at 1 d.p.i. as a function of phage density measured at 1 d.p.i. The proportion of resistant clones present after 1 day of infection is higher in populations in which phage spread rapidly during the first 24h. In A and B, individual replicates are shown (n=6 per treatment). (C) The proportion of clones resistant to DMS3vir per treatment over time.

Strikingly, the type of phage resistance that had evolved was strongly dependent on the presence and the type of antibiotic. In the absence of antibiotics, or in the presence of bactericidal antibiotics, only a minority of bacteria evolved CRISPR-Cas immunity, whereas a large proportion of the bacterial population evolved CRISPR-Cas immunity in the presence of bacteriostatic antibiotics (Figure 34A). These data, and the fact that Chl,

Tc, Erm, and Tm have different modes of action, suggest that bacteriostatic antibiotics promote the evolution of CRISPR-Cas immunity because they limit bacterial growth rates.

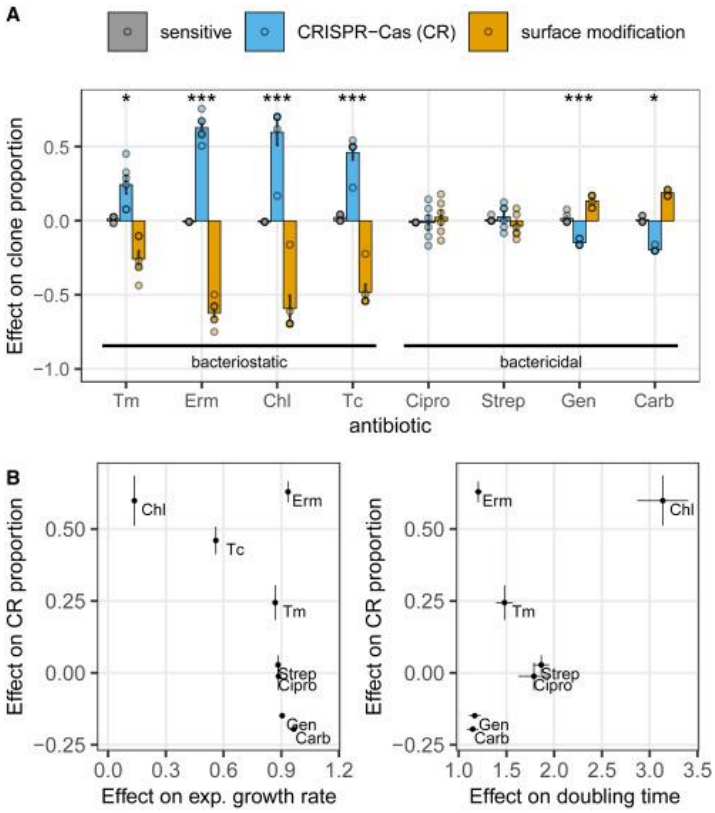


Figure 34. (A) Effect of each tested antibiotic on the proportion of phage sensitive clones (gray), and clones resistant to phage due to acquisition of CRISPR-Cas (blue) or SM (surface modification) (yellow) resistance at 3 d.p.i., calculated as the change in the proportion of each type of clone upon antibiotic treatment relative to the average proportion of the same type in no-antibiotic control. Bars and error bars show mean ± SEM, and individual biological replicates are plotted as dots (n = 6). Asterisks indicate treatments where the proportion of CRISPR-Cas resistant clones is significantly different from the associated control (∗ 0.01 < p < 0.05; ∗∗∗ p < 0.001). Antibiotics are ordered from left to right by decreasing MBC/MIC ratio, a measure of their bacteriostatic versus bactericidal activity. (B) Antibiotic effect at 3 d.p.i. evolved CRISPR-Cas resistance (CR), shown as a function of antibiotic effect on exponential growth rate (measured by OD600 change in 96-well plates, left) and as a function of antibiotic effect on doubling time measured in microfluidics (right). Dots and error bars show mean ± SEM, respectively.

To better understand the relationship between bacterial growth and evolution of CRISPR-Cas immunity, we first measured bacterial growth rates in the presence of each antibiotic at concentrations used in our evolution experiments. Analysis of exponential growth rates in batch culture (based on the optical density, OD600, of the cultures) and doubling times of individual cells in a microfluidics device showed that Chl and Tc cause particularly slow growth, and this is associated with the increased proportion of CRISPR-Cas immune bacteria in the population (Figure 34B). More generally, this analysis revealed a correlation between exponential growth rate and the evolution of CRISPR-Cas immunity, except for Erm. Erm is known, similarly to other macrolides, to mostly affect bacterial growth and gene expression at high cell densities, instead of early exponential phase (Tateda et al., 2007) (consistent with our OD600 growth curves, in which growth is affected mostly after 10 h, and not measurable with our microfluidics setup). In our evolution experiments, phages undergo epidemic spread during the first 24 h, and the majority of phage infections occur after 12 h (Figure 35). Thus, in the presence of all bacteriostatic antibiotics, most phage infection events will happen in cells experiencing reduced growth rates.

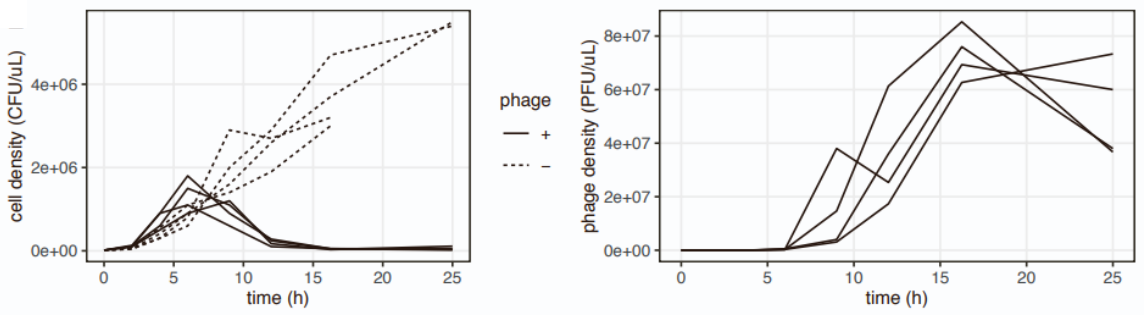


Figure 35. Population dynamics in the first 24h in the absence of antibiotics. Cell density (left) and phage density (right) are shown as a function of time (n=4).

4.4.2.Bacteriostatic antibiotics promote spacer acquisition

The effect of bacteriostatic antibiotics on phage resistance could be potentially explained by a change in mutation rates toward SM (surface modification) or toward CRISPR-Cas immunity (by insertion of phage-derived spacers into the host CRISPR loci),

or the fitness cost of these alternative phage resistance mechanisms. To measure the effects of antibiotics on the frequency of CRISPR adaptation events, we performed short-term (3 h) infection assays and measured the proportion of bacteria that had acquired CRISPR-Cas immunity in the presence or absence of each antibiotic. The time was chosen to limit the possible effects of natural selection on the frequencies of CRISPR-Cas immune bacteria in the population. Short-term infection experiments revealed that in the presence of all bacteriostatic antibiotics, more cells acquired CRISPR-Cas immunity, whereas bactericidal antibiotics had no effect (Figure 36A). Across antibiotics, the frequency of CRISPR adaptation in these short-term experiments was significantly correlated to the levels of CRISPR-Cas immunity that evolved at 3 d.p.i. (Figure 36B, Pearson's correlation $t_{1,6} = 3.9$, $p = 0.008$, $\rho = 0.85$).

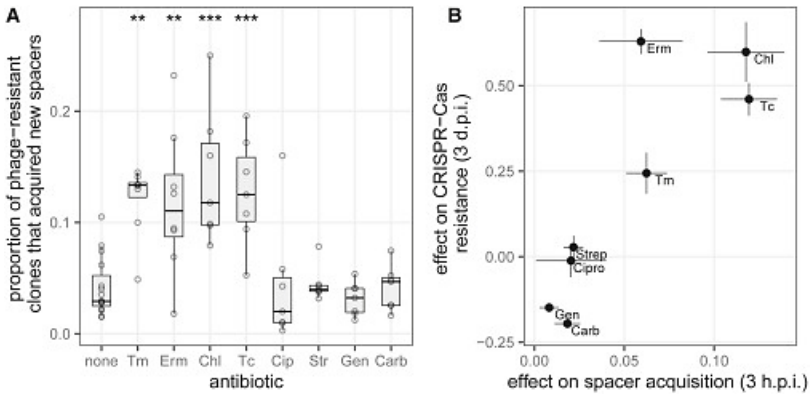


Figure 36. (A) The proportion of resistant clones that are CRISPR-Cas immune after 3 h phage infection. The center value of the boxplots, boxes, and whiskers represent the median, respectively, first and third quartile, and 1.5 times the interquartile range; dots show individual data points (n= 6). Asterisks show treatments significantly different from the no-antibiotic control (Tukey HSD; $**0.001 < p < 0.01$; $***p < 0.001$). (B) Average change in proportion per treatment plotted against the average increase in the proportion of CRISPR-Cas immune clones in the evolution experiments, error bars showing SEM (N = 6).

In contrast, none of the antibiotics except Cip affected the rates at which bacteria with SM are generated (Figure 37A). Thus, bacteriostatic antibiotics have no effect on SM resistance acquisition.

We next tested whether antibiotics impact the way selection acts on clones with CRISPR-Cas immunity and SM mutants. Competition experiments between a clone with CRISPR-Cas immunity (BIM2) and an SM-resistant clone (3A) showed that the presence of bacteriostatic antibiotics had either no impact or reduced the fitness of CRISPR-Cas immune bacteria relative to receptor mutants (Figure 37B). As some treatments led to an absolute decline in the abundance of the CRISPR-Cas resistant clone, we used selection rate rather than relative fitness as a measure of antibiotic effects on fitness (Lenski et al., 1991).

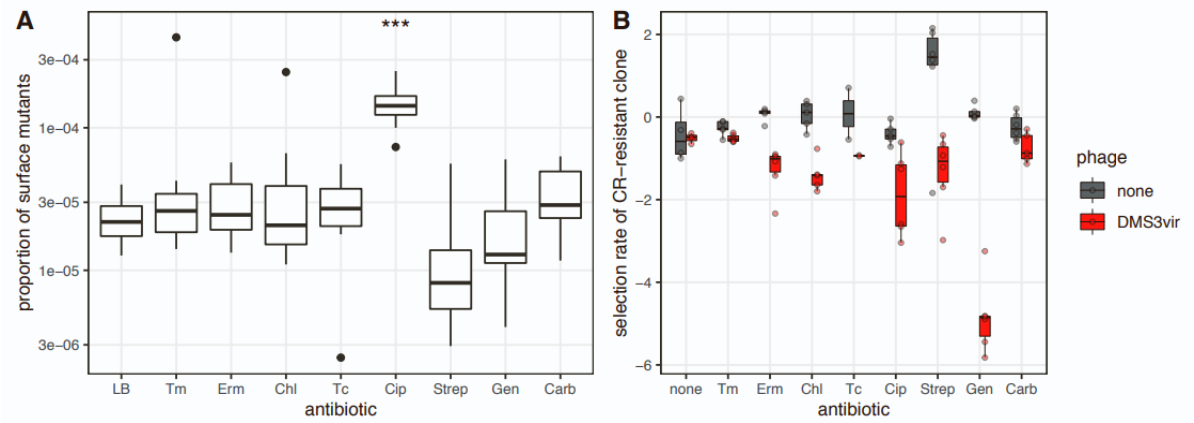


Figure 37. (A) The proportion of SM clones in populations grown in the absence of phage. The center value of the boxplots shows the median, boxes the first and third quartile, whiskers represent 1.5 times the interquartile range and dots are outliers (n=18 from 3 independent experiments). Asterisks show treatments significantly different from the no-antibiotic controls (***, $p<0.001$). (B) Resistant clones' fitness effects. The selection rate of the CRISPR-Cas resistant clone BIM2 against the surface mutant 3A is plotted as a function of antibiotic treatment, in the absence (gray) or presence (red) of phage DMS3vir. The center value of the boxplots shows the median, boxes the first and third quartile, and whiskers represent 1.5 times the interquartile range; individual data points are shown as dots (n=6).

These results show that bacteriostatic antibiotics likely increase phage resistance due to CRISPR-Cas immunity by an increase in the frequency of CRISPR adaptation.

4.4.3.Bacteriostatic antibiotics slow down progeny phage production

Bacteriostatic and bactericidal antibiotics impact cell metabolism differently and lead, respectively, to decreased and increased cell metabolic rates (Lobritz et al., 2015). Because phage production is dependent on the metabolism and protein synthesis machinery of the host (Hadas et al., 1997; L. You et al., 2002), we hypothesized that bacteriostatic antibiotics may slow down phage development, providing a larger window of time for the CRISPR-Cas immune system to acquire spacers from the phage prior to irreversible cell damage or cell death. To test this hypothesis, we performed one-step phage growth assays to detect when mature intracellular phages are produced. First, we inoculated phages to cell cultures grown in each antibiotic treatment for 12 h (the time at which the majority of infections take place in our evolution experiments). We found that all bacteriostatic antibiotics caused a strong reduction in phage production compared with cells cultured in the absence of antibiotics (Figure 38A). Under those conditions, we were unable to analyze the effect of bactericidal antibiotics on phage production, due to high rates of cell death in these treatments. To directly compare the effects of bacteriostatic and bactericidal antibiotics on phage progeny formation, we repeated these assays using a shorter period of antibiotics exposure (30 min, at which point bactericidal antibiotics do not affect viable cell density). Interestingly, we found that Erm had no effect on phage production (Figure 38B), consistent with its minor effect on the exponential growth rate of the bacteria. All other bacteriostatic antibiotics (Chl, Tc, and Tm) delayed or decreased the formation of infectious progeny phages (Figure 38B). Bactericidal antibiotics had more variable effects: Cip and Carb showed no interference with the production of infectious phages, in agreement with their mechanism of action and known synergy with phage therapy (Comeau et al., 2007), whereas the presence of Strep or Gen (both aminoglycosides) resulted in very little phage production (Figure 38B).

Finally, to understand whether some of these antibiotics might inhibit phage production altogether, we carried out highly replicated experiments (96 bacterial cultures per treatment) of bacterial populations, each infected with a very small initial dose of phage (around 5 phage particles per infection experiment). After 24 h of infection, we measured

how the proportion of successful phage amplification depends on the antibiotics present, hence providing an estimate of their impact on the probability of successful phage infection (Figure 38C). In the absence of antibiotics or in the presence of Tm, Chl, Cip, and Carb, all infected populations ultimately produced new phages. However, phage amplification was abolished in almost all infection experiments in the presence of Gen, and in the majority of infection experiments in the presence of Strep. Thus, Strep and Gen totally inhibit phage production (Figure 38C) (Kever et al., 2022). Interestingly, we also observed a small but significant increase in failed infections in the presence of Erm and Tc. While this may contribute to the evolution of CRISPR-Cas immunity (Hynes et al., 2014), it is insufficient to explain the effects of bacteriostatic antibiotics in general, since these had only small (Erm and Tc) or no (Tm, Chl) effect on the proportion of unproductive infections (Figure 38C). With most infections being ultimately productive, the delayed production of infectious phage particles observed for all static antibiotics (Figure 38A) is thus due to a delay in the phage eclipse period.

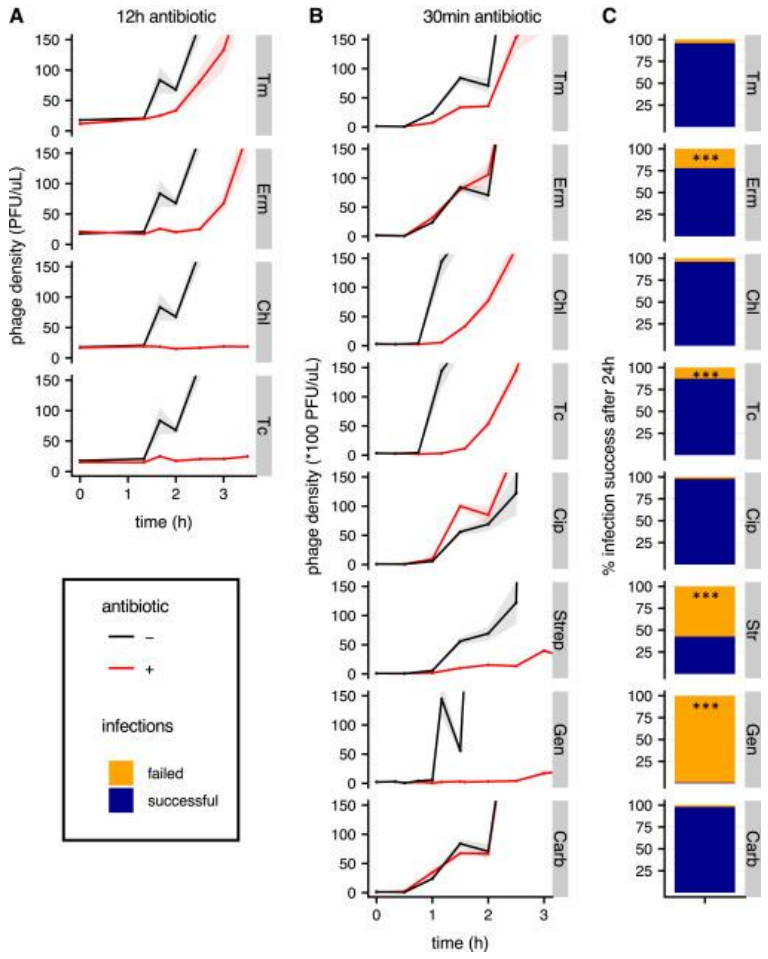


Figure 38. (A and B) Effect of antibiotics on phage production dynamics. Phage density over time during infection of cells that were pre-exposed to antibiotics for 12 h (A) or 30 min (B) is shown in red and no-antibiotic controls in black. The y-axis is cropped to focus on antibiotic treatment dynamics. Lines and shaded areas are, respectively, mean and SEM (n= 4). (C) Effect of antibiotics on the frequency of failed phage infections. 96 parallel populations were infected with a low number of phages per population (~5) and grown for 24 h. Bar plots show the percentage of populations with failed infection (no phages detected after 24 h) or successful infection (infectious phages present after 24 h). Asterisks indicate antibiotics with a significant increase in the number of populations with no phages (chi-square tests, ***$p < 0.001$).

Collectively, these data show that bacteriostatic antibiotics cause phage to replicate more slowly and increase the frequency at which initially sensitive cells acquire CRISPR-

Cas immunity. We propose that there is a causal relationship between these two phenomena.

4.4.4.Bactericidal antibiotics inhibit CRISPR-Cas immunity in M9 medium

In contrast to bacteriostatic antibiotics, which inhibit cell growth, bactericidal ones accelerate cell metabolism (Dwyer et al., 2007; Kohanski et al., 2007, 2008; Lobritz et al., 2015) and eventually lead to cell death. While bacteriostatic antibiotics promoted CRISPR-Cas immunity in rich LB medium, bactericidal antibiotics Gen and Carb demonstrated the opposite effects (Figure 34A). Still, in the absence of antibiotics the vast majority of cells acquired receptor mutations during infection in LB, so the effects of bactericidal antibiotics were not prominent in LB. We conducted similar evolution experiments with the same set of antibiotics in a minimal M9 medium and calculated the proportions of SM and CRISPR-Cas resistant clones after 5 d.p.i. (Figure 39).

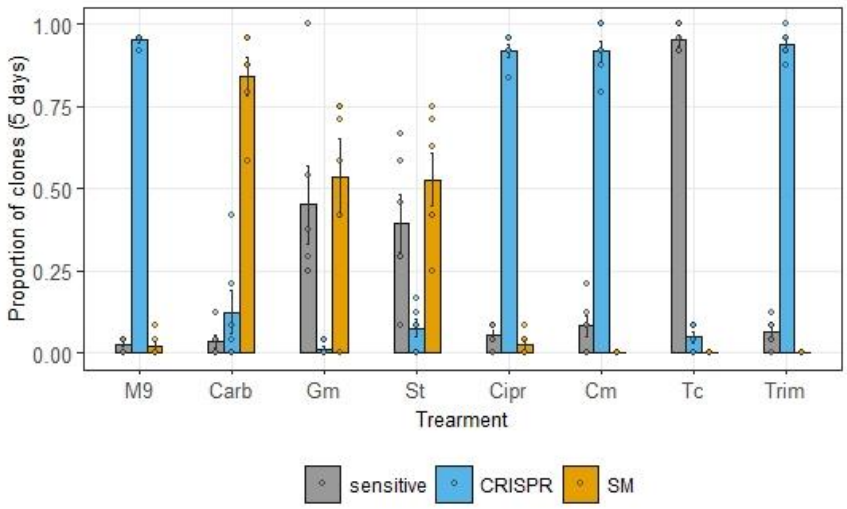


Figure 39. Proportion of phage-sensitive, CRISPR-Cas and SM phage-resistant clones at 5 d.p.i. in M9 medium. Bars and error bars show mean ± SEM, individual biological replicates are plotted as dots (n = 6).

Consistently with previous results, the majority of cells acquired CRISPR-Cas resistance in M9 medium without antibiotics (Westra et al., 2015) or with added Cipr, Cm,

Tc, and Trim. On the contrary, in the presence of bactericidal Carb, Gm and St a small number of clones acquired CRISPR-Cas immunity, while a large proportion of bacteria acquired receptor mutations (Figure 39). Interestingly, many clones remained phage sensitive in the presence of Gm and St, and almost all clones appeared to be sensitive at 5 d.p.i. in samples with added Tc. Further experiments are required to investigate the phenomena.

To check if antibiotics could influence the relative fitness cost of CRISPR-Cas vs. SM immunity in M9, competition experiments with CRISPR-Cas resistant and resistant SM clones were conducted in this medium with added antibiotics. Similar to the results obtained in LB (Figure 37B), most antibiotics did not influence the fitness of CRISPR-Cas resistant clones, while there was a pronounced effect for Gm (Figure 40).

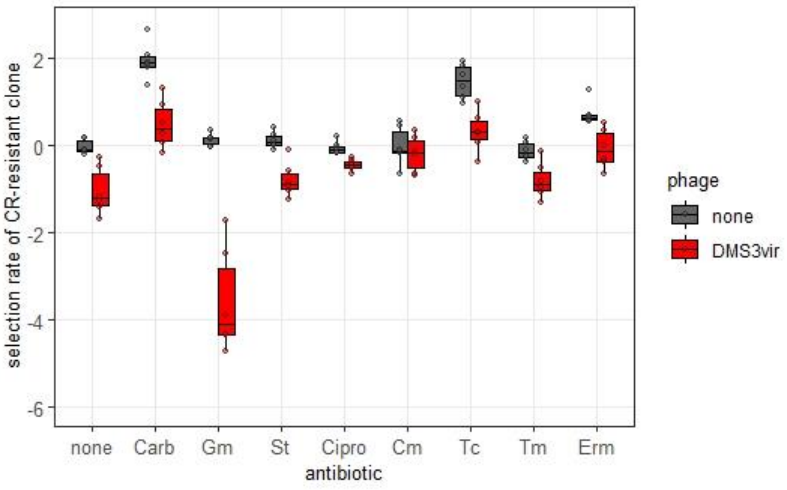


Figure 40. Resistant clone fitness effects. The selection rate of the CRISPR-Cas resistant clone BIM2 against the surface mutant 3A in M9 is plotted as a function of antibiotic treatment, in the absence (gray) or presence (red) of phage DMS3vir. The center value of the boxplots shows the median, boxes show the first and third quartile, and whiskers represent 1.5 times the interquartile range; individual data points are shown as dots (n=6).

While the observed fitness drop could influence the resistance type shift in the case of Gm treatment, it is insufficient to explain the effects of other bactericidal antibiotics.

One-step phage growth assays in LB demonstrated that St and Gm abolish phage production, while Cip and Carb do not affect phage production. Thus, possible influence of bactericidal antibiotics on phage production is unlikely to explain the change of the preferred resistance type.

Chapter 5. Discussion

5.1. The role of genome maintenance proteins in primed CRISPR adaptation by the Type I-E CRISPR-Cas system

Naïve and primed CRISPR adaptation should involve the generation of prespacers and their incorporation into the CRISPR array. In naïve CRISPR adaptation, the efficiency of spacer acquisition varies for different genomic regions and is stimulated by the breaks in genomic DNA (Levy et al., 2015). No preferences in the selection of new spacers with specific orientation of the PAM sequence have been reported so far during naïve CRISPR adaptation. The stark difference of primed adaptation is that newly acquired spacers mostly target protospacers located upstream of the PPS and flanked by the PAM sequence in the same orientation as for the PPS (AAG sequence in the NT-strand; CTT sequence in the T-strand) (Datsenko et al., 2012; Savitskaya et al., 2013; Swarts et al., 2012). While prespacer integration performed solely by the Cas1-Cas2 complex should be similar for both naïve and primed adaptation, the mechanism responsible for prespacer generation must be the source of the differences observed.

The RecBCD complex affects naïve CRISPR adaptation, which proceeds in the absence of Cascade and Cas3, presumably by providing substrates for spacer acquisition (Levy et al., 2015). The helicase but not the nuclease activity of RecBCD is required for naïve adaptation and it was suggested that Cas1-Cas2 is targeted to DNA end structures (Radovčić et al., 2018). For primed adaptation, it was proposed that during CRISPR interference Cas3 cleaves target DNA into fragments used by Cas1-Cas2 as spacer precursors (Swarts et al., 2012). If that were the case, generation of Cas3-dependent fragments associated with either AAG or CTT sequence depending on a strand would be expected. *In vitro*, Cas3 cleaves a PPS-containing plasmid bound by a Cascade-crRNA complex into double-stranded fragments that can be bound by Cas1-Cas2, processed and integrated into CRISPR array (Künne et al., 2016). The Cas3-generated fragments are enriched with T in their 3’ ends that might contribute to a higher number of fragments

containing the 3'-end CTT sequence (Künne et al., 2016). However, the *in vitro* system does not recuperate the orientation bias of PAM sequences observed *in vivo* (Datsenko et al., 2012; Savitskaya et al., 2013; Swarts et al., 2012), since the enrichment with T in the 3' ends was detected for both strands. Moreover, recently it was demonstrated that *in vivo* prespacer ends are enriched with CTTNN rather than CTT sequences (Shiriaeva et al., 2019). This discrepancy indicates that the mechanism of primed adaptation *in vivo* might be much more complex than the simple cleavage of both strands by Cas3 followed by the binding of these fragments by Cas1-Cas2.

Due to the degradation of the target DNA, host DNA repair and recombination proteins can be recruited to the ends of the broken DNA in the PPS-region and these proteins might affect CRISPR interference and primed adaptation. RecG and PriA proteins were shown to be important in a model system targeting a phage genome (Ivančić-Baće et al., 2015). In this work, we addressed the effect of genome maintenance proteins on primed adaptation that occurs when the CRISPR-Cas system targets the host chromosome, a situation that has been observed in natural settings (Shmakov et al., 2017). A critical advantage of the self-targeting model system is that it allows one to avoid complications of popular systems that follow spacer acquisition from plasmids, as plasmids become unstable in Rec mutants, which severely affects the yields of acquired spacers (Biek & Cohen, 1986; A. Cohen & Clark, 1986; Silberstein & Cohen, 1987). We show that single deletions of *recB, recC, recD, recJ* or simultaneous deletions of *recB* and *recJ*, *recB* and *sbcD, recB* and *recQ, xseA* and *recJ* significantly and to varying degrees diminish primed adaptation. Moreover, the pattern of acquired spacers was affected in the absence of RecBCD and SbcCD. Thus, at least in conditions of self-targeting, primed adaptation is not an autonomous process and depends on the non-CRISPR machinery of the cell.

Both direct and indirect mechanisms of DNA maintenance machinery involvement in CRISPR adaptation may be at play. None of the proteins studied here affect strand specificity or high preference for AAG PAM of spacers originating from the NT-strand upstream of the PPS and on the T-strand downstream of it. This suggests that Cas3 recruited to PPS by the bound Cascade along with the Cas1-Cas2 adaptation complex are

the main players, which determine characteristic features of primed spacer acquisition (Dillard et al., 2018; Künne et al., 2016). Single-molecule experiments with the purified Cas proteins from *Thermobifida fusca* demonstrated that Cas1, Cas2, Cas3, and Cascade form a primed acquisition complex (PAC) that translocates along the DNA (Dillard et al., 2018). The assembly of PAC was also shown *in vivo* using a bimolecular fluorescence complementation assay (Dillard et al., 2018). The structure of this complex is not known, but it is tempting to assume that the arrangement of the PAC components enables recognition of the AAG PAM in only one strand.

Analysis of acquired spacers patterns in single mutants shows that the genomic region around the PPS, from which spacers originate, is expanded in the Δ*sbcD*, Δ*recB*, Δ*recC*, and Δ*recD* mutants. Consistently, we show that prespacers are also generated at greater distances from the PPS in Δ*recB*, Δ*recC*, and Δ*recD* cells. RecBCD displaces roadblock proteins immediately after encountering them or after pushing them for up to ~20 kb (Eggleston et al., 1995; Finkelstein et al., 2010). The interaction of RecBCD with moving Cas3 or PAC has not been studied to date. The reported velocity of Cas3 and PAC is ~80-90 bp/s (Dillard et al., 2018), while the reported RecBCD velocity ranges from 300 bp/s to 1400 bp/s (Finkelstein et al., 2010; Taylor & Smith, 1980, 2003; Terakawa et al., 2017). We hypothesize that RecBCD can catch up with PAC moving away from the PPS and displace it (Figure 41). Such a mechanism would explain the observed wider area from which prespacers and spacers are selected in the absence of RecBCD. Our data suggest that RecBC is unable to displace PAC, although this requires further investigation. SbcCD is less studied, but its ability to remove proteins from DNA was reported (J. C. Connelly et al., 2003).

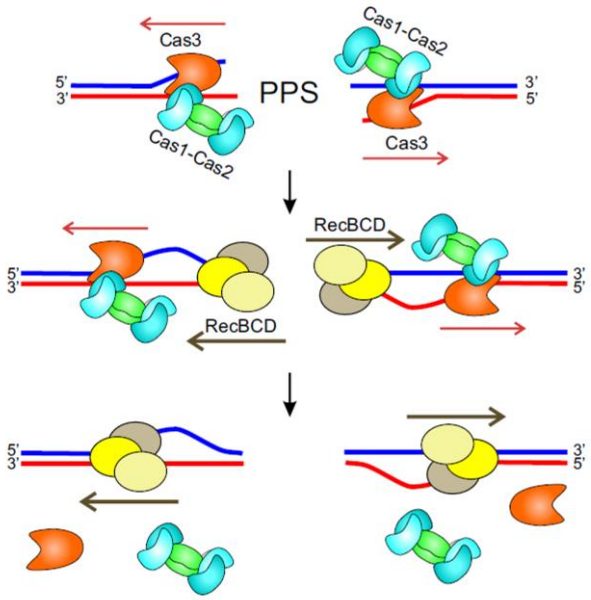


Figure 41. The proposed mechanism of RecBCD interacting with PAC. RecBCD displaces PAC which is moving away from the PPS while searching for protospacers. Thus, the presence of RecBCD limits the area from which protospacers can be selected.

Once the Cas1-Cas2 complex binds the prespacer precursor, further processing of protruding DNA ends may be required to form a 33-bp prespacer with a 3'-end overhang on a PAM-derived end (Shiriaeva et al., 2019). It was reported that 3' ends of prespacers are cleaved by Cas1 (J. Wang et al., 2015) or by DnaQ-like domain of Cas2 in the type I-E system of *Streptococcus thermophilus* (Finkelstein et al., 2010). How 5' ends of prespacers are trimmed was never addressed. The analysis of prespacers in Rec mutants showed that RecBCD participates in the generation of prespacer 5' ends, though it cannot produce prespacers alone. *In vitro*, the RecBCD nuclease was unable to process the ends of the tested prespacer precursors to the length of mature prespacers. Consistently, no changes in the structure of prespacer ends were detected in the nuclease-deficient Δ*recD* mutant. However, when *recB* or *recC* was deleted, prespacers with 1-nt shorter 5' ends were formed. This suggests the direct involvement of the RecBCD helicase activity in prespacer generation. Spacer acquisition was decreased by ~36% in Δ*recD* cells and by

70% in Δ*recB* or Δ*recC* mutants. Prespacer generation in the 1-Mb PPS region was decreased by ~35% in Δ*recB* or Δ*recC* but was not significantly affected in Δ*recD* cells. When a narrower 0.1-Mb region was assessed, prespacer generation was decreased by 83% in Δ*recB*, 68% in Δ*recC*, and 49% in Δ*recD* cells.

Our results demonstrate that RecJ is a major contributor to 5' end trimming. Deletion of *recJ* decreases spacer acquisition and prespacer generation by ~90% and 70%, respectively. Prespacers formed in the absence of RecJ have one additional nucleotide on their 5' ends, suggesting a direct involvement of RecJ in 5'-end processing. The analysis of HIPs also showed the predominance of 1-nt longer adaptation intermediates in the *recJ* mutant. This conclusion is further supported by *in vitro* results, demonstrating that RecJ trims long single-stranded 5' ends of dual-forked DNA substrates up to the protospacer sequence protected by Cas1-Cas2. Interestingly, the absence of RecJ also affects the precision of spacer incorporation, increasing the number of incorrectly incorporated spacers by 30%. In addition, more spacers with internal AAG trinucleotides are acquired in Δ*recJ* mutant, compared to the *wt*.

Our results suggest that ExoVII (the product of *xseAB*) partially substitutes for RecJ activity during prespacer 5' end trimming. First, ExoVII produces slightly longer ends than RecJ in biochemical experiments, similar to those observed *in vivo* in Δ*recJ* cells. Second, while primed adaptation efficiency is not affected by the *xseA* deletion, a ~700-fold decrease in adaptation efficiency is detected in Δ*recJ* Δ*xseA* cells, suggesting that ExoVII is mainly responsible for the residual level of adaptation in Δ*recJ* cells. Additionally, we observed a largely altered HIPs distribution and length in Δ*recJ* Δ*xseA* strain. These results are consistent with functional redundancy between ExoVII and RecJ reported in earlier studies of methyl-directed mismatch repair (Cooper et al., 1993; M. Viswanathan et al., 2001) and repair of UV-induced damages (M. Viswanathan & Lovett, 1998). Furthermore, ExoVII partially compensates for RecJ activity during recombination and repair of gamma radiation-induced DNA damage in Δ*recD* cells (Dermić et al., 2006).

We suggest that RecBCD unwinds double-stranded regions in prespacer precursors and thus provides access to 5' ends for single-strand specific RecJ (Figure 42). Indeed,

when a Cas1-Cas2-bound prespacer precursor with a terminal double-stranded extension was treated with a combination of RecBCD and RecJ, or RecBC and RecJ, proper 5' ends were generated.

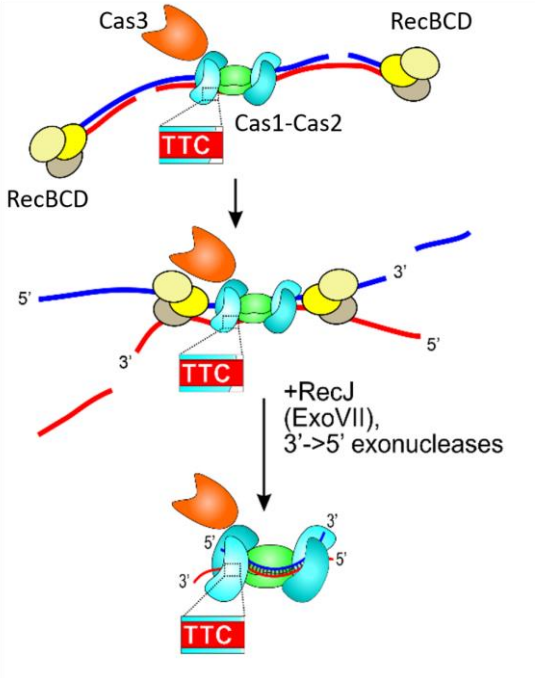


Figure 42. The proposed model of prespacer precursor trimming. RecBCD (or RecBC) unwinds double-stranded regions of prespacer precursors bound by Cas1-Cas2, ensuring processing of ends by single-strand specific exonucleases. RecJ trims the 5' ends of the prespacer precursor. ExoVII partially compensate when RecJ is absent.

A slight decrease in prespacer generation and spacer acquisition efficiency in the Δ*recD* mutant is likely caused by the fact that the RecBC helicase is slower than RecBCD (Korangy & Julin, 1994; Taylor & Smith, 2003). Since Δ*recBC* mutants are not fully compromised in spacer acquisition, we suggest that the RecBCD helicase activity can be partially substituted by helicases that remain to be identified. One such helicase could be RecQ, since a double Δ*recB* Δ*recQ* mutant is more strongly affected in spacer acquisition, than the corresponding single mutants (Figure 28). However, we did not see any effect on

HIPs distribution and length in Δ*recB* Δ*recQ* strain, compared to Δ*recB* (Figure 30, Figure 31). The reason(s) for the extra shortening of prespacers in the absence of RecB or RecC remain unknown. One possible explanation is that the helicase that unwinds prespacer ends in the absence of RecBCD partially displaces the Cas1-Cas2 complex allowing the removal of additional nucleotides by RecJ/ExoVII.

In Figure 43, we present a model of prespacer generation during primed adaptation that is consistent with available published data and results obtained in our work. A primed acquisition complex (PAC) composed of Cas1-Cas2, Cas3, and Cascade is assembled on the PPS (Dillard et al., 2018). The complex moves along DNA due to the Cas3 translocase activity that provides Cas1-Cas2 with single-stranded substrates for binding. When Cas1-Cas2 recognizes a 3'-TTC-5' sequence, it binds to the adjacent protospacer sequence, captures its complementary strand generated by the Cas3 translocase/nuclease, and forms duplex DNA. Since during primed adaptation, prespacers are selected in a specific orientation relative to the PPS (so that the 3'-TTC-5' motif is in the T-strand upstream of the PPS and in the NT-strand downstream of the PPS), we speculate that the Cas1-Cas2 positioning within PAC allows sampling of 3'-TTC-5' in one strand only. Primed adaptation is accompanied by DNA cleavage by Cas3 and, possibly, other cellular nucleases, generating prespacer precursors of various lengths and structures. Cas1-Cas2 bound to the selected protospacer sets the boundaries of a future prespacer by protecting it from degradation. The 5' ends of double-stranded sequences flanking the protospacer are processed by RecJ and RecBCD (or other helicases). ExoVII can perform the 5'-end trimming and partially compensate when RecJ is absent. A set of enzymes generating prespacer 3' ends *in vivo* is yet to be determined. According to our recent data, Cas3 remains associated with Cas1-Cas2 at least until the mature 33/37-nt prespacer is formed.

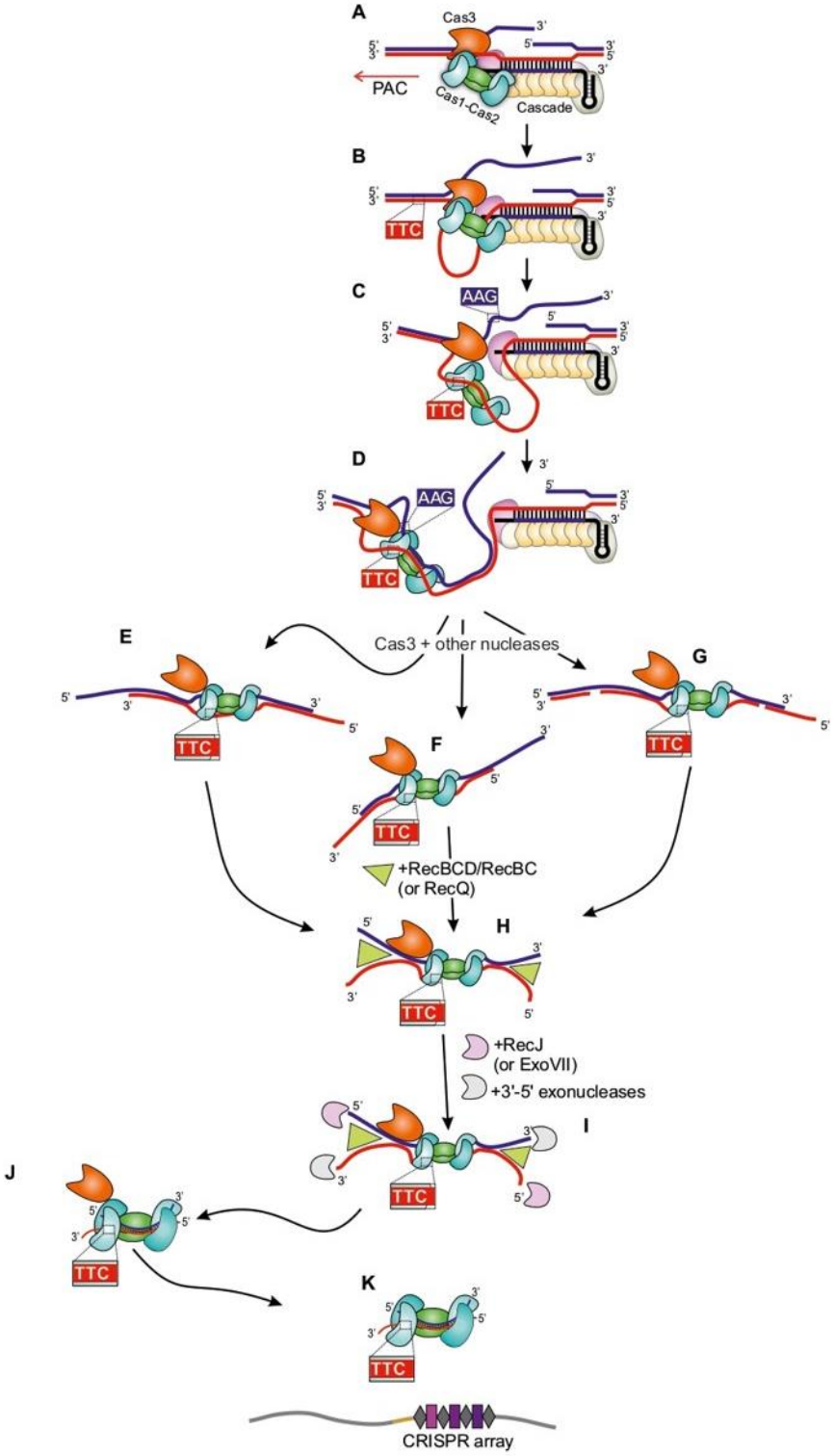


Figure 43. A model of primed adaptation in the type I-E CRISPR-Cas system.

5.2. Bacteriostatic antibiotics promote CRISPR adaptation during phage infection in *P. aeruginosa*

Previous studies have identified several environmental variables that shape the evolution of CRISPR-Cas immunity by affecting the fitness of CRISPR-Cas immune *P. aeruginosa* clones relative to those with mutated phage receptors (Alseth et al., 2019; van Houte et al., 2016; Westra et al., 2015). Here, we identify clinically relevant environmental factors, which increase the frequency at which sensitive *P. aeruginosa* clones acquire CRISPR-Cas immunity during phage infection. Cells that just acquired spacers might still be irrevocably damaged before clearing the infection (Horvath & Barrangou, 2010; S. A. Jackson et al., 2017; Modell et al., 2017). The factors that determine whether or not bacteria acquire spacers from infecting phages remain unclear. Previously identified determinants of CRISPR adaptation include the density of defective phages (Hynes et al., 2014), and bacterial nucleases, including crRNA-Cas ribonucleoprotein effector complexes, that inactivate the phage and generate phage DNA degradation products that serve as substrates for the adaptation machinery (S. A. Jackson et al., 2017). Here, we show that the speed of phage development is another key determinant of CRISPR adaptation and that bacteriostatic antibiotics promote CRISPR adaptation by slowing bacterial growth, which in turn delays phage development. Our results suggest that any environmental factor or stress that slows down cell growth might also promote CRISPR adaptation, as phage development time is directly linked to bacterial host growth rate (Hadas et al., 1997; Jin & Yin, 2021; Rabinovitch et al., 2002).

In the absence of priming, slowing down phage growth might be even more crucial for the acquisition of spacers that can lead to phage resistance. However, while our data suggest that slower bacterial and phage growth rates will generally favour the evolution of CRISPR-Cas immunity, future studies will be needed to experimentally examine how growth rate reductions due to antibiotics and other factors affect primed and naive CRISPR adaptation in other species with other CRISPR-Cas immune system types.

Interestingly, some CRISPR-Cas systems induce dormancy following target recognition (Meeske et al., 2019; Rostøl et al., 2021; Rostøl & Marraffini, 2019) or are

coupled to genes that induce dormancy (E. V. Koonin & Zhang, 2017; Makarova et al., 2012). A dormancy response of infected cells with CRISPR-Cas immunity can benefit neighbouring cells by eliminating phage from the environment and by limiting the invasion of phage mutants that overcome CRISPR-Cas immunity (Meeske et al., 2019). Our data suggest a possibility that another advantage of a dormancy response could be that it may lead to more efficient CRISPR adaptation during infections. In addition, previous studies have shown that targeting early-injected genome sequences of a phage provides more robust levels of CRISPR-Cas immunity compared with targeting of late-injected sequences (Modell et al., 2017; Strotskaya et al., 2017), presumably because this gives CRISPR-Cas more time to detect and destroy the phage genome. However, our competition experiments between CRISPR-Cas immune bacteria and those with SM-based resistance did not show a stronger selection for CRISPR-Cas immune bacteria in the presence of bacteriostatic antibiotics, suggesting that they specifically affect the rates of CRISPR adaptation and not the levels of immunity.

The finding that slower phage development facilitates CRISPR adaptation may also help to explain why the acquisition of CRISPR-Cas immunity is relatively rare under laboratory conditions (Westra & Levin, 2020), in which bacteria commonly grow at rates much higher than in natural or clinical contexts (B. Gibson et al., 2018). *P. aeruginosa* displays slow growth rates in biofilms (Werner et al., 2004) and in the lungs of cystic fibrosis patients (Yang et al., 2008). This will be compounded by exposure to bacteriostatic antibiotics widely used in the clinic and becoming more common in the environment. Promoting CRISPR-Cas immunity might have unfortunate consequences, specifically in the context of phage therapy: the rise in antibiotic resistance among bacterial pathogens has caused a resurgence of interest in phage therapy to treat bacterial infections (Kortright et al., 2019). Over the past few years, interest in phage-antibiotic combination therapies has increased in the hope that phage and antibiotics can act in concert to control infections (Kortright et al., 2019; Segall et al., 2019; Tagliaferri et al., 2019; Torres-Barceló & Hochberg, 2016). However, CRISPR-Cas systems are common in *P. aeruginosa* (van Belkum et al., 2015), and *P. aeruginosa* clones that acquire CRISPR-Cas immunity can

escape these trade-offs and retain virulence (Alseth et al., 2019). Thus, a combination of bacteriostatic antibiotics and phage therapy might steer pathogens toward maintaining both virulence and phage resistance.

Our results suggest that the choice of antibiotics to use for phage-antibiotic combination therapy should consider the possibility of increased evolution of CRISPR-Cas immunity when using bacteriostatic antibiotics, and the potential for pathogenic strains to retain their virulence if they evolve CRISPR-Cas-based resistance to therapeutic phages. Moreover, antibiotics can act on non-target bacteria, including other pathogens in multi-species communities, and also on commensal species. Non-pathogenic commensals will be exposed to antibiotics during antibiotic treatment, potentially more widely promoting CRISPR-Cas immunity in the microbiome.

This study specifically examined the effect of antibiotics on the acquisition of CRISPR-Cas immunity against phage and demonstrated how bacteriostatic antibiotics can tip the balance in favour of the host immune system by enabling it to launch an immune response before the phage causes irrevocable damage to the host. However, CRISPR-Cas systems can provide defense against a broad range of mobile genetic elements, including plasmids (Marraffini & Sontheimer, 2008), which play a key role in the transmission of antimicrobial resistance (AMR) (Brockhurst et al., 2019). The ability to trigger CRISPR adaptation from plasmids could therefore have important implications for constraining the spread of AMR genes (Pursey et al., 2018). Recent studies have demonstrated that the evolution of CRISPR-Cas immunity against plasmids can be enhanced using quorum-sensing autoinducers, which activate the expression of the CRISPR-Cas machinery (Høyland-Kroghsbo et al., 2017; Patterson et al., 2016). Although the precise mechanism remains unclear, the evolution of CRISPR-Cas immunity against plasmids can also be higher at low temperatures or during the late exponential growth phase (Amlinger et al., 2017; Høyland-Kroghsbo et al., 2018). Identifying ways to simultaneously promote spacer acquisition from plasmids and limiting the acquisition of spacers from phage could provide a powerful means to control pathogen abundance, virulence, and resistance to antibiotics.

Conclusions

1. Primed adaptation efficiency in self-targeting *E. coli* strain is dramatically inhibited in double Δ*recB* Δ*sbcD* mutant but not in single mutants, suggesting independent involvement and redundancy of RecBCD and SbcCD pathways in spacer acquisition.

2. RecBCD and SbcCD affect the pattern of acquired spacers, suggesting the enzymes could interfere with the processivity of PAC.

3. The deletion of *recJ* decreases adaptation efficiency and affects accuracy of spacers incorporation into CRISPR array.

4. RecJ 5'-3' exonuclease and RecBCD helicase are involved in trimming of prespacer 5' ends during primed adaptation *in vivo*.

5. ExoVII can partially substitute for RecJ activity during prespacer generation but produces slightly longer ends.

6. Exposure to bacteriostatic antibiotics promotes primed CRISPR adaptation by slowing down phage development in *P. aeruginosa* PA14 infected by the DMS3vir phage.

7. Bactericidal antibiotics inhibit CRISPR adaptation during phage infection by a yet unknown mechanism.

Bibliography

1. Abudayyeh, O. O., Gootenberg, J. S., Konermann, S., Joung, J., Slaymaker, I. M., Cox, D. B. T., Shmakov, S., Makarova, K. S., Semenova, E., Minakhin, L., Severinov, K., Regev, A., Lander, E. S., Koonin, E. V., & Zhang, F. (2016). C2c2 is a single-component programmable RNA-guided RNA-targeting CRISPR effector. *Science*, *353*(6299), aaf5573.

2. Almendros, C., Nobrega, F. L., McKenzie, R. E., & Brouns, S. J. J. (2019). Cas4–Cas1 fusions drive efficient PAM selection and control CRISPR adaptation. *Nucleic Acids Res.*, *47*(10), 5223–5230.

3. Alseth, E. O., Pursey, E., Luján, A. M., McLeod, I., Rollie, C., & Westra, E. R. (2019). Bacterial biodiversity drives the evolution of CRISPR-based phage resistance. *Nature*, *574*, 549–552.

4. Amlinger, L., Hoekzema, M., Wagner, E. G. H., Koskiniemi, S., & Lundgren, M. (2017). Fluorescent CRISPR Adaptation Reporter for rapid quantification of spacer acquisition. *Sci. Rep.*, *7*, 10392.

5. Amundsen, S. K., Sharp, J. W., & Smith, G. R. (2016). RecBCD enzyme "Chi recognition" mutants recognize Chi recombination hotspots in the right DNA context. *Genetics*, *204*(1), 139–152.

6. Amundsen, S. K., Taylor, A. F., Chaudhury, A. M., & Smith, G. R. (1986). recD: the gene for an essential third subunit of exonuclease V. *Proc. Natl. Acad. Sci. U. S. A.*, *83*(15), 5558–5562.

7. Anderson, D. G., Churchill, J. J., & Kowalczykowski, S. C. (1997). Chi-activated RecBCD enzyme possesses 5'-3' nucleolytic activity, but RecBC enzyme does not: evidence suggesting that the alteration induced by Chi is not simply ejection of the RecD subunit. *Genes to Cells*, *2*(2), 117–128.

8. Appasani, K., Thaler, D. S., & Goldberg, E. B. (1999). Bacteriophage T4 gp2 interferes with cell viability and with bacteriophage lambda Red recombination. *J. Bacteriol.*, *181*(4), 1352–1355.

9. Baba, T., Ara, T., Hasegawa, M., Takai, Y., Okumura, Y., Baba, M., Datsenko, K. A.,

Tomita, M., Wanner, B. L., & Mori, H. (2006). Construction of Escherichia coli K-12 in-frame, single-gene knockout mutants: the Keio collection. *Mol. Syst. Biol.*, *2*, 2006.0008.

10. Babu, M., Beloglazova, N., Flick, R., Graham, C., Skarina, T., Nocek, B., Gagarinova, A., Pogoutse, O., Brown, G., Binkowski, A., Phanse, S., Joachimiak, A., Koonin, E. V, Savchenko, A., Emili, A., Greenblatt, J., Edwards, A. M., & Yakunin, A. F. (2011). A dual function of the CRISPR-Cas system in bacterial antivirus immunity and DNA repair. *Mol. Microbiol.*, *79*(2), 484–502.

11. Baharoglu, Z., & Mazel, D. (2011). Vibrio cholerae triggers SOS and mutagenesis in response to a wide range of antibiotics: a route towards multiresistance. *Antimicrob. Agents Chemother.*, *55*, 2438–2441.

12. Bamford, R. A., Smith, A., Metz, J., Glover, G., Titball, R. W., & Pagliara, S. (2017). Investigating the physiology of viable but non-culturable bacteria by microfluidics and time-lapse microscopy. *BMC Biol.*, *15*, 121.

13. Bandyopadhyay, P. K., Studier, F. W., Hamilton, D. L., & Yuan, R. (1985). Inhibition of the type I restriction-modification enzymes EcoB and EcoK by the gene 0.3 protein of bacteriophage T7. *J. Mol. Biol.*, *182*(4), 567–578.

14. Beguin, P., Charpin, N., Koonin, E. V., Forterre, P., & Krupovic, M. (2016). Casposon integration shows strong target site preference and recapitulates protospacer integration by CRISPR-Cas systems. *Nucleic Acids Res.*, *44*(21), 10367.

15. Behme, M. T., Lilley, G. D., & Ebisuzaki, K. (1976). Postinfection control by bacteriophage T4 of Escherichia coli recBC nuclease activity. *J. Virol.*, *18*(1), 20.

16. van Belkum, A., Soriaga, L. B., LaFave, M. C., Akella, S., Veyrieras, J.-B., Barbu, E. M., Shortridge, D., Blanc, B., Hannum, G., Zambardi, G., Miller, K., Enright, M. C., Mugnier, N., Brami, D., Schicklin, S., Felderman, M., Schwartz, A. S., Richardson, T. H., Peterson, T. C., Hubby, B., Cady, K. C. (2015). Phylogenetic distribution of CRISPR-Cas systems in antibiotic-resistant Pseudomonas aeruginosa. *MBio*, *6*(6), e01796-15.

17. Beloglazova, N., Brown, G., Zimmerman, M. D., Proudfoot, M., Makarova, K. S., Kudritska, M., Kochinyan, S., Wang, S., Chruszcz, M., Minor, W., Koonin, E. V., Edwards, A. M., Savchenko, A., & Yakunin, A. F. (2008). A novel family of sequence-

specific endoribonucleases associated with the clustered regularly interspaced short palindromic repeats. *J. Biol. Chem.*, *283*(29), 20361–20371.

18. Benzinger, R., Enquist, L. W., & Skalka, A. (1975). Transfection of Escherichia coli spheroplasts. V. Activity of recBC nuclease in rec+ and rec minus spheroplasts measured with different forms of bacteriophage DNA. *J. Virol.*, *15*(4), 861.

19. Bernheim, A., Bikard, D., Touchon, M., & Rocha, E. P. C. (2019). A matter of background: DNA repair pathways as a possible cause for the sparse distribution of CRISPR-Cas systems in bacteria. *Philos. Trans. R. Soc. B Biol. Sci.*, *374*(1772).

20. Bernheim, A., Calvo-Villamañán, A., Basier, C., Cui, L., Rocha, E. P. C., Touchon, M., & Bikard, D. (2017). Inhibition of NHEJ repair by type II-A CRISPR-Cas systems in bacteria. *Nat. Commun. 2017 81*, *8*(1), 1–9.

21. Bianco, P. R., Brewer, L. R., Corzett, M., Balhorn, R., Yeh, Y., Kowalczykowski, S. C., & Baskin, R. J. (2001). Processive translocation and DNA unwinding by individual RecBCD enzyme molecules. *Nature*, *409*(6818), 374–378.

22. Biek, D. P., & Cohen, S. N. (1986). Identification and characterization of recD, a gene affecting plasmid maintenance and recombination in Escherichia coli. *J. Bacteriol.*, *167*(2), 594–603.

23. Bobay, L.-M., Touchon, M., & Rocha, E. P. C. (2013). Manipulating or superseding host recombination functions: a dilemma that shapes phage evolvability. *PLoS Genet.*, *9*(9), e1003825.

24. Briner, A. E., Donohoue, P. D., Gomaa, A. A., Selle, K., Slorach, E. M., Nye, C. H., Haurwitz, R. E., Beisel, C. L., May, A. P., & Barrangou, R. (2014). Guide RNA functional modules direct Cas9 activity and orthogonality. *Mol. Cell*, *56*(2), 333–339.

25. Brockhurst, M. A., Harrison, E., Hall, J. P. J., Richards, T., McNally, A., & MacLean, C. (2019). The ecology and evolution of pangenomes. *Curr. Biol.*, *29*, R1094–R1103.

26. Brouns, S. J. J., Jore, M. M., Lundgren, M., Westra, E. R., Slijkhuis, R. J. H., Snijders, A. P. L., Dickman, M. J., Makarova, K. S., Koonin, E. V., & van der Oost, J. (2008). Small CRISPR RNAs guide antiviral defense in prokaryotes. *Science*, *321*(5891), 960–964.

27. Burstein, D., Harrington, L. B., Strutt, S. C., Probst, A. J., Anantharaman, K., Thomas,

B. C., Doudna, J. A., & Banfield, J. F. (2017). New CRISPR–Cas systems from uncultivated microbes. *Nature*, *542*(7640), 237–241.

28. Cady, K. C., Bondy-Denomy, J., Heussler, G. E., Davidson, A. R., & O'Toole, G. A. (2012). The CRISPR/Cas adaptive immune system of Pseudomonas aeruginosa mediates resistance to naturally occurring and engineered phages. *J. Bacteriol.*, *194*, 5728–5738.

29. Cady, K. C., & O'Toole, G. A. (2011). Non-identity-mediated CRISPR-bacteriophage interaction mediated via the Csy and Cas3 proteins. *J. Bacteriol.*, *193*(14), 3433–3445.

30. Cady, K. C., White, A. S., Hammond, J. H., Abendroth, M. D., Karthikeyan, R. S. G., Lalitha, P., Zegans, M. E., & O'Toole, G. A. (2011). Prevalence, conservation and functional analysis of Yersinia and Escherichia CRISPR regions in clinical Pseudomonas aeruginosa isolates. *Microbiology*, *157*(Pt 2), 430.

31. Cama, J., Voliotis, M., Metz, J., Smith, A., Iannucci, J., Keyser, U. F., Tsaneva-Atanasova, K., & Pagliara, S. (2020). Single-cell microfluidics facilitates the rapid quantification of antibiotic accumulation in Gram-negative bacteria. *Lab Chip*, *20*, 2765–2775.

32. Chase, J. W., & Richardson, C. C. (1974). Exonuclease VII of Escherichia coli. Purification and properties. *J. Biol. Chem.*, *249*(14), 4545–4552.

33. Chayot, R., Montagne, B., Mazel, D., & Ricchetti, M. (2010). An end-joining repair mechanism in Escherichia coli. *Proc. Natl. Acad. Sci. U. S. A.*, *107*(5), 2141.

34. Chen, H. W., Ruan, B., Yu, M., Wang, J. d, & Julin, D. A. (1997). The RecD subunit of the RecBCD enzyme from Escherichia coli is a single-stranded DNA-dependent ATPase. *J. Biol. Chem.*, *272*(15), 10072–10079.

35. Chen, Z., Yang, H., & Pavletich, N. P. (2008). Mechanism of homologous recombination from the RecA-ssDNA/dsDNA structures. *Nature*, *453*(7194), 489–494.

36. Cheng, K., Wilkinson, M., Chaban, Y., & Wigley, D. B. (2020). A conformational switch in response to Chi converts RecBCD from phage destruction to DNA repair. *Nat. Struct. Mol. Biol.*, *27*(1), 71–77.

37. Churchill, J. J., Anderson, D. G., & Kowalczykowski, S. C. (1999). The RecBC enzyme loads RecA protein onto ssDNA asymmetrically and independently of chi, resulting in

constitutive recombination activation. *Genes Dev.*, *13*(7), 901–911.

38. Cohen, A., & Clark, A. J. (1986). Synthesis of linear plasmid multimers in Escherichia coli K-12. *J. Bacteriol.*, *167*(1), 327–335.

39. Cohen, D., Melamed, S., Millman, A., Shulman, G., Oppenheimer-Shaanan, Y., Kacen, A., Doron, S., Amitai, G., & Sorek, R. (2019). Cyclic GMP–AMP signalling protects bacteria against viral infection. *Nature*, *574*(7780), 691–695.

40. Comeau, A. M., Tétart, F., Trojet, S. N., Prère, M. F., & Krisch, H. M. (2007). Phage-antibiotic synergy (PAS): β-lactam and quinolone antibiotics stimulate virulent phage growth. *FEMS One*, *2*, e799.

41. Cong, L., Ran, F. A., Cox, D., Lin, S., Barretto, R., Habib, N., Hsu, P. D., Wu, X., Jiang, W., Marraffini, L. A., & Zhang, F. (2013). Multiplex genome engineering using CRISPR/Cas systems. *Science*, *339*(6121), 819–823.

42. Connelly, J. (1999). DNA cleavage and degradation by the SbcCD protein complex from Escherichia coli. *Nucleic Acids Res.*, *27*(4), 1039–1046.

43. Connelly, J. C., de Leau, E. S., & Leach, D. R. F. (2003). Nucleolytic processing of a protein-bound DNA end by the E. coli SbcCD (MR) complex. *DNA Repair (Amst).*, *2*(7), 795–807.

44. Connelly, J. C., & Leach, D. R. F. (1996). The sbcC and sbcD genes of Escherichia coli encode a nuclease involved in palindrome inviability and genetic recombination. *Genes to Cells*, *1*(3), 285–291.

45. Cooper, D. L., Lahue, R. S., & Modrich, P. (1993). Methyl-directed mismatch repair is bidirectional. *J. Biol. Chem.*, *268*(16), 11823–11829.

46. Cui, L., Wang, X., Huang, D., Zhao, Y., Feng, J., Lu, Q., Pu, Q., Wang, Y., Cheng, G., Wu, M., & Dai, M. (2020). CRISPR-cas3 of Salmonella upregulates bacterial biofilm formation and virulence to host cells by targeting quorum-sensing systems. *Pathogens*, *9*(1), 53.

47. Datsenko, K. A., Pougach, K., Tikhonov, A., Wanner, B. L., Severinov, K., & Semenova, E. (2012). Molecular memory of prior infections activates the CRISPR/Cas adaptive bacterial immunity system. *Nat. Commun.*, *3*(1), 945.

48. Datsenko, K. A., & Wanner, B. L. (2000). One-step inactivation of chromosomal genes in Escherichia coli K-12 using PCR products. *Proc. Natl. Acad. Sci. U. S. A.*, *97*(12), 6640–6645.

49. Deatherage, D. E., & Barrick, J. E. (2014). Identification of mutations in laboratory evolved microbes from next-generation sequencing data using breseq. *Methods Mol. Biol.*, *1151*, 165.

50. Delmas, S., Shunburne, L., Ngo, H.-P., & Allers, T. (2009). Mre11-Rad50 promotes rapid repair of DNA damage in the polyploid archaeon Haloferax volcanii by restraining homologous recombination. *PLoS Genet.*, *5*(7), e1000552.

51. Deltcheva, E., Chylinski, K., Sharma, C. M., Gonzales, K., Chao, Y., Pirzada, Z. A., Eckert, M. R., Vogel, J., & Charpentier, E. (2011). CRISPR RNA maturation by trans-encoded small RNA and host factor RNase III. *Nature*, *471*(7340), 602–607.

52. Dermić, D. (2006). Functions of multiple exonucleases are essential for cell viability, DNA repair and homologous recombination in recD mutants of Escherichia coli. *Genetics*, *172*(4), 2057–2069.

53. Dermić, D., Zahradka, D., & Petranović, M. (2006). Exonuclease requirements for recombination of lambda-phage in recD mutants of Escherichia coli. *Genetics*, *173*(4), 2399–2402.

54. Deveau, H., Barrangou, R., Garneau, J. E., Labonté, J., Fremaux, C., Boyaval, P., Romero, D. A., Horvath, P., & Moineau, S. (2008). Phage response to CRISPR-encoded resistance in Streptococcus thermophilus. *J. Bacteriol.*, *190*(4), 1390–1400.

55. Deveau, H., Garneau, J. E., & Moineau, S. (2010). CRISPR/Cas system and its role in phage-bacteria interactions. *Annu. Rev. Microbiol.*, *64*, 475–493.

56. Dillard, K. E., Brown, M. W., Johnson, N. V., Xiao, Y., Dolan, A., Hernandez, E., Dahlhauser, S. D., Kim, Y., Myler, L. R., Anslyn, E. V., Ke, A., & Finkelstein, I. J. (2018). Assembly and translocation of a CRISPR-Cas primed acquisition complex. *Cell*, *175*(4), 934-946.e15.

57. Dillingham, M. S., & Kowalczykowski, S. C. (2008). RecBCD enzyme and the repair of double-stranded DNA breaks. *Microbiol. Mol. Biol. Rev.*, *72*(4), 642–671.

58. Dillingham, M. S., Spies, M., & Kowalczykowski, S. C. (2003). RecBCD enzyme is a bipolar DNA helicase. *Nature*, *423*(6942), 893–897.

59. Dixon, D. A., & Kowalczykowski, S. C. (1995). Role of the Escherichia coli recombination hotspot, χ, in RecABCD- dependent homologous pairing. *J. Biol. Chem.*, *270*(27), 16360–16370.

60. Dolan, A. E., Hou, Z., Xiao, Y., Gramelspacher, M. J., Heo, J., Howden, S. E., Freddolino, P. L., Ke, A., & Zhang, Y. (2019). Introducing a spectrum of long-range genomic deletions in human embryonic stem cells using Type I CRISPR-Cas. *Mol. Cell*, *74*(5), 936-950.e5.

61. Dong, C., Fontana, J., Patel, A., Carothers, J. M., & Zalatan, J. G. (2018). Author Correction: Synthetic CRISPR-Cas gene activators for transcriptional reprogramming in bacteria. *Nat. Commun.*, *9*(1).

62. Dong, D., Guo, M., Wang, S., Zhu, Y., Wang, S., Xiong, Z., Yang, J., Xu, Z., & Huang, Z. (2017). Structural basis of CRISPR-SpyCas9 inhibition by an anti-CRISPR protein. *Nature*, *546*(7658), 436–439.

63. Drabavicius, G., Sinkunas, T., Silanskas, A., Gasiunas, G., Venclovas, Č., & Siksnys, V. (2018). DnaQ exonuclease-like domain of Cas2 promotes spacer integration in a type I-E CRISPR-Cas system. *EMBO Rep.*, *19*(7).

64. Dryden, D. T. F. (2004). Reeling in the bases. *Nat. Struct. Mol. Biol.*, *11*(9), 804–806.

65. Dryden, D. T. F., Murray, N. E., & Rao, D. N. (2001). Nucleoside triphosphate-dependent restriction enzymes. *Nucleic Acids Res.*, *29*(18), 3728.

66. Dwyer, D. J., Kohanski, M. A., Hayete, B., & Collins, J. J. (2007). Gyrase inhibitors induce an oxidative damage cellular death pathway in Escherichia coli. *Mol. Syst. Biol.*, *3*, 91.

67. East-Seletsky, A., O'Connell, M. R., Knight, S. C., Burstein, D., Cate, J. H. D., Tjian, R., & Doudna, J. A. (2016). Two distinct RNase activities of CRISPR-C2c2 enable guide-RNA processing and RNA detection. *Nature*, *538*(7624), 270–273.

68. Eggleston, A. K., & Kowalczykowski, S. C. (1993). Biochemical characterization of a mutant recBCD enzyme, the recB2109CD enzyme, which lacks chi-specific, but not non-

specific, nuclease activity. *J. Mol. Biol.*, *231*(3), 605–620.
69. Eggleston, A. K., O'Neill, T. E., Bradbury, E. M., & Kowalczykowski, S. C. (1995). Unwinding of nucleosomal DNA by a DNA helicase. *J. Biol. Chem.*, *270*(5), 2024–2031.
70. Eichler, D. C., & Lehman, I. R. (1977). On the role of ATP in phosphodiester bond hydrolysis catalyzed by the recBC deoxyribonuclease of Escherichia coli. *J. Biol. Chem.*, *252*(2), 499–503.
71. Emmerson, P. T. (1968). Recombination deficient mutants of Escherichia coli K12 that map between thyA and argA. *Genetics*, *60*(1), 19–30.
72. Ershova, A. S., Karyagina, A. S., Vasiliev, M. O., Lyashchuk, A. M., Lunin, V. G., Spirin, S. A., & Alexeevski, A. V. (2012). Solitary restriction endonucleases in prokaryotic genomes. *Nucleic Acids Res.*, *40*(20), 10107.
73. Eykelenboom, J. K., Blackwood, J. K., Okely, E., & Leach, D. R. F. (2008). SbcCD causes a double-strand break at a DNA palindrome in the Escherichia coli chromosome. *Mol. Cell*, *29*(5), 644–651.
74. Faure, G., Shmakov, S. A., Yan, W. X., Cheng, D. R., Scott, D. A., Peters, J. E., Makarova, K. S., & Koonin, E. V. (2019). CRISPR–Cas in mobile genetic elements: counter-defence and beyond. *Nat. Rev. Microbiol.*, *17*(8), 513–525.
75. Fineran, P. C., & Charpentier, E. (2012). Memory of viral infections by CRISPR-Cas adaptive immune systems: Acquisition of new information. *Virology*, *434*(2), 202–209.
76. Fineran, P. C., Gerritzen, M. J. H., Suárez-Diez, M., Künne, T., Boekhorst, J., van Hijum, S. A. F. T., Staals, R. H. J., & Brouns, S. J. J. (2014). Degenerate target sites mediate rapid primed CRISPR adaptation. *Proc. Natl. Acad. Sci. USA*, *111*, E1629–E1638.
77. Finkelstein, I. J., Visnapuu, M. L., & Greene, E. C. (2010). Single-molecule imaging reveals mechanisms of protein disruption by a DNA translocase. *Nature*, *468*(7326), 983–987.
78. Fonfara, I., Richter, H., Bratovič, M., Le Rhun, A., & Charpentier, E. (2016). The CRISPR-associated DNA-cleaving enzyme Cpf1 also processes precursor CRISPR RNA. *Nature*, *532*(7600), 517–521.
79. Gao, N. J., Al-Bassam, M. M., Poudel, S., Wozniak, J. M., Gonzalez, D. J., Olson, J.,

Zengler, K., Nizet, V., & Valderrama, J. A. (2019). Functional and proteomic analysis of streptococcus pyogenes virulence upon loss of its native cas9 nuclease. *Front. Microbiol.*, *10*(AUG), 476691.

80. Garneau, J. E., Dupuis, M.-È., Villion, M., Romero, D. a, Barrangou, R., Boyaval, P., Fremaux, C., Horvath, P., Magadán, A. H., & Moineau, S. (2010). The CRISPR/Cas bacterial immune system cleaves bacteriophage and plasmid DNA. *Nature*, *468*(7320), 67–71.

81. Garrett, R. A., Vestergaard, G., & Shah, S. A. (2011). Archaeal CRISPR-based immune systems: Exchangeable functional modules. *Trends Microbiol.*, *19*(11), 549–556.

82. Gasiunas, G., Barrangou, R., Horvath, P., & Siksnys, V. (2012). Cas9-crRNA ribonucleoprotein complex mediates specific DNA cleavage for adaptive immunity in bacteria. *Proc. Natl. Acad. Sci. U. S. A.*, *109*(39), E2579-86.

83. Gerdes, K., Christensen, S. K., & Løbner-Olesen, A. (2005). Prokaryotic toxin–antitoxin stress response loci. *Nat. Rev. Microbiol.*, *3*(5), 371–382.

84. Gerdes, K., Rasmussen, P. B., & Molin, S. (1986). Unique type of plasmid maintenance function: postsegregational killing of plasmid-free cells. *Proc. Natl. Acad. Sci. U. S. A.*, *83*(10), 3116.

85. Gibson, B., Wilson, D. J., Feil, E., & Eyre-Walker, A. (2018). The distribution of bacterial doubling times in the wild. *Proc. Biol. Sci.*, *285*, 20180789.

86. Gibson, F. P., Leach, D. R. F., & Lloyd, R. G. (1992). Identification of sbcD mutations as cosuppressors of recBC that allow propagation of DNA palindromes in Escherichia coli K-12. *J. Bacteriol.*, *174*(4), 1222–1228.

87. Goldfarb, T., Sberro, H., Weinstock, E., Cohen, O., Doron, S., Charpak-Amikam, Y., Afik, S., Ofir, G., & Sorek, R. (2015). BREX is a novel phage resistance system widespread in microbial genomes. *EMBO J.*, *34*(2), 169–183.

88. Gong, B., Shin, M., Sun, J., Jung, C.-H., Bolt, E. L., van der Oost, J., & Kim, J.-S. (2014). Molecular insights into DNA interference by CRISPR-associated nuclease-helicase Cas3. *Proc. Natl. Acad. Sci. U. S. A.*, *111*(46), 16359–16364.

89. Gordeeva, J., Morozova, N., Sierro, N., Isaev, A., Sinkunas, T., Tsvetkova, K.,

Matlashov, M., Truncaite, L., Morgan, R. D., Ivanov, N. V., Siksnys, V., Zeng, L., & Severinov, K. (2019). BREX system of Escherichia coli distinguishes self from non-self by methylation of a specific DNA site. *Nucleic Acids Res.*, *47*(1), 253–265.

90. Greenfield, T. J., Ehli, E., Kirshenmann, T., Franch, T., Gerdes, K., & Weaver, K. E. (2000). The antisense RNA of the par locus of pAD1 regulates the expression of a 33-amino-acid toxic peptide by an unusual mechanism. *Mol. Microbiol.*, *37*(3), 652–660.

91. Hadas, H., Einav, M., Fishov, I., & Zaritsky, A. (1997). Bacteriophage T4 development depends on the physiology of its host Escherichia coli. *Microbiol.*, *143*, 179–185.

92. Hadley Wickham. (2016). *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York. https://ggplot2.tidyverse.org

93. Haft, D. H., Selengut, J., Mongodin, E. F., & Nelson, K. E. (2005). A guild of 45 CRISPR-associated (Cas) protein families and multiple CRISPR/Cas subtypes exist in prokaryotic genomes. *PLoS Comput. Biol.*, *1*(6), e60.

94. Han, E. S. (2006). RecJ exonuclease: substrates, products and interaction with SSB. *Nucleic Acids Res.*, *34*(4), 1084–1091.

95. Handa, N., Yang, L., Dillingham, M. S., Kobayashi, I., Wigley, D. B., & Kowalczykowski, S. C. (2012). Molecular determinants responsible for recognition of the single-stranded DNA regulatory sequence, χ, by RecBCD enzyme. *Proc. Natl. Acad. Sci. U. S. A.*, *109*(23), 8901–8906.

96. Harrington, L. B., Doxzen, K. W., Ma, E., Liu, J.-J., Knott, G. J., Edraki, A., Garcia, B., Amrani, N., Chen, J. S., Cofsky, J. C., Kranzusch, P. J., Sontheimer, E. J., Davidson, A. R., Maxwell, K. L., & Doudna, J. A. (2017). A broad-spectrum inhibitor of CRISPR-Cas9. *Cell*, *170*(6), 1224-1233.e15.

97. Hickson, I. D., Robson, C. N., Atkinson, K. E., Hutton, L., & Emmerson, P. T. (1985). Reconstitution of RecBC DNase activity from purified Escherichia coli RecB and RecC proteins. *J. Biol. Chem.*, *260*(2), 1224–1229.

98. Ho, W. S., Ou, H. Y., Yeo, C. C., & Thong, K. L. (2015). The dnd operon for DNA phosphorothioation modification system in Escherichia coli is located in diverse genomic islands. *BMC Genomics*, *16*(1).

99. Hochstrasser, M. L., Taylor, D. W., Bhat, P., Guegler, C. K., Sternberg, S. H., Nogales, E., & Doudna, J. A. (2014). CasA mediates Cas3-catalyzed target degradation during CRISPR RNA-guided interference. *Proc. Natl. Acad. Sci. U. S. A.*, *111*(18), 6618–6623.

100. Horvath, P., & Barrangou, R. (2010). CRISPR/Cas, the immune system of bacteria and archaea. *Science*, *327*(5962), 167–170.

101. van Houte, S., Ekroth, A. K. E., Broniewski, J. M., Chabas, H., Ashby, B., Bondy-Denomy, J., Gandon, S., Boots, M., Paterson, S., Buckling, A., & Westra, E. R. (2016). The diversity-generating benefits of a prokaryotic adaptive immune system. *Nature*, *532*(7599), 385–388.

102. Howard-Flanders, P., & Theriot, L. (1966). Mutants of Escherichia coli K-12 defective in DNA repair and in genetic recombination. *Genetics*, *53*(6), 1137–1150.

103. Howard, J. A. L., Delmas, S., Ivančić-Baće, I., & Bolt, E. L. (2011). Helicase dissociation and annealing of RNA-DNA hybrids by Escherichia coli Cas3 protein. *Biochem. J.*, *439*(1), 85–95.

104. Høyland-Kroghsbo, N. M., Muñoz, K. A., & Bassler, B. L. (2018). Temperature, by controlling growth rate, regulates CRISPR-Cas activity in Pseudomonas aeruginosa. *MBio*, *9*(6).

105. Høyland-Kroghsbo, N. M., Paczkowski, J., Mukherjee, S., Broniewski, J., Westra, E., Bondy-Denomy, J., & Bassler, B. L. (2017). Quorum sensing controls the Pseudomonas aeruginosa CRISPR-Cas adaptive immune system. *Proc. Natl. Acad. Sci. USA*, *114*, 131–135.

106. Hudaiberdiev, S., Shmakov, S., Wolf, Y. I., Terns, M. P., Makarova, K. S., & Koonin, E. V. (2017). Phylogenomics of Cas4 family nucleases. *BMC Evol. Biol.*, *17*(1).

107. Hunt, E. A., Evans, T. C., & Tanner, N. A. (2018). Single-stranded binding proteins and helicase enhance the activity of prokaryotic argonautes in vitro. *PLoS One*, *13*(8), e0203073.

108. Hutvagner, G., & Simard, M. J. (2008). Argonaute proteins: key players in RNA silencing. *Nat. Rev. Mol. Cell Biol.*, *9*(1), 22–32.

109. Hynes, A. P., Villion, M., & Moineau, S. (2014). Adaptation in bacterial CRISPR-

Cas immunity can be driven by defective phages. *Nat. Commun.*, *5*, 4399.

110. Iida, S., Streiff, M. B., Bickle, T. A., & Arber, W. (1987). Two DNA antirestriction systems of bacteriophage P1, darA, and darB: characterization of darA- phages. *Virology*, *157*(1), 156–166.

111. Isaev, A., Drobiazko, A., Sierro, N., Gordeeva, J., Yosef, I., Qimron, U., Ivanov, N. V., & Severinov, K. (2020). Phage T7 DNA mimic protein Ocr is a potent inhibitor of BREX defence. *Nucleic Acids Res.*, *48*(10), 5397–5406.

112. Ivančić-Baće, I., Cass, S. D., Wearne, S. J., & Bolt, E. L. (2015). Different genome stability proteins underpin primed and naïve adaptation in E. coli CRISPR-Cas immunity. *Nucleic Acids Res.*, *43*(22), 10821–10830.

113. Jackson, R. N., Golden, S. M., van Erp, P. B. G., Carter, J., Westra, E. R., Brouns, S. J. J., van der Oost, J., Terwilliger, T. C., Read, R. J., & Wiedenheft, B. (2014). Crystal structure of the CRISPR RNA-guided surveillance complex from Escherichia coli. *Science*, *345*(6203), 1473–1479.

114. Jackson, S. A., McKenzie, R. E., Fagerlund, R. D., Kieper, S. N., Fineran, P. C., & Brouns, S. J. J. (2017). CRISPR-Cas: Adapting to change. *Science*, *356*(6333), eaal5056.

115. Janscak, P., MacWilliams, M. P., Sandmeier, U., Nagaraja, V., & Bickle, T. A. (1999). DNA translocation blockage, a general mechanism of cleavage site selection by type I restriction enzymes. *EMBO J.*, *18*(9), 2638–2647.

116. Jensen, R. B., & Gerdes, K. (1995). Programmed cell death in bacteria: proteic plasmid stabilization systems. *Mol. Microbiol.*, *17*(2), 205–210.

117. Jin, T., & Yin, J. (2021). Patterns of virus growth across the diversity of life. *Integr. Biol. (Camb)*, *13*, 44–59.

118. Jinek, M., Chylinski, K., Fonfara, I., Hauer, M., Doudna, J. A., & Charpentier, E. (2012). A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science*, *337*(6096), 816–821.

119. Jore, M. M., Lundgren, M., Van Duijn, E., Bultema, J. B., Westra, E. R., Waghmare, S. P., Wiedenheft, B., Pul, Ü., Wurm, R., Wagner, R., Beijer, M. R., Barendregt, A., Zhou, K., Snijders, A. P. L., Dickman, M. J., Doudna, J. A., Boekema, E. J., Heck, A. J. R., Van

Der Oost, J., & Brouns, S. J. J. (2011). Structural basis for CRISPR RNA-guided DNA recognition by Cascade. *Nat. Struct. Mol. Biol.*, *18*(5), 529–536.

120. Kazlauskiene, M., Kostiuk, G., Venclovas, Č., Tamulaitis, G., & Siksnys, V. (2017). A cyclic oligonucleotide signaling pathway in type III CRISPR-Cas systems. *Science*, *357*(6351), 605–609.

121. Kever, L., Hardy, A., Luthe, T., Hünnefeld, M., Gätgens, C., Milke, L., Wiechert, J., Wittmann, J., Moraru, C., Marienhagen, J., & Frunzke, J. (2022). Aminoglycoside antibiotics inhibit phage infection by blocking an early step of the infection cycle. *MBio*, *13*(3), e0078322.

122. Kim, S., Loeff, L., Colombo, S., Jergic, S., Brouns, S. J. J., & Joo, C. (2020). Selective loading and processing of prespacers for precise CRISPR adaptation. *Nature*, *579*(7797), 141–145.

123. Kim, T.-Y., Shin, M., Huynh Thi Yen, L., & Kim, J.-S. (2013). Crystal structure of Cas1 from Archaeoglobus fulgidus and characterization of its nucleolytic activity. *Biochem. Biophys. Res. Commun.*, *441*(4), 720–725.

124. Klompe, S. E., Vo, P. L. H., Halpin-Healy, T. S., & Sternberg, S. H. (2019). Transposon-encoded CRISPR–Cas systems direct RNA-guided DNA integration. *Nature*, *571*(7764), 219–225.

125. Kohanski, M. A., DePristo, M. A., & Collins, J. J. (2010). Sublethal antibiotic treatment leads to multidrug resistance via radical-induced mutagenesis. *Mol. Cell*, *37*, 311–320.

126. Kohanski, M. A., Dwyer, D. J., Hayete, B., Lawrence, C. A., & Collins, J. J. (2007). A common mechanism of cellular death induced by bactericidal antibiotics. *Cell*, *130*(5), 797–810.

127. Kohanski, M. A., Dwyer, D. J., Wierzbowski, J., Cottarel, G., & Collins, J. J. (2008). Mistranslation of membrane proteins and two-component system activation trigger aminoglycoside-mediated oxidative stress and cell death. *Cell*, *135*(4), 679.

128. Koonin, E. V., & Zhang, F. (2017). Coupling immunity and programmed cell suicide in prokaryotes: life-or-death choices. *Bioessays*, *39*, 1–9.

129. Koonin, E. V. (2017). Evolution of RNA- and DNA-guided antivirus defense systems in prokaryotes and eukaryotes: common ancestry vs convergence. *Biol. Direct 2017 121*, *12*(1), 1–14.

130. Koonin, E. V., & Krupovic, M. (2015). Evolution of adaptive immunity from transposable elements combined with innate immune systems. *Nat. Rev. Genet.*, *16*(3), 184.

131. Koonin, E. V., & Makarova, K. S. (2019). Origins and evolution of CRISPR-Cas systems. *Philos. Trans. R. Soc. B Biol. Sci.*, *374*(1772), 20180087.

132. Korangy, F., & Julin, D. A. (1994). Efficiency of ATP hydrolysis and DNA unwinding by the RecBC enzyme from Escherichia coli. *Biochemistry*, *33*(32), 9552–9560.

133. Kortright, K. E., Chan, B. K., Koff, J. L., & Turner, P. E. (2019). Phage therapy: a renewed approach to combat antibiotic-resistant bacteria. *Cell Host Microbe*, *25*, 219–232.

134. Kranzusch, P. J., Wilson, S. C., Lee, A. S. Y., Berger, J. M., Doudna, J. A., & Vance, R. E. (2015). Ancient origin of cGAS-STING reveals mechanism of universal 2′,3′ cGAMP signaling. *Mol. Cell*, *59*(6), 891–903.

135. Krivoy, A., Rutkauskas, M., Kuznedelov, K., Musharova, O., Rouillon, C., Severinov, K., & Seidel, R. (2018). Primed CRISPR adaptation in Escherichia coli cells does not depend on conformational changes in the Cascade effector complex detected in vitro. *Nucleic Acids Res.*, *46*(8), 4087–4098.

136. Krüger, D. H., & Bickle, T. A. (1983). Bacteriophage survival: multiple mechanisms for avoiding the deoxyribonucleic acid restriction systems of their hosts. *Microbiol. Rev.*, *47*(3), 345.

137. Krupovic, M., Béguin, P., & Koonin, E. V. (2017). Casposons: mobile genetic elements that gave rise to the CRISPR-Cas adaptation machinery. *Curr. Opin. Microbiol.*, *38*, 36.

138. Künne, T., Kieper, S. N., Bannenberg, J. W., Vogel, A. I. M., Miellet, W. R., Klein, M., Depken, M., Suarez-Diez, M., & Brouns, S. J. J. (2016). Cas3-derived target DNA degradation fragments fuel primed CRISPR adaptation. *Mol. Cell*, *63*, 852–864.

139. Kushner, S. R., Nagaishi, H., Templin, A., & Clark, A. J. (1971). Genetic recombination in Escherichia coli: The role of Exonuclease I. *Proc. Natl. Acad. Sci. U. S.*

A., *68*(4), 824–827.

140. Kuzmenko, A., Yudin, D., Ryazansky, S., Kulbachinskiy, A., & Aravin, A. A. (2019). Programmable DNA cleavage by Ago nucleases from mesophilic bacteria Clostridium butyricum and Limnothrix rosea. *Nucleic Acids Res.*, *47*(11), 5822–5836.

141. Kuzminov, A. (2001). Single-strand interruptions in replicating chromosomes cause double-strand breaks. *Proc. Natl. Acad. Sci. U. S. A.*, *98*(15), 8241.

142. Lam, S. T., Stahl, M. M., McMilin, K. D., & Stahl, F. W. (1974). Rec mediated recombinational hot spot activity in bacteriophage lambda. II. A mutation which causes hot spot activity. *Genetics*, *77*(3), 425–433.

143. Langendonk, R. F., Neill, D. R., & Fothergill, J. L. (2021). The building blocks of antimicrobial resistance in Pseudomonas aeruginosa: implications for current resistance-breaking therapies. *Front. Cell. Infect. Microbiol.*, *11*, 665759.

144. Łapińska, U., Glover, G., Capilla-Lasheras, P., Young, A. J., & Pagliara, S. (2019). Bacterial ageing in the absence of external stressors. *Philos. Trans. R. Soc. Lond. B Biol. Sci.*, *374*, 20180442.

145. Lenski, R. E., Rose, M. R., Simpson, S. C., & Tadler, S. C. (1991). Long-term experimental evolution in Escherichia coli. I. Adaptation and divergence during 2,000 generations. *Am. Nat.*, *138*(6), 1315–1341.

146. Levin, B. R., Moineau, S., Bushman, M., & Barrangou, R. (2013). The population and evolutionary dynamics of phage and bacteria with CRISPR-mediated immunity. *PLoS Genet.*, *9*(3), e1003312.

147. Levy, A., Goren, M. G. M. G., Yosef, I., Auster, O., Manor, M., Amitai, G., Edgar, R., Qimron, U., & Sorek, R. (2015). CRISPR adaptation biases explain preference for acquisition of foreign DNA. *Nature*, *520*(7548), 505–510.

148. Li, M., Liu, H., Han, J., Liu, J., Wang, R., Zhao, D., Zhou, J., & Xiang, H. (2013). Characterization of CRISPR RNA biogenesis and Cas6 cleavage-mediated inhibition of a provirus in the haloarchaeon Haloferax mediterranei. *J. Bacteriol.*, *195*(4), 867–875.

149. Li, R., Fang, L., Tan, S., Yu, M., Li, X., He, S., Wei, Y., Li, G., Jiang, J., & Wu, M. (2016). Type I CRISPR-Cas targets endogenous genes and regulates virulence to evade

mammalian host immunity. *Cell Res. 2016 2612*, *26*(12), 1273–1287.

150. Li, Y., & Bondy-Denomy, J. (2021). Anti-CRISPRs go viral: The infection biology of CRISPR-Cas inhibitors. *Cell Host Microbe*, *29*(5), 704–714.

151. Lingel, A., & Izaurralde, E. (2004). RNAi: Finding the elusive endonuclease. *RNA*, *10*(11), 1675.

152. Liu, J., Carmell, M. A., Rivas, F. V., Marsden, C. G., Thomson, J. M., Song, J. J., Hammond, S. M., Joshua-Tor, L., & Hannon, G. J. (2004). Argonaute2 is the catalytic engine of mammalian RNAi. *Science*, *305*(5689), 1437–1441.

153. Liu, T., Li, Y., Wang, X., Ye, Q., Li, H., Liang, Y., She, Q., & Peng, N. (2015). Transcriptional regulator-mediated activation of adaptation genes triggers CRISPR de novo spacer acquisition. *Nucleic Acids Res.*, *43*(2), 1044.

154. Liu, T., Liu, Z., Ye, Q., Pan, S., Wang, X., Li, Y., Peng, W., Liang, Y., She, Q., & Peng, N. (2017). Coupling transcriptional activation of CRISPR–Cas system and DNA repair genes by Csa3a in Sulfolobus islandicus. *Nucleic Acids Res.*, *45*(15), 8978.

155. Liu, Z., Sun, M., Liu, J., Liu, T., Ye, Q., Li, Y., & Peng, N. (2020). A CRISPR-associated factor Csa3a regulates DNA damage repair in Crenarchaeon Sulfolobus islandicus. *Nucleic Acids Res.*, *48*(17), 9681.

156. Lloyd, R. G., & Buckman, C. (1985). Identification and genetic analysis of sbcC mutations in commonly used recBC sbcB strains of Escherichia coli K-12. *J. Bacteriol.*, *164*(2), 836–844.

157. Lobritz, M. A., Belenky, P., Porter, C. B. M., Gutierrez, A., Yang, J. H., Schwarz, E. G., Dwyer, D. J., Khalil, A. S., & Collins, J. J. (2015). Antibiotic efficacy is linked to bacterial cellular respiration. *Proc. Natl. Acad. Sci. U. S. A.*, *112*(27), 8173–8180.

158. Lopatina, A., Tal, N., & Sorek, R. (2020). Abortive infection: bacterial suicide as an antiviral immune strategy. *Annu. Rev. Virol.*, *7*(1), 371–384.

159. Lovett, S. T., & Kolodner, R. D. (1989). Identification and purification of a single-stranded-DNA-specific exonuclease encoded by the recJ gene of Escherichia coli. *Proc. Natl. Acad. Sci. U. S. A.*, *86*(8), 2627–2631.

160. Ma, K., Cao, Q., Luo, S., Wang, Z., Liu, G., Lu, C., & Liu, Y. (2018). cas9 enhances

bacterial virulence by repressing the regR transcriptional regulator in Streptococcus agalactiae. *Infect. Immun.*, *86*(3).

161. Makarova, K. S. (2011). Evolution and classification of the CRISPR–Cas systems. *Nat. Publ. Gr.*, *9*(6), 467–477.

162. Makarova, K. S., Anantharaman, V., Aravind, L., & Koonin, E. V. (2012). Live virus-free or die: coupling of antivirus immunity and programmed suicide or dormancy in prokaryotes. *Biol. Direct*, *7*, 40.

163. Makarova, K. S., Aravind, L., Grishin, N. V., Rogozin, I. B., & Koonin, E. V. (2002). A DNA repair system specific for thermophilic archaea and bacteria predicted by genomic context analysis. *Nucleic Acids Res.*, *30*(2), 482–496.

164. Makarova, K. S., Grishin, N. V., Shabalina, S. A., Wolf, Y. I., & Koonin, E. V. (2006). A putative RNA-interference-based immune system in prokaryotes: computational analysis of the predicted enzymatic machinery, functional analogies with eukaryotic RNAi, and hypothetical mechanisms of action. *Biol. Direct*, *1*(1), 7.

165. Makarova, K. S., Wolf, Y. I., Iranzo, J., Shmakov, S. A., Alkhnbashi, O. S., Brouns, S. J. J., Charpentier, E., Cheng, D., Haft, D. H., Horvath, P., Moineau, S., Mojica, F. J. M., Scott, D., Shah, S. A., Siksnys, V., Terns, M. P., Venclovas, Č., White, M. F., Yakunin, A. F., Yan, W., Zhang, F., Garrett, R. A., Backofen, R., van der Oost, J., Barrangou, R., Koonin, E. V. (2020). Evolutionary classification of CRISPR–Cas systems: a burst of class 2 and derived variants. *Nat. Rev. Microbiol.*, *18*(2), 67–83.

166. Makarova, K. S., Wolf, Y. I., van der Oost, J., & Koonin, E. V. (2009). Prokaryotic homologs of Argonaute proteins are predicted to function as key components of a novel system of defense against mobile genetic elements. *Biol. Direct*, *4*, 29.

167. Malone, R. E., Chattoraj, D. K., Faulds, D. H., Stahl, M. M., & Stahl, F. W. (1978). Hotspots for generalized recombination in the Escherichia coli chromosome. *J. Mol. Biol.*, *121*(4), 473–491.

168. Marians, K. J. (2004). Mechanisms of replication fork restart in Escherichia coli. *Philos. Trans. R. Soc. London. Ser. B Biol. Sci.*, *359*(1441), 71–77.

169. Marraffini, L. A., & Sontheimer, E. J. (2008). CRISPR interference limits horizontal

gene transfer in Staphylococci by targeting DNA. *Science*, *322*(5909), 1843–1845.

170. Masterson, C., Boehmer, P. E., McDonald, F., Chaudhuri, S., Hickson, I. D., & Emmerson, P. T. (1992). Reconstitution of the activities of the RecBCD holoenzyme of Escherichia coli from the purified subunits. *J. Biol. Chem.*, *267*(19), 13564–13572.

171. Meeske, A. J., Jia, N., Cassel, A. K., Kozlova, A., Liao, J., Wiedmann, M., Patel, D. J., & Marraffini, L. A. (2020). A phage-encoded anti-CRISPR enables complete evasion of type VI-A CRISPR-Cas immunity. *Science*, *369*(6499), 54–59.

172. Meeske, A. J., Nakandakari-Higa, S., & Marraffini, L. A. (2019). Cas13-induced cellular dormancy prevents the rise of CRISPR-resistant bacteriophage. *Nature*, *570*(7760), 241–245.

173. Meisel, A., Bickle, T. A., Kriiger, D. H., & Schroeder, C. (1992). Type III restriction enzymes need two inversely oriented recognition sites for DNA cleavage. *Nature*, *355*(6359), 467–469.

174. Miezner, G., Turgeman-Grott, I., Zatopek, K. M., Gardner, A. F., Reshef, L., Choudhary, D. K., Alstetter, M., Allers, T., Marchfelder, A., & Gophna, U. (2023). An archaeal Cas3 protein facilitates rapid recovery from DNA damage. *MicroLife*, *4*, 1–9.

175. Mitić, D., Bolt, E. L., & Ivančić-Baće, I. (2023). CRISPR-Cas adaptation in Escherichia coli. *Biosci. Rep.*, *43*(3).

176. Modell, J. W., Jiang, W., & Marraffini, L. A. (2017). CRISPR–Cas systems exploit viral DNA injection to establish and maintain adaptive immunity. *Nature*, *544*, 101–104.

177. Mojica, F. J. M., Díez-Villaseñor, C., García-Martínez, J., & Almendros, C. (2009). Short motif sequences determine the targets of the prokaryotic CRISPR defence system. *Microbiology*, *155*(3), 733–740.

178. Molineux, I. J., & Gefter, M. L. (1975). Properties of the Escherichia coli DNA-binding (unwinding) protein interaction with nucleolytic enzymes and DNA. *J. Mol. Biol.*, *98*(4), 811–825.

179. Molineux, I. J., Schmitt, C. K., & Condreay, J. P. (1989). Mutants of bacteriophage T7 that escape F restriction. *J. Mol. Biol.*, *207*(3), 563–574.

180. Moore, S. D. (2011). Assembling new Escherichia coli strains by transduction using

phage P1. In *Methods in Molecular Biology* (pp. 155–169).

181. Morgan, M., Anders, S., Lawrence, M., Aboyoun, P., Pagès, H., & Gentleman, R. (2009). ShortRead: a bioconductor package for input, quality assessment and exploration of high-throughput sequence data. *Bioinformatics*, *25*(19), 2607–2608.

182. Mulepati, S., & Bailey, S. (2013). In vitro reconstitution of an Escherichia coli RNA-guided immune system reveals unidirectional, ATP-dependent degradation of DNA Target. *J. Biol. Chem.*, *288*(31), 22184–22192.

183. Murphy, K. C. (1991). Lambda Gam protein inhibits the helicase and chi-stimulated recombination activities of Escherichia coli RecBCD enzyme. *J. Bacteriol.*, *173*(18), 5808–5821.

184. Murugan, K., Babu, K., Sundaresan, R., Rajan, R., & Sashital, D. G. (2017). The revolution continues: Newly discovered systems expand the CRISPR-Cas toolkit. *Mol. Cell*, *68*(1), 15–25.

185. Musharova, O., Medvedeva, S., Klimuk, E., Guzman, N. M., Titova, D., Zgoda, V., Shiriaeva, A., Semenova, E., Severinov, K., & Savitskaya, E. (2021). Prespacers formed during primed adaptation associate with the Cas1–Cas2 adaptation complex and the Cas3 interference nuclease–helicase. *Proc. Natl. Acad. Sci. U. S. A.*, *118*(22).

186. Musharova, O., Sitnik, V., Vlot, M., Savitskaya, E., Datsenko, K. A., Krivoy, A., Fedorov, I., Semenova, E., Brouns, S. J. J., & Severinov, K. (2019). Systematic analysis of Type I-E Escherichia coli CRISPR-Cas PAM sequences ability to promote interference and primed adaptation. *Mol. Microbiol.*, *111*(6), 1558.

187. Musharova, O., Vyhovskyi, D., Medvedeva, S., Guzina, J., Zhitnyuk, Y., Djordjevic, M., Severinov, K., & Savitskaya, E. (2018). Avoidance of trinucleotide corresponding to consensus Protospacer Adjacent Motif controls the efficiency of prespacer selection during primed adaptation. *MBio*, *9*(6).

188. Nam, K. H., Ding, F., Haitjema, C., Huang, Q., DeLisa, M. P., & Ke, A. (2012). Double-stranded endonuclease activity in Bacillus halodurans clustered regularly interspaced short palindromic repeats (CRISPR)-associated Cas2 protein. *J. Biol. Chem.*, *287*(43), 35943–35952.

189. Nam, K. H., Haitjema, C., Liu, X., Ding, F., Wang, H., Delisa, M. P., & Ke, A. (2012). Cas5d protein processes Pre-crRNA and assembles into a cascade-like interference complex in subtype I-C/Dvulg crispr-cas system. *Structure*, *20*(9), 1574–1584.

190. Neylon, C., Kralicek, A. V., Hill, T. M., & Dixon, N. E. (2005). Replication termination in Escherichia coli: structure and antihelicase activity of the Tus-Ter complex. *Microbiol. Mol. Biol. Rev.*, *69*(3), 501–526.

191. Niewoehner, O., Garcia-Doval, C., Rostøl, J. T., Berk, C., Schwede, F., Bigler, L., Hall, J., Marraffini, L. A., & Jinek, M. (2017). Type III CRISPR–Cas systems produce cyclic oligoadenylate second messengers. *Nat. 2017 5487669*, *548*(7669), 543–548.

192. Nuñez, J. K., Bai, L., Harrington, L. B., Hinder, T. L., & Doudna, J. A. (2016). CRISPR immunological memory requires a host factor for specificity. *Mol. Cell*, *62*(6), 824–833.

193. Nuñez, J. K., Harrington, L. B., Kranzusch, P. J., Engelman, A. N., & Doudna, J. A. (2015). Foreign DNA capture during CRISPR–Cas adaptive immunity. *Nature*, *527*(7579), 535–538.

194. Nuñez, J. K., Kranzusch, P. J., Noeske, J., Wright, A. V., Davies, C. W., & Doudna, J. A. (2014a). Cas1-Cas2 complex formation mediates spacer acquisition during CRISPR-Cas adaptive immunity. *Nat. Struct. Mol. Biol.*, *21*(6), 528–534.

195. Nuñez, J. K., Kranzusch, P. J., Noeske, J., Wright, A. V, Davies, C. W., & Doudna, J. A. (2014b). Cas1 – Cas2 complex formation mediates spacer acquisition during CRISPR – Cas adaptive immunity. *Nat. Publ. Gr.*, *21*(6), 528–534.

196. Nuñez, J. K., Lee, A. S. Y., Engelman, A., & Doudna, J. A. (2015). Integrase-mediated spacer acquisition during CRISPR–Cas adaptive immunity. *Nature*, *519*(7542), 193–198.

197. Nussenzweig, P. M., McGinn, J., & Marraffini, L. A. (2019). Cas9 cleavage of viral genomes primes the acquisition of new immunological memories. *Cell Host Microbe*, *26*(4), 515-526.e6.

198. Nygaard, V., Rødland, E. A., & Hovig, E. (2016). Methods that remove batch effects while retaining group differences may lead to exaggerated confidence in downstream analyses. *Biostatistics*, *17*(1), 29–39.

199. O'Toole, G. A., & Kolter, R. (1998). Flagellar and twitching motility are necessary for Pseudomonas aeruginosa biofilm development. *Mol. Microbiol.*, *30*, 295–304.

200. Ogura, T., & Hiraga, S. (1983). Mini-F plasmid genes that couple host cell division to plasmid proliferation. *Proc. Natl. Acad. Sci. U. S. A.*, *80*(15), 4784–4788.

201. Olovnikov, I., Chan, K., Sachidanandam, R., Newman, D. K., & Aravin, A. A. (2013). Bacterial Argonaute samples the transcriptome to identify foreign DNA. *Mol. Cell*, *51*(5), 594–605.

202. Pagès, H., Aboyoun, P., Gentleman, R., & DebRoy, S. (2008). *Biostrings: Efficient manipulation of biological strings.* (R Package Version 2.50.2.). https://bioconductor.org/packages/Biostrings

203. Palas, K. M., & Kushner, S. R. (1990). Biochemical and physical characterization of exonuclease V from Escherichia coli. Comparison of the catalytic activities of the RecBC and RecBCD enzymes. *J. Biol. Chem.*, *265*(6), 3447–3454.

204. Pandey, D. P., & Gerdes, K. (2005). Toxin–antitoxin loci are highly abundant in free-living but lost from host-associated prokaryotes. *Nucleic Acids Res.*, *33*(3), 966.

205. Pankey, G. A., & Sabath, L. D. (2004). Clinical relevance of bacteriostatic versus bactericidal mechanisms of action in the treatment of gram-positive bacterial infections. *Clin. Infect. Dis.*, *38*, 864–870.

206. Patterson, A. G., Jackson, S. A., Taylor, C., Evans, G. B., Salmond, G. P. C. C., Przybilski, R., Staals, R. H. J., & Fineran, P. C. (2016). Quorum sensing controls adaptive immunity through the regulation of multiple CRISPR-Cas systems. *Mol. Cell*, *64*(6), 1102–1108.

207. Persky, N. S., & Lovett, S. T. (2008). Mechanisms of recombination: lessons from E. coli. *Crit. Rev. Biochem. Mol. Biol.*, *43*(6), 347–370.

208. Peters, J. E., Makarova, K. S., Shmakov, S., & Koonin, E. V. (2017). Recruitment of CRISPR-Cas systems by Tn7-like transposons. *Proc. Natl. Acad. Sci. U. S. A.*, *114*(35), E7358–E7366.

209. Petzoldt, T. (2022). *Estimate growth rates from experimental data [R package growthrates version 0.8.4]*. https://cran.r-project.org/package=growthrates

210. Philippe, C., Morency, C., Plante, P. L., Zufferey, E., Achigar, R., Tremblay, D. M., Rousseau, G. M., Goulet, A., & Moineau, S. (2022). A truncated anti-CRISPR protein prevents spacer acquisition but not interference. *Nat. Commun. 2022 131*, *13*(1), 1–8.

211. Pingoud, A., Wilson, G. G., & Wende, W. (2014). Type II restriction endonucleases — a historical perspective and more. *Nucleic Acids Res.*, *42*(12), 7489.

212. Ponticelli, A. S., Schultz, D. W., Taylor, A. F., & Smith, G. R. (1985). Chi-dependent DNA strand cleavage by RecBC enzyme. *Cell*, *41*(1), 145–151.

213. Pougach, K., Semenova, E., Bogdanova, E., Datsenko, K. A., Djordjevic, M., Wanner, B. L., & Severinov, K. (2010). Transcription, processing and function of CRISPR cassettes in Escherichia coli. *Mol. Microbiol.*, *77*(6), 1367–1379.

214. Powell, L. M., Dryden, D. T. F., Willcock, D. F., Pain, R. H., & Murray, N. E. (1993). DNA recognition by the EcoK methyltransferase. The influence of DNA methylation and the cofactor S-adenosyl-L-methionine. *J. Mol. Biol.*, *234*(1), 60–71.

215. Przybilski, R., Richter, C., Gristwood, T., Clulow, J. S., Vercoe, R. B., & Fineran, P. C. (2011). Csy4 is responsible for CRISPR RNA processing in Pectobacterium atrosepticum. *RNA Biol.*, *8*(3), 517–528.

216. Pul, Ü., Wurm, R., Arslan, Z., Geißen, R., Hofmann, N., & Wagner, R. (2010). Identification and characterization of E. coli CRISPR-cas promoters and their silencing by H-NS. *Mol. Microbiol.*, *75*(6), 1495–1512.

217. Pursey, E., Sünderhauf, D., Gaze, W. H., Westra, E. R., & van Houte, S. (2018). CRISPR-Cas antimicrobials: challenges and future prospects. *PLoS Pathog.*, *14*, e1006990.

218. Qi, L. S., Larson, M. H., Gilbert, L. A., Doudna, J. A., Weissman, J. S., Arkin, A. P., & Lim, W. A. (2013). Repurposing CRISPR as an RNA-guided platform for sequence-specific control of gene expression. *Cell*, *152*(5), 1173–1183.

219. Rabinovitch, A., Fishov, I., Hadas, H., Einav, M., & Zaritsky, A. (2002). Bacteriophage T4 development in Escherichia coli is growth rate dependent. *J. Theor. Biol.*, *216*, 1–4.

220. Radovčić, M., Killelea, T., Savitskaya, E., Wettstein, L., Bolt, E. L., & Ivančić-Baće,

I. (2018). CRISPR–Cas adaptation in Escherichia coli requires RecBCD helicase but not nuclease activity, is independent of homologous recombination, and is antagonized by 5′ ssDNA exonucleases. *Nucleic Acids Res.*, *46*(19), 10173–10183.

221. Ramachandran, A., Summerville, L., Learn, B. A., DeBell, L., & Bailey, S. (2020). Processing and integration of functionally oriented prespacers in the Escherichia coli CRISPR system depends on bacterial host exonucleases. *J. Biol. Chem.*, *295*(11), 3403–3414.

222. Rasband, W. (2012). ImageJ. *U. S. Natl. Institutes Heal. Bethesda, Maryland, USA*.

223. Rath, D., Amlinger, L., Hoekzema, M., Devulapally, P. R., & Lundgren, M. (2015). Efficient programmable gene silencing by Cascade. *Nucleic Acids Res.*, *43*(1), 237.

224. Redding, S., Sternberg, S. H., Marshall, M., Gibb, B., Bhat, P., Guegler, C. K., Wiedenheft, B., Doudna, J. A., & Greene, E. C. (2015). Surveillance and processing of foreign DNA by the Escherichia coli CRISPR-Cas system. *Cell*, *163*(4), 854–865.

225. Reich, S., Gössl, I., Reuter, M., Rabe, J. P., & Krüger, D. H. (2004). Scanning force microscopy of DNA translocation by the Type III restriction enzyme EcoP15I. *J. Mol. Biol.*, *341*(2), 337–343.

226. Richter, C., Dy, R. L., McKenzie, R. E., Watson, B. N. J., Taylor, C., Chang, J. T., McNeil, M. B., Staals, R. H. J., & Fineran, P. C. (2014). Priming in the Type I-F CRISPR-Cas system triggers strand-independent spacer acquisition, bi-directionally from the primed protospacer. *Nucleic Acids Res.*, *42*(13), 8516.

227. Richter, C., Gristwood, T., Clulow, J. S., & Fineran, P. C. (2012). In vivo protein interactions and complex formation in the Pectobacterium atrosepticum subtype I-F CRISPR/Cas System. *PLoS One*, *7*(12), e49549.

228. Rollie, C., Schneider, S., Brinkmann, A. S., Bolt, E. L., & White, M. F. (2015). Intrinsic sequence specificity of the Cas1 integrase directs new spacer acquisition. *Elife*, *4*(AUGUST2015), 1–19.

229. Rostøl, J. T., & Marraffini, L. A. (2019). Non-specific degradation of transcripts promotes plasmid clearance during type III-A CRISPR–Cas immunity. *Nat. Microbiol.*, *4*, 656–662.

230. Rostøl, J. T., Xie, W., Kuryavyi, V., Maguin, P., Kao, K., Froom, R., Patel, D. J., & Marraffini, L. A. (2021). The Card1 nuclease provides defence during type III CRISPR immunity. *Nature*, *590*, 624–629.

231. Saha, C., Horst-Kreft, D., Kross, I., van der Spek, P. J., Louwen, R., & van Baarlen, P. (2020). Campylobacter jejuni Cas9 modulates the transcriptome in Caco-2 intestinal epithelial cells. *Genes (Basel).*, *11*(10), 1193.

232. Samai, P., Smith, P., & Shuman, S. (2010). Structure of a CRISPR-associated protein Cas2 from Desulfovibrio vulgaris. *Acta Crystallogr. Sect. F Struct. Biol. Cryst. Commun.*, *66*(12), 1552–1556.

233. Sampson, T. R., Saroj, S. D., Llewellyn, A. C., Tzeng, Y.-L., & Weiss, D. S. (2013). A CRISPR/Cas system mediates bacterial innate immune evasion and virulence. *Nature*, *497*(7448), 254–257.

234. Sashital, D. G., Jinek, M., & Doudna, J. A. (2011). An RNA-induced conformational change required for CRISPR RNA cleavage by the endoribonuclease Cse3. *Nat. Struct. Mol. Biol.*, *18*(6), 680–687.

235. Sashital, D. G., Wiedenheft, B., & Doudna, J. A. (2012). Mechanism of foreign DNA selection in a bacterial adaptive immune system. *Mol. Cell*, *46*(5), 606–615.

236. Savitskaya, E., Semenova, E., Dedkov, V., Metlitskaya, A., & Severinov, K. (2013). High-throughput analysis of type I-E CRISPR/Cas spacer acquisition in E. coli. *RNA Biol.*, *10*(5), 716–725.

237. Schmitt, C. K., Kemp, P., & Molineux, I. J. (1991). Genes 1.2 and 10 of bacteriophages T3 and T7 determine the permeability lesions observed in infected cells of Escherichia coli expressing the F plasmid gene pifA. *J. Bacteriol.*, *173*(20), 6507–6514.

238. Schmitt, C. K., & Molineux, I. J. (1991). Expression of gene 1.2 and gene 10 of bacteriophage T7 is lethal to F plasmid-containing Escherichia coli. *J. Bacteriol.*, *173*(4), 1536–1543.

239. Schneider, C. A., Rasband, W. S., & Eliceiri, K. W. (2012). NIH Image to ImageJ: 25 years of image analysis. *Nat. Methods*, *9*, 671–675.

240. Schultz, D. W., Taylor, A. F., & Smith, G. R. (1983). Escherichia coli RecBC

pseudorevertants lacking Chi recombinational hotspot activity. *J. Bacteriol.*, *155*(2), 664–680.

241. Segall, A. M., Roach, D. R., & Strathdee, S. A. (2019). Stronger together? Perspectives on phage-antibiotic synergy in clinical applications of phage therapy. *Curr. Opin. Microbiol.*, *51*, 46–50.

242. Seidel, R., Van Noort, J., Van Der Scheer, C., Bloom, J. G. P., Dekker, N. H., Dutta, C. F., Blundell, A., Robinson, T., Firman, K., & Dekker, C. (2004). Real-time observation of DNA translocation by the type I restriction modification enzyme EcoR124I. *Nat. Struct. Mol. Biol.*, *11*(9), 838–843.

243. Semenova, E., Jore, M. M., Datsenko, K. a, Semenova, A., Westra, E. R., Wanner, B., van der Oost, J., Brouns, S. J. J., & Severinov, K. (2011). Interference by clustered regularly interspaced short palindromic repeat (CRISPR) RNA is governed by a seed sequence. *Proc. Natl. Acad. Sci. U. S. A.*, *108*(25), 10098–10103.

244. Severin, G. B., Ramliden, M. S., Hawver, L. A., Wang, K., Pell, M. E., Kieninger, A.-K., Khataokar, A., O'Hara, B. J., Behrmann, L. V., Neiditch, M. B., Benning, C., Waters, C. M., & Ng, W.-L. (2018). Direct activation of a phospholipase by cyclic GMP-AMP in El Tor Vibrio cholerae. *Proc. Natl. Acad. Sci. U. S. A.*, *115*(26), E6048–E6055.

245. Severinov, K., Ispolatov, I., & Semenova, E. (2016). The influence of copy-number of targeted extrachromosomal genetic elements on the outcome of CRISPR-Cas defense. *Front. Mol. Biosci.*, *3*(AUG), 45.

246. Shabbir, M. A. B., Tang, Y., Xu, Z., Lin, M., Cheng, G., Dai, M., Wang, X., Liu, Z., Yuan, Z., & Hao, H. (2018). The involvement of the Cas9 gene in virulence of Campylobacter jejuni. *Front. Cell. Infect. Microbiol.*, *8*(AUG), 362795.

247. Sharma, S., Javadekar, S. M., Pandey, M., Srivastava, M., Kumari, R., & Raghavan, S. C. (2015). Homology and enzymatic requirements of microhomology-dependent alternative end joining. *Cell Death Dis.*, *6*(3), e1697.

248. Shen, B. W., Doyle, L. A., Werther, R., Westburg, A. A., Bies, D. P., Walter, S. I., Luyten, Y. A., Morgan, R. D., Stoddard, B. L., & Kaiser, B. K. (2013). Structure, substrate binding and activity of a unique AAA+ protein: the BrxL phage restriction factor. *Nucleic*

Acids Res., *1*(1256879), 13–14.

249. Shiimori, M., Garrett, S. C., Graveley, B. R., & Terns, M. P. (2018). Cas4 nucleases define the PAM, length, and orientation of DNA fragments integrated at CRISPR loci. *Mol. Cell*, *70*(5), 814-824.e6.

250. Shinedling, S., Gayle, M., Pribnow, D., & Gold, L. (1987). Mutations affecting translation of the bacteriophage T4 rIIB gene cloned in Escherichia coli. *MGG Mol. Gen. Genet.*, *207*(2–3), 224–232.

251. Shipman, S. L., Nivala, J., Macklis, J. D., & Church, G. M. (2016). Molecular recordings by directed CRISPR spacer acquisition. *Science*, *353*(6298), aaf1175.

252. Shiriaeva, A. A., Savitskaya, E., Datsenko, K. A., Vvedenskaya, I. O., Fedorova, I., Morozova, N., Metlitskaya, A., Sabantsev, A., Nickels, B. E., Severinov, K., & Semenova, E. (2019). Detection of spacer precursors formed in vivo during primed CRISPR adaptation. *Nat. Commun.*, *10*(1), 4603.

253. Shmakov, S. A., Abudayyeh, O. O., Makarova, K. S., Wolf, Y. I., Gootenberg, J. S., Semenova, E., Minakhin, L., Joung, J., Konermann, S., Severinov, K., Zhang, F., & Koonin, E. V. (2015). Discovery and functional characterization of diverse Class 2 CRISPR-Cas systems. *Mol. Cell*, *60*(3), 385–397.

254. Shmakov, S. A., Makarova, K. S., Wolf, Y. I., Severinov, K. V., & Koonin, E. V. (2018). Systematic prediction of genes functionally linked to CRISPR-Cas systems by gene neighborhood analysis. *Proc. Natl. Acad. Sci. U. S. A.*, *115*(23), E5307–E5316.

255. Shmakov, S. A., Savitskaya, E., Semenova, E., Logacheva, M. D., Datsenko, K. A., & Severinov, K. V. (2014). Pervasive generation of oppositely oriented spacers during CRISPR adaptation. *Nucleic Acids Res.*, *42*(9), 5907–5916.

256. Shmakov, S. A., Sitnik, V., Makarova, K. S., Wolf, Y. I., Severinov, K. V., & Koonin, E. V. (2017). The CRISPR spacer space is dominated by sequences from species-specific mobilomes. *MBio*, *8*(5).

257. Shuman, S., & Glickman, M. S. (2007). Bacterial DNA repair by non-homologous end joining. *Nat. Rev. Microbiol.*, *5*(11), 852–861.

258. Silberstein, Z., & Cohen, A. (1987). Synthesis of linear multimers of OriC and

pBR322 derivatives in Escherichia coli K-12: role of recombination and replication functions. *J. Bacteriol.*, *169*(7), 3131–3137.

259. Silverstein, J. L., & Goldberg, E. B. (1976). T4 DNA injection. II. Protection of entering DNA from host exonuclease V. *Virology*, *72*(1), 212–223.

260. Singleton, M. R., Dillingham, M. S., Gaudier, M., Kowalczykowski, S. C., & Wigley, D. B. (2004). Crystal structure of RecBCD enzyme reveals a machine for processing DNA breaks. *Nature*, *432*(7014), 187–193.

261. Sinha, A. K., Possoz, C., & Leach, D. R. F. (2020). The roles of bacterial DNA double-strand break repair proteins in chromosomal DNA replication. *FEMS Microbiol. Rev.*, *44*(3), 351–368.

262. Sinkunas, T., Gasiunas, G., Waghmare, S. P., Dickman, M. J., Barrangou, R., Horvath, P., & Siksnys, V. (2013). In vitro reconstitution of Cascade-mediated CRISPR immunity in Streptococcus thermophilus. *EMBO J.*, *32*(3), 385–394.

263. Sistla, S., & Rao, D. N. (2004). S-Adenosyl-L-methionine-dependent restriction enzymes. *Crit. Rev. Biochem. Mol. Biol.*, *39*(1), 1–19.

264. Slavcev, R. A., & Hayes, S. (2003). Stationary phase-like properties of the bacteriophage λ Rex exclusion phenotype. *Mol. Genet. Genomics*, *269*(1), 40–48.

265. Snyder, L. (1995). Phage-exclusion enzymes: a bonanza of biochemical and cell biology reagents? *Mol. Microbiol.*, *15*(3), 415–420.

266. Spencer, B. L., Deng, L., Patras, K. A., Burcham, Z. M., Sanches, G. F., Nagao, P. E., & Doran, K. S. (2019). Cas9 contributes to group b streptococcal colonization and disease. *Front. Microbiol.*, *10*(AUG), 478424.

267. Spies, M., Amitani, I., Baskin, R. J., & Kowalczykowski, S. C. (2007). RecBCD enzyme switches lead motor subunits in response to chi recognition. *Cell*, *131*(4), 694–705.

268. Stachler, A. E., Turgeman-Grott, I., Shtifman-Segal, E., Allers, T., Marchfelder, A., & Gophna, U. (2017). High tolerance to self-targeting of the genome by the endogenous CRISPR-Cas system in an archaeon. *Nucleic Acids Res.*, *45*(9), 5208.

269. Stella, S., Alcón, P., & Montoya, G. (2017). Class 2 CRISPR–Cas RNA-guided

endonucleases: Swiss Army knives of genome editing. *Nat. Struct. Mol. Biol.*, *24*(11), 882–892.

270. Stewart, F. J., Panne, D., Bickle, T. A., & Raleigh, E. A. (2000). Methyl-specific DNA binding by McrBC, a modification-dependent restriction enzyme. *J. Mol. Biol.*, *298*(4), 611–622.

271. Strotskaya, A., Savitskaya, E., Metlitskaya, A., Morozova, N., Datsenko, K. A. K. A., Semenova, E., & Severinov, K. (2017). The action of Escherichia coli CRISPR–Cas system on lytic bacteriophages with different lifestyles and development strategies. *Nucleic Acids Res.*, *45*(4), 1946–1957.

272. Strutt, S. C., Torrez, R. M., Kaya, E., Negrete, O. A., & Doudna, J. A. (2018). RNA-dependent RNA targeting by CRISPR-Cas9. *Elife*, *7*.

273. Studier, F. W., & Rao Movva, N. (1976). SAMase gene of bacteriophage T3 is responsible for overcoming host restriction. *J. Virol.*, *19*(1), 136.

274. Subramaniam, S., & Smith, G. R. (2022). RecBCD enzyme and Chi recombination hotspots as determinants of self vs. non-self: Myths and mechanisms. *Adv. Genet.*, *109*, 1–37.

275. Swarts, D. C., & Jinek, M. (2019). Mechanistic insights into the cis- and trans-acting DNase activities of Cas12a. *Mol. Cell*, *73*(3), 589-600.e4.

276. Swarts, D. C., Jore, M. M., Westra, E. R., Zhu, Y., Janssen, J. H., Snijders, A. P., Wang, Y., Patel, D. J., Berenguer, J., Brouns, S. J. J., & van der Oost, J. (2014). DNA-guided DNA interference by a prokaryotic Argonaute. *Nature*, *507*(7491), 258–261.

277. Swarts, D. C., Makarova, K., Wang, Y., Nakanishi, K., Ketting, R. F., Koonin, E. V., Patel, D. J., & Van Der Oost, J. (2014). The evolutionary journey of Argonaute proteins. *Nat. Struct. Mol. Biol.*, *21*(9), 743.

278. Swarts, D. C., Mosterd, C., van Passel, M. W. J., & Brouns, S. J. J. (2012). CRISPR interference directs strand specific spacer acquisition. *PLoS One*, *7*(4), e35888.

279. Swarts, D. C., Szczepaniak, M., Sheng, G., Chandradoss, S. D., Zhu, Y., Timmers, E. M., Zhang, Y., Zhao, H., Lou, J., Wang, Y., Joo, C., & van der Oost, J. (2017). Autonomous generation and loading of DNA guides by bacterial Argonaute. *Mol. Cell*, *65*(6), 985-

998.e6.

280. Tagliaferri, T. L., Jansen, M., & Horz, H. P. (2019). Fighting pathogenic bacteria on two fronts: phages and antibiotics as combined strategy. *Front. Cell. Infect. Microbiol.*, *9*, 22.

281. Tateda, K., Ishii, Y., Kimura, S., Yamaguchi, K., Horikawa, M., & Miyairi, S. (2007). Suppression of Pseudomonas aeruginosa quorum-sensing systems by macrolides: a promising strategy or an oriental mystery? *J. Infect. Chemother.*, *13*(6), 357–367.

282. Taylor, A. F., Schultz, D. W., Ponticelli, A. S., & Smith, G. R. (1985). RecBC enzyme nicking at chi sites during DNA unwinding: Location and orientation-dependence of the cutting. *Cell*, *41*(1), 153–163.

283. Taylor, A. F., & Smith, G. R. (1980). Unwinding and rewinding of DNA by the RecBC enzyme. *Cell*, *22*(2 Pt 2), 447–457.

284. Taylor, A. F., & Smith, G. R. (1985). Substrate specificity of the DNA unwinding activity of the RecBC enzyme of Escherichia coli. *J. Mol. Biol.*, *185*(2), 431–443.

285. Taylor, A. F., & Smith, G. R. (1992). RecBCD enzyme is altered upon cutting DNA at a chi recombination hotspot. *Proc. Natl. Acad. Sci. U. S. A.*, *89*(12), 5226–5230.

286. Taylor, A. F., & Smith, G. R. (1995a). Monomeric RecBCD enzyme binds and unwinds DNA. *J. Biol. Chem.*, *270*(41), 24451–24458.

287. Taylor, A. F., & Smith, G. R. (1995b). Strand specificity of nicking of DNA at Chi sites by RecBCD enzyme. Modulation by ATP and magnesium levels. *J. Biol. Chem.*, *270*(41), 24459–24467.

288. Taylor, A. F., & Smith, G. R. (2003). RecBCD enzyme is a DNA helicase with fast and slow motors of opposite polarity. *Nature*, *423*(6942), 889–893.

289. Terakawa, T., Redding, S., Silverstein, T. D., & Greene, E. C. (2017). Sequential eviction of crowded nucleoprotein complexes by the exonuclease RecBCD molecular motor. *Proc. Natl. Acad. Sci. U. S. A.*, *114*(31), E6322–E6331.

290. Thavalingam, A., Cheng, Z., Garcia, B., Huang, X., Shah, M., Sun, W., Wang, M., Harrington, L., Hwang, S., Hidalgo-Reyes, Y., Sontheimer, E. J., Doudna, J., Davidson, A. R., Moraes, T. F., Wang, Y., & Maxwell, K. L. (2019). Inhibition of CRISPR-Cas9

ribonucleoprotein complex assembly by anti-CRISPR AcrIIC2. *Nat. Commun.*, *10*(1).

291. Thoms, B., & Wackernagel, W. (1998). Interaction of RecBCD enzyme with DNA at double-strand breaks produced in UV-irradiated Escherichia coli : Requirement for DNA end processing. *J. Bacteriol.*, *180*(21), 5639–5645.

292. Tock, M. R., & Dryden, D. T. F. (2005). The biology of restriction and anti-restriction. *Curr. Opin. Microbiol.*, *8*(4), 466–472.

293. Toothman, P., & Herskowitz, I. (1980). Rex-dependent exclusion of lambdoid phages II. Determinants of sensitivity to exclusion. *Virology*, *102*(1), 147–160.

294. Torres-Barceló, C., Gurney, J., Gougat-Barberá, C., Vasse, M., & Hochberg, M. E. (2018). Transient negative effects of antibiotics on phages do not jeopardise the advantages of combination therapies. *FEMS Microbiol. Ecol.*, *94*, fiy107.

295. Torres-Barceló, C., & Hochberg, M. E. (2016). Evolutionary rationale for phages as complements of antibiotics. *Trends Microbiol.*, *24*, 249–256.

296. Umezu, K., Nakayama, K., & Nakayama, H. (1990). Escherichia coli RecQ protein is a DNA helicase. *Proc. Natl. Acad. Sci. U. S. A.*, *87*(14), 5363–5367.

297. Vestergaard, G., Garrett, R. A., & Shah, S. A. (2014). CRISPR adaptive immune systems of Archaea. *RNA Biol.*, *11*(2), 156.

298. Viswanathan, M., Burdett, V., Baitinger, C., Modrich, P., & Lovett, S. T. (2001). Redundant exonuclease involvement in Escherichia coli methyl-directed mismatch repair. *J. Biol. Chem.*, *276*(33), 31053–31058.

299. Viswanathan, M., & Lovett, S. T. (1998). Single-strand DNA-specific exonucleases in Escherichia coli: Roles in repair and mutation avoidance. *Genetics*, *149*(1), 7–16.

300. Viswanathan, P., Murphy, K., Julien, B., Garza, A. G., & Kroos, L. (2007). Regulation of dev, an operon that includes genes essential for Myxococcus xanthus development and CRISPR-associated genes and repeats. *J. Bacteriol.*, *189*(10), 3738–3750.

301. Wagih, O. (2017). ggseqlogo: a versatile R package for drawing sequence logos. *Bioinformatics*, *33*(22), 3645–3647.

302. Wang, J., Chen, R., & Julin, D. A. (2000). A single nuclease active site of the Escherichia coli RecBCD enzyme catalyzes single-stranded DNA degradation in both

directions. *J. Biol. Chem.*, *275*(1), 507–513.

303. Wang, J., Li, J., Zhao, H., Sheng, G., Wang, M., Yin, M., & Wang, Y. (2015). Structural and mechanistic basis of PAM-dependent spacer acquisition in CRISPR-Cas systems. *Cell*, *163*(4), 840–853.

304. Wang, Y., Yin, X., Zhou, Z., Hu, S., Li, S., Liu, M., Wang, X., Xiao, Y., Shi, D., Bi, D., & Li, Z. (2019). Cas9 regulated gene expression and pathogenicity in Riemerella anatipestifer. *Microb. Pathog.*, *136*, 103706.

305. Warren, R. A. (1980). Modified bases in bacteriophage DNAs. *Annu. Rev. Microbiol.*, *34*(1), 137–158.

306. Wei, Y., Chesne, M. T., Terns, R. M., & Terns, M. P. (2015). Sequences spanning the leader-repeat junction mediate CRISPR adaptation to phage in Streptococcus thermophilus. *Nucleic Acids Res.*, *43*(3), 1749.

307. Werner, E., Roe, F., Bugnicourt, A., Franklin, M. J., Heydorn, A., Molin, S., Pitts, B., & Stewart, P. S. (2004). Stratified growth in Pseudomonas aeruginosa biofilms. *Appl. Environ. Microbiol.*, *70*, 6188–6196.

308. Westra, E. R., van Houte, S., Oyesiku-Blakemore, S., Makin, B., Broniewski, J. M., Best, A., Bondy-Denomy, J., Davidson, A., Boots, M., & Buckling, A. (2015). Parasite exposure drives selective evolution of constitutive versus inducible defense. *Curr. Biol.*, *25*, 1043–1049.

309. Westra, E. R., & Levin, B. R. (2020). It is unclear how important CRISPR-Cas systems are for protecting natural populations of bacteria against infections by mobile genetic elements. *Proc. Natl. Acad. Sci. U. S. A.*, *117*(45), 27777–27785.

310. Westra, E. R., Pul, U., Heidrich, N., Jore, M. M., Lundgren, M., Stratmann, T., Wurm, R., Raine, A., Mescher, M., Van Heereveld, L., Mastop, M., Wagner, E. G. H., Schnetz, K., van der Oost, J., Wagner, R., & Brouns, S. J. J. (2010). H-NS-mediated repression of CRISPR-based immunity in Escherichia coli K12 can be relieved by the transcription activator LeuO. *Mol. Microbiol.*, *77*(6), 1380–1393.

311. Westra, E. R., Semenova, E., Datsenko, K. A., Jackson, R. N., Wiedenheft, B., Severinov, K., & Brouns, S. J. J. (2013). Type I-E CRISPR-cas systems discriminate target

from non-target DNA through base pairing-independent PAM recognition. *PLoS Genet.*, *9*(9), e1003742.

312. Westra, E. R., van Erp, P. B. G., Künne, T., Wong, S. P., Staals, R. H. J., Seegers, C. L. C., Bollen, S., Jore, M. M., Semenova, E., Severinov, K. V., de Vos, W. M., Dame, R. T., de Vries, R., Brouns, S. J. J., & van der Oost, J. (2012). CRISPR immunity relies on the consecutive binding and degradation of negatively supercoiled invader DNA by Cascade and Cas3. *Mol. Cell*, *46*(5), 595–605.

313. Whiteley, A. T., Eaglesham, J. B., de Oliveira Mann, C. C., Morehouse, B. R., Lowey, B., Nieminen, E. A., Danilchanka, O., King, D. S., Lee, A. S. Y., Mekalanos, J. J., & Kranzusch, P. J. (2019). Bacterial cGAS-like enzymes synthesize diverse nucleotide signals. *Nature*, *567*(7747), 194–199.

314. Wiedenheft, B., Lander, G. C., Zhou, K., Jore, M. M., Brouns, S. J. J., van der Oost, J., Doudna, J. A., & Nogales, E. (2011). Structures of the RNA-guided surveillance complex from a bacterial immune system. *Nature*, *477*(7365), 486–489.

315. Wiedenheft, B., Van Duijn, E., Bultema, J., Waghmare, S., Zhou, K., Barendregt, A., Westphal, W., Heck, A., Boekema, E., Dickman, M., & Doudna, J. A. (2011). RNA-guided complex from a bacterial immune system enhances target recognition through seed sequence interactions. *Proc. Natl. Acad. Sci. U. S. A.*, *108*(25), 10092–10097.

316. Wiedenheft, B., Zhou, K., Jinek, M., Coyle, S. M., Ma, W., & Doudna, J. A. (2009). Structural basis for DNase activity of a conserved protein implicated in CRISPR-mediated genome defense. *Structure*, *17*(6), 904–912.

317. Wilke, C. O. (2017). *cowplot: Streamlined plot theme and plot annotations for "ggplot2". R package version 0.8.0*. https://cran.r-project.org/package=cowplot

318. Willetts, N. S., & Mount, D. W. (1969). Genetic analysis of recombination-deficient mutants of Escherichia coli K-12 carrying rec mutations cotransducible with thyA. *J. Bacteriol.*, *100*(2), 923–934.

319. Williams, J. G., & Radding, C. M. (1981). Partial purification and properties of an exonuclease inhibitor induced by bacteriophage Mu-1. *J. Virol.*, *39*(2), 548–558.

320. Willkomm, S., Oellig, C. A., Zander, A., Restle, T., Keegan, R., Grohmann, D., &

Schneider, S. (2017). Structural and mechanistic insights into an archaeal DNA-guided Argonaute protein. *Nat. Microbiol.*, *2*(6), 17035.

321. Wilson, G. G. (1991). Organization of restriction-modification systems. *Nucleic Acids Res.*, *19*(10), 2539–2566.

322. Wilson, J. E., & Chin, A. (1991). Chelation of divalent cations by ATP, studied by titration calorimetry. *Anal. Biochem.*, *193*(1), 16–19.

323. Wörtz, J., Smith, V., Fallmann, J., König, S., Thuraisingam, T., Walther, P., Urlaub, H., Stadler, P. F., Allers, T., Hille, F., & Marchfelder, A. (2022). Cas1 and Fen1 display equivalent functions during archaeal DNA repair. *Front. Microbiol.*, *13*.

324. Xiao, Y., Luo, M., Hayes, R. P., Kim, J., Ng, S., Ding, F., Liao, M., & Ke, A. (2017). Structure basis for directional R-loop formation and substrate handover mechanisms in Type I CRISPR-Cas system. *Cell*, *170*(1), 48-60.e11.

325. Yang, L., Haagensen, J. A. J., Jelsbak, L., Johansen, H. K., Sternberg, C., Høiby, N., & Molin, S. (2008). In situ growth rates and biofilm development of Pseudomonas aeruginosa populations in chronic lung infections. *J. Bacteriol.*, *190*, 2767–2776.

326. Yosef, I., Goren, M. G., & Qimron, U. (2012). Proteins and DNA elements essential for the CRISPR adaptation process in Escherichia coli. *Nucleic Acids Res.*, *40*(12), 5569–5576.

327. Yosef, I., Shitrit, D., Goren, M. G., Burstein, D., Pupko, T., & Qimron, U. (2013). DNA motifs determining the efficiency of adaptation into the Escherichia coli CRISPR array. *Proc. Natl. Acad. Sci. U. S. A.*, *110*(35), 14396–14401.

328. You, D., Wang, L., Yao, F., Zhou, X., & Deng, Z. (2007). A novel DNA modification by sulfur: DndA is a NifS-like cysteine desulfurase capable of assembling DndC as an iron-sulfur cluster protein in Streptomyces lividans. *Biochemistry*, *46*(20), 6126–6133.

329. You, L., Suthers, P. F., & Yin, J. (2002). Effects of Escherichia coli physiology on growth of phage T7 in vivo and in silico. *J. Bacteriol.*, *184*, 1888–1894.

330. Yuan, Y. R., Pei, Y., Ma, J. B., Kuryavyi, V., Zhadina, M., Meister, G., Chen, H. Y., Dauter, Z., Tuschl, T., & Patel, D. J. (2005). Crystal structure of A. aeolicus argonaute, a site-specific DNA-guided endoribonuclease, provides insights into RISC-mediated mRNA

cleavage. *Mol. Cell*, *19*(3), 405–419.

331. Zabeau, M., Friedman, S., Van Montagu, M., & Schell, J. (1980). The ral gene of phage lambda. I. Identification of a non-essential gene that modulates restriction and modification in E. coli. *Mol. Gen. Genet.*, *179*(1), 63–73.

332. Zander, A., Willkomm, S., Ofer, S., van Wolferen, M., Egert, L., Buchmeier, S., Stöckl, S., Tinnefeld, P., Schneider, S., Klingl, A., Albers, S.-V., Werner, F., & Grohmann, D. (2017). Guide-independent DNA cleavage by archaeal Argonaute from Methanocaldococcus jannaschii. *Nat. Microbiol.*, *2*(6), 17034.

333. Zhu, Y., Gao, A., Zhan, Q., Wang, Y., Feng, H., Liu, S., Gao, G., Serganov, A., & Gao, P. (2019). Diverse mechanisms of CRISPR-Cas9 inhibition by Type IIC anti-CRISPR proteins. *Mol. Cell*, *74*(2), 296-309.e7.