



Skolkovo Institute of Science and Technology

EFFECTS OF TARGETING BY THE *ESHERICHIA COLI* I-E CRISPR-CAS SYSTEM
ON INFECTION BY VARIOUS PHAGES

Doctoral Thesis

by

ALEKSANDRA STROTSKAYA

DOCTORAL PROGRAM IN LIFE SCIENCES

Supervisor
Professor Konstantin Severinov

Moscow - 2017

© Aleksandra Strotskaya 2017

Abstract

The enormous variety of bacteriophages and their attack strategies led their prokaryotic hosts to develop various defense systems. CRISPR-Cas is a recently discovered protective system with a unique mechanism of action that can successfully cope with viral infections when other defense systems fail. However, there has been no systematic study of how a bacterium with CRISPR-Cas system can prevent infection by phages with different attack strategies. In this work, we investigated the effects of type I-E CRISPR-Cas system of model organism *Escherichia coli* on the infection by diverse bacteriophages - M13, λ , T5, T7, T4 and R1-37. To achieve this goal, a collection of *Escherichia coli* strains was created that targeted different genomic regions of these viruses through CRISPR-Cas. The infection process at conditions of CRISPR-Cas counteraction was analyzed for each phage. Distinct consequences of infection in the presence of active CRISPR-Cas system were observed for cells infected with different phages. Based on the data, we propose that type I-E CRISPR-Cas system acts not as a true immune system but causes altruistic death of cells infected with lytic viruses.

Key words: *Escherichia coli*; CRISPR, Cas proteins, bacteriophage, spacers, strain construction.

Publications

1. Strotskaya A, Semenova E, Savitskaya E, Severinov K. Rapid Multiplex Creation of Escherichia coli Strains Capable of Interfering with Phage Infection Through CRISPR. *Methods in molecular biology*. 2015. 1311:147-59
2. Semenova E, Savitskaya E, Musharova O, Strotskaya A, Vorontsova D, Datsenko KA, Logacheva MD, Severinov K. Highly efficient primed spacer acquisition from targets destroyed by the Escherichia coli type I-E CRISPR-Cas interfering complex. *Proc. Natl. Acad. Sci. U.S.A.* 2016. 113:7626–7631
3. Strotskaya A, Savitskaya E, Metlitskaya A, Morozova N, Datsenko KA, Semenova E, Severinov K. The action of Escherichia coli CRISPR-Cas system on lytic bacteriophages with different lifestyles and development strategies. *Nucleic Acids Res.* 2017. 45(4):1946-1957

Acknowledgements

I would like to express my gratitude to my supervisor Konstantin Severinov, to Skoltech and to our collaborators: Ekaterina Semenova (Rutgers University), Ekaterina Savitskaya, Anastasia Metlitskaya (Institute of Molecular Genetics), Natalia Morozova (St. Petersburg Polytechnical University), Maria Logacheva (Institute of Physico-Chemical Biology), Kirill A. Datsenko (Purdue University).

Table of Contents

Abstract.....	2
Publications.....	3
Acknowledgements.....	4
Table of Contents.....	5
List of Symbols, Abbreviations	9
List of Figures.....	11
List of Tables	12
Chapter 1. Introduction	13
1.1 Significance of the work	13
1.2 Project Objectives	15
1.3 Novelty and Practical Use	15
1.4 Personal Contribution	16
Chapter 2. Review of the Literature	17
2.1 Bacteriophages, Their Diversity, and Life Cycles.....	17
2.1.1 Adsorption	19
2.1.2 Penetration of Viral Genetic Material	21
2.1.3 Transition from Normal Cellular Metabolism to a Metabolism Controlled by the Virus	23
2.1.4 Assembly of Phage Particles	25
2.1.5 Lysis of the Cell	27
2.2 Defense Strategies of Bacteria.....	30

2.2.1 Blocking or Modification of the Receptor	30
2.2.2 Superinfection Exclusion	33
2.2.3 Restriction-Modification Systems	34
2.2.4 Abortive Infection Systems.....	36
2.2.5 BREX Systems.....	38
2.3 Mode of Action of CRISPR-Cas Systems	39
2.3.1 CRISPR Adaptation.....	42
2.3.2 Transcription and Maturation of crRNA.....	43
2.3.3 CRISPR Interference	44
2.4 Interaction of Bacteriophages with Bacteria Containing the CRISPR-Cas System .	45
Chapter 3. Materials and Methods	52
3.1 Nutrient Media	52
3.2 Bacterial Strains	52
3.3 Bacteriophages	55
3.4 Plasmids	55
3.5 Strain Construction	55
3.6 Isolation of Bacteriophage DNA	58
3.7 PCR.....	58
3.8 Restriction	61
3.9 Extraction of DNA Fragments from Agarose.....	61
3.10 DNA Ligation	61
3.11 Preparation of Competent <i>E.coli</i> Cells and Chemical Transformation.....	62

3.12 Isolation of Plasmids.....	63
3.13 Sequencing of Recombinant Plasmids and PCR Fragments.....	63
3.14 Preparation of Electrocompetent <i>E.coli</i> Cells and Electroporation	63
3.15 Induction of Components of The CRISPR-Cas System and Selection of Clones that Incorporate a new Spacer.....	64
3.16 CRISPR Interference Assays	66
3.17 Monitoring of Phage Infection.....	67
3.18 Total DNA Extraction.....	67
3.19 CRISPR Adaptation Assay	68
3.20 High Throughput Sequencing and Data Analysis.....	68
3.21 Fluorescence Microscopy	71
Chapter 4. Results and Discussion	73
4.1 Using primed adaptation to generate <i>E. coli</i> cells carrying CRISPR spacers targeting different phages	73
4.2 Bacteriophage M13	77
4.3 Bacteriophage λ	80
4.4 Bacteriophage T5	84
4.5 Bacteriophage T7	92
4.6 Phages with modified genomes: T4 and R1-37	97
Chapter 5. Conclusions.....	100
Bibliography	105

List of Symbols, Abbreviations

Abi: Abortive infection

ATP: Adenosine TriPhosphate

bp: Base Pairs

BREX: Bacteriophage exclusion

cAMP: Cyclic Adenosine MonoPhosphate

Cascade: CRISPR-Associated Complex for Antiviral Defence

cfu: Colony Forming Units

CRISPR-Cas: CRISPR Associated

CRISPR: Clustered Regular Interspaced Short Palindromic Repeats

crRNA: CRISPR RNA

CTAB: CetylTrimethylAmmonium Bromide

Da: Dalton

DMSO: DiMethylSulfOxide

DNA: DeoxyriboNucleic Acid

dNTPs: mixture of deoxyribonucleotides (dGTP, dATP, dTTP and dCTP)

dsDNA: Double-Stranded DNA

DTT: DiThioTreitol

EDTA: EthyleneDiamineAetraacetic Acid

EOP: Efficiency Of Plaquing

EOT: Efficiency Of Transformation

h: Hour

hmdC: HydroxyMethylDeoxyCytidine

HTS: High Throughput Sequencing

ICTV: International Committee on Taxonomy of Viruses

kbp: KiloBase Pairs

LPS: LipoPolySaccharide

MOI: Multiplicity Of Infection

mRNA: matrix RNA

NA: Not Available

ND: No Data

NI: No Interference

OD: Optical Density units

PAM: Protospacer Adjacent Motif

PCR: Polymerase Chain Reaction

pfu: Plaque Forming Units

Pgl: Phage Growth Limiting

pre-crRNA: pre-CRISPR RNA

Protospacer: target sequence complementary to a spacer

RNA: RiboNucleic Acid

SAR: Signal Anchor-Release

SDS: Sodium Dodecyl Sulphate

sgRNA: single guide RNA

Spoligotyping: Spacer Oligonucleotide typing technique

tracRNA: *trans*-acting CRISPR RNA

UV: UltraViolet

List of Figures

Figure 1. Stages of bacteriophage life cycle and defense strategies of bacteria.	29
Figure 2. Creation of <i>E.coli</i> strains with CRISPR spacers targeting various phages using primed adaptation.	57
Figure 3. Screening for CRISPR array expansion.	66
Figure 4. CRISPR interference and primed CRISPR adaptation against matching and partially matching protospacer targets located on plasmids.	76
Figure 5. Long-term cultivation experiment.	78
Figure 6. Gradients of spacers acquired during primed adaptation by M13 infected cells .	79
Figure 7. The effect of CRISPR-Cas targeting on λ_{vir} infection.	81
Figure 8. The effect of CRISPR-Cas targeting on T5 infection.	86
Figure 9. The effect of CRISPR-Cas targeting on T7 infection.	93
Figure 10. The effect of CRISPR-Cas targeting on T4 and R1-37 infections.	99
Figure 11. The overall scheme of the effects of targeting by the <i>E. coli</i> type I-E CRISPR-Cas system on infection by various phages.	104

List of Tables

Table 1. Details of strain and phage collection.....	53
Table 2. Primers for T4 and R1-37 cloning.....	60
Table 3. HTS statistics.....	69

Chapter 1. Introduction

1.1 Significance of the work

Bacteriophages are viruses that infect prokaryotic organisms. In order to survive, phages must constantly improve their mechanisms of infection and cells must develop more sophisticated defense systems to protect themselves from infection. Bacteriophage-bacterial interactions are an invaluable source of interesting and unexpected discoveries due to their great abundance and diversity.

Since the virus cannot reproduce itself outside the bacterial cell, the viral attack most commonly begins by attaching of the bacteriophage to the surface of the cell through a specific receptor, after which a viral genetic material (DNA or RNA) is injected into the cytoplasm. Active transcription and translation of viral proteins then begin, followed by replication of viral genomes and assembly of viral particles. Dozens to hundreds of new phage particles are formed and accumulate inside the cell during the infection cycle, which culminates in lysis of the cell, releasing new viral particles into the environment to start a new cycle of infection.

However, the bacterial cell can protect itself at each stage of the infection and, in turn, the virus develops its own mechanisms to evade the host defense systems. Study of the molecular mechanisms at work in the course of this evolutionary "arms race" is of interest not only to fundamental science, but also for the practical purposes of biotechnology, genetic engineering, microbiology and the food industry. The CRISPR-Cas system is the most promising and practically significant bacterial defense strategy to have been discovered to date. A prokaryotic cell carrying the CRISPR-Cas system is

capable of “remembering” information about previously infecting phages and destroys their genetic material upon reinfection. Information of previous infections can be transmitted to daughter cells, which ensures long-term immunity in new generations of cells. However, efficiency of the protective action of CRISPR-Cas systems against different viruses is bound to vary in view of the enormous variety of strategies used by phages to overcome their hosts. Thus, studies of model systems where the host is a bacterium and bacteriophage is a parasite with modern interdisciplinary methods of analysis, such as high throughput sequencing, bioinformatics, mathematical modeling, allows studying the molecular basis for CRISPR-interference and CRISPR-adaptation processes during phage infection and tracing the results of this interaction at the level of individual viruses and cells, and at the population level under conditions of long-term cultivation. However, no such studies have been conducted so far. Systematic assay of the infection by bacteriophages which developed different strategies for efficient host takeover under conditions of CRISPR-Cas targeting developed in this work, will lead to a more complete understanding of microevolutionary processes during bacteria-bacteriophage interactions and macroevolutionary processes of horizontal gene transfer. We therefore created a collection of *Escherichia coli* strains with the type I-E CRISPR-Cas system targeting various positions in the genomes of bacteriophages M13, λ , T5, T7, T4 and R1-37 and investigated the ability of these strains to resist infection and acquire additional CRISPR spacers from the infecting phage. We find that the efficiency of CRISPR-Cas targeting by the host is determined by the phage life style, the positions of the targeted protospacer within the genome, and the state of phage DNA. The results also

suggest that during infection by lytic phages that are susceptible to CRISPR interference, CRISPR-Cas does not act as a true immunity system that saves the infected cell but rather enforces an abortive infection pathway leading to infected cell death with no release of phage progeny.

Our studies have thus enabled a deeper understanding of the co-evolution of microbiological systems.

1.2 Project Objectives

The goal of this work is to assess how different dsDNA phages are affected by the action of the type I-E CRISPR-Cas system of *E. coli*.

Several specific aims were addressed for this purpose:

- performing a systematic study of CRISPR-Cas response to phage infection at either the adaptation or interference stages;
- investigating the infection of *E. coli* cells containing CRISPR arrays with spacers targeting different locations in the genomes of classical bacteriophages M13, λ , T5, T7 and T4 and a giant phage R1–37;
- identifying how phages can limit the efficiency of CRISPR-Cas immunity by host take-over strategies.

1.3 Novelty and Practical Use

In the course of this work, the method of creation of *E. coli* strains targeting various phages was first developed and experimentally tested (Strotksaya et al., 2015). The availability of such strains opens up new interesting areas of research, including determination of requirements of efficient CRISPR-Cas interference other than spacer-

protospacer match and the presence of a functional PAM sequence. The procedure can be made even more powerful by using libraries of spacer capture plasmids containing the entire genome of a phage or an episome.

A comprehensive study of CRISPR response to infection by various well-studied bacteriophages was conducted and it was shown, for the first time, that CRISPR-Cas system could function as an abortive infection mechanism (Strotskaya et al., 2017).

Practical application of the results of this work lies in the use of technique for the rapid construction of strains with spacers against various, even hard-to-reach targets and also in the case of further discovery of anti-CRISPR proteins in bacteriophages T5 or T7 it is possible to use these proteins to control or regulate CRISPR interference and adaptation.

1.4 Personal Contribution

Most of the research presented here was carried out by the author. The rapid multiplex method to obtain *E. coli* strains capable of targeting specific DNA sequences was developed and tested with single plasmids containing cloned fragments of different phages after PCR amplification or with plasmid libraries containing various fragments of phage DNA. Experiments monitoring CRISPR interference and adaptation with different bacteriophages were all carried out by the author. High Throughput Sequencing (HTS) data of samples prepared by the author were obtained by Dr. Maria Logacheva and analyzed by the author under the guidance of Dr. Ekaterina Savitskaya. Live fluorescent microscopy were conducted under direction of Ms. Natalia Morozova. The author analyzed and processed the results and prepared material for publication.

Chapter 2. Review of Literature

2.1 Bacteriophages, their diversity, and life cycles

The bacteriophage, a virus that destroys bacteria ("phagos" in Ancient Greek means "eater"), is an extracellular life form defined as an obligate parasite of prokaryotes. Bacteriophages were first discovered by the scientist Hankin in 1896 when studying antibacterial activity in the rivers of India (Hankin, 1896), and his discovery was subsequently confirmed by Gamaleya in 1897 in the study of lysis of a *Bacillus anthracis* due to its exposure to a transplantable agent. But the term "bacteriophages" was only coined to define these organisms as bacterial viruses 20 years later by d'Herelle (d'Herelle, 1917).

Information on the biology of bacteriophages has expanded enormously over the past 100 years. The taxonomy of the viruses of bacteria and archaea has been substantially enlarged and reissued by the International Committee on Taxonomy of Viruses based on new data, most recently in 2015. To date, 29 families of viruses have been divided into 7 orders, although there are 82 families that have not yet been assigned to any of the orders (Krupovic *et al.*, 2016). The most extensive, comprising about 95% of all known bacteriophages, is the *Caudovirales* order, which includes the 3 families of *Myoviridae*, *Podoviridae*, *Siphoviridae*, 6 subfamilies, 80 genera and 441 species (Krupovic *et al.*, 2016). This classification is being constantly updated and by 2017 the *Myoviridae* family had been expanded to 5 subfamilies, so that the *Caudovirales* order now has 9 subfamilies (Adriaenssens *et al.*, 2017, <http://www.ictvonline.org/virustaxonomy.asp>).

Extreme abundance of viral particles on Earth, which number about 10^{31} (Whitman *et al.*, 1998) has led to rapid growth of data on phages in last two centuries. It is assumed that bacteriophages can exist in almost all ecosystems (Maranger *et al.*, 1995; Hendrix, 2002), including ecosystems with extremely high temperatures (hot springs (Breitbart *et al.*, 2004), the Sahara desert (Prigent *et al.*, 2005)), and ecosystems with extremely low temperatures (polar ice (Sawstrom *et al.*, 2008)).

Study of the morphological and physiological parameters of a large collection of bacteriophages from urban wastewater revealed a high degree of biodiversity among viruses. The phages tested in the laboratory were able to survive in the most extreme conditions: they can form plaques at low (4 °C) and high (95 °C) temperatures, and they can retain infectious activity even in the presence of organic solvents (ethanol, acetone, DMSO, chloroform) or detergents (SDS, CTAB) (Jurczak-Kurek *et al.*, 2016).

The number of phages in seawater is as high as 10^7 viral particles per mL (Breitbart, 2012), and viral particles in the soil number 10^8 - 10^9 per cm^3 (Williamson *et al.*, 2005). Bacteriophages play a key role in maintaining the balance of all microbial ecosystems that have been studied.

The structure of bacteriophages is also very diverse. Most of them have a head or capsid, which consists of a protein or lipoprotein with different organization and structure, and genetic information in the form of single- or double-stranded RNA or DNA (Brussow *et al.*, 2002). The sizes of phage genomes also vary considerably and can range from 4,000 to 500,000 base pairs (Brussow *et al.*, 2002).

Bacteriophages of the *Caudovirales* order combine the presence of an icosahedral protein capsid, a protein tail with tail fibers, by which they attach to the receptors of the bacterial cell, and double-stranded DNA as genetic information. Classical and well-studied viruses from this order are the phages of the *Siphoviridae* family (T1, λ , T5, which have a long, mobile but non-contractile tail), of the *Myoviridae* family (P1, P2, T4, AR1 with a long, contractile tail), and of the *Podoviridae* family (N4, Sp6, T7 with a short, non-contractile tail). However, despite the fact that these viruses belong to the same order or family and may have a similar structure of their virions, the individual phages may differ significantly in their interaction with the bacterial cell.

For each phage, only a specific range of bacteria can be its host, and these are most often closely related strains of bacteria within one species (Kutter, 2009). The infection cycle, which can occur when a phage particle and a sensitive bacterium are present in the environment, consists of several main stages: adsorption, penetration of the genetic material, the transition from normal cellular metabolism to a metabolism controlled by the virus, the assembly of viral particles and cell lysis, which is the destruction of the bacterial cell wall. At the end of the infection cycle, new phage particles are released and can re-infect other sensitive cells. Other forms of infection, namely lysogeny and variants of chronic infection (Kutter, Sulakvelidze, 2004), can also occur.

2.1.1 Adsorption

Adsorption is the attachment of a bacteriophage through a series of interactions with the surface receptors on a bacterial cell (Figure 1). The virus recognizes susceptible

cells by their receptors. The following stages of adsorption are distinguished: initial contact, reversible (when the phage particle can be de-adsorbed) and irreversible binding (Duckworth, 1987).

Phages are able to bind to the surface of either Gram-negative cells, on which any protein, oligosaccharide or lipopolysaccharide can serve as a receptor, or Gram-positive cells, the murine layer of which is also rich in receptors. Most bacteriophages require a set of specific molecules (Kutter, Sulakvelidze, 2004). For example, in the case of phage T5, which is a virus of *Escherichia coli*, reversible adsorption occurs as a result of binding of L-like phage fibers to the oligosaccharide fragment of mannose in the lipopolysaccharide on the host cell wall, while irreversible adsorption can be achieved only by attaching the phage tail protein Pb5 to the membrane receptor FhuA (Heller K.J. *et al.*, 1982; Heller K.J., 1984; Heller K.J., 1992).

The phage T4 has 2 types of fibers: long and short. Reversible adsorption occurs as a result of binding of at least three of the six long fibers with LPS on the *E. coli* membrane, after which the components of the basal plate are reorganized and six short fibers interact with the LPS heptose fragment, leading to irreversible adsorption (Riede, 1987).

The number of receptors available on the cell surface should be as high as possible for irreversible adsorption. However, for some phages, such as phage N4, the receptor is formed from NfrA, NfrB and NfrC proteins, which are represented on the cell membrane in only a few copies. This complicates the search for an intact receptor by a viral particle (Kiino *et.al.*, 1993).

Information on the receptors studied to date is presented in the collection of phage receptor data (Phage Receptor Database - <https://phred.herokuapp.com>).

2.1.2 Penetration of viral genetic material

Once successful adsorption has occurred and the phage is firmly attached to the cell surface, the virus must enter its genetic material into the cytoplasm of the cell (Figure 1). During infection by tailed phages of the *Caudovirales* order, the capsid and tail remain outside the cell after DNA injection into the cell. Inside the cytoplasm the DNA can remain either as a linear or a circular form if terminal redundant ends are present. In the case of some filamentous phages, whose genomes are single-stranded circular RNA, structural proteins can remain on the surface of the cell after injection of the genetic material and can be re-used to form new phage particles (Molineux *et.al.*, 2013).

Usually, the DNA inside the capsid is very compactly packaged, under pressure of about 6 MPa (~ 60 atm) (Smith *et al.*, 2001). Such pressure should ensure rapid penetration of DNA into the cell but should not lead to deformation of the DNA during packaging inside the capsid when a new phage particle is formed. Studies of the injection of DNA at single particle level showed that the phage λ injects its DNA continuously with speeds of up to 60 kbp per second, but the rate of injection is disproportionate to the amount of the genome remaining in the capsid, so that, after the introduction of about half of the genome, the rate increases dramatically and then decreases almost to zero to the end of the genome (Grayson *et al.*, 2007).

The injection of genetic material by some viruses can occur by few steps; for example, bacteriophage T5 injects its DNA in two stages. The reason for two-step

transfer of T5 DNA observed in bulk experiments is that the phage needs to start transcription of its pre-early genes at the initial stage of infection. However, experiments, carried out at the single viral particle level, to assess the rate of DNA penetration, showed results that were significantly different from the results of bulk experiments in cultures. In single-molecule measurements (the same were conducted for the phage λ) it was observed that about 25% of viruses inject their DNA from the capsid completely within the first minute in the presence of a high concentration of FhuA receptor, with which phage fibers bind irreversibly, but most phage particles introduce their DNA with several brief pauses, coinciding with areas of nicks on the T5 genome. The first nicks in the T5 genome are detected at a distance of about 10,000 bp, at the border of the pre-early and early regions of the genome (Lanni, 1968; Davison, 2015).

One can conclude that T5 DNA injection is regulated not only by pauses in the region of nicks but also by the transcription of pre-early genes before the rest of the genome enters the cell. *In vivo* experiments show, that at the first stage of injection, at which the pre-early region is introduced, the rate is about 100 bp per second, and at the second stage, during the introduction of the remaining part of the genome, the rate can reach about 500 bp per second (Filali Maltouf *et al.*, 1983). The pause after DNA insertion of the first 8% of the genome can last for up to 5 minutes (Lanni, 1968). To inject the remaining 92% of T5 phage DNA into the cell, the virus needs DNA-binding proteins A1, A2-3 which are synthesized from the pre-early region of the T5 phage genome (Lanni, 1968; McCorquodale, Warner, 1988; Davison, 2015).

There are other viruses, whose DNA injection depends on their own or host proteins. Phage T7 injects its DNA independently of transcription at the first stages of infection, and when first promoters of the virus penetrate into the cell, they attract host RNA polymerase, which, initiates transcription and thus helps the phage to introduce more of its DNA into the cell. Injection of the remaining part of the genome (middle and late genes) occurs with the accumulation of T7 RNA polymerase expressed by the phage. The whole process of entering viral DNA takes from 9 to 12 minutes at 30 °C (Garcia *et al.*, 1995, Kemp *et al.*, 2004).

In the case of phage N4, the entire DNA is introduced with only the help of the N4 RNA polymerase II (Choi *et al.*, 2008).

2.1.3 Transition from normal cellular metabolism to a metabolism controlled by the virus

In most phages, it is possible to identify certain groups of genes transcribed at different stages of infection. Quite often they are grouped and arranged sequentially within the genome. For example, in bacteriophage T7 early, middle and late genes are isolated, and in the phage T5 genome is separated into pre-early, early and late regions, which partly correspond by their functions to the early, middle, and late regions of the T7 virus. However, there are also viruses in which genes are arranged in mosaic fashion and their genomes cannot be divided into regions grouped by function. For example, phage T4 has early, middle and late genes that are randomly located within the genome (Calendar, Abedon, 2006).

Different groups of genes are responsible for distinct events during infection. The products of early genes are synthesized at the first stages of infection and inhibit

intracellular restriction enzymes, protect the genome of the virus and inhibit host molecular processes (Figure 1). The products of middle genes are mainly responsible for the replication of phage DNA, and the products of late genes are responsible for assembly of new virions (Figure 1) (Kutter, Sulakvelidze, 2004).

Some genes or groups of genes are responsible for the start of the so-called lysogenic state of the virus, in which the bacteriophage penetrates the cell, but does not lyse it. As a result of this alternative pathway, the bacteriophage DNA is inserted into the genome of the host and exists in the form of a prophage, in which state it replicates together with the host DNA and is transmitted to the daughter cells. Insertion into the genome can occur at any random place in the genome or at a specific site. For example, phage λ chooses *att λ* as the insertion point in the *E. coli* genome. The incorporation of a virus into the genome of a bacterium can sometimes provide the resulting bacterial lysogen with certain advantages, for example, it may become resistant to superinfection by the phage, acquire resistance to antibiotics (Haaber *et al.*, 2016) or develop certain pathogenic properties (Davies *et al.*, 2016). The decision to switch from a virulent to a lysogenic pathway depends on the metabolic state of the cell (Lieb, 1953). Very often the phage decides to integrate into the genome if the cell is in a state of starvation, since, if the virus continued to develop along the lytic path, this would lead to a lack of resources for the successful development of new phage particles (Stewart *et al.*, 1984). In order to maintain its prophage status, the virus has to repress lytic pathway proteins; for example, the λ prophage expresses protein CI, which is simultaneously an inducer for its own

promoter and suppresses the phage promoters responsible for the development of new viral particles and bacterial cell lysis (Ptashne, 2004).

Activation of the prophage can occur in an individual cell in the population as a result of exposure to SOS-response proteins upon damage to the host DNA. Exposure of lysogens to certain antibiotics or critical temperatures can also induce prophages (Refardt, Rainey, 2010).

2.1.4 Assembly of phage particles

When there is induction of a prophage or common infection with a lytic virus, the assembly of new viral particles begins. Each viral particle must be formed in a certain way and consist of a capsid, tail, and genetic material packaged into the capsid (Figure 1). Accordingly, when the molecular apparatus of the infected cell is redirected to the needs of the virus, the replication of the new viral DNA begins. The replicative apparatus can either be "borrowed" from the host or synthesized independently during the infection process. When the cell is infected with phage λ , all of the host proteins that take part in replication (helicase, DNA polymerase, Ssb-proteins and others) are mobilized by the virus. The replication mechanism of this virus in the initial stages of infection occurs as a "theta"-replication, where the replication originates from two replicative forks, but on the later stages of infection, replication passes over to a "rolling circle" mechanism that generates concatemeric DNA (Baranska *et al.*, 2001).

T4 virus needs both its own proteins (helicase Gp41, primase Gp61, polymerase Gp43 and others (Nossal, 1992; Noble *et al.*, 2015)) as well as some host proteins. In particular, the host RNA polymerase synthesizes primers for replication of the leading

strand, and the virus's own primase Gp61 synthesizes primers for the lagging strand. As a result, many recombination intermediates of T4 DNA are formed. Their number equals the number of forming capsids in the late stages of infection, and the replication of the intermediates can be completed (Mosig *et al.*, 1995).

During the replication and recombination of the genome of another lytic bacteriophage, T7, linear multicopy DNA concatemers are formed, which are subsequently cut with special enzymes for packing into the capsid (Kulczyk *et al.*, 2016; Agirrezabala *et al.*, 2007).

The formation of virion structures is also a very complex and multi-stage process. Viral capsids can consist of simply-organized proteins and of proteins that require helper proteins for proper folding, which are synthesized and only participate in assembly of the virus particle but do not enter in the mature virus (Liljas, 1999).

In the case of bacteriophage T4, a big particle with a long, contractile tail, the virion consists of more than 2,000 protein subunits, which are the products of about 50 genes. About 20 genes participate in the capsid assembly, more than 20 - in the tail and tail fibers assembly, and 10 more genes encode catalytic, chaperonin-like proteins that help to assemble the basic protein structures that form the virion (Mesyanzhinov *et al.*, 2004). The production of phage proteins necessary for the assembly of capsids occurs in the second half of the infection cycle and begins with the production of major capsid protein Gp23, which, matures with the help of the viral co-chaperonin Gp31, and the bacterial chaperonins GroEL and GroES. The capsid matures on the inner surface of the cell membrane (Hunt *et al.*, 1997). When the capsid structure forms, DNA begins to be

packed into it, and only then the capsid is completed (Ishii et al., 1978; Keller et al., 1986; Richardson et al., 1998).

Separately, the tail of the phage is formed by the products of 22 viral genes and the basal plate is assembled as a multi-protein complex structure consisting of 16 different proteins, to which a few proteins are bound to form long and short tail fibers (Makhov *et al.*, 1993; Leiman *et al.*, 2004; Kostyuchenko *et al.*, 2005).

When all the structural units of the phage are organized and the recombinant DNA intermediates have completed the replication and are packed into the capsid, virions remain attached to the cell's internal membrane for some time. When most of the phage particles are successfully formed, the cell wall is destroyed and the new phages are released.

2.1.5 Lysis of the cell

Previously, it was assumed that cell lysis occurs as a result of the accumulation of phage particles inside the cell and the mechanical destruction of the cell wall, but it later became clear that, in most cases, the cell lysis is strictly programmed and occurs due to the action of specific signals. Several lysis systems have now been identified: a choline-endolysin system, a signal anchor-release system (SAR) and a system of spanins.

The choline-endolysin system has been well described for phage λ , in which the expression of late genes begins about 10 minutes after the start of infection. At this stage the choline S105 is expressed and is distributed on the inner surface of the cell membrane (Altman *et al.*, 1983). The number of S105 protein molecules approaches 1000-3000 per cell (Zagotta *et al.*, 1990). During the formation of phage particles, endolysin R is

produced, which has transglycosylase activity. It accumulates in the cell in an active form and, in the late stages of infection, causes cell lysis process after being released outside the cell through holes made by choline (Adhya *et al.*, 1971; Garret *et al.*, 1982). The process of activation of choline is not completely clear, but studies using cryoelectron microscopy showed that extensive holes appear after the choline action on the cytoplasmic membrane, reaching $>1\mu\text{m}$ in size (Dewey *et al.*, 2010). T protein of phage T4 has a function similar to choline of phage λ (Savva *et al.*, 2014), and E protein is the analogue of endolysin, otherwise called lysozyme, which is capable of hydrolyzing the MurNac-GlcNac glycosidic bond of the cell wall (Matthews *et al.*, 1981).

The signal anchor-release system has been described for the phage P1, in which the endolysin has a SAR (signal-anchor-release) domain on the N-terminal endolysin part, exported to the periplasm of the cell, while choline remains bound to the membrane in an inactive form. Lysis occurs only when the SAR-endolysin protein is detached from bilayer of the membrane and spontaneously transformed into an active form, interacting with choline through the SAR domain. So the mureynolytic effect is impossible until the interaction between choline and SAR-endolysin (Xu *et al.*, 2005).

Other proteins were discovered in the phage λ , destroying both the peptidoglycan and the outer membrane. These proteins were called spanins, and their genes are often referred to as Rz/Rz1. Bioinformatically, spanins are found in most phages affecting Gram-negative bacteria (T7, T4, P1, T1, Mu) (Summer *et al.*, 2007). These proteins generate transmembrane complexes and through the fusion between the outer and inner membranes form vesicles, which probably destabilize the outer membrane and serve to

rapidly release the phage progeny where Ca^{2+} ions are present in the environment (Berry *et al.*, 2012).

Successful infection and favorable conditions for the lytic pathway enable bacteriophages to inhibit the life processes of the bacteria and reproduce viral progeny, after which lysis of the bacterial cell occurs and new viral particles can infect new sensitive cells, repeating the cycle (Figure 1). At each of the listed stages of infection, the virus may encounter resistance to infection on the part of the host. To survive, the phage develops new and improved attack mechanisms, in turn, the bacteria must counter these mechanisms. Such is the logic of the “arms race” between viruses and bacteria.

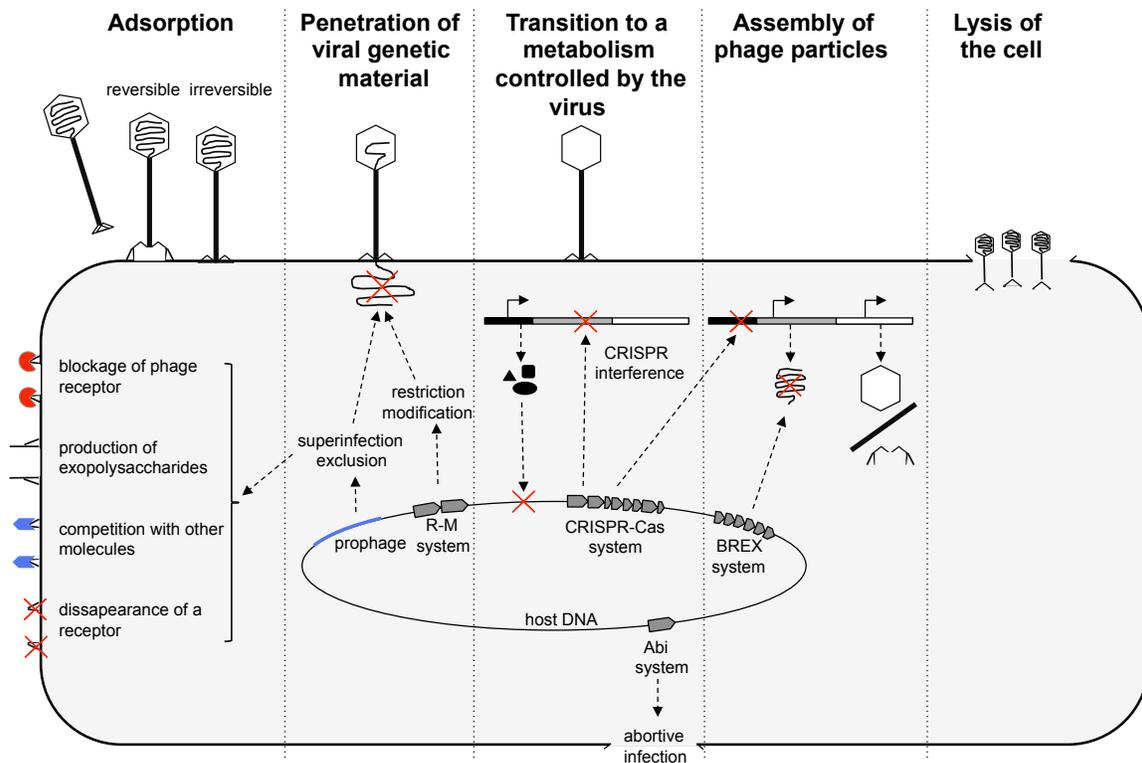


Figure 1. Stages of bacteriophage life cycle and defense strategies of bacteria. Phage genomes are represented as linear DNA with different genome regions colored in black, gray, and white. Bacterium genome is represented as circular DNA with different defense systems that act during various stages of phage life cycle. Phage enzymes and

molecules that block or compete with phage receptors are shown as red, blue and black figures.

2.2 Defense strategies of bacteria

To date, there are 6 major bacterial defense strategies that are effective at specific stages of infection. Not all of them have been fully studied, some were discovered relatively recently and their mechanisms remain a mystery.

2.2.1 Blocking or modification of the receptor

Receptors play a crucial role in the emergence of bacterial resistance to phage attack (Labrie *et al.*, 2010). There are several mechanisms that can prevent the stage of irreversible adsorption during infection (Figure 1):

- a. blockage of phage receptors (formation of a capsule shell, production of competitive inhibitors), by which the bacterium hides its receptors and makes them inaccessible to the phage;
- b. production of exopolysaccharides that impede phage access to the cell surface (this strategy does not protect the cell from viruses, which use LPS for binding (Lindberg, 1973));
- c. competition between phage fibers and molecules that are also able to bind to receptors of the cell (nutrients can bind to receptors and block access to them of the virus) (Wayne *et al.*, 1975);
- d. disappearance of a specific receptor necessary for phage adsorption from the surface of the cell (Labrie *et al.*, 2010).

Actual presence or absence of a receptor on the surface of the bacterium may be associated with factors such as the availability of certain nutrients, density of the bacterial

population, and species diversity in the population. Some bacteria, for example, *Vibrio anguillarum*, which are able to attach to substrate and form a biofilm, rely on quorum sensing mechanisms to decrease receptor production as population density increases, protecting the bacteria from phages (Tan *et al.*, 2015). However, loss of the receptor is not always beneficial to the bacterium, since it can lead to the loss of certain vital functions, hindering not only the intake of nutrients, but also the formation of intercellular contacts and transfer of the genetic material. *Escherichia coli* can carry F pili on their surface, through which horizontal transfer of genetic material, for example, plasmids, can be carried out. However, an F pilus can also serve as a receptor for bacteriophage M13 binding. This phage causes a chronic infection of the cell, where the bacterium does not lyse, but constantly produces new M13 phage particles, expending its own resources in production of the virus (Smeal *et al.*, 2017; Kick *et al.*, 2017). With increasing population density, bacteria can lose F pili which saves them from infection, but also prevents the acquisition of plasmids that can carry important genes necessary for survival in an unfavorable environment, including antibiotic resistance genes.

Some phages, for example, prophage D3 of *Pseudomonas aeruginosa*, are able to prevent the injection of the genetic material of other viruses into the infected cell by affecting the cell receptor and changing its structure. Prophage D3 changes the structure of the O-antigen of the surface LPS, which is required as a receptor by a wide range of viruses (Newton *et al.*, 2001). However, even if irreversible adsorption occurs, in which the virus attaches securely to the cell surface and injection of the viral genetic material begins, the bacterium can still mount resistance to the virus.

Two models have been proposed to describe the process of injection: mechanical and hydrodynamic. They differ fundamentally in respect to the DNA injection mechanism, and depend on the diameter of the channel in the cell membrane (Molineux *et al.*, 2013). The mechanical model applies when the diameter of the DNA injection channel is only sufficient for injection of the dsDNA, in which case the entry of DNA into the cell is preferably carried out using the energy of certain enzymes, for example, host RNA polymerase during T7 phage infection (Molineux, 2006). The hydrodynamic model is applicable if the channel for DNA injection is wider than the diameter of the dsDNA helix, in which case the space between the channel and the DNA can be used for transfer of K^+ and some other ions, that are present in the cytoplasm in a higher concentration and will flow into the medium. However, the osmotic imbalance will result in water entering the cytoplasm along with the viral DNA (Panja *et al.*, 2010). The consequence of the hydrodynamic model is that the cell, which needs to preserve a high concentration of salts to maintain a certain turgor, may be defenseless against the introduction of the phage genome. But, if resources of the cell decrease, for example when the culture reaches a stationary growth phase, the difference in salt concentration between the bacterium and the environment will also decrease. Accordingly, the cell will become less susceptible to infection, because the viral DNA will penetrate into the cytoplasm more slowly (Schrader *et al.*, 1997).

Another obstacle to infection at the stage of viral DNA entry into the cell can be studied on the example of the *Clostridium difficile*, which has the CwpV protein in its cell wall, the C-terminal domain of which has anti-phage activity. The phage is adsorbed on

the surface of the cell, but there is no further development of infection, as shown by the lack of new replicative forms of the virus, presumably caused by the inability of the phage to inject its DNA (Sekulovic *et al.*, 2015).

2.2.2 Superinfection exclusion

The phenomenon where infection by two phages occurs with some delay in time and the secondary infection does not lead to the production of phage progeny has been called “superinfection exclusion” (Dulbecco, 1952) (Figure 1). Some of the receptors may disappear from the cell surface as a stress response to penetration by the phage (Karlsson *et al.*, 2005). If the phage λ is inserted into the genome of the bacterium as a prophage, then it actively expresses the CI protein which inhibits the lytic cycle and also inhibits infection with a new phage, i.e., provides immunity to superinfection. In the case of attempted superinfection, the CI protein acts on the alien DNA as well as on prophage and prevents the development of the lytic pathway and destruction of the bacterial cell (Labrie *et al.*, 2010).

Some prophages mask or alter the receptor on the cell surface in a way that prevents binding to new phage particles. Phages with mechanisms for superinfection exclusion include the T4 phage, which encodes an Imm protein that provides about 80% immunity to re-infection by a new phage particle and is expressed only after 1-2 minutes of the infection. This protein binds to receptors of T4 phage adsorption on the inner cell membrane and blocks the entry of new phage DNA (Anderson *et al.*, 1971; Lu *et al.*, 1989). T4 also has the Sp protein that provides the remaining 20% of immunity. The Sp protein is also attached to the inner side of the membrane at the point where the assembly

of the phage procapsid begins but, instead of receptor blocking, it inhibits the lysozyme of the phage, which is a part of the basal plate and is needed to create a channel in the membrane for the introduction of viral DNA (Emrich, 1968; Obringer *et al.*, 1988; Lu *et al.*, 1994).

Investigation of the mechanism of superinfection exclusion in *P.aureginosa* lysogens has shown that it is based on inhibition by prophages of formation of pili necessary for binding of a new phage particle, for example, due to the production of the Tip protein by the D3112 prophage. However, other phages, which do not require binding with this particular receptor, are still capable of infecting the bacterium. The dependence of the level of resistance to reinfection on production of the Tip protein has been shown. Secondary infection was significantly less prevalent among lysogens, in which expression of the Tip protein was higher (Bondy-Denomy *et al.*, 2016).

Another form of receptor blocking is exemplified by the Ltp protein of the temperate phage TP-J34 of *Streptococcus thermophilus*, which is expressed when the phage DNA is inserted into the host genome. Ltp is a lipoprotein, released outside the cytoplasmic membrane, and it blocks the entry of a new streptococcal phage or even of the lactococcal phage P008 (Sun *et al.*, 2006). Similar proteins are expressed by prophages of *Lactococcus lactis* strains which inhibit the entry of DNA into the bacterial cell. However, they are unable to provide the cell with full immunity at high multiplicity of infection, apparently due to limited production of these proteins (Mahony *et al.*, 2008).

2.2.3 Restriction-modification systems

Restriction-modification systems protect the bacterium if the injection of the phage DNA has already occurred or is occurring and none of protection mechanisms described above has stopped the introduction of foreign genetic information into the cell (Figure 1) . Restriction-modification systems can destroy DNA that has penetrated into cells and they are based on two enzymes: restriction endonuclease and methyltransferase (Tock *et al.*, 2005). These enzymes are found in 75% of bacterial genomes that have been sequenced to date (Rocha *et al.*, 2001). They protect host DNA by adding methyl groups to adenine or cytosine at specific positions. If foreign DNA, which does not have methyl groups in the genome, enters the cytoplasm of the cell, the nucleotide sequences are recognized by the restriction endonuclease and the foreign DNA is cut in these regions. There are four types of restriction-modification systems (I, II, III, IV) based on the composition of their subunits, the recognition of certain sites in the genome and their mode of action (Roberts *et al.*, 2003).

While the restriction enzymes of the first three types recognize unmodified nucleotide bases, type IV restriction enzymes are sensitive to sites with modified bases. An example of a microorganism with the latter system is the *E. coli* strain McrBC. The effectiveness of these systems against bacteriophages is extremely high (about 10^8 , Tock *et al.*, 2005).

However, some bacteriophages are able to overcome the protective effect of this system, avoiding recognition sites of restriction endonucleases in their genomes (Luria *et al.*, 1952). The genomes of at least two phages - staphylococcus phage K (Tock *et al.*, 2005) and coliphage N4 - have entirely dispensed with sites that are recognized by

restriction enzymes present in their hosts. Viruses can also use other modifications in their genome, which prevent the restriction enzyme binding to a specific site. For example, in the genome of bacteriophage T4, the cytosine is modified to hmcC, so that it avoids the action of most restriction-modification systems. However, the restriction enzymes of the IV type system can recognize this modification of the T4 virus (Bair *et al.*, 2007).

Restriction-modification systems are not limited to bacteria. A type I functional system has been found in phage P1 that presumably enables stabilization of phage replicons during infection, but the exact mechanism of action of this system has not yet been clarified (Tyndall *et al.*, 1997).

2.2.4 Abortive infection systems

Abortive infection (Abi) systems are widespread and highly diverse in their mode of action. They can affect processes of phage replication, transcription, and translation, and lead to the death of a cell infected with a phage (Figure 1) (Forde *et al.*, 1999). In the context of a bacterial population, systems that lead the cell to a programmed death can be called altruistic, since the infected cell cannot produce a virus, so that neighboring sensitive bacteria are protected.

Abi systems were first discovered in *Escherichia coli* containing λ prophage. This system was initially taken to be a variant of a superinfection exclusion system but it was subsequently isolated as a separate defense system, as it resulted in the death of a cell infected with T4 virus. The λ prophage carries the *rexAB* gene that produces the intracellular RexA protein, which activates the membrane-binding protein RexB, and

these proteins are able to prevent infection by phage T4 with the *rII* mutation. When replication of the T4rII phage begins, RexA binds to RexB, which opens the ion channels in the membrane and causes its depolarization, reducing the ATP level inside the cell and eventually causing it to die (Landsmann *et al.*, 1982; Snyder *et al.*, 1989; Nawroz *et al.*, 1992).

Another similar example was found in the infection cycle with bacteriophage T7. *E. coli* bacteria containing the F plasmid on which the *pifABC* operon is encoded, express the PifA protein, which, in response to the expression of phage genes Gp1.2 and Gp10, binds to the cell membrane and leads to membrane depolarization (Schmitt, Molineux, 1991; Schmitt *et al.*, 1991).

To date, Abi systems have been most studied in *Lactobacilli* strains. Various strains of *L. lactis* are widely used for the fermentation of dairy products, and a sudden phage infection in the cultures of these bacteria can lead to significant economic losses. A detailed study of these microorganisms and their defense systems led to the discovery of more than 20 types of Abi systems (Hill *et al.*, 1990; Forde *et al.*, 1999). In most cases, Abi systems consist of one or two genes. The AbiA and AbiK system types prevent replication of the lactobacilli phage P335, although mutations in genes responsible for replication lead to viruses that are able to avoid the effects of these systems (Boucher *et al.*, 2000). The effect of the AbiK system was found to be different against phage 936: in the case of this virus, phage DNA replication was not disrupted, but the formation of viral procapsid and cell lysis did not occur (Boucher *et al.*, 2000). In cells containing the AbiB

system, the decay of phage mRNA transcripts was observed, preventing the synthesis of viral proteins (Parreira *et al.*, 1996).

Toxin-antitoxin systems are also often compared with Abi systems. They consist of a toxin that can act on various intracellular mechanisms and an antitoxin that represses the toxin. In the absence of a stress factor, the toxin is successfully inhibited by the antitoxin without affecting the cell processes, but when certain stimuli occur, the antitoxin is destroyed and the toxin acts on the cellular targets, resulting in a bacteriostatic effect or cell death (Dy *et al.*, 2014). Some toxin-antitoxin systems can be activated in response to penetration of the bacteriophage, since the host DNA is destroyed during the lytic infection, and transcription of the host genes is suspended, which leads to release of the toxin (due to rapid degradation of the antitoxin), causing death of the bacterial cell from the toxic effect (Fineran *et al.*, 2009).

2.2.5 BREX systems

In early 1980s, a system was discovered that limited the growth of viruses of *Streptomyces coelicolor* and it was supposed to be a restriction-modification system. Genes encoding Pgl proteins could affect infection by the phage ϕ C31 in such a way that the virus was incapable of subsequent infection of a strain containing *pgl* genes, but remained able to reproduce on cells not having *pgl* genes (Sumby *et al.*, 2003). The system comprised two operons *pglWX* and *pglYZ*, and the supposed mechanism of action was modification of the viral progeny in the first cycle of infection and destruction of the modified viruses during the secondary infection by means of restriction, since no viral DNA replication was observed in the infected cell (Sumby *et al.*, 2002). The Pgl locus

was predicted bioinformatically in 10% of sequenced genomes of bacteria and archaea. The *pglZ* gene was assigned to the family of alkaline phosphatases, *pglW* to serine-threonine kinases, and *pglX* contained a DNA methyltransferase motif (Sumbly *et al.*, 2003).

The Pgl system has not been studied extensively, but scientists at Sorek laboratory have recently resumed studies of the similar system, that they called bacterial exclusion (BREX). The BREX operon is composed of *brxABCXZL* genes, of which *brxX* and *brxZ* are homologs of the *pglZ* and *pglX* genes. It has been found that the BREX system does not inhibit phage adsorption on the cells, but can inhibit the replication of new phage genomes (Figure 1). It is believed that the protective mechanism of the BREX system is activated in the late stages of infection, when expression of phage proteins occurs, which may be susceptible to modifications by the BrxZ and/or BrxL proteins. A significant delay in cell lysis was also noted in the infected culture (Goldfarb *et al.*, 2015). However, there is still insufficient data for confident assertions about the mode of action of the BREX system.

2.3 Mode of action of CRISPR-Cas systems

CRISPR-Cas (Clustered Regular Interspaced Short Palindromic Repeats and CRISPR associated genes) systems represent another prokaryotic defense system and consist of short palindrome repeats separated by variable spacers and *cas* (CRISPR-associated system) genes (Makarova *et al.*, 2006).

Various approaches were undertaken to understand the purpose of these systems, discovered back in 1987 (Ishino *et al.*, 1987; Makarova *et al.*, 2002; Jansen *et al.* 2002),

but interest in these systems was first aroused by a publication of Makarova and co-authors (Makarova *et al.*, 2006), reporting how a series of palindrome repeats and conserved protein domains located near to them could participate in DNA and RNA metabolism, with some of these proteins being variable, and some remaining conserved and found in the genomes of both archaea and bacteria. The possibility that these systems intervene when the cells of bacteria or archaea are invaded by foreign RNA was proposed (Makarova *et al.*, 2002; Makarova *et al.*, 2006). The relationship between CRISPR systems and protection from phage infection was tested and confirmed experimentally in the course of work with *Streptococcus thermophilus*. Scientists found that the presence of this system in bacteria reduces their sensitivity to viruses (Barrangou *et al.*, 2007). Experimentally, a correlation was found between the presence of a spacer and resistance to virus with corresponding sequence (Barrangou *et al.*, 2007; Marraffini *et al.*, 2008).

Another remarkable property of CRISPR-Cas systems, which has been discovered, is that the presence of palindromic repeats and spacers (the CRISPR array) not only makes bacteria resistant to phages, but, in case of infection, new spacers can be acquired in the CRISPR array, providing subsequent resistance to the corresponding phage. It was also confirmed that the *cas* genes located nearby are necessary for this process (Barrangou *et al.*, 2007). More detailed study of these systems showed that they are extremely diverse, probably due to rapid evolution in response to the pressure exerted by mobile genetic elements (bacteriophages and plasmids) (Van der Oost *et al.*, 2014, Makarova *et al.*, 2015).

Adaptive immune systems are present in the genomes of most bacteria and archaea species. The most conserved proteins are Cas1 and Cas2 and they likely developed independently of the other Cas proteins and were responsible for adaptation of foreign DNA as mechanisms of CRISPR immunity evolved. The classification of CRISPR-Cas systems is based on a comparison of *cas* genes, and divides these systems into two classes and six types, within which subtypes are distinguished. The systems are defined depending on the composition and structure of the variable Cas proteins that perform the protective effector function (Makarova *et al.*, 2015). Class I (types I, III and IV) is characterized by the presence of a multisubunit effector complex, while in class II (types II, V, VI) the effector complex is represented by a large single multidomain protein. Effector complex in class I CRISPR-Cas system is divided into genes responsible for recognition and binding to crRNA and the target, and gene(s) that responsible for cleavage of the target (Koonin *et al.*, 2017). CRISPR-Cas systems of several types can be present in the genome of the same species (Grissa *et al.*, 2007a; Yosef *et al.*, 2012). However, there are still many unclassified versions of CRISPR-Cas systems, and their number is expected to increase significantly with analysis of rare and unexplored bacterial and archaeal genomes and metagenomes from natural communities.

The mechanism by which CRISPR-Cas systems provide immunity against mobile genetic elements can be divided into 3 stages:

- a. the insertion of a fragment of foreign genetic material into the CRISPR array (CRISPR-adaptation);
- b. transcription and maturation of crRNA;

c. recognition of the DNA (RNA) target and its destruction by the Cas-protein complex (CRISPR interference) (Van der Oost *et al.*, 2014).

In this thesis, the mode of action of the E subtype of type I CRISPR-Cas system, which provides immunity in the laboratory strain of the bacterium *Escherichia coli* K12, will be discussed in detail. This system includes *cas1*, *cas2*, *cas3* genes and genes which form the Cascade (CRISPR-Associated Complex for Antiviral Defence) effector complex - *cas8e* (*cse1*), *cse2*, *cas7*, *cas5* and *cas6* (Makarova *et al.*, 2015).

2.3.1 CRISPR adaptation

CRISPR-Cas systems have a unique ability to direct the incorporation of the fragments of foreign genetic material into the CRISPR array, which then serve as new spacers. CRISPR adaptation provides inherited resistance to re-infection by viruses containing corresponding protospacers in a process that resembles the Lamarckian acquisition of traits through interaction with the environment (genetic parasites) (Makarova *et al.*, 2015).

Two modes of CRISPR adaptation - non-primed and primed - have been identified. Non-primed CRISPR adaptation in *E. coli* occurs extremely rarely when a foreign DNA (virus or plasmid) enters the cell, or when new spacers from the bacterial genome are accidentally inserted into the CRISPR array, and such bacteria are eliminated from the population due to the autoimmune response (Vercoe *et al.*, 2013).

The mechanism of non-primed CRISPR adaptation of the I-E type system is provided by the Cas1 protein with DNase activity and by the Cas2 protein. Both proteins are necessary for the acquisition of a new spacer (Yosef *et al.*, 2012). Insertion of a new

spacer occurs into the proximal part of the CRISPR array immediately after the AT-rich leader region (Diez-Villasenor *et al.*, 2013). In the course of adaptation, one new repeat is also embedded into the CRISPR array (Yosef *et al.*, 2012).

Although one spacer is sufficient for target degradation, one CRISPR locus can contain multiple spacers from a single genetic element. This discovery revealed the phenomenon of "primed" insertion of a spacer, where the presence of one spacer can drive the insertion of additional spacers from the same target (Datsenko *et al.*, 2012). The phenomenon of primed CRISPR adaptation can be observed when there is penetration into the cell of a target with a protospacer that is completely or partially identical to the spacer in the CRISPR array (Semenova *et al.*, 2016). In this case, priming uses all protein components of the CRISPR locus and efficiency of the new spacer acquisition in the array increases by a factor of 10-20 (Datsenko *et al.*, 2012; Savitskaya *et al.*, 2013; Semenova *et al.*, 2016).

2.3.2 Transcription and maturation of crRNA

Transcription of CRISPR-Cas elements is necessary for complete recognition of a protospacer on foreign DNA. A CRISPR array of I-E type consists of palindromic repeats with length of 29 base pairs, separated by varied sequences with length of 32 base pairs (spacers). The CRISPR array is transcribed into a long non-coding RNA, and then the transcript precursor is processed by the family of Cas6 metal-independent endoribonucleases to form short crRNAs, each containing one spacer and a fragment of flanking repeats (Garside *et al.*, 2012).

The processed crRNA functions by binding to the Cascade protein complex, which is a distinctive feature of CRISPR-Cas systems of the type I. This large, multi-subunit complex of Cas proteins, in combination with crRNA, is able to recognize the DNA target (Brouns *et al.*, 2008; Jore *et al.*, 2011).

2.3.3 CRISPR interference

The *E. coli* Cascade complex contains one copy of Cse1 protein, two copies of Cse2, six copies of Cas7, one Cas5 and one Cas6 copy, which together with 61 nucleotide-long crRNAs represent the 405-kDa Cascade complex (Jore *et al.*, 2011). In this complex, crRNA is "fixed" on both sides with proteins Cas5, Cse1 (from the 5'-end) and Cas6 (from the 3'-end), and, in the middle, there are six Cas7 subunits (Zhao *et al.*, 2014). Such configuration of the complex allows the RNA-DNA hybrid to be retained when a target is encountered (Mulepati *et al.*, 2014). A Cse1 protein located at the 5'-end of the crRNA-Cascade complex is responsible for non-specific interactions with DNA during the primary scanning of a target (Jore *et al.*, 2011).

A requirement for CRISPR interference is full complementarity between the nucleotide sequences of the spacer part of the crRNA and the protospacer on the target molecule, and the presence of a PAM sequence immediately adjacent to the protospacer. The PAM trinucleotide is at the 3'-end of the protospacer, and the Cse1 protein in the Cascade complex is responsible for recognition and binding to the PAM sequence (Mojica *et al.*, 2009). Presence of the PAM allows the system to distinguish the foreign DNA protospacer from the spacer in the CRISPR array of the bacterium and, therefore, to avoid an autoimmune reaction (Deveau *et al.*, 2008; Horvath *et al.*, 2008; Mojica *et al.*,

2009). The preferred sequence of the PAM trinucleotide for a CRISPR-Cas system of I-E type is AAG (Savitskaya *et al.*, 2013). In CRISPR interference, the target protospacer is recognized and an R-loop is formed containing a heteroduplex between the spacer part of the crRNA and one of the chains of the protospacer on the foreign DNA.

The Cascade complex, together with the crRNA, recognizes a complementary protospacer on foreign DNA and attracts the Cas3 nuclease that causes degradation of the foreign DNA (Figure 1). Cas3 has exo- and endonuclease activities and it destroys DNA in the 3' → 5' direction. Cas3 protein also has a helicase activity and is capable of unwinding double-stranded DNA for subsequent degradation. The biological substrate for the Cas3 protein is the R-loop formed between the crRNA-DNA duplex upon binding of the crRNA to the protospacer on the DNA target (Beloglazova *et al.*, 2011; Sinkunas *et al.*, 2012; Westra *et al.*, 2013).

2.4 Interaction of bacteriophages with bacteria containing the CRISPR-Cas system

The bacterium can incorporate the viral genetic material into its own genome in the form of spacers. These spacer sequences can be mapped to the genomes of viruses from the same community and this would identify the infections that have occurred. In laboratory conditions, under phage pressure, it was shown that active CRISPR loci could acquire new spacers even daily with constant cultivation (Paez-Espino *et al.*, 2013). It is therefore possible to mimic interactions between bacteria and their viruses in natural habitats and, using this approach, interactions in the human microbiota (Minot *et al.*, 2013), in samples of glacial waters (Sanguino *et al.*, 2015) and from marine ecosystems (Cassman *et al.*, 2012) have been charted. Taking into account the feature of the

incorporation of new spacers, when new spacers are inserted predominantly into the leader-proximal region of the CRISPR array (Barrangou *et al.*, 2007; Deveau *et al.*, 2008; Andersson *et al.*, 2008; Tyson *et al.*, 2008), it is possible to detect both recent infectious events and older events that characterize the more steady state of CRISPR-mediated immunity. However, some microorganisms and their phages are impossible to grow under laboratory conditions because it is difficult to recreate an environment for them with an identical set of nutrients and microelements (Clokic, Kropinski, 2008). In addition, about 60% of microorganisms that make up the natural microflora of animals and humans have been found to be uncultivated (Peterson *et al.*, 2009). It is believed that, when certain strains of archaea are cultivated in the laboratory, they lose CRISPR loci due to the absence of invasion by extrachromosomal elements (Lillestol *et al.*, 2006). Although in the natural habitat the CRISPR-Cas systems are extremely important elements for the survival of archaea and are found in 80% of them, while only 40% of bacteria species have the CRISPR-Cas system (Grissa *et al.*, 2007a). To date, use of high-tech research methods, such as high-throughput sequencing or shotgun sequencing, has enabled the analysis of microbiological communities without the need for microbiological and cultural approaches (Staden *et al.*, 2016). So the expansion of information on metagenomic data from various habitats and their comparison makes it possible to detect new bacterial-viral interactions in the course of CRISPR immunity, as well as to search for new phages and their hosts, and to add to the existing and predicted information about their interactions. For example, metagenomic analysis of biofilm samples of the drainage shaft restored viral genomes and compared them with their

archaeal and bacterial hosts, using homology with the sequences of CRISPR spacers, and at the same time for the *Leptospirillum* microorganism, distinguishing two subpopulations based on preserved ancestral spacers in its CRISPR arrays (Simmons *et al.*, 2008). The abundance of various spacers observed in a CRISPR array within a single strain and in only one of the CRISPR-Cas system types indicates the complexity and variety of interactions in communities of microorganisms. So an in-depth study of microbial communities gives access to an unprecedented variety of genomes, enhancing the capabilities of biomedical biotechnology techniques.

Bacteria that carry a functional CRISPR-Cas system are not always able to withstand a viral attack, despite the high efficiency of the system. Thus, bacteriophages, which have mutations in the PAM or nearby “seed” (proximal 8 nucleotide-long part of the protospacer) regions of the protospacer, are not susceptible to CRISPR interference (Semenova *et al.*, 2011). Further, the presence of two or more substitutions in a less significant distal area of the protospacer also makes the virus insensitive to interference by the CRISPR-Cas system. In the short term, the accumulation of mutations in the protospacer enables the phage to overcome the protective effect of CRISPR interference. However, it can eventually lead to a "mutational catastrophe", when an excessive number of nucleotide substitutions in the phage genome cause loss of viability (Paez-Espino *et al.*, 2013). From the side of the bacterium, response to phage attempts to avoid interference by the accumulation of mutations in the protospacer is primed adaptation, where the bacterium acquires multiple spacers against target DNA after complete or incomplete recognition of the target (Datsenko *et al.*, 2012; Semenova *et al.*, 2016).

In response to primed adaptation phenomenon in the host, bacteriophages developed additional strategies. For example, the ICP1 phage, infecting *Vibrio cholerae*, acquired its own CRISPR-Cas I-F type system, capable of interfering with the host defense. The system contains 2 spacers in the CRISPR array, completely complementary to the host genome, and CRISPR interference by this system facilitates cell lysis (Seed *et al.*, 2013).

Another interesting example was shown for *Campylobacter jejuni* carrying the type II-A CRISPR-Cas system, which requires the Cas4 protein for the CRISPR adaptation process. One of the *C. jejuni* phages encodes a Cas4-like protein, which, in an infected cell, actively acquires new spacers directed against the genome of the host cell, leading to degradation of the bacterial DNA and inactivating bacterial CRISPR-Cas system (Hooton *et al.*, 2015).

There are also examples of lysogenic phages containing their own CRISPR-Cas system, which ensures the super-immunity of lysogens against other viruses. Among such prophages there is an *E. coli* virus, CY1_63_1964 of the *Caudovirales* family, found in the virome of glacial waters and carrying spacers of the I-E type CRISPR array against another phage, CY1_24_10782, from the same virome (Bellás *et al.*, 2015).

Given the significant variety of viral strategies for avoiding the CRISPR-Cas system, the next question is: which protection strategy enables bacteria to fight against the viral attack most effectively. Several studies have compared CRISPR-Cas systems with other protective systems, such as receptor modification. Dutch scientists carried out an experiment on a model system including *P. aeruginosa* and the phage DMS3vir. The

experiment was conducted in different types of media, some rich and some poor in nutrients. It was noted that the response by the CRISPR-Cas system might be more effective in conditions where there is only a minimal amount of nutrients (Westra *et al.*, 2015).

Co-evolution processes have also been described, in which the genome of the host cell is enriched with new spacers corresponding to the phage genome regions, and mutant bacteriophages are selected, which can avoid CRISPR recognition because of changes in protospacers. Such co-evolutionary experiments using *S. thermophiles* and bacteriophage D2972 found that the greatest resistance to the virus came from the activity of CRISPR-Cas systems, and not from other protection systems. The analysis of acquired spacers made it possible to detect the optimal target sites in the phage genomes, against which the action of the CRISPR-Cas system was mostly directed (Paez-Espino *et al.*, 2015).

In other co-evolutionary experiments in cultures with large species diversity, the number of escape mutants of viruses infecting particular species of bacteria was smaller. But CRISPR-Cas system presented a strong obstacle to the spread of phage infection in less diverse mixed cultures (Lundgren, 2016).

This effective defense is probably what has led to the development of other viral devices that interfere with CRISPR-Cas systems. Such devices include anti-CRISPR proteins. The presence of anti-CRISPR proteins was first detected in the phage DMS3vir of *P. aeruginosa*, and enabled the virus to survive in both a mixed bacterial population and a mono-population, despite the presence of active CRISPR-Cas system in the cells targeting the virus. The bacteria, in turn, could respond to this by the emergence of a wide

variety of CRISPR-Cas systems, as well as the acquisition of several types of these systems with diverse modes of action (Bondy-Denomy *et al.*, 2013). Anti-CRISPR proteins are found in anti-CRISPR operons (Bondy-Denomy *et al.*, 2013; Pawluk *et al.*, 2014).

Anti-CRISPR proteins are extremely diverse in their structure and function, and it is difficult to predict bioinformatically similar protein sequences. For example, proteins, which were found in the I-F system, named AcrF1 and AcrF2, bind to the Csy complex, which is similar to the Cascade complex in the I-E system, and block the binding of crRNA to the target DNA. In addition, these proteins are directed against different parts of the Csy complex: AcrF2 binds to Csy1-Csy2 proteins before they have attached to the DNA target, while AcrF1 binds to Csy3 and then prevents its interaction with the target, even if the recognition and binding of the Csy complex have already occurred. AcrF1 subsequently binds to the Csy complex very firmly, changing its conformation, so that adequate binding to the target is impossible. A third protein, AcrF3, interacts with the Cas3 nuclease, preventing attraction of the nuclease to Csy complex bound DNA target (Bondy-Denomy *et al.*, 2015). The existence of extremely different modes of action of anti-CRISPR proteins reflects their independent development during evolution.

In some *Pseudomonas* strains, homologs of anti-CRISPR proteins are encoded in bacterial genome. Such proteins could contribute to the penetration of mobile genetic elements or target mobile genetic elements that encode their own CRISPR-Cas systems (Bondy-Denomy *et al.*, 2013; Pawluk *et al.*, 2014).

Expression of CRISPR-Cas system components could be costly and/or toxic and should be tightly regulated (Vale *et al.*, 2015; Westra *et al.*, 2015). In *E. coli*, expression of Cas proteins is activated in response to stress (Perez-Rodriguez *et al.*, 2011). The CRISPR-Cas system of the *Pectobacterium atrosepticum* is repressed by glucose but is activated by excess *cAMP*, which induces starvation (Patterson *et al.*, 2015). These data correlate with the findings that have shown that CRISPR-Cas system perform better when the culture is grown in media with minimal nutrient content (Westra *et al.*, 2015).

In denser cell cultures, bacteria are more susceptible to infection by viruses (Papenfort *et al.*, 2016). At high densities quorum sensing systems, based on accumulation of specific signaling molecules - autoinducers produced and released by each cell – are activated, causing all cells in the population to maintain a certain physiological state (Høyland-Kroghsbo *et al.*, 2016). Experiments conducted with *P. aeruginosa* suggest that quorum sensing reduces the probability of infection of sensitive cells by a virus by reducing the number of surface receptors (Høyland-Kroghsbo *et al.*, 2017). It was further noted that cells could huddle together, thereby blocking phage access to the receptors (Tan *et al.*, 2015). Finally, quorum sensing leads to the activation of CRISPR-Cas system components providing additional protection from infection (Høyland-Kroghsbo *et al.*, 2017).

Chapter 3. Materials and methods

3.1 Nutrient media

The following components were used to constitute the "Luria-Bertani" (LB) liquid nutrient medium (tryptone (Amresco, Spain), 10 g/L; yeast extract (Difco, USA), 5 g/L; NaCl, 5 g/L. To obtain a solid medium, 15 g/L of bacto-agar (Difco, USA) was added; to obtain a soft medium, 5.5 g/L of bacto-agar was added.

3.2 Bacterial strains

The following strains were used in the study: *E. coli* DH5alpha (Invitrogen, F- Φ 80lacZ Δ M15 Δ (lacZYA-argF) U169 recA1 endA1 hsdR17 (rK-, mK +) phoA supE44 λ -thi-1 gyrA96 relA1), *E. coli* BW25113 (KEIO-collection, F-, DE (araD-araB) 567, lacZ4787 (del) :: rrnB-3, LAM-, rph-1, DE (rhaD-rhaB) 568, hsdR514): *E. coli* KD263 (F + , AraB8p-casA-uv5-cas3, one CRISPR array (R-M13g8-R), Δ (CRISPR II + III)), *E. coli* KD587 (F-, araB8p-casA-uv5-cas3, one CRISPR array (R -M13g8-R), Δ (CRISPR II + III)). These strains have an inducible CRISPR-Cas system containing *araB8p* and *uv5* promoters, and an introduced spacer g8 against gene 8 of bacteriophage M13. The strains are based on the recombination protocol using Red recombinase (Datsenko *et al.*, 2000) and were provided by Kirill Datsenko, Purdue University, USA. A collection of strains, listed in Table 1, was created, by a method that will be covered in the further sections of the "Materials and Methods" chapter.

Table 1. Details of strain and phage collection. PAM = protospacer adjacent motif, EOP = efficiency of plating, EOT = efficiency of transformation, ND = no data, NI = no interference, NA = not available

Spacer	PAM	Position of protospacer	Sequence of spacer	EOP	Escape-phages	EOT
λ phage						
λ-E1-F	AAG	25592-25623	TTCTCACCGAATGTCTCAATATCCGGACGGAT	2.67 x 10 ⁻⁶	C23DEL	4.32 x 10 ⁻³
λ-E2-F	AAG	25742-25773	TCGCTTAATGTTTCGTA AAAAAGCAGAGAGCAA	2.89 x 10 ⁻⁷	A20DEL	2.80 x 10 ⁻²
λ-E3-F	AAG	25776-25807	GTGGATGCAGATGAACCTCTGGTTCATCGAAT	8.61 x 10 ⁻⁷	ND	1.77 x 10 ⁻³
λ-E4-R	AAG	29829-29798	ATAACGCTTGTGAAAATGCTGAATTCGCGTC	5.33 x 10 ⁻⁶	T11-T26DEL	2.05 x 10 ⁻³
λ-E5-F	AAG	42248-42279	ACGTCGTGCGAGGAAAACAAGGTGATTGACCA	5.33 x 10 ⁻⁶	ND	ND
λ-E6-F	AAG	42429-42460	ATGATGGCTAAACCAGCGCGAAGACGATGTAA	4.67 x 10 ⁻⁶	A22DEL	6.26 x 10 ⁻²
λ-L1-R	AAG	2859-2828	GGTGGGAATGGTGGGCGTTTTTCATACATAAAA	1.00 x 10 ⁻⁹	NA	ND
λ-L2-R	AAG	3324-3293	CCGCGACGAACTGGTATCCAGGTGGCCTGAA	1.00 x 10 ⁻⁹	NA	8.07 x 10 ⁻³
λ-L3-F	AAG	3361-3392	CGCATCAGCAACCCGAACAATACCGGCGACAG	1.00 x 10 ⁻⁹	NA	4.06 x 10 ⁻²
λ-L4-R	AAG	3951-3920	CTGCTCATACGAGACACCCAGCCCGCAGCGA	1.46 x 10 ⁻⁷	G-1T, G25T A28C	ND
λ-L5-R	AAG	4019-3988	TACGCCACGACTCGTTCGACTGGCCCGTGC	1.00 x 10 ⁻⁹	NA	3.09 x 10 ⁻²
λ-L6-R	AAG	4347-4316	ACCGGCTGCACGGCGCTCCATCGTTTCACGGA	1.00 x 10 ⁻⁹	NA	ND
λ-L7-R	AAG	4891-4860	CGCCCATACCGGTTTTATGTCACGCACACGGG	1.00 x 10 ⁻⁹	NA	ND
λ-L8-F	AAG	5072-5103	GTGGATGGCAACCCCTACAGCCATCTCCGGA	1.00 x 10 ⁻⁹	NA	ND
λ-L9-R	AAG	5263-5232	TTCATCAGCCAGTCCGGCATCAATGGCCTCCT	1.48 x 10 ⁻⁸	G-1T, A4C, T20G A22C	4.11 x 10 ⁻³
T5 phage						
T5-PE1-F	AAG	621-652	TATTTGCGTCATTCATGCACTGGCGCATATAT	7.78 x 10 ⁻²	ND	3.10 x 10 ⁻¹
T5-PE2-R	AAG	1381-1350	CTAGCAAATAAATTAATCCCTTATTAACCTA	2.25 x 10 ⁻⁴	T2C, C14DEL	5.90 x 10 ⁻¹
T5-PE3-F	AAG	1403-1434	CGAGATTTGTTTCTTCCATACGTCAGGAATA	1.58 x 10 ⁻¹	ND	3.14 x 10 ⁻¹
T5-PE4-R	AAG	1414-1383	AAACAAATCTCGCTTTTGTGTAAAGTCTTGAA	3.36 x 10 ⁻⁴	ND	1.38 x 10 ⁻¹
T5-PE5-R	AAG	1552-1521	ATAATAATTGTTATATTTATCAAAATGAACTA	9.50 x 10 ⁻³	ND	1.00 x 10 ⁻¹
T5-PE6-R	AAG	1726-1695	AATGGTAAATCTTTGCGTAATTAATTTATTC	6.10 x 10 ⁻⁴	A2G, T3C, G5A, A7G, A8G	8.62 x 10 ⁻²
T5-PE7-R	AAG	1900-1869	AAATGGGCAAATTAACACCGGACACACCCA	7.54 x 10 ⁻⁶	G-1A	1.81 x 10 ⁻³
T5-PE8-R	AAG	1975-1944	ATTGCAGGCATTGAATGCGGGCTACACGGTCA	2.53 x 10 ⁻⁶	A1G, T2C, C5G, G7A	4.60 x 10 ⁻¹
T5-PE9-R	AAG	2029-1998	GTTGCAGGCTGTGAATTTAACGCCATTTTCCT	1.00 x 10 ⁻⁸	NA	3.18 x 10 ⁻²
T5-E1-F	AAG	25953-25984	TTACCTGTCTATATATAGGAAAGCTAGTTGTT	1.00	NI	ND
T5-E2-R	AAG	32238-32207	GGTTGCTAGCTCAACTGGTTTAGAGCACTGGT	9.46 x 10 ⁻¹	NI	1.64 x 10 ⁻³
T5-E3-F	AAG	48106-48137	TAGGTAATGTATCACCTATCATTGATATGGGA	1.03	NI	3.64 x 10 ⁻²
T5-E4-F	CAG	48757-48788	ATCTTTGTCAAACCTGCTGGAACATAATCTGGA	1.01	NI	2.26 x 10 ⁻²
T5-E5-R	AAG	61219-61188	CGTGTGCTTTTGACAGCATTTTACGGTGATGA	9.09 x 10 ⁻¹	NI	8.85 x 10 ⁻³
T5-L1-R	AAG	86896-86865	GTGTTGATGGCAGGCCGGGGCTACAGGACAA	8.00 x 10 ⁻¹	NI	ND
T5-L2-R	AAG	87054-87023	TCCTGCATACCCACCTTCTGGTTGGTCTAAAA	9.06 x 10 ⁻¹	NI	7.78 x 10 ⁻²
T5-L3-F	AAG	94686-94717	AGCATCATCTAGGTATCCAAATCTACGAGTAG	8.70 x 10 ⁻¹	NI	1.09 x 10 ⁻²
T5-L4-R	AAG	95345-95314	GGTTTAATGACACTGCTGGTGTCTGGTAG	1.03	NI	1.43 x 10 ⁻³

T5-L5-R	AAG	101360-101329	TTCTCCCTCAGCTACTTCAGTAACCAAAATGA	9.62×10^{-1}	NI	ND
T5-L6-R	AAG	102506-102475	CAGTGAATCAGTCTTCTCCGTAGAAAGTTTCT	1.00	NI	ND
T7 phage						
T7-E1-R	AAG	2796-2765	CGCCCCCTGTGACGTTCTTTACGGGCCTTAGC	1.59×10^{-2}	large deletions 59 bp including 12 first bp of protospacer	1.94×10^{-1}
T7-E2-F	AAG	3793-3824	ACTCTATTCATGTAGGAGTACGCTGCATCGAG	2.50×10^{-6}	A-2C, A1G	6.83×10^{-3}
T7-E3-F	AAG	3889-3920	ACTCTGAGACTATCGAACTCGCACCTGAATAC	3.57×10^{-5}	A-1C, C4A	ND
T7-E4-F	AAG	3999-4030	CCGTGGACTGGCATTACTGGTGGTGGCTATTG	2.11×10^{-4}	C1T, C1A, G-1A, GGT triplet insertion into 24 position	1.91×10^{-1}
T7-M1-F	AAG	9519-9550	AGACCAAGCACATCAATCTGGTGTGGTTGAC	1.00×10^{-1}	G-1A, A1G, A1C	4.36×10^{-1}
T7-M2-F	AAG	9893-9924	GTGACTTCCTCTGGGTGTTGGGAGTGGCAGGG	1.32×10^{-1}	G1T, G-1A	3.12×10^{-2}
T7-M3-F	AAG	9977-10008	GTTGTCTACTGTATCGCGTAATGTCTAATGC	1.60×10^{-2}	NA	1.39×10^{-1}
T7-M4-F	AAG	10017-10048	GTTCTACCGCTCGACTCCTGTGATAATCCA	1.60×10^{-2}	ND	ND
T7-L1-F	AAG	26076-26107	AGTTACTCATCTGGTCCGATGAAGCACAATTC	1.00×10^{-9}	ND	1.18×10^{-3}
T7-L2-F	AAG	26100-26131	CACAATTCGTCCTGACTGCCTCGGGTACTCTC	1.84×10^{-4}	A-2G, C1T, A5G, A-3C	5.10×10^{-1}
T7-L3-F	AAG	26141-26172	TCGGTTGAGTTGAACCTAACGACCCAGTTTGA	7.97×10^{-6}	A-2C, T1G	1.42×10^{-2}
T4 phage						
T4-E1-F	AAG	44001-44032	CGATAACACAGGTTTTACAGCAGGAGTCAATA	8.00×10^{-1}	NI	2.29×10^{-3}
T4-M1-R	AAG	21661-21630	AAAATGCCGGATGTTGGTGCTATTCCTGACAT	5.00×10^{-1}	NI	5.00×10^{-2}
T4-L1-F	AAG	84327-84358	AAAGAAGACGTATTCAACCCGGATATGCGAAT	3.00×10^{-1}	NI	ND
T4-L2-R	AAG	84389-84358	ACAGAGAAACTAATTTACCGTCAGTTGACATA	1.00	NI	3.57×10^{-1}
T4-L3-R	AAG	85600-85569	CAAATCCTTTCTTTAAACCCACGAATAATTT	5.00×10^{-1}	NI	3.36×10^{-2}
T4-L4-R	AAG	86502-86471	GAGCTGAGTTACACACTACAATATCGTTAATT	2.50×10^{-1}	NI	1.77×10^{-3}
T4-L5-R	AAG	86781-86750	GAACCACGATATATTCATTCGTGCATCTATT	4.67×10^{-1}	NI	8.86×10^{-2}
T4-L6-F	AAG	86806-86837	ACCCGACTAGATGGGGATATGAAGATAATCTC	1.00×10^{-1}	NI	ND
T4-L7-R	AAG	151008-150977	CGGTTATCATAAATGATTACAAAATCTTTTAA	3.90×10^{-1}	NI	ND
T4-L8-R	AAG	151290-151259	TTTACAATACTTTGTACTGGAGCTACAATTAA	9.70×10^{-1}	NI	8.10×10^{-3}
T4-L9-R	AAG	151595-151564	TCGTTTCATCGTATGTAGTAACAATTTGATGA	3.00×10^{-1}	NI	ND
T4-L10-R	AAG	151913-151882	CAATTTTATCTTCATCAGCAGCTTTGATTTA	3.70×10^{-1}	NI	ND
R1-37 phage						
R1_37-E1-R	AAG	46357-46326	ATTTGACTACTGATTACAGATTCATGCCAGCT	1.00	NI	1.40×10^{-2}
R1_37-E2-R	AAG	46637-46606	TATCCGGGAGCATCGACTGCATAATGAATGG	9.00×10^{-1}	NI	2.03×10^{-2}

R1_37-E3-R	AAG	46922-46953	GGATTCTTCATTCATCCAGTTATATGGTTGAA	1.01	NI	2.93×10^{-3}
R1_37-L1-F	AAG	178732-178763	ATATTGATGGTGATCCGATGCTTAATATCGGT	1.02	NI	6.17×10^{-2}
R1_37-L2-F	AAG	178586-178617	GACACATTTAATGGGTGTTTTAACGAAAAGCT	8.00×10^{-1}	NI	4.32×10^{-2}
M13 phage						
M13_g8R	AAG	1360-1329	ACAGCATCGGAACGAGGGTAGCAACGGCTACA	8.33×10^{-7}	G-1T	4.18×10^{-2}
M13_g8	ATG	1357-1388	CTGTCTTTCGCTGCTGAGGGTGACGATCCCGC	1.67×10^{-5}	C1T	4.05×10^{-3}

3.3 Bacteriophages

Bacteriophages T4, T5 and T7 were used in the study, as well as a virulent mutant of phage λ bearing a mutation in the *cl* gene and a giant bacteriophage that infects *Yersinia pseudotuberculosis* R1-37 (Skurnik *et al.*, 2012).

3.4 Plasmids

A plasmid pG8 and pG8mut based on the vector pT7blue (Novagen, UK) and containing a region identical to the 209-nucleotide fragment of the bacteriophage M13 gene (positions 1311-1519 in the phage genome) was used in the study; this fragment carries the protospacer g8 (in the case of pG8), and protospacer g8 with a single C1T mutation in the first position of the protospacer (in the case of pG8mut) and was provided by Ekaterina Semenova (Waksman Institute, Rutgers University, Piscataway, USA).

In the work with phage R1-37, a plasmid that constitutively expressed a receptor, with which the virus could contact on the surface of the bacterial membrane, was used.

3.5 Strain construction

The preparation of strains bearing new spacers against bacteriophages in their CRISPR array consists of several steps:

- 1) isolation of the bacteriophage DNA (this step was used for phages λ , T5, T7);
- 2) PCR of different regions of the genome of bacteriophage DNA (this step was used for phages T4, R1 -37);
- 3) restriction of PCR fragments or bacteriophage DNA to obtain a library;
- 4) restriction of pG8mut plasmid, ligation of restriction fragments with pG8mut plasmid;
- 5) screening for recombinant clones on a selective medium;
- 6) selection of recombinant plasmids by PCR;
- 7) sequencing of recombinant plasmids;
- 8) transformation of these plasmids into KD263 and KD578 strains;
- 9) induction of the CRISPR-Cas system;
- 10) selection of clones with insertion of a new spacer on a non-selective medium;
- 11) selection of clones which have lost the plasmid on the selective medium;
- 12) PCR of the expanded CRISPR array of selected clones;
- 13) sequencing of PCR products (Figure 1).

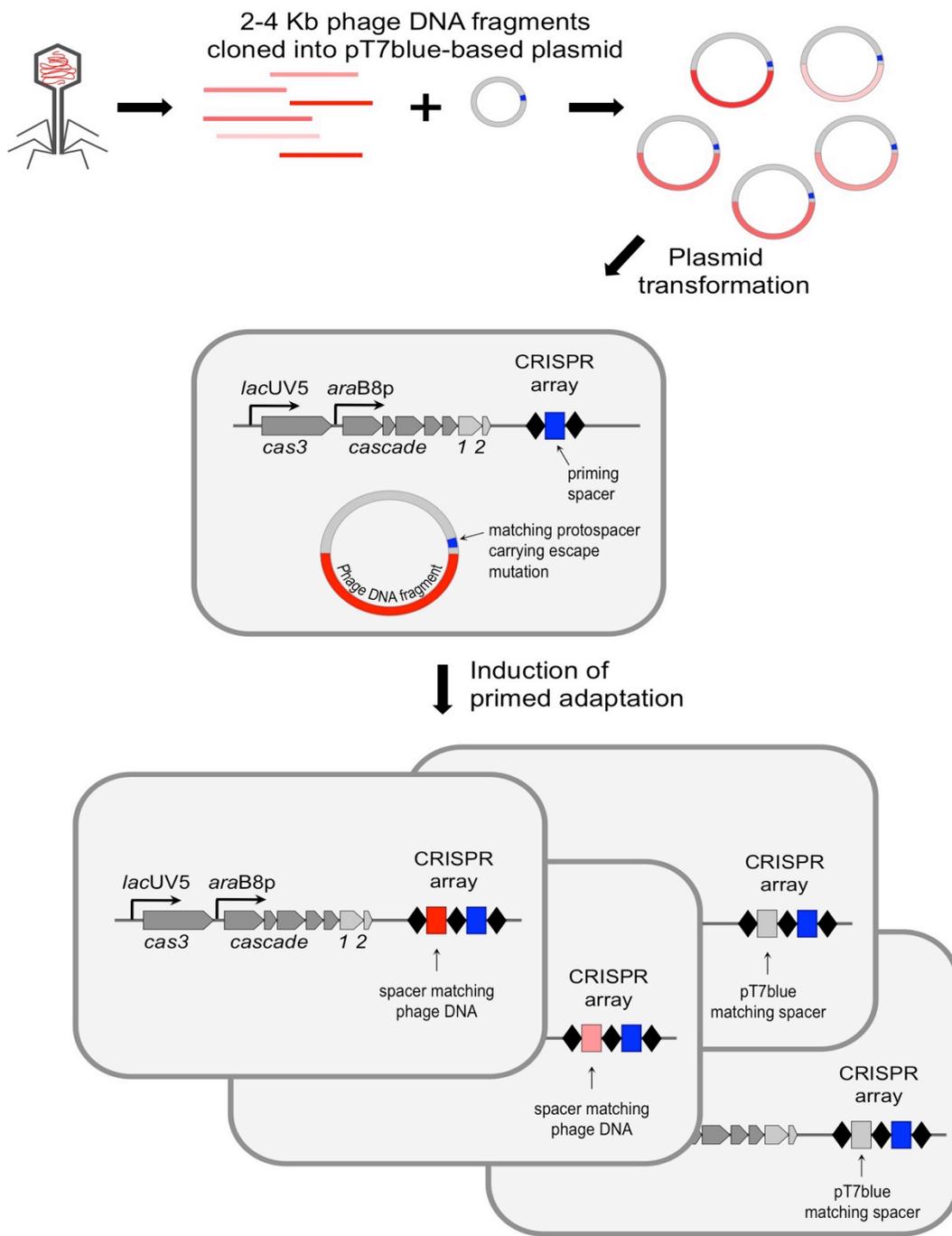


Figure 2. Creation of *E. coli* strains with CRISPR spacers targeting various phages using primed adaptation.

Fragments of phage DNA (different shades of red) are cloned into a plasmid vector carrying an ampicillin resistance gene and a priming protospacer (blue) with a mismatch (escape mutation) to a resident CRISPR spacer in an available *E. coli* strain. The strain is

transformed with a resulting plasmid library, *cas* gene expression is induced and cells that lost ampicillin resistance are selected. Such cells contain new spacers acquired into their CRISPR array through primed adaptation. A new spacer can be acquired either from the plasmid background (grey) or from phage-derived inserts cloned into the plasmid (red).

3.6 Isolation of bacteriophage DNA

While PCR amplification directly from cell lysates obtained after phage infection is efficient in most cases, we find that longer fragments of phage DNA (5 kbp or more) are best amplified from purified phage genomic DNA. A method developed for plasmid purification and described by Lee and Rasheed (Lee, Rasheed, 1990) works well for most phage lysates. Isolation for further digestion of the DNA of phages λ , T5 and T7 was performed by the same method. 50 μ L of 0.5 M EDTA (Amresco, Spain) and 400 μ L of buffer-saturated phenol pH = 8 (Sigma-Aldrich, USA) was added to 450 μ L of phage lysate (a titer of 10^9 pfu/mL or higher), then the sample was centrifuged at 13,400 rpm for 10 min at room temperature and the upper aqueous phase was transferred into a fresh 1.5 mL microcentrifuge tube. 200 μ L of saturated phenol pH = 8 and 200 μ L of chloroform (Khimmed Synthesis, Russia) were added to the upper fraction, and the sample was centrifuged at 13,400 rpm for 10 min. The aqueous phase was transferred to a fresh microcentrifuge tube and extracted with 400 μ L chloroform as above. The upper fraction was transferred to a new tube and 1 mL of ice-cold 95% ethanol was added and the precipitate phage DNA was incubated at -20 °C for 15 min, then the sample was centrifuged by 10 min at 13,400 rpm. The supernatant was removed, the pellet was washed with 100 μ L of ice-cold 70% ethanol, the sample was centrifuged for 1 min at 13,400 rpm, and the pellet was air-dried before being dissolved in 50 μ L of mQ H₂O.

3.7 PCR

A polymerase chain reaction (PCR) was used to obtain DNA fragments that were subsequently used for cloning, to select clones containing inserts and to observe the expansion of the CRISPR array.

The PCR mix was prepared in 20 μ L with the following components: a thin-wall Eppendorf, 0.2 mL; 10x Taq buffer (ThermoFisher Scientific, USA); 0.2 μ L of a mixture of dNTP (10 mM of each deoxyribonucleotide); forward and reverse primers (5 pmol each; Evrogen, Russia); 5 ng of phage DNA or a single colony from a plate; 0.2 μ L of Taq polymerase (ThermoFisher Scientific, USA); mQ H₂O up to 20 μ L.

The PCR program included up to 30 cycles, with the following steps:

- 1 cycle: preliminary denaturation at 95 °C (3 min)
- 30 cycles: denaturation at 95 °C (20 seconds);
annealing of primers at 55 °C (15 seconds);
elongation at 72 °C for n min (n depends on the length of the amplified fragment)
- 31 cycles: final elongation at 72 °C (1-5 min)
- storage at 4 °C.

The reaction with each pair of primers was carried out at an annealing temperature calculated using the OligoAnalyzer program (<http://eu.idtdna.com/analyzer/applications/oligoanalyzer/>). The time of amplification of the fragment depended on the fragment length and the DNA polymerase that was used (elongation time was calculated on the basis that 1 minute was required for every 1000 bp using Taq polymerase). Aliquots of the samples obtained were visualized in 1% agarose

gel (Helicon, Russia) for large-size fragments (1000-4000 bp), and in 2% agarose gel to separate the PCR fragments amplified from the expanded CRISPR array. Further, horizontal electrophoresis was performed, the results of which were recorded in a Fusion FX transilluminator (Vilber, France).

A list of primers used to clone fragments of bacteriophages T4 and R1-37 is given below. The underlining indicates the restriction sites.

Table 2. Primers for T4 and R1-37 cloning.

Phage	Primer	Sequence
T4	T4B_7 - F	TTT TTG <u>GAT CCG</u> CGA CTT TAC CAG CGA ATG
T4	T4S_7 - R	TTT TTG <u>AGC TCG</u> GTA ATG CAG CTT CAG GAA AA
T4	T4_20600_F	AAA AAG <u>GAT CCG</u> CCA AAG CAT CTA ACT GAG C
T4	T4_21700_R	AAA AAG <u>GAT CCG</u> CGA TGT TTA ATG CTA CGT C
T4	T4_150800_F	AAA AAG <u>GAT CCG</u> AGC AGA AGA TGG TCT GGA C
T4	T4_151900_R	AAA AAG <u>GAT CCC</u> CAT TCA GCT TCA GGT GGA T
T4	T4_43500_F	AAA AAG <u>GAT CCC</u> GTT CGG ATT TCC CAA ATA AC
T4	T4_44700_R	AAA AAG <u>GAT CCG</u> GAA CTG ACT GGA CTG AAC G
R1-37	R_45100_ear_F	AAA AAG <u>GAT CCG</u> GTG TTT CTG GAT TCG TAT CTG
R1-37	R_47500_ear_R	AAA AAG <u>GAT CCC</u> CAT GAA TAA TTA CAC CAT CAT CTG
R1-37	R_178300_1_F	AAA AAG <u>GAT CCC</u> TCA GAA TGA CTC TTC CAG C
R1-37	R_179600_1_R	AAA AAG <u>GAT CCC</u> ATT TCC TGG AAC TCA GTG TC

The following primers were used to evaluate the success of cloning and the production of recombinant plasmids:

T7pro TAA TAC GAC TCA CTA TAG G

U-19mer CGC CAG GGT TTT CCC AGT CAC GAC

The following primers were used to evaluate the expansion of the CRISPR cassette:

Ec-LDR_F AAG GTT GGT GGG TTG TTT TTA TGG

M13_g8 GGA TCG TCA CCC TCA GCA GCG

3.8 Restriction

The plasmid was treated with *NdeI* and ligated with phage DNA digested with *MseI* in order to construct pG8mut-based plasmids containing λ , T5 or T7 fragments. Plasmids containing phage DNA inserts of desired size (2–4 kbp) were selected for further use. 2–4 kbp regions of phage genomes were amplified using appropriate primers and individually cloned in pG8mut in order to construct pG8mut-based plasmids containing fragments of T4 or R1–37 DNA. The sequences of primers are listed in Table 2.

Treatment with restriction enzymes was carried out according to the following protocol: 50-100 ng of pG8mut plasmid; 200 ng of phage DNA or PCR fragment; 2 μ L of 10x reaction buffer FastDigest Green (ThermoFisher Scientific, USA); 0.2 μ L of restriction enzyme FastDigest (ThermoFisher Scientific, USA); mQ H₂O up to 20 μ L. The reaction mixture was incubated at 37 °C for 20-40 min.

3.9 Extraction of DNA fragments from agarose

Restricted DNA fragments of bacteriophages λ , T5 and T7 were separated in 1% agarose gel and a piece of gel containing DNA fragments of 2-4 kbp was cut with a scalpel, then it was placed in the eppendorf and isolated with the help of the GeneJET Extraction Kit (ThermoFisher Scientific, USA), following the manufacturer's protocol. Restriction fragments of plasmid pG8mut and phages T4 and R1-37 were also purified using the GeneJET Extraction Kit (ThermoFisher Scientific, USA).

3.10 DNA ligation

Ligation was prepared by combination of the restriction fragments, the vector and μL of T4 DNA ligase (ThermoFisher Scientific, USA) in the 10x T4 ligation buffer (ThermoFisher Scientific, USA); the final volume of the ligase mixture was 10 μL . Reactions were incubated at 22 °C for 30 min. The insert-vector molar ratio under these conditions was 3:1.

3.11 Preparation of competent *E. coli* cells and chemical transformation

This protocol was used to produce transform recombinant plasmids. 1 mL of a night culture of *E. coli* DH5alpha cells was diluted in a flask with 100 mL of liquid medium LB 100. The culture was incubated with agitation at 37 °C until the OD_{600} reached 0.5; the culture was then chilled on ice for 10 min. The cells were collected by centrifuging at 5,000 rpm for 10 min at 4 °C. The supernatant was removed and the cell pellet was gently resuspended in 50 mL of ice-cold calcium buffer (100 mM CaCl_2 , 10 mM Tris-HCl pH 8.0) and left on ice for 30 min. The cell culture was centrifuged at 5,000 rpm for 10 min at 4 °C and the supernatant was removed. The cell pellet was resuspended in 2 mL of calcium buffer and 500 μL of sterile glycerol (PanReac AppliChem, Germany) and was distributed 50 μL aliquots into sterile prechilled 1.5 mL tubes, and then flash-frozen in liquid nitrogen for subsequent transformation or storage at -80 °C.

The chemical transformation was carried out according to the following protocol. 1–2 μL of ligation reaction (above) was added to a 50 μL aliquot of thawed ice-cold competent cells. The mixture was incubated on ice for 5 min, then a heat-shock was applied at 42 °C for 30 seconds, and the mixture was returned immediately to ice for 2

min. 250 µL of pre-warmed (37 °C) liquid LB was added and the cells were incubated for 60–90 min at 37 °C with agitation. After incubation, 10 and 100 µL of the cells were plated on two separate LB agar plates containing 50 µg/mL carbenicillin or 50 µg/mL ampicillin and the plates were incubated overnight at 37 °C. Clones containing recombinant plasmids were tested by PCR with primers to pT7blue vector T7pro and U-19mer, and the cloning results were analyzed and detected in 1% agarose gel.

3.12 Isolation of plasmids

Recombinant plasmids obtained by cloning and containing inserts of the required length were isolated from 5 mL of the overnight cultures using the GeneJET Plasmid Miniprep kit (ThermoFisher Scientific, USA), following the manufacturer's instructions.

3.13 Sequencing of recombinant plasmids and PCR fragments

Sequencing by the Sanger method was used to analyze insertions in the plasmid pG8mut after cloning, as well as for the analysis of PCR fragments of the expanded CRISPR array. Sequencing was performed at the Eurogen laboratory. The results of sequencing were analyzed using the Snapgene software (<http://www.snapgene.com>).

3.14 Preparation of electrocompetent *E. coli* cells and electroporation

Electrocompetent cells were prepared using a standard protocol (Miller *et al.*, 1995), and were used to transform the sequenced recombinant plasmids into KD263 and KD578 strains, as well as to analyze the transformation efficiency. 1 mL of a night culture of *E. coli* KD263 or KD578 cells was diluted in 100 mL of LB medium, with or without inducers (1 mM arabinose (DiaM, Russia) and 1 mM IPTG (Helicon, Russia)). The culture was incubated with agitation at 37 °C until the OD₆₀₀ reached 0.5, then the

culture was chilled on ice for 10 min. The cells were collected by centrifugation at 5,000 rpm for 10 min at 4 °C. The supernatant was removed and the cell pellet was gently resuspended in 20 mL of ice-cold mQ H₂O; this step was repeated twice. Then the cell pellet was resuspended in 2 mL of mQ H₂O and 500 µL of sterile glycerol and was distributed 50 µL aliquots into sterile prechilled 1.5 mL tubes, and then flash-frozen in liquid nitrogen for subsequent transformation or storage at -80 °C.

Electroporation was carried out according to the following protocol: the solution containing ~1 ng of pure plasmid was added to 50 µL of competent cells, the suspension was transferred to the electroporator cell, electroporation was performed at parameters 2.5 kV, 25µF, 200 Ω, cell clearance 0.1 cm (pulse time 5.4-5.9 mseconds). After the pulse, the suspension was eluted from the cell in 500 µL LB and incubated for 1 hour at 37 °C. After incubation, 10 and 100 µL of the cells were plated on two separate LB agar plates containing 50 µg/mL carbenicillin or 50 µg/mL ampicillin and the plates were incubated overnight at 37 °C.

3.15 Induction of components of the CRISPR-Cas system and selection of clones that incorporate a new spacer

Generally, the chance of obtaining an *E. coli* strain with a CRISPR spacer derived by primed adaption from the phage DNA segment of a spacer capture plasmid increases together with increase of the size of the phage insert. We routinely use fragments of 1–5 kbp. Since the size of pT7blue is 2,922 bp, screening of strains that acquired spacers from the phage becomes very easy.

An individual colony of KD263 and KD578 cells transformed with recombinant plasmids were grown overnight in 5 mL of liquid LB in the presence of carbenicillin or ampicillin at 37 °C. 100 µL of the overnight culture were transferred into 5 mL of liquid LB and after 1 h of growth in the incubator shaker at 37 °C the culture was induced by the addition of 1 mM arabinose and 1 mM IPTG.

When needed, for CRISPR interference and adaptation assays, half of the culture was induced, and another half of the culture remained uninduced. The cultivation was continued, and aliquots were withdrawn at various times postinduction. Serial dilutions of withdrawn culture aliquots were plated on LB agar with or without 50 µg/mL ampicillin to obtain colony forming unit (cfu) numbers.

The culture was grown for 8 h or overnight. Aliquots of serial dilutions of induced culture were plated on LB agar plates to obtain individual colonies. A dozen individual colonies were selected and plated, each using tooth-picks on LB agar plates with and without antibiotic. After overnight growth at 37°C, clones that had lost ampicillin resistance were identified (usually, less than 10% of colonies remain ampicillin-resistant and, therefore, harbor the plasmid after growth in the presence of inducers).

The expansion of the CRISPR array was confirmed by the PCR method with the primers Ec-LDR_F and M13_g8 (Figure 1), and the origin of a new spacer was determined by Sanger sequencing. All new strains obtained as a result of this experiment are listed in Table 1.

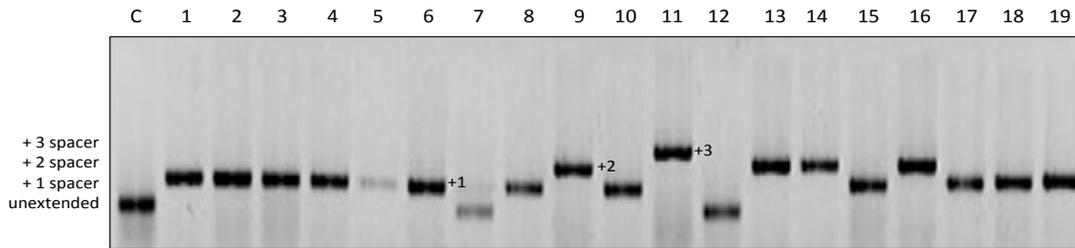


Figure 3. Screening for CRISPR array expansion. The results of analysis using agarose gel electrophoresis of 18 clones obtained from an induced culture of KD263 *E. coli* cells containing pG8 cloned with a PCR fragment from a T4 phage, which lost ampicillin-resistance. The lane labeled “unextended” is a control lane (amplification of a CRISPR array from KD263 *E. coli* that has not been transformed with any plasmid). Clones that have undergone array expansion by 1, 2, or 3 repeat-spacer units are indicated.

3.16 CRISPR interference assays

To assay for plasmid interference, cells were grown in LB medium at 37 °C in the presence or in the absence of 1 mM arabinose and 1 mM IPTG until the culture OD₆₀₀ reached 0.6. Electrocompetent cells were used from the step below “Preparation of electrocompetent *E. coli* cells and electroporation” and transformed with 5 ng of plasmids containing protospacers. After 1 h of outgrowth at 37 °C in 500 µL LB without antibiotic, 100 µL aliquots of serial dilutions of transformation mixtures were plated onto LB agar plates containing 100 µg/mL ampicillin. The plates were incubated at 37 °C overnight. The transformation was repeated three times for each plasmid. Efficiency of transformation was determined as a ratio of transformants formed by induced and uninduced cells.

To determine interference with phage infection, aliquots of induced and uninduced cultures grown as described above were combined with serial dilutions of

phage lysates, combined with soft LB agar and overlaid on LB plates with or without 1 mM arabinose and 1 mM IPTG. Each experiment was repeated at least three times. After drying, the plates were incubated at 37 °C overnight. The efficiency of plaquing was determined as a ratio of phage plaques formed by induced and uninduced cells. The data are provided in Table 1.

3.17 Monitoring of phage infection

The growth of induced and uninduced cultures with or without the phage was continuously monitored using EnSpire Multimode Plate Reader (PerkinElmer, USA). At least three growth curves were determined for each strain at every condition. Induced (1 mM arabinose and 1 mM IPTG) and uninduced cells were grown in LB at 37 °C to $OD_{600} \approx 0.4$ and infected with phages at various multiplicities. Growth of the cultures at 37 °C was continued for 2–3 h with intensive aeration.

To determine phage burst size, infections were carried out essentially as described in (Hyman *et al.*, 2009). Cultures grown to OD_{600} 0.6 were infected at MOI of 0.1. After 3–7 min at 37 °C to allow phage adsorption, cells were separated from unbound phage by centrifugation, resuspended in fresh medium with or without inducers and incubation at 37 °C was continued. At various times, aliquots were taken and the phage titer was determined.

3.18 Total DNA extraction

Induced or uninduced cultures were grown to OD_{600} of 0.6 and infected with a phage at MOI of 2. Cells were collected shortly before lysis, centrifuged in a Minispin centrifuge (Eppendorf, Germany) at 13,400 rpm, the supernatant was removed, and the

cell pellet was resuspended in 450 μ L of Tris buffer (10 mM Tris-HCl pH8.0). DNA was isolated by the phenol-chloroform method from the cell pellet as described (Pickard, 2009). The DNA was next digested with restriction endonucleases recognizing multiple sites in phage DNA (*BspHI* for cultures infected with phage λ , *SalI* and *MluI* for phage T5, *SspI* for phage T7) and analyzed by agarose electrophoresis.

3.19 CRISPR adaptation assay

Induced cells cultures ($OD_{600} = 0.6$) were infected at the MOI of 0.001-1.0 with wild-type or escape phages and grown for 16 h at 37 °C. Culture aliquots were withdrawn and cells were subjected to PCR with primers annealing to the CRISPR array leader sequence and to the phage specific spacer. Amplification reaction products were analyzed as described in “Induction of components of the CRISPR-Cas system and selection of clones that incorporate a new spacer”.

Expansion of the CRISPR array in cultures was also observed by the PCR method with primers Ec-LDR_F and M13_g8. Further, the amplified PCR products indicative of the presence of new spacers were separated in a 2% agarose gel, cut out from the gel with a scalpel and extracted as described in "Extraction of DNA fragments from agarose", The samples were collected and subjected to high throughput sequencing.

3.20 High throughput sequencing and data analysis

High throughput sequencing (HTS) data was prepared by Maria Logacheva at the Institute of Physico-Chemical Biology, MSU, Russia. PCR-products corresponding to the expanded CRISPR array were subjected to HTS using the MiSeq Illumina system. The resulting data were analyzed using ShortRead (Morgan *et al.*, 2009) and BioStrings

(Pages *et al.*, 2012) Bioconductor packages. Sequences located between two CRISPR repeats were considered as spacers. They were mapped onto phage genomes with no mismatches allowed. R scripts were used for statistical analysis and Circos (Krzywinski *et al.*, 2009) was used for graphical representation of the data. The results of the HTS analysis are presented in Table 3.

Table 3. HTS statistics.

Here are represented the amounts of spacers acquired during CRISPR adaptation calculated for different strands of DNA, as well as the number of spacers with canonical AAG PAM, which is typical for primed adaptation

*Upstream and downstream values for calculated for, respectively, $\frac{3}{4}$ of circular (λ or M13) viral genomes upstream (clockwise in Fig.3 or Fig.7B) of the priming protospacer and $\frac{1}{4}$ of the genome downstream (counterclockwise) of it.

λ adaptation statistics		
Phage derived spacers	112 987	87.1%
Upstream spacers from non-target strand	85 936	89.1%
Upstream spacers from target strand	10 470	10.9%
Downstream spacers from non-target strand	4 948	29.8%
Downstream spacers from target strand	11 633	70.2%
Spacers from non-target strand	90 884	79.3%
Spacers from target strand	22 103	20.7%
Spacers with AAG PAM	103 948	92%
T5 adaptation statistics		
Phage derived spacers	193 169	85.6%
Upstream spacers from non-target strand	172 684	99.5%
Upstream spacers from target strand	1 072	0.5%
Downstream spacers from non-target strand	975	5.4%
Downstream spacers from target strand	18 438	94.6%
Spacers from non-target strand	173 659	89.1%

Spacers from target strand	19 510	10.1%
Spacers with AAG PAM	192 590	99.7%
T7 genome coverage statistics in control infection		
Total reads count	4 441 188	100%
Reads mapped on T7 genome	3 399 953	76.56%
T7 genome coverage statistics in cells targeting early genes		
Total reads count	2 135 170	100%
Reads mapped on T7 genome	922 694	43.21%
T7 genome coverage statistics in cells targeting middle genes		
Total reads count	1 178 698	100%
Reads mapped on T7 genome	325 326	27.60%
T7 genome coverage statistics in cells targeting late genes		
Total reads count	4 478 266	100%
Reads mapped on T7 genome	2 865 244	63.98%
M13 adaptation statistics for M13_g8R targeting cells		
Phage derived spacers	159 616	77.6%
Upstream spacers from “+” strand	26 324	20.8%
Upstream* spacers from “-” strand	100 445	79.2%
Downstream spacers from “+” strand	25 906	78.7%
Downstream* spacers from “-” strand	6 941	21.1%
Total spacers from “+” strand	52 230	32.7%
Total spacers from “-” strand	107 386	67.3%
Spacers with AAG PAM	151 801	95.1%
M13 adaptation statistics for M13_g8 targeting cells		
Phage derived spacers	193 216	97.7%
Upstream* spacers from “+” strand	178 510	96.8%
Upstream spacers from “-” strand	5 761	3.1%

Downstream* spacers from “-” strand	5 915	65.5%
Downstream spacers from “+” strand	3 030	34.5%
Total spacers from “+” strand	184 425	95.5%
Total spacers from “-” strand	8 791	4.5%
Spacers with AAG PAM	152 655	79%

To sequence T7 phage genome in infected cells, total DNA was extracted from infected cells as described in “Total DNA extraction” several minutes before expected lysis time and treated with restriction endonucleases *BamHI*, *EcoRV*, *HindIII*, *PstI*, which do not recognize T7 DNA, in order to destroy host DNA. Samples were loaded onto 0.8% agarose gel, resolved by electrophoresis and a high-molecular weight band of phage genomic DNA was extracted from the gel using a GeneJET Extraction Kit (ThermoFisher Scientific, USA) and sequenced with a MiSeq Illumina system. Reads were trimmed and mapped on the phage genome with no mismatches allowed.

3.21 Fluorescence microscopy

These experiments were performed at the Nanobiotechnology research department of St. Petersburg Polytechnical University with assistance from Natalia Morozova using a Zeiss Axiolmager.Z1 upright microscope equipped with a custom-made incubation system to maintain cells at 37°C. A Semrock mCherry-40LP filter set was used for propidium iodide fluorescence detection. Images were collected using a Cascade II 1024 back-illuminated EM-CCD camera (Photometrics, USA). Autofocusing, multichannel and multifield time-lapse acquisition were implemented using MicroManager (Edelstein *et al.*, 2010) with custom scripts. Image analyses were performed using Fiji (ImageJ) (Schneider *et al.*, 2012).

Microscope chambers were prepared using the described procedure in order to provide time-lapse imaging (Morozova *et al.*, 2016). 1.5% agarose diluted in LB was used as a medium to monitor bacterial growth during microscopy. One microscope chamber allowed simultaneous cultivation and monitoring of cells with and without *cas* gene induction by using two or more separate agarose pads with and without inducers. Cells were mixed with phage and propidium iodide (“Molecular Probes”, USA), which was added to the final concentration of 20 μ M and placed in the microscope chamber (0.5 μ L on each pad). When cells had been absorbed onto the agarose pads the chamber was sealed and installed in the microscope. Typically 4–10 fields of view were visualized in a single experiment. Each field was imaged every 5–10 min during 1–4 h in a transmitted light channel and propidium iodide fluorescence channel.

Chapter 4. Results and discussion

4.1 Using primed adaptation to generate *E. coli* cells carrying CRISPR spacers targeting different phages

In order to study the CRISPR interference response directed to a specific mobile genetic element, a CRISPR array with a spacer targeting a protospacer in foreign DNA is required in addition to sufficient levels of *cas* gene expression. So the first part of this work was focused on developing a simple method of creating model *E. coli* strains for studying the CRISPR-Cas system response towards a variety of DNA targets. In practice, the experimentallist has to test multiple different spacers in order to observe strong interference, since the magnitude of CRISPR interference appears to vary from one spacer to another. The reasons for this spacer-to-spacer variation are not yet understood, but they may be connected not only with the strength of the crRNA-Cascade interaction with the target DNA but also with the location of the protospacer in the phage or plasmid genome. Until recently, specific “targeting” *E. coli* strains had to be created by genetic engineering of genomic CRISPR arrays using a modified version of the recombineering technique (Datsenko *et al.*, 2000). For various reasons (mostly having to do with the presence of multiple identical repeats in the cassette) the procedure for engineering expanded CRISPR arrays has proved technically challenging, time-consuming, and did not lend itself to multiplexing.

While studying the CRISPR adaptation process in *E. coli* the priming phenomenon was discovered in our laboratory (Datsenko *et al.*, 2012). When foreign DNA contained a mutated protospacer that rendered CRISPR interference inactive (due to a single-

nucleotide substitution in the PAM or the seed region (Semenova *et al.*, 2011)), very efficient acquisition of additional spacers derived from DNA located in *cis* with respect to the mutated protospacer was observed (Datsenko *et al.*, 2012). The molecular mechanism of the priming phenomenon is not yet fully understood; however, it is clear that it must be biologically significant, because it provides molecular “memory” of prior encounters with foreign DNA that allows specific adaptive response to targets that managed to escape initial interference.

It was believed that the phenomenon of priming could be observed only in response to the entry into the bacterium of a foreign DNA with a mutant protospacer (Figure 4D). The following experiments provide the data that DNA target with fully matching protospacer degrades faster than a target with mutated protospacer, but new spacers are acquired mostly in the second case. To assess the rate of CRISPR interference, ampicillin-resistant transformants were grown in the absence of ampicillin until $OD_{600} = 0.5$, after which *cas* gene expression was induced. Induced cultures grew at the same rate as uninduced controls. Cells transformed with pG8 rapidly lost ampicillin resistance upon induction (Figure 4C), whereas control cells transformed with pT7Blue remained antibiotic-resistant (Figure 4B). Induced cultures of cells transformed with pG8mut remained antibiotic-resistant for most of the experiment, with some loss becoming apparent at later times (Figure 4D).

Plasmid loss was also monitored directly, by purifying plasmids from aliquots of induced cultures and electrophoretic analysis. The results showed that the yield of pT7Blue increased along with culture density (Figure 4B). The pG8 plasmid was

effectively purged from the culture 1 hour postinduction (Figure 4C). The amount of pG8mut plasmid was close to that in control pT7Blue harboring cultures 1 hour postinduction and then slowly decreased (Figure 4D). Thus, a mismatch between the g8 crRNA spacer and the C1T protospacer does not completely inactivate CRISPR interference.

Aliquots of induced cultures were also subjected to PCR amplification to monitor the CRISPR array expansion. In cultures transformed with pT7Blue or pG8, only a PCR product corresponding to the unexpanded CRISPR array was observed (Figure 4B and C). In contrast, in cultures of cells transformed with pG8mut, time-dependent accumulation of longer PCR product corresponding to an insertion of an additional repeat-spacer unit in the array was observed (Figure 4D). Later it was also proven that presence of a fully matching as well as partially matching protospacer lead to efficient primed adaptation (Semenova *et al.*, 2016).

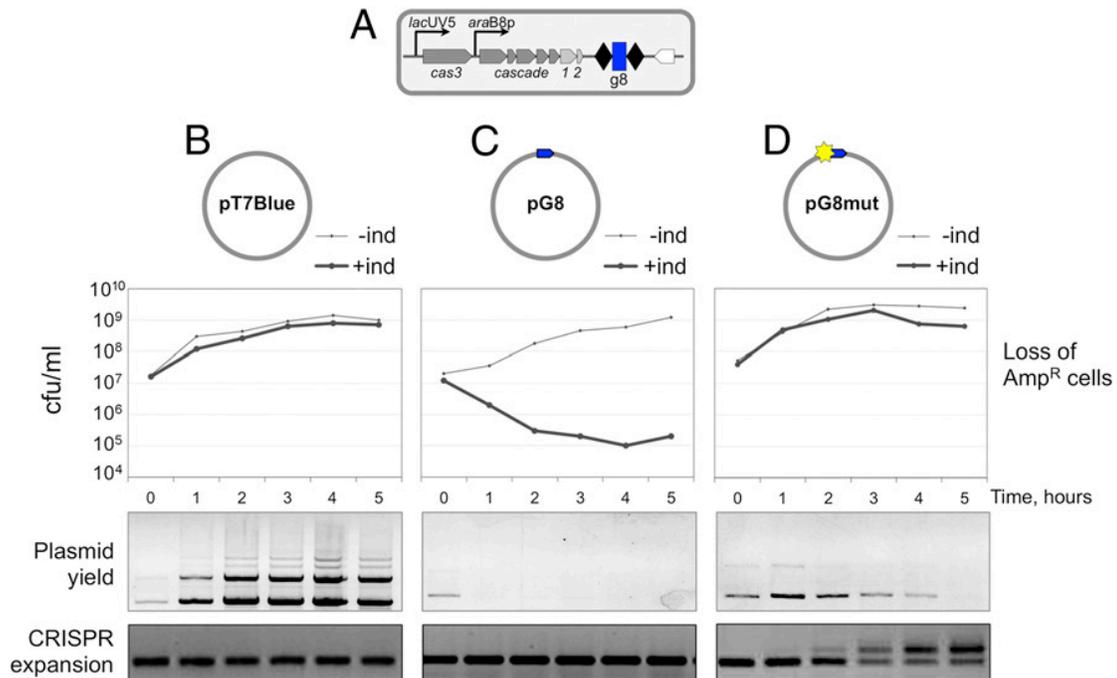


Figure 4. CRISPR interference and primed CRISPR adaptation against matching and partially matching protospacer targets located on plasmids. (A) The *E. coli* KD263 cells capable of inducible *cas* gene expression and carrying an engineered CRISPR array with a single *g8* spacer are schematically shown. *E. coli* KD263 was transformed with ampicillin-resistant pT7Blue vector (B) or pT7Blue-based plasmid pG8 (C) with a fully matching *g8* protospacer with a functional PAM or pG8mut (D), carrying, respectively, a protospacer with a functional PAM and a C1T mutation in the protospacer. Protospacers are shown as blue arrows with the arrowpoint directed away from the 5'-ATG-3' PAM. The C1T mutation is shown as a yellow star. Graphs show the number of ampicillin-resistant cfu in cultures transformed with each plasmid with or without induction of *cas* gene expression. Plasmid DNA was purified from cells collected at the times indicated and resolved by agarose gel electrophoresis. The products of PCR amplification with primers annealing at the *g8* spacer and further upstream in the CRISPR array leader are shown at the bottom.

We used the priming phenomenon for easy construction of strains targeting various regions of phage and/or plasmid DNA for studies of CRISPR-Cas function directed against various mobile genetic elements. The priming-mediated strain construction method that we developed involves the creation of pT7Blue-based plasmids

containing a protospacer corresponding to the g8 CRISPR spacer but bearing a mismatch at position +1 (an escape) and cloned fragments of different genomic regions of *E. coli* bacteriophages λ , T5, T7, T4 and R1-37, a giant phage infecting *Yersinia pseudotuberculosis* (Skurnik *et al.*, 2012). Since cloned phage DNA fragments were between 2 and 4 kbp in length, which is comparable with ~3 kbp length of the pT7Blue vector, after growth in conditions of *cas* genes expression approximately half of ampicillin-sensitive clones with expanded CRISPR arrays acquired spacers from phage DNA inserts. After sequencing, clones that acquired spacers from the DNA of interest were identified and used for downstream applications. Strains harboring phage-specific spacers in our collection are listed in Table 1.

We have successfully used the procedure described above to obtain numerous *E. coli* strains containing CRISPR spacers targeting multiple bacteriophages as explained in the following sections.

4.2 Bacteriophage M13

In previous studies the single-stranded M13 phage was extensively used as a model object to study defensive function of type I-E *E. coli* CRISPR-Cas system (Semenova *et al.*, 2011; Datsenko *et al.*, 2012). M13 infected cells do not lyse but preserve the replicative dsDNA form of the virus and continue to grow while continuously producing viral progeny. During infection by this phage, escape viruses that avoid CRISPR-interference are selected easily, to which cells respond by acquiring new spacers. As a result, the viral population dies (Figure 5), presumably because it is unable

to bare the mutational load of multiple escape mutations for each of the numerous acquired spacers.

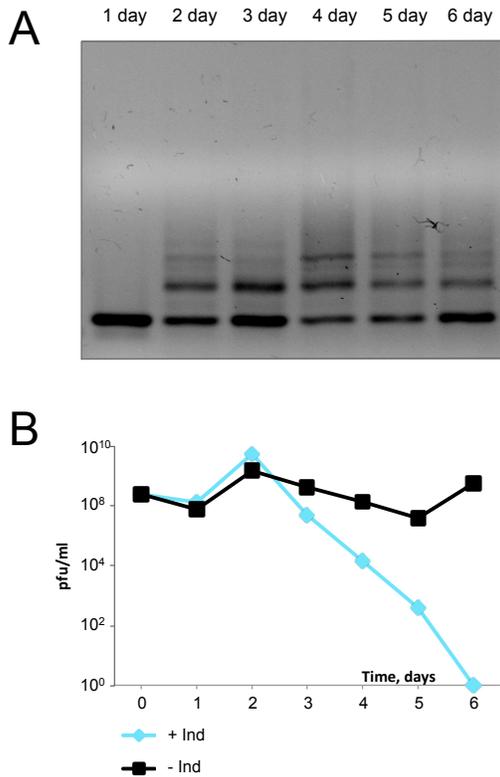


Figure 5. Long-term cultivation of KD263 *E. coli* cells with crRNA targeting M13 phage in the presence of the phage. (A) CRISPR array expansion in the course of several days of cultivation of induced KD263 culture is shown. (B) The dynamics of M13 plaque forming units (pfu) during the infection of induced ('+Ind', blue lines) or uninduced ('-Ind', black lines) KD263 cultures cultivated over several days.

Primed CRISPR adaptation for M13 infected cells was extensively analyzed previously (Datsenko *et al.*, 2012). A strand bias but no gradient of spacer acquisition efficiency as a function of distance from the priming site was revealed. The lack of observable gradients is likely caused by the smaller sizes of these circular genomes. This results in extensive overlaps of upstream and downstream spacers selection gradients and insufficient depth of sequencing of newly acquired spacers. In our study, spacers acquired

during the M13 phage infection was subjected to HTS and the results of their analysis are presented in Figure 6, which reveals gradients of spacer efficiency acquisition and change in the direction of strand bias that were undetected previously (Datsenko *et al.*, 2012).

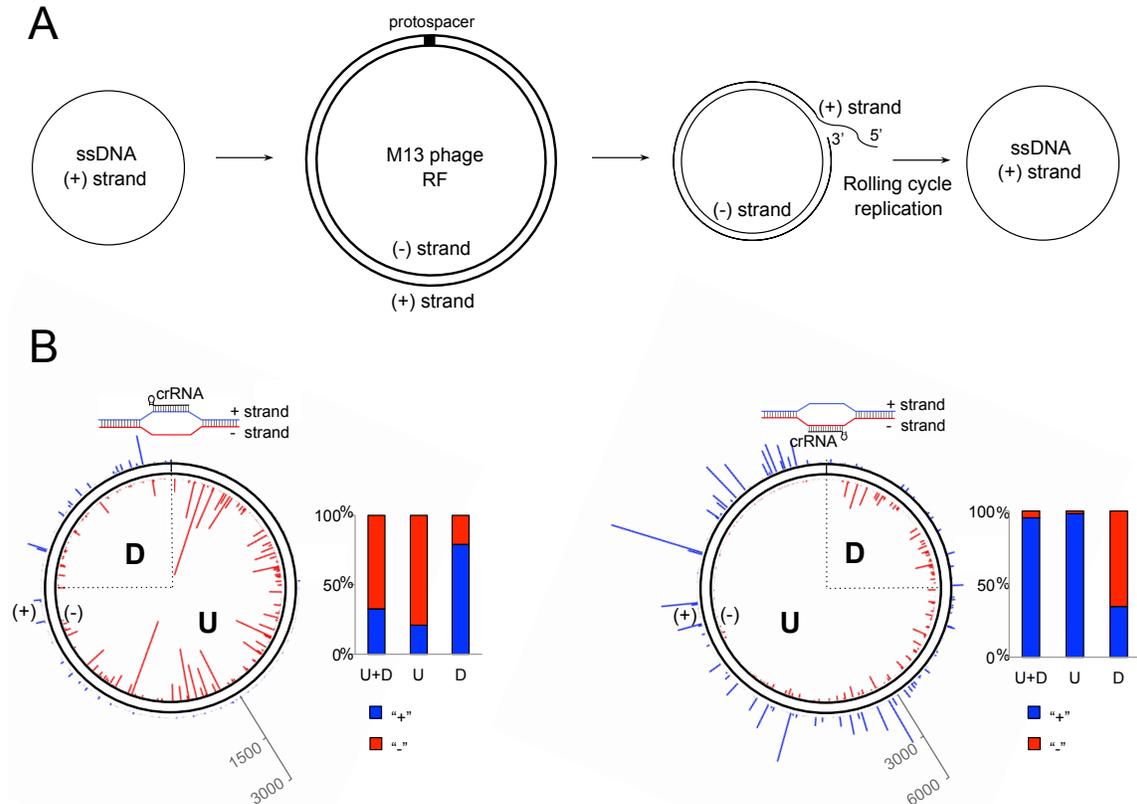


Figure 6. Gradients of spacers acquired during primed adaptation by M13 infected cells. (A) The life cycle of bacteriophage M13 is schematically shown. The location of priming protospacer is indicated on the double-stranded replicative form (RF) of the genome. (B) Graphical representation of HTS analysis of spacers acquired during the infection of M13_g8 (right) and M13_g8R (left) cells infected by bacteriophage M13. The two strains target different strands of M13 RF. The priming protospacers are indicated at the top and a structure of the R-loop formed by each crRNA is shown. Acquired spacers are mapped on the circular viral genome (scale bars indicate numbers of Illumina reads). Red and blue lines show spacers matching, respectively, the - and + strands of phage genome. Line heights indicate relative frequency of reads corresponding to different spacers. Rectangular bars show the proportion of spacers matching the - (red) and + (blue) strands for the entire genome (“U+D”); in the downstream (“D”) quarter of the viral genome, and in the remaining upstream part (“U”).

4.3 Bacteriophage λ

Fifteen *E. coli* strains targeting λ early or late genes (Table 1) were tested for susceptibility to infection with λ_{vir} , a phage λ mutant that is only capable of lytic cycle. The number of plaques on induced (CRISPR-Cas system ON) and uninduced (CRISPR-Cas system OFF) cell lawns were recorded and the efficiency of plaquing (EOP) was calculated as the ratio of the two numbers. All strains interfered strongly (10^{-5} - 10^{-9}) with λ_{vir} infection in the plaque-forming assay (Table 1 and Figure 7B). For seven strains, no phage plaques were observed, so the calculated EOP of 10^{-9} presents a lower estimate of CRISPR interference efficiency. Phages that formed plaques on remaining strains contained escape mutations, which included: point substitutions in the PAM or the seed of the targeted protospacer; a 15-bp deletion in the protospacer (the affected protospacer is in an intergenic region of the phage); and a double mutation outside the seed region (positions +20/+22 and +25/+28) but with intact PAM/seed segments, or single nucleotide deletions in positions 23 and 20 (Table 1).

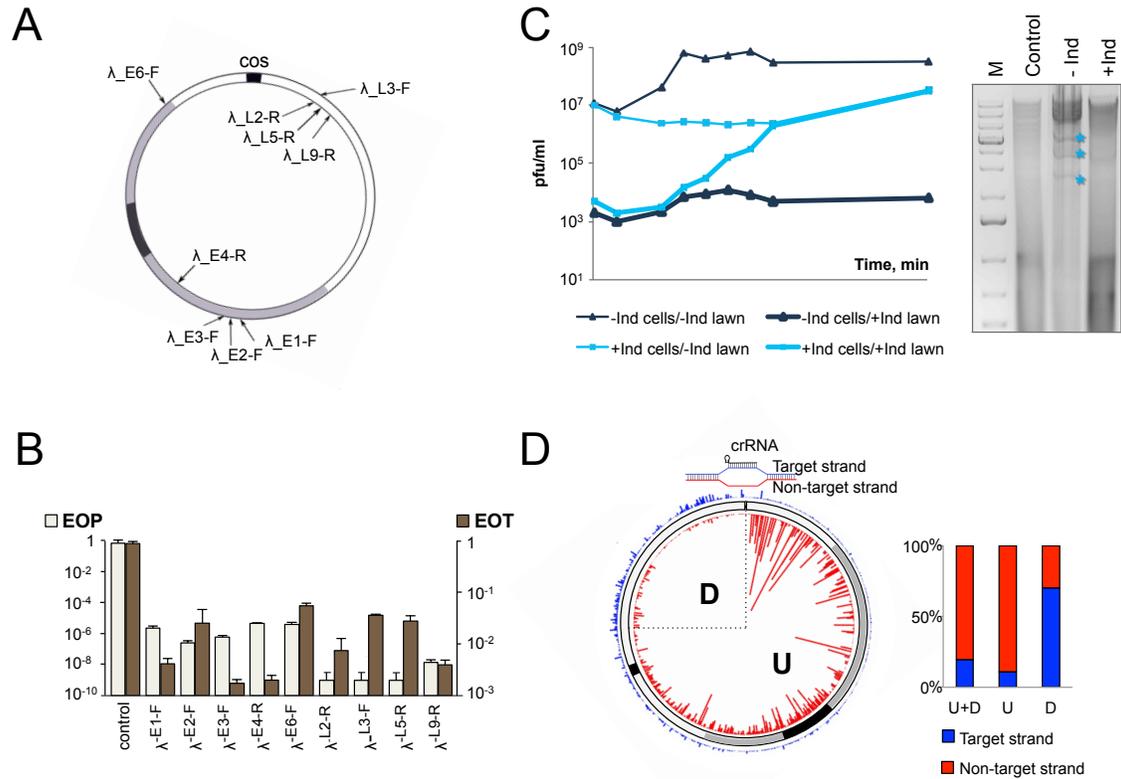


Figure 7. The effect of CRISPR–Cas targeting on λ vir infection. (A) The circular genome of bacteriophage λ formed after annealing at the *cos* sites is shown. The region containing immediate early genes is colored black; the region containing delayed early genes is gray; and the late gene region is white. Black arrows indicate the positions of protospacers targeted by crRNA spacers from different strains. Protospacers located at different strands of the genome are shown by arrows located inside and outside the circle representing the phage genome. (B) The efficiency of plaque formation (EOP) by λ_{vir} and efficiency of transformation (EOT) by cognate protospacer-containing plasmids into indicated strains. Mean values and standard deviations from three independent experiments with each strain are presented. EOP values for strains targeting additional protospacers can be found in Table 1. (C) On the left, the dynamics of plaque forming units (pfu) during the infection of induced (‘+Ind’, blue lines) or uninduced (‘–Ind’, black lines) λ -L9-R culture are shown. The phage from each infection was plated on lawns of uninduced (thin lines) and induced (thick lines) λ -L9-R cells to determine the total number of infectious phages and the number of escape phages, respectively. On the right, an agarose gel electrophoresis showing the products of *BspHI* digestion of DNA prepared from induced (‘+Ind’) and uninduced (‘–Ind’) λ -L9-R cells infected with λ_{vir} and collected 60 min post-infection is presented. ‘M’ is a molecular weight marker lane. The lane labeled ‘Control’ shows DNA from uninfected cells. Asterisks indicate restriction

fragments of phage DNA. (D) Graphical representation of HTS analysis of spacers acquired during the infection of λ -L9-R cells by a λ_{vir} G-1T escape mutant. The position of the priming protospacer is indicated at the top and a structure of the R-loop formed by crRNA at this protospacer is shown. Acquired spacers are mapped on the circular viral genome. Red and blue lines show spacers matching, respectively, non-target strands and target strands. Line heights indicate relative frequency of reads corresponding to different spacers. Rectangular bars show the proportion of spacers matching the non-target (red) and target (blue) strands for the entire genome ('U+D'); in the downstream ('D') quarter of the viral genome; and in the remaining upstream part ('U').

The efficiency of CRISPR interference could be affected by spacer-protospacer sequences. In order to account for such effects, competent cells from induced and uninduced targeting strains targeting the phage protospacers shown in Figure 7A were prepared, transformed with plasmids containing matching phage protospacers, and the efficiency of transformation was determined by calculating the ratio of numbers of transformants formed by induced and uninduced cells. EOT values for the tested λ targeting strains varied by as much as 2–3 orders of magnitude (Figure 7B), suggesting that the sequences of crRNA spacer and cognate protospacer strongly affect CRISPR interference. There was no correlation between the EOP and EOT values (Figure 7B).

λ_{vir} infected liquid cultures lysed poorly in our conditions, even when infected at high multiplicity of infection, since the phage receptor requires glucose for maximal expression (Schwartz, 1976), while *cas* gene expression requires arabinose and the absence of glucose. It was not therefore possible to monitor the effect of CRISPR interference on the growth of infected λ targeting strains. The total amount of infectious phage particles in infected liquid cultures was determined by counting plaques formed on lawns of uninduced cells. The amount of escape phages in infected cultures was determined by plating on induced cell lawns. A representative result is shown in Figure

7C. In the absence of *cas* gene induction the phage titer in infected cultures increased about 100-fold ~80 min post-infection (thin black line). In the presence of an active CRISPR-Cas system (thin blue line) the increase in phage titer was delayed to ~160 min. The late increase in phage titer was due to the accumulation of escape phages, which were rare (10^{-4}) at the start of infection but completely overtook the population at late stages (Figure 7C, left, thick blue line). There was no such increase of escape mutants in the course of infection of uninduced cells (thick black line).

To monitor accumulation of phage DNA in infected cells, total DNA was extracted from infected cells, digested with restriction endonuclease *BspHI* and examined using agarose electrophoresis. A representative experiment showing accumulation of λ DNA in the presence and in the absence of *cas* gene expression inducers is shown on the right of Figure 7C. Distinct bands corresponding to fragments of λ genome were clearly visible on the smeary background of host DNA in the absence of *cas* genes expression. In contrast, only cellular DNA was present when *cas* genes were induced. So CRISPR interference prevents accumulation of replicated λ_{vir} DNA in infected cells.

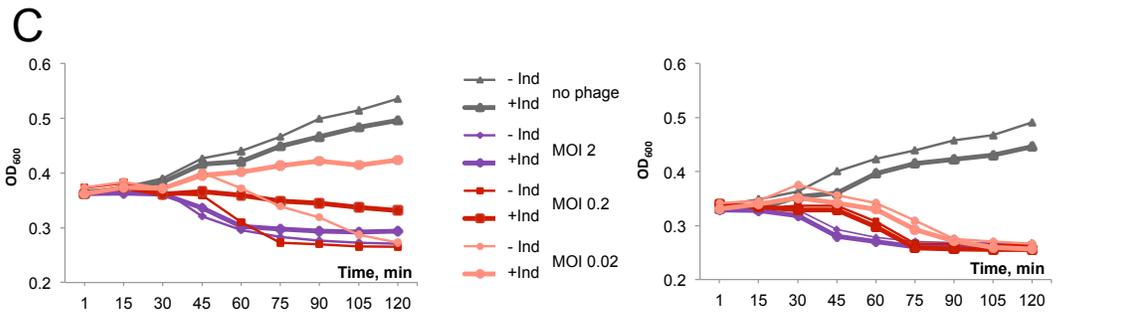
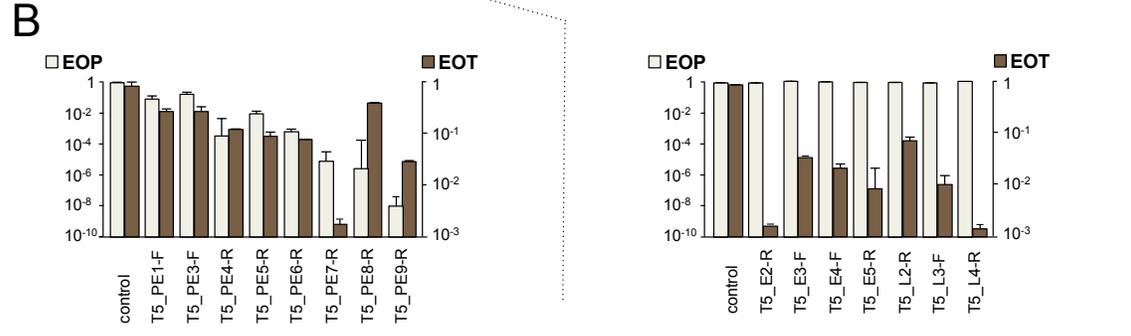
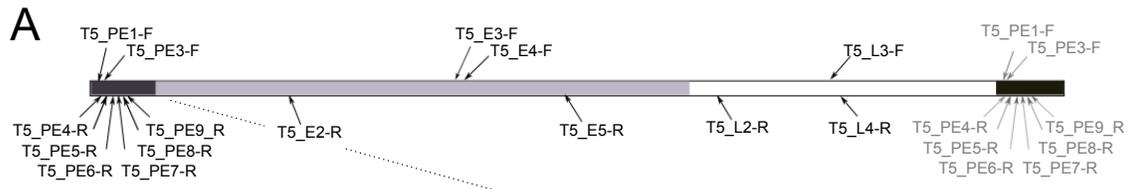
CRISPR array expansion was detected in the course of infection of λ targeting cells with escape phages. Several bacterial clones that survived the infection of λ -L9-R strain with the G-1T escape phage carrying a G to T substitution in the PAM of the λ -L9-R protospacer and harboring an expanded CRISPR array were isolated. These clones were resistant to both wild-type and G-1T phage infection in conditions of *cas* genes induction. So newly acquired λ -derived spacers are competent for interference. PCR amplified material corresponding to the expanded λ -L9-R array from the infected culture was

subjected to Illumina high-throughput sequencing (HTS) and spacers were extracted from sequencing reads and mapped on the circular phage genome (Figure 7D). It was found that 92% of new spacers were acquired from protospacers associated with canonical AAG PAM and there was an overall 79% bias towards the selection of spacers from the non-targeted strand (Table 3). These values are typical for primed adaptation (Savitskaya *et al.*, 2013). However, the extent and even the direction of strand bias was not uniform throughout the viral genome (Figure 7D, Table 3). 89% of spacers acquired from the upstream (clockwise in Figure 7D) three quarters of the genome matched protospacers in the non-targeted strand. In the downstream (counterclockwise) quarter of the phage genome the direction of strand bias was reversed and most (~70%) of spacers matched protospacer in the opposite, targeted strand (Figure 7D). Efficiency of spacer selection decreased as the distance from the priming protospacer increased and overall spacer acquisition was most efficient close to and upstream of the priming protospacer (the bias towards the non-targeted strand was also the strongest in this area, approaching 99%).

4.4 Bacteriophage T5

Bacteriophage T5 is a strictly lytic virus with a linear ~121,300 bp genome containing terminal 10,160 bp direct repeats (Figure 8A). The proximal repeat carries pre-early genes that are injected into the host cell at the first stage of infection followed by a pause to allow the synthesis of pre-early proteins that inactivate several essential cellular processes and destroy host DNA (Calendar, Abedon 2006; Davison, 2015). The remaining phage DNA, from which early and late genes are expressed, is then injected

into the infected cell. Strains targeting unique protospacers in pre-early, early and late T5 genes are listed in Table 1 and some are shown in Figure 5A. In the plaque formation assay, induced strains carrying spacers targeting late or early T5 genes were infected by the phage as efficiently as uninduced or the parental non-targeting strain (Figure 8B, right and Table 1). The morphology of phage plaques was also unaffected. In contrast, all strains targeting pre-early genes exhibited some level of interference with the infection (Figure 8B, left and Table 1). The EOP values ranged from 10^{-1} for T5-PE3-F to $<10^{-8}$ for T5-PE9-R strain (Figure 8B, left; in the latter case, no phage plaques was observed, so the given interference efficiency value is a lower estimate). Phages were recovered from several individual randomly picked plaques formed on lawns of induced strains targeting the pre-early region and were analyzed further. Phages recovered from strains that gave low (10^{-1} - 10^{-2}) levels of interference continued to be restricted at the same level upon reinfection and, when sequenced through the protospacer region, were found to contain a wild-type sequence. Phages recovered from strains that efficiently interfered with infection (10^{-4} - 10^{-6}) demonstrated an EOP of one when re-plated on the same strain expressing *cas* genes and therefore behaved as escape mutants. Indeed, such a phage contained mutations in the PAM or seed (positions +1 to +8) segments of targeted protospacers (Table 1).



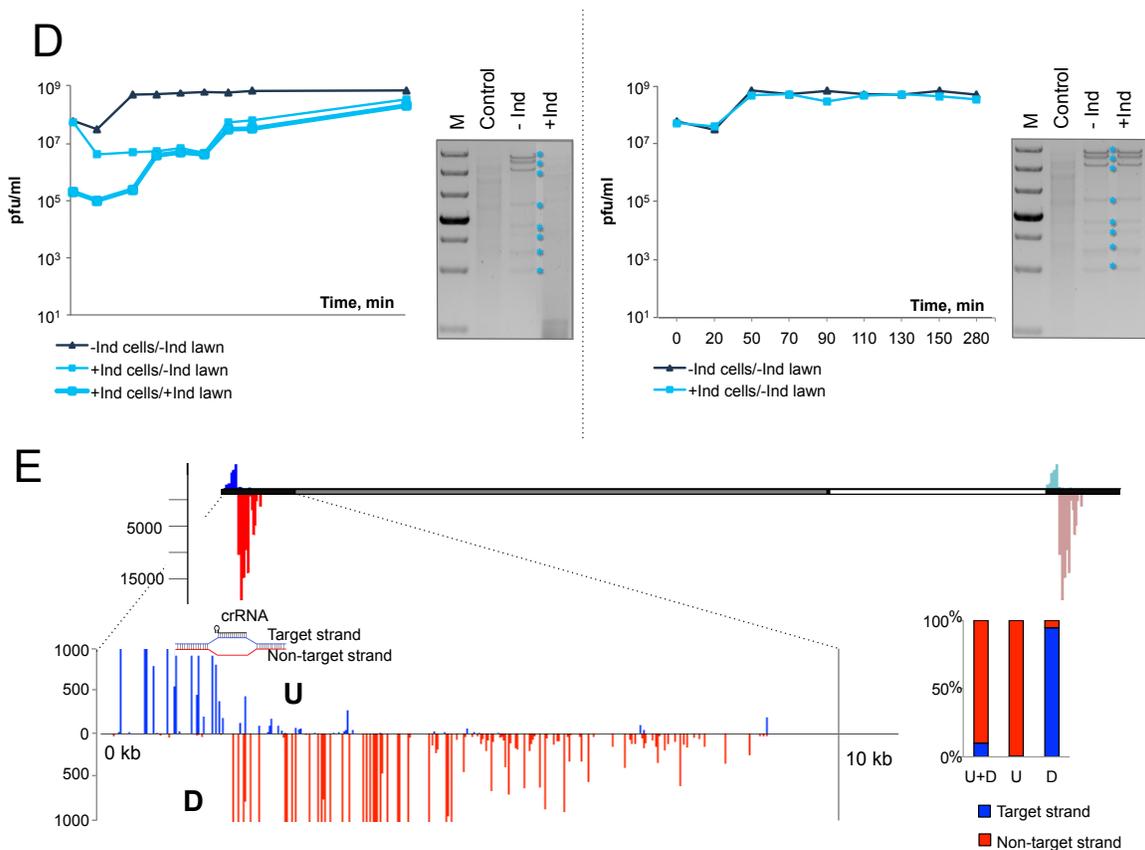


Figure 8. The effect of CRISPR-Cas targeting on T5 infection. (A) The linear genome of bacteriophage T5 and positions of protospacers targeted by crRNA spacers of strains from our collection are schematically shown. Pre-early genes are colored black, the region containing early genes is gray, and the late gene region is white. The T5 genome is terminally redundant with the entire pre-early area present at either end of the genome. Spacers targeting the pre-early region thus have two matching sites. Protospacers at the right end are shown as grey arrows. Protospacers originating from different strands of the genome are shown above or below the genome scheme. (B) EOP by T5 and EOT by cognate protospacer-containing plasmids into indicated strains. Results obtained with pre-early targeting strains are shown on the left. Results obtained with early and late targeting strains are on the right. Mean values and standard deviations from three independent experiments with each strain are presented. EOP values for strains targeting additional protospacers can be found in Table 1. (C) Growth curves of induced ('+Ind') and uninduced ('-Ind') cultures of T5-PE-6-R (left) and T5-E4-F (right) strains infected with T5 at indicated MOIs. (D) Phage progeny production during the infection of T5-PE-6-R (left) or T5-E4-F (right) cultures. See Figure 7C legend for details. At the right of each diagram with growth curves, agarose gels are presented showing the products of *SaI* and *MluI* digestion of DNA prepared from cells from an

uninfected culture, or induced and uninduced infected (MOI = 2) cultures collected 45 min post-infection. Phage DNA restriction fragments are marked by asterisks. (E) Graphic representation of HTS results of spacers acquired by the T5-PE-6-R cultures infected with the A2G escape T5 phage. Red vertical lines indicate spacers matching the non-target strand, blue lines indicate spacers matching the target strand. Line heights indicate relative frequency of reads corresponding to each acquired spacer. The scale bar at the left shows numbers of Illumina reads. The pre-early region at the left end of the genome is expanded below to show mapping results on a larger scale. The position of the priming protospacer is indicated. A vertical black line shows the position up to which phage DNA is initially inserted into the infected cell (Davison, 2015). Rectangular bars show the proportion of spacers matching non-target (red) and target (blue) strands upstream ('U') and downstream ('D') of the priming protospacer and the total strand bias of acquired spacers ('U+D').

Plasmid transformation experiments indicated that CRISPR-Cas was able to recognize and interfere with target protospacers from pre-early, early, and late T5 genes (Figure 8B), suggesting that genomic location is responsible for lack of protection from phage infection of cells with early and late-gene targeting spacers.

Cultures of cells targeting early or late genes behaved indistinguishably from uninduced cultures and were lysed by the phage at all MOIs tested (Figure 8C, right). Induced cultures targeting pre-early genes infected at low MOI of ~ 0.02 were able to partially withstand the infection, with cell density rising slightly over the course of the experiment (Figure 8C, left, thick pink line). At the intermediate MOI of ~ 0.2 the optical density of induced cultures declined slowly (Figure 8C, left, thick red line). Induced cultures infected at the MOI of ~ 2 lysed as fast as the uninduced control (Figure 8C compare thick and thin purple lines).

In single-burst experiments, no difference in phage progeny yield or eclipse time was observed between induced and uninduced cells targeting early or late T5

genes (Figure 8D, right). For pre-early targeting cells the increase in phage titer in infected cultures was delayed and was due to accumulation of escape phages (Figure 8D, left).

Total DNA was extracted from infected cells and digested with *SalI* and *MluI* restriction endonucleases. No changes in the amounts or profile of phage DNA restriction fragments was observed when DNA purified from cells bearing spacers against early or late T5 genes was compared with DNA prepared from uninduced infected cells and host DNA was degraded (Figure 8D, right). In the case of pre-early targeting no viral DNA was present. Host DNA was also absent, apparently degraded by phage pre-early proteins (Figure 8D, left).

Individual T5 phage isolates that behaved as escape mutants were tested for ability to cause primed adaptation during infection of corresponding targeting cells. CRISPR array expansion was observed only after infection of induced pre-early targeting cells. Several bacterial clones that survived the infection of the T5-PE6-R culture with an A2G escape phage bearing an A to G substitution in the second position of the targeted protospacer and that expanded their CRISPR array were isolated. All clones were resistant to both the wild-type and the A2G phage in conditions of cas gene induction. So newly acquired T5-derived spacers are competent for interference.

PCR amplified fragments corresponding to expanded CRISPR arrays in the T5-PE6-R culture infected with the A2G phage were subjected to Illumina sequencing. The results of mapping of acquired spacers are presented in Figure 8E. Almost all (99.7%) of spacers were acquired from proto-spacers associated with canonical AAG PAM (Table 3) and an overall 89% bias towards acquisition of spacers from the non-targeted strand was observed (Figure 8E, Table 3). More than 99% of spacers matching protospacers located upstream of the priming site had the same orientation as the priming protospacer. Conversely, 95% of spacers that originated downstream had the opposite orientation. The efficiency of spacer acquisition declined as the distance from the priming site increased in both upstream and downstream directions, resulting in a gradient of protospacer use.

Our experiments do not directly address the question whether effector complexes that preexist in infected cells are present in infected cells, but even if they do, they clearly cannot prevent phage progeny production. The strategy used by T5 and its relatives thus appears to be an effective anti-CRISPR mechanism, with less than 10% of the genome being susceptible to CRISPR interference and only for a short period of time.

Analysis of spacers acquired during primed adaptation by cells targeting the pre-early region of T5 reveals a gradient of spacer acquisition efficiency and a switch in the direction of strand bias of acquired spacers upstream and downstream of the priming site. Most acquired spacers match protospacers in the non-targeted strand upstream of the priming protospacer and target strand protospacers in the downstream direction. Since

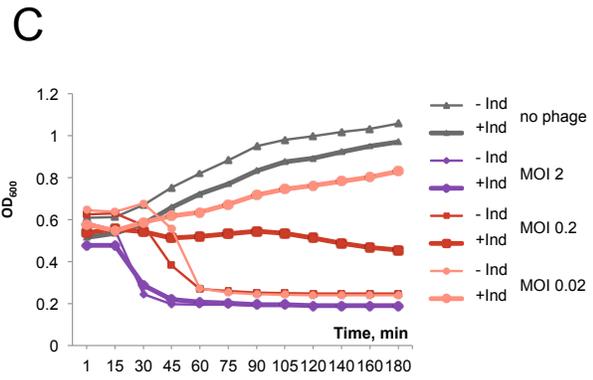
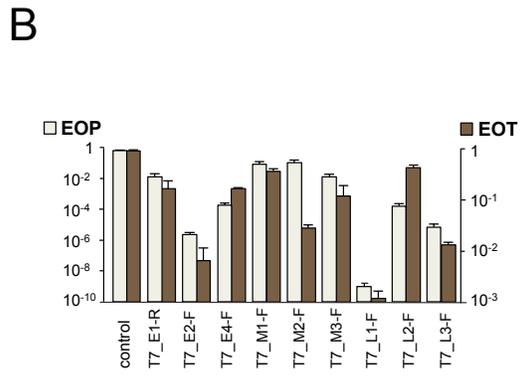
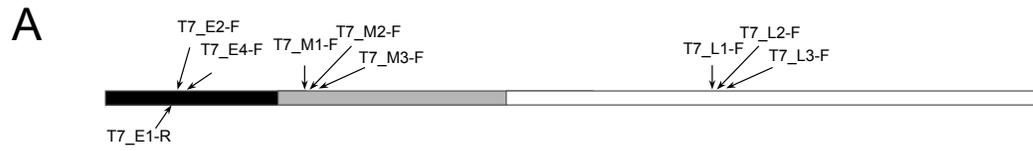
there are ~10 times more spacers acquired from the upstream region, there is a ~90% bias towards the non-targeted strand when all spacers are considered.

Spacers acquired from the circular λ genome exhibit an 80% overall bias towards the non-targeted strand. Unlike the situation with T5, a pronounced gradient of spacer acquisition efficiency is observed only upstream of the priming site and the strand bias disappears as the distance from the priming site is increased in the upstream direction (Figure 7D). This behavior is probably due to an ‘overlap’ of spacers acquired from different strands of the circular genome coupled with overall lower efficiency of spacer selection downstream of the priming site. The difference in overall efficiency of spacer acquisition upstream and downstream of priming sites revealed in our work is consistent with the preferential cleavage by the *E. coli* Cas3 of the non-targeted strand in the R-loop complex and its 3'-, 5'- helicase activity (Mulepati *et al.*, 2013). It is reasonable to suppose that Cas3 processively unwinds the priming protospacer and, after the initial cleavage of its single-stranded non-target strand, moves upstream (in the 3'-, 5'- direction with respect to the non-target strand) generating substrates that are channeled for integration into the CRISPR array by Cas1 and Cas2. The strand bias in spacer acquisition would be consistent with recent in vitro data showing that Cas3 generated fragments that fuel primed adaption are partially single-stranded (Kunne *et al.*, 2016). The gradient in spacer acquisition efficiency must be related to either the processivity of Cas3 nuclease or the life times of Cas3 generated DNA fragments. Acquisition of spacers downstream from the priming site must be caused by Cas3 induced cleavage of the target strand followed by its subsequent degradation. In vitro, target strand cleavage proceeds

with lower efficiency (Mulepati *et al.*, 2013), which can explain the smaller number of spacers acquired in the downstream direction.

4.5 Bacteriophage T7

The linear genome of bacteriophage T7 gradually enters the host cell starting from the left (early) end in a process that is coupled to transcription of early T7 genes by the host RNA polymerase (Calendar, Abedon 2006). Our collection contained 11 strains targeting four protospacers in T7 early genes, four in middle genes and three in late genes (Figure 9A, Table 1). All strains interfered with T7 infection, decreasing EOP between one to at least nine orders of magnitude (Figure 9B). Each strain also interfered with plasmid transformation (Figure 9B). The EOT and EOP efficiencies for different spacers appeared to be correlated, suggesting that the contribution of genomic context to CRISPR interference against phage T7 infection is not significant.



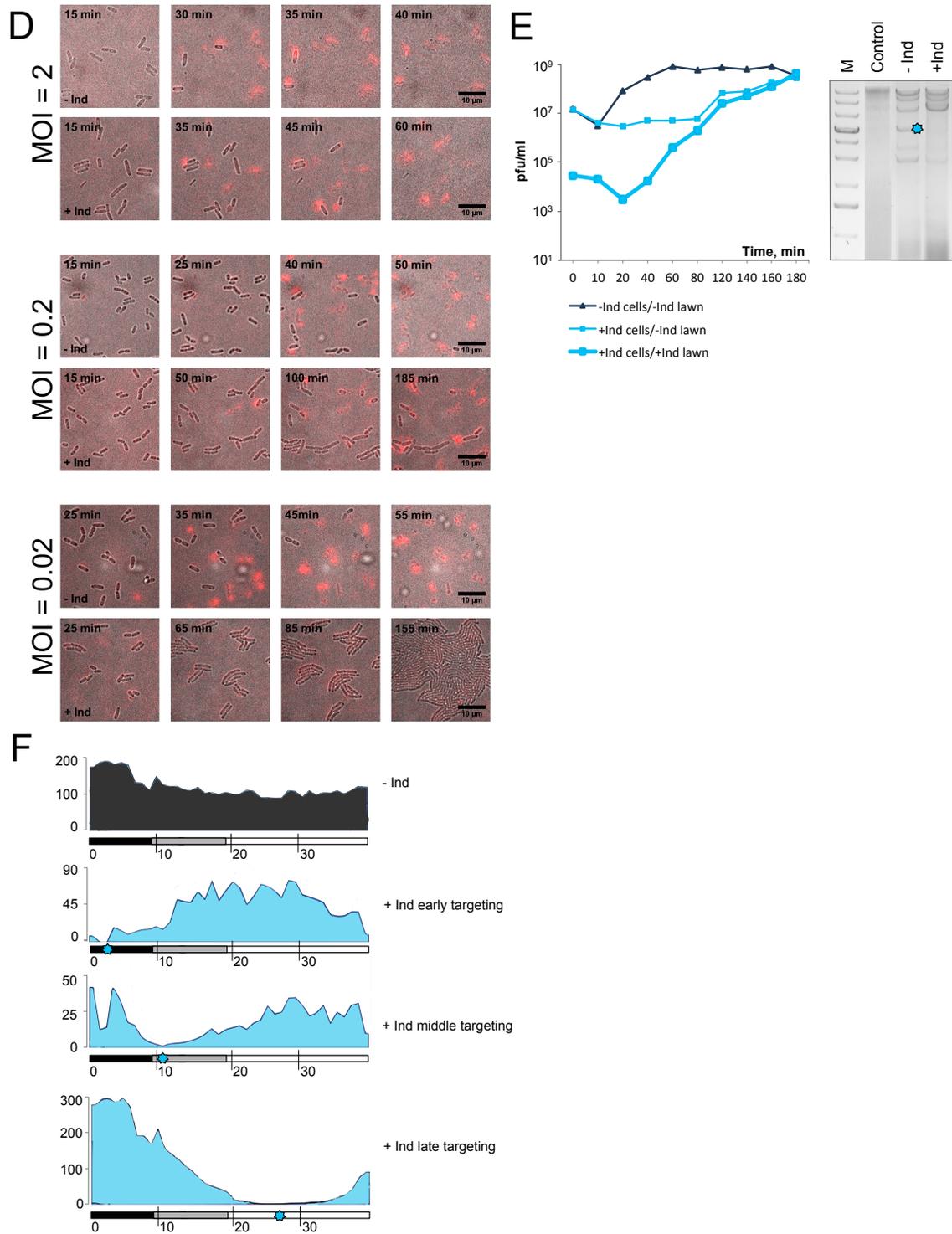


Figure 9. The effect of CRISPR-Cas targeting on T7 infection. (A) The linear genome of bacteriophage T7 and positions of protospacers targeted by crRNA spacers of

strains from our collection are schematically shown. The early genes region is colored black, the region containing middle genes is gray and the late gene region is white. The area that contains both early and middle genes is striped. (B) EOP by T7 and EOT by cognate protospacer-containing plasmids into indicated strains. Mean values and standard deviations from three independent experiments with each strain are shown. EOP values for strains targeting additional protospacers can be found in Table 1. (C) Growth curves of induced and uninduced cultures of T7-E2-F infected with T7 at indicated MOIs. (D) Images of induced and uninduced T7-E2-F cultures infected with T7 at different MOIs. Images were taken at the times indicated. Depolarized or lysed cells appear red in the images. For each MOI, images in the top row show the progress of infection in the absence of *cas* gene expression; images in the lower row show infection of induced cells. (E) On the left, phage progeny production during the infection of induced and uninduced T7-L2-F cells. See Figure 7C legend for details. On the right, an agarose gel showing the products of *SspI* digestion of DNA prepared from uninfected or infected (MOI = 2) induced and uninduced T7-L2-F cells collected 18 min post-infection. (F) Graphical representation of mapping of Illumina reads from total DNA prepared from T7-infected uninduced (top) or induced (bottom) cells targeting early (T7-E2-F), middle (T7-M1-F) and late (T7-L2-F) phage genes onto the T7 genome. Scale bars at the left show numbers of Illumina reads. The positions of protospacers targeted by T7-E2-F, T7-M1-F or T7-L2-F crRNAs is shown at the bottom with blue asterisks.

Escape mutants were recovered for 8 strains targeting all temporal classes of T7 genes. In addition to point escape mutants changing the PAM of the protospacer seed, a deletion removing 59 nucleotides from early gene 0.7 was obtained. *gp0.7* codes for a protein kinase that is dispensable for growth in laboratory conditions (Nguyen *et al.*, 2014), so the deletion did not interfere with phage viability. Another escape mutant contained a GGT insertion between positions 24 and 25 of protospacer T7-E4-F (Table 1). This protospacer is part of gene 1, which codes for T7 RNA polymerase. The insertion occurred between glycine-coding codons GGT and GGC. The mutation may have arisen due to slippage during replication of a G-rich stretch of phage genome, resulting in insertion of an additional glycine residue in phage RNA polymerase.

When growth of cultures infected at different MOIs was monitored, similar results were obtained with cells targeting any one of the three gene expression classes.

Representative results obtained with an early gene targeting T7-E2-F strain are shown in Figure 9C. In the absence of *cas* gene lysis of induction cells occurred shortly after the addition of the phage (MOI of 2) or after a delay (MOIs of 0.2 and 0.002) needed for progeny phage to appear, infect and lyse remaining uninfected cells. At high MOI, induced cultures lysed as efficiently as uninduced control cultures, indicating that the phage was able to overrun the defensive action of CRISPR-Cas. Induced cultures infected at MOI of 0.2 maintained constant OD₆₀₀ at first and then started to gradually decline. Induced cultures infected at MOI of 0.02 withstood the infection and grew almost as well as the uninfected control.

Cells in infected T7-E2-F cultures were monitored by live microscopy in the presence of propidium iodide, which selectively stains cells with depolarized membrane. Membrane depolarization occurs shortly before cell lysis by the phage (Calendar, Abedon 2006; Young, 1992). The results showed that all cells in cultures infected at MOI of 2 were depolarized and lysed irrespective of whether *cas* gene expression was induced or not (Figure 9D). At the intermediate MOI of 0.2 cells in uninduced cultures lysed completely, while in induced cultures a dynamic behavior including lysis of some cells and continued growth of others was observed during the course of the experiment. At MOI of 0.02 some cells lysed between 20 and 60 min, but later no lysis was observed and the culture grew as well as the uninfected control.

In single burst experiments, progeny phage appearance during infection of induced T7 targeting cells was delayed and was due to the accumulation of escape phages (Figure 9E, left).

Phage DNA accumulated in infected cells that grew without induction of *cas* genes as judged by the appearance of restriction fragments of expected size (Figure 9E, right). Restriction fragments corresponding to T7 genomic DNA were also observed in induced cells. However, the pattern of these fragments was different and some fragments were missing (Figure 9E). To better understand the state of T7 DNA accumulating in conditions of CRISPR interference, DNA from induced and uninduced cultures was subjected to HTS (Figure 9F, Table 3). Mapping of reads from uninduced cultures revealed uniform coverage over the entire length of the T7 genome. By contrast, there was a dramatic decrease in coverage for DNA prepared from targeting cells. The area of decreased coverage was extensive and centered at the targeted protospacer (Figure 9F, Table 3). The overall profiles of genome coverage were different and very uneven - in all three cases of targeting – including areas which were seemingly unaffected by CRISPR interference. This may occur due to damage of phage genes responsible for expression of proteins involved in the replication of its DNA. Overall, we conclude that in conditions of CRISPR interference large portions of the T7 genome are lost, preventing production of progeny phage.

The primed adaptation experiment was attempted with every T7 targeting strain from our collection with wild-type and appropriate escape phages. In no case could we detect CRISPR array expansion in infected cultures.

4.6 Phages with modified genomes: T4 and R1-37

Bacteriophage T4 contains hydroxymethylcytosine instead of cytosine in its large (170,000 kbp) circularly permuted genome. The unusual nucleotide is additionally

modified by glycosylation (Calendar, Abedon, 2006; York *et al.*, 2003). Bacteriophage R1-37 contains uracil instead of thymine in its 260,000 kbp genome (Skurnik *et al.*, 2012). Both phages have genes belonging to different expression classes scattered throughout their genomes (Figure 10A and C, respectively). *Escherichia coli* is naturally resistant to R1-37; however, the phage can infect *E. coli* carrying a plasmid expressing its receptor (Schindelin *et al.*, 2012). Accordingly, a plasmid expressing the R1-37 receptor was introduced in strains containing R1-37-specific spacers. When T4 and R1-37 targeting strains were tested for ability to interfere with cognate phage infection, no resistance was observed (Figure 10B and D). In contrast, when these cells were transformed by plasmids carrying protospacers, a decrease of transformation efficiency by one to three orders of magnitude was observed. Neither phage was able to induce CRISPR array expansion in infected induced target cultures.

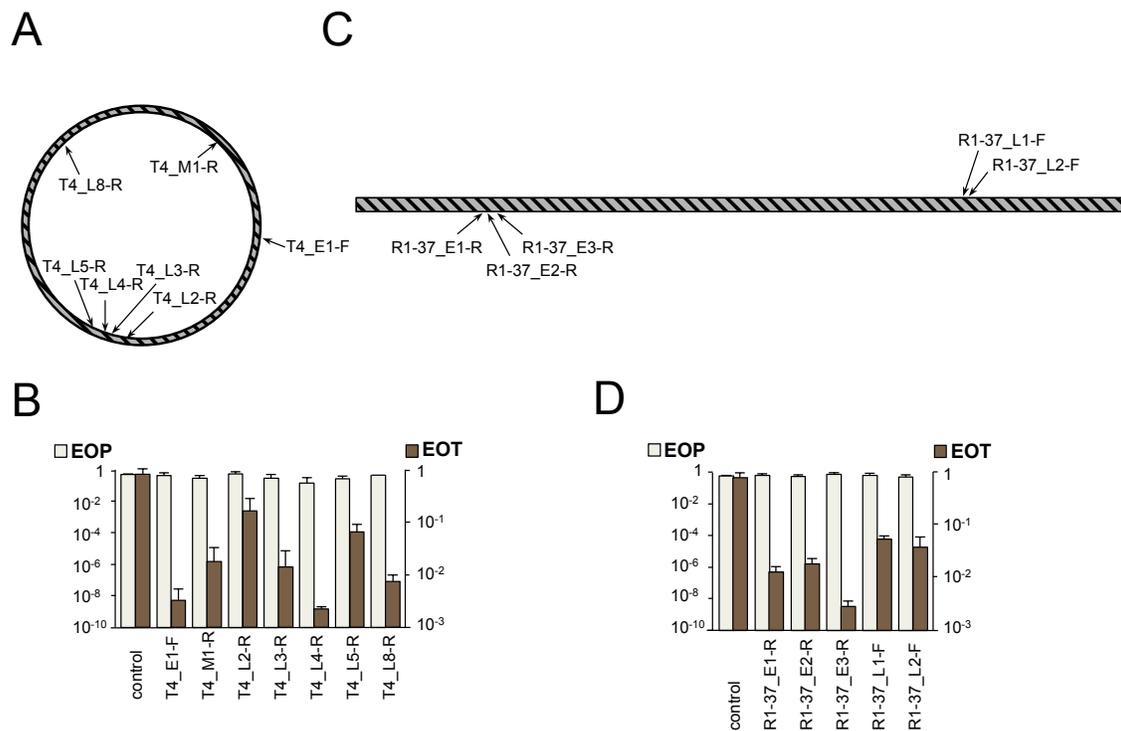


Figure 10. The effect of CRISPR-Cas targeting on T4 and R1-37 infections. (A) The circular genome of bacteriophage T4 is shown. Genes belonging to different expression classes are scattered throughout the genome. Arrows indicate the positions of protospacers targeted by crRNA spacers of strains from our collection. (B) EOP by T4 and EOT by cognate protospacer-containing plasmids into indicated strains. Mean values and standard deviations from three independent experiments with each strain are shown. EOP values for strains targeting additional protospacers can be found in Table 1. (C) The linear genome of bacteriophage R1-37 and positions of protospacers targeted by crRNA spacers of strains from our collection. Genes belonging to different expression classes are scattered throughout the genome. (D) EOP by R1-37 and EOT by cognate protospacer-containing plasmids into indicated strains. Mean values and standard deviations from three independent experiments with each strain are shown. EOP values for strains targeting additional protospacers can be found in Table 11.

Chapter 5. Conclusions

The goal of this work was to assess how different DNA phages are affected by the action of the type I-E CRISPR-Cas system of *E. coli*. Whenever phage DNA can be effectively targeted by crRNA-effector complex, a delay in the appearance of phage progeny is observed in infected cultures. Eventually, phage titers recover, and this recovery is due to accumulation of escape mutants, which must either preexist in the phage population used for infection, or are rapidly selected in the course of infection. The results thus show that CRISPR-Cas interference provides a strong selective pressure on diverse lytic phages, which, however, are able to rapidly overcome it. CRISPR-Cas systems can thus be regarded as a potent factor in generating phage variety. If the effector complex targets phage protospacers where no escape mutants can be selected, no phage progeny is observed. So there will be strong selection in the bacterial population in favor of cells that have acquired spacers targeting phage sequences that cannot be easily mutated, to allow long-lasting protection.

From the ‘point of view’ of a clonal bacterial culture, the outcome of infections by ‘professional’ lytic viruses T5 and T7 in the face of CRISPR defense very much depends on the multiplicity of infection. When multiplicity of infection (MOI) is higher than one, all infected cells succumb to infection even though no progeny phage is released. When MOI is significantly less than one, most cells remain uninfected, while infected cells die from the infection. However, since no or very little phage progeny is released, the infected cultures continue to grow and the infection process dies off unless an escape

phage appears. So it appears that, in the cases studied here, CRISPR immunity functions similarly to abortive infection mechanisms (Chopin *et al.*, 2005; Labrie *et al.*, 2010), not curing individual infected cells but preventing the spread of the virus through the population.

Targeting of phage λ or the pre-early region of T5 DNA prevents accumulation of viral DNA in infected cells, presumably by destroying templates for replication of circular λ DNA or preventing the entry of stage II early and late T5 genes by cleavage of the pre-early region. This, however, does not save T5-infected cells, which die because of cytotoxic pre-early proteins (Davison, 2015). In the case of T7, replication of phage DNA occurs even in conditions of ongoing CRISPR interference. However, extensive regions in both directions from the targeted protospacer are destroyed, presumably due to the function of Cas3 nuclease-helicase (Westra *et al.*, 2012). T7 is nevertheless able to replicate parts of its DNA in these conditions due to active recombination of phage genomes during replication (Nossal *et al.*, 1997; Sun *et al.*, 2015).

CRISPR interference is ineffective against early and late T5 genes, which are inserted in the cells after the point when cellular DNA is degraded.

Since most cells infected by the phages used here die even when they have CRISPR-Cas targeting a phage, only low levels of primed adaptation are observed. The situation contrasts with that seen in M13 phage infections, where infected cells continue to grow and very high levels of spacer acquisition in infected cultures are observed (Datsenko *et al.*, 2012).

Gradients of spacer acquisition efficiency were observed during primed adaptation by the I-F and I-B CRISPR-Cas systems (Li *et al.*, 2014; Richter *et al.*, 2014; Vorontsova *et al.*, 2015; Westra *et al.*, 2015). Spacer acquisition efficiency gradients observed in this work are extensive and much longer than those reported for type I-B system (Li *et al.*, 2014) but comparable to those observed in the I-F system (Vorontsova *et al.*, 2015). Interestingly, the length of the gradient also differs between λ and T5, indicating that phage genome structure/replication affects adaptation. In the case of T5, the gradient may also be physically limited by the two-stage viral DNA injection mechanism. Since no early or late DNA is present in the cell during the first stage of injection, spacers cannot be acquired from these areas. In the upstream direction, the length of the gradient is limited by the end of the linear genome. Another possibility is that the length of the time window when adaptation can happen in an infected cell differs for different phages (it should be much shorter during T5 infection).

We were unable to detect primed adaptation during T7 infection. Given our success with T5 and λ , the result may indicate that T7 has a mechanism to inhibit primed adaptation machinery. This putative mechanism must be targeting the adaptation proteins only, since CRISPR interference against T7 is highly efficient. Such a system could be biologically advantageous for the phage, given the ease of appearance of escape phage mutants, which promote primed adaptation.

The results with the two phages with modified DNA studied here, T4 and R1-37, show that the type I-E CRISPR-Cas system of *E. coli* is totally ineffective against these viruses. This may indicate that DNA modifications provide protection against a type I

system effector complex. It was recently reported that T4 can be successfully targeted by a type II Cas9 effector (Yaung *et al.*, 2014), which would mean that these systems are more robust. Alternatively, given the complexity of T4 and R1-37, it is entirely possible that the lack of CRISPR protection is due to other, as yet unknown phage mechanisms that are unrelated to DNA modification but specifically inactivate type I CRISPR-Cas systems.

In this paper, a comparative analysis of CRISPR response to the infection by various bacteriophages has been highlighted, but given the significant differences in the interaction of the *E.coli* cells and various phages, it is necessary to conduct a more detailed and in-depth analysis of these interactions for each virus separately. The study with phage M13 suggests experiments on the long-term cultivation of bacteria and virus in targeting conditions to identify the results of the battle of the phage and the cell at the population level. Work with T5 phage involves the search for anti-CRISPR proteins or other reasons for preventing any CRISPR response after injection of the early and late regions of the phage genome. Work with T7 phage requires a more detailed study of the bioinformatic data of products of the viral DNA degradation during CRISPR interference, as well as the search for proteins or mechanisms that inhibit CRISPR adaptation. Mathematical modeling of the data found during the work with fluorescence microscopy is also necessary. Work with bacteriophages T4 and R1-37 requires in vitro experiments that will reflect the presence or absence of binding of crRNA complexes to the target.

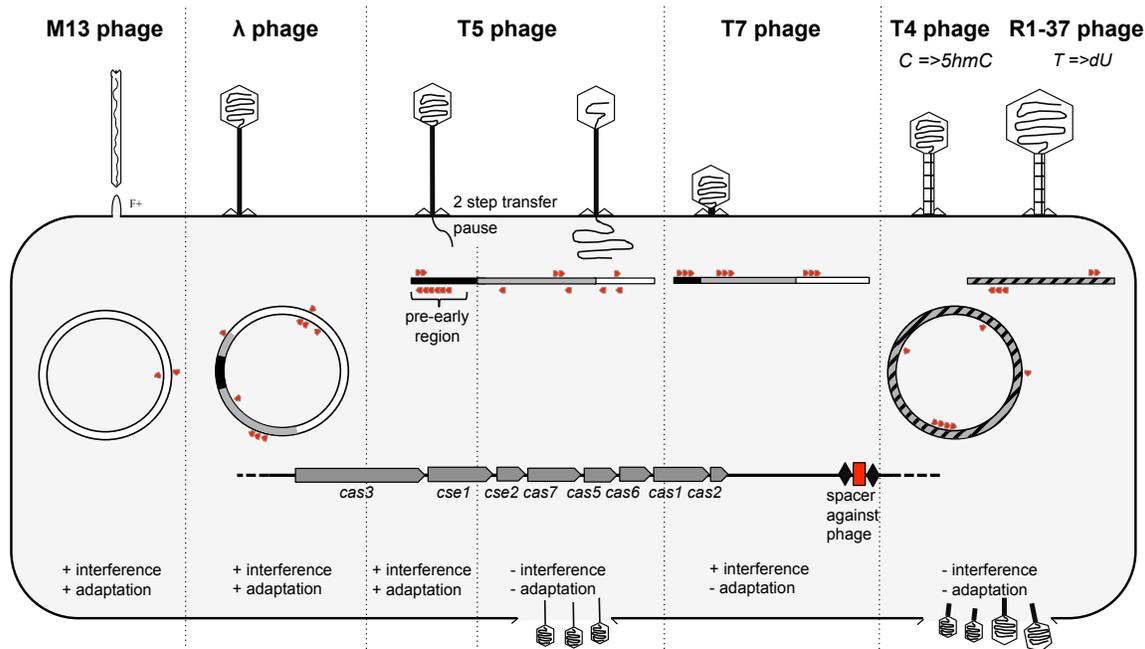


Figure 11. The overall scheme of effects of targeting by the *E. coli* type I-E CRISPR-Cas system on infection by various phages.

Phage genomes are represented as circular or linear dsDNA with different genome regions colored in black, gray, and white (hatching is used to show phage genomes where genes of different temporal classes are mixed together in the genome). Red arrows indicate the positions of protospacers targeted by crRNA spacers of strains from our collection.

Overall, our results establish certain commonalities but also reveal differences that may reflect specific viral adaptations used to overcome CRISPR-Cas defense. The outcome of infections in the presence of a functional CRISPR-Cas system targeting viral DNA is determined not by the intrinsic efficiency of target protospacer recognition by the effector complex but by the phage development strategy, genomic positions of targeted protospacers, and, possibly, DNA modification state. These factors can allow phages to overcome or at least limit the damaging effects of the CRISPR-Cas systems even in the absence of dedicated anti-CRISPR proteins.

Bibliography

- Adhya S., Sen A., Mitra S. The role of gene s In: The Bacteriophage Lambda. // Cold Spring Harbor. 1971. p. 743-746.
- Adriaenssens E.M., Krupovic M., Dutilh B.E., Adriaenssens E.M., Wittmann J., Vogensen F.K., Sullivan M.B., Rumnieks J., Prangishvili D., Lavigne R., Kropinski A.M., Klumpp J., Gillis A., Enault F., Edwards R.A., Duffy S., Clokie M.R., Barylski J., Ackermann H.W., Kuhn J.H. Taxonomy of prokaryotic viruses: 2016 update from the ICTV bacterial and archaeal viruses subcommittee. // Arch. Virol. 2017. 162(4):1153-1157
- Agirrezabala X., Velázquez-Muriel J.A., Gómez-Puertas P., Scheres S.H., Carazo J.M., Carrascosa J.L. Quasi-atomic model of bacteriophage t7 procapsid shell: insights into the structure and evolution of a basic fold. // Structure. 2007 15(4):461-72.
- Altman E., Altman R.K., Garrett J.M., Grimaila R.J., Young R. S gene product: identification and membrane localization of a lysis control protein. // J.Bacteriology. 1983. 155:1130–1137
- Anderson C.W., Eigner, J. Breakdown and exclusion of superinfecting T-even bacteriophage in Escherichia coli. // J. Virol. 1971. 8, 869-886
- Andersson A.F., Banfield J.F. Virus population dynamics and acquired virus resistance in natural microbial communities. // Science. 2008. 320:1047–50
- Bair C.L., Black L.W. A type IV modification dependent restriction nuclease that targets glucosylated hydroxymethyl cytosine modified DNAs. // J. Mol. Biol. 2007. 366(3):768-78.

Baranska S., Gabig M., Wegrzyn A., Konopa G., Herman-Antosiewicz A., Hernandez P., Schvartzman J.B., Helinski D.R., Wegrzyn G. Regulation of the switch from early to late bacteriophage lambda DNA replication. // *Microbiology*. 2001. 147(Pt 3):535-47

Barrangou R., Fremaux C., Deveau H., Richards M., Boyaval P., Moineau S., Romero D.A., Horvath P. CRISPR provides acquired resistance against viruses in prokaryotes. // *Science*. 2007. 315(5819):1709-12

Bellas C.M., Anesio A.M., Barker G. Analysis of virus genomes from glacial environments reveals novel virus groups with unusual host interactions. // *Front. Microbiol.* 2015. 6:656

Beloglazova N., Petit P., Flick R., Brown G., Savchenko A., Yakunin A. F. Structure and activity of the Cas3 HD nuclease MJ0384, an effector enzyme of the CRISPR interference. // *EMBO J.* 2011. 30; 4616–4627

Berry J., Rajaure M., Pang T., Young R. The Spanin Complex Is Essential for Lambda Lysis // *J. Bacteriol.* 2012. 194(20):5667-74.

Bondy-Denomy J., Garcia B., Strum S., Du M., Rollins M. F., Hidalgo-Reyes Y. Multiple mechanisms for CRISPR-Cas inhibition by anti-CRISPR proteins. // *Nature*. 2015. 526; 136–139

Bondy-Denomy J., Pawluk A., Maxwell K.L., Davidson A.R. Bacteriophage genes that inactivate the CRISPR/Cas bacterial immune system. // *Nature*. 2013. 493(7432):429-32

Bondy-Denomy J., Qian J., Westra E.R., Buckling A., Guttman D.S., Davidson A.R., Maxwell K.L. Prophages mediate defense against phage infection through diverse mechanisms. // *ISME J.* 2016. 10(12):2854-2866

Boucher I., Emond E., Dion E., Montpetit D., Moineau S. Microbiological and molecular impacts of AbiK on the lytic cycle of *Lactococcus lactis* phages of the 936 and P335 species. // *Microbiology*. 2000. 146:445-453.

Breitbart M. Marine viruses: truth or dare. // *Annu. Rev. Mar. Sci.* 2012. 4:425–448

Breitbart M., Wegley L., Leeds S., Schoenfeld T., Rohwer F. Phage community dynamics in hot springs. // *Appl. Environ. Microbiol.* 2004. 70: 1633–1640

Brouns S.J., Jore M.M., Lundgren M., Westra E.R., Slijkhuis R.J., Snijders A.P., Dickman M.J., Makarova K.S., Koonin E.V., van der Oost J. Small CRISPR RNAs guide antiviral defense in prokaryotes. // *Science*. 2008. 321(5891):960-4

Brüssow H., Hendrix R.W. Phage genomics: small is beautiful. // *Cell*. 2002. 108, 13–16

Calendar R., Abedon S.T. *The Bacteriophages*. 2nd edn. // Oxford University Press. 2006

Calendar R., Abedon S.T. *The Bacteriophages*. 2nd edn. // Oxford University Press. New York. 2006.

Cassman N., Prieto-Davo A., Walsh K. Oxygen minimum zones harbour novel viral communities with low diversity. // *Environ. Microbiol.* 2012 14:3043–65

Choi K.H., McPartland J., Kaganman I., Bowman V.D., Rothman-Denes L.B., Rossmann M.G. Insight into DNA and protein transport in double-stranded DNA viruses: the structure of bacteriophage N4. // *J. Mol. Biol.* 2008. 378, 726–736

Chopin M.C., Chopin A., Bidnenko E. Phage abortive infection in lactococci: variations on a theme. // *Curr. Opin. Microbiol.* 2005. 8:473–479. Labrie S.J., Samson J.E., Moineau S. Bacteriophage resistance mechanisms. // *Nat. Rev. Microbiol.* 2010. 8:317–327

Clokie M.R.J., Kropinski A. Bacteriophages: Methods and Protocols, Volume 1: Isolation, Characterization, and Interactions. // New York: Humana Press. 2008.

d'Herelle F. On an invisible microbe antagonistic to dysentery bacilli. 1917. // Res. Microbiol. 2007. 158(7):553-4

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

Datsenko K.A., Wanner B.L. One-Step inactivation of chromosomal genes in Escherichia coli K-12 using PCR products. // Proc. Natl. Acad. Sci. USA. 2000. 97:6640-6645

Davies E.V., James C.E., Williams D., O'Brien S., Fothergill J.L., Haldenby S., Paterson S., Winstanley C., Brockhurst M.A. Temperate phages both mediate and drive adaptive evolution in pathogen biofilms. // Proc. Natl. Acad. Sci. U S A. 2016.113(29):8266-71

Davison J. Pre-early functions of bacteriophage T5 and its relatives. // Bacteriophage. 2015. 5:e1086500

Deveau H., Barrangou R., Garneau J.E., Labonté J., Fremaux C., Boyaval P., Romero D.A., Horvath P., Moineau S. Phage response to CRISPR-encoded resistance in Streptococcus thermophiles. // J. Bacteriol. 2008. 190(4):1390-400

Dewey J.S., Savva C.G., White R.L., Vitha S., Holzenburg A., Young R. Micron-scale holes terminate the phage infection cycle. // Proc. Natl. Acad. Sci. U S A. 2010. 107:2219–2223

Diez-Villasenor C., Guzman N.M., Almendros C., Garcia-Martinez J., Mojica F.J. CRISPR-spacer integration reporter plasmids reveal distinct genuine acquisition

specificities among CRISPR-Cas I-E variants of *Escherichia coli*. // *RNA Biol.* 2013. 10; 792–802

Duckworth D.H. History and basic properties of bacterial viruses. *Phage ecology*. // New York: John Wiley and Sons. 1987. 1–43

Dulbecco R. Mutual exclusion between related phages. // *J. Bacteriol.* 1952. 63, 209-217

Dy R.L., Przybilski R., Semeijn K., Salmond G.P., Fineran P.C. A widespread bacteriophage abortive infection system functions through a Type IV toxin-antitoxin mechanism. // *Nucleic Acids Res.* 2014. 42(7):4590-605

Edelstein A., Amodaj N., Hoover K., Vale R., Stuurman N. Computer control of microscopes using microManager. // *Curr. Protoc. Mol. Biol.* 2010 Chapter 14: p. Unit 14 20.

Emrich J. // *Virology*. 1968. 35,158-165 Obring J., McCreary, Bernstein H. // *J. Virol.* 1988. 62,3043-3045

Filali Maltouf A., Labedan B. Host cell metabolic energy is not required for injection of bacteriophage T5 DNA. // *J Bacteriol.* 1983. 153(1):124-33

Fineran P.C., Blower T.R., Foulds I.J., Humphreys D.P., Lilley K.S., Salmond G.P. The phage abortive infection system, ToxIN, functions as a protein-RNA toxin-antitoxin pair. // *Proc. Natl. Acad. Sci. USA.* 2009. 106(3):894-9

Forde A., Fitzgerald G.F. Bacteriophage defense systems in lactic acid bacteria. // *Antonie Van Leeuwenhoek.* 1999. 76:89-113.

García L.R., Molineux I. J. Rate of translocation of bacteriophage T7 DNA across the membranes of *Escherichia coli*. // *J. Bacteriol.* 1995. 177, 4066–4076

Garrett J.M., Young R. Lethal action of bacteriophage lambda s gene. // Journal of virology. 1982. 44:886–892

Garside E.L., Schellenberg M.J., Gesner E.M., Bonanno J.B., Sauder J.M., Burley S.K., Almo S.C., Mehta G., MacMillan A.M. Cas5d processes pre-crRNA and is a member of a larger family of CRISPR RNA endonucleases. // RNA. 2012. 18(11):2020-8

Goldfarb T., Sberro H., Weinstock E., Cohen O., Doron S., Charpak-Amikam Y., Afik S., Ofir G., Sorek R. BREX is a novel phage resistance system widespread in microbial genomes. // EMBO J. 2015. 34(2):169-83

Gong B., Shin M., Sun J., Jung C.H., Bolt E.L., van der Oost J., Kim J.S. Molecular insights into DNA interference by CRISPR-associated nuclease-helicase Cas3. // Proc. Nat. Acad. Sci. USA. 2011. 111; 16359–16364

Grayson P., Han L., Winther T., Phillips R. Realtime observations of single bacteriophage λ DNA ejections in vitro. // Proc. Natl. Acad. Sci. USA. 2007. 104, 14652–14657

Grissa I., Vergnaud G., Pourcel C. The CRISPRdb database and tools to display CRISPRs and to generate dictionaries of spacers and repeats. // BMC Bioinform. 2007a. 8:172

Haaber J., Leisner J.J., Cohn M.T., Catalan-Moreno A., Nielsen J.B., Westh H., Penadés J.R., Ingmer H. Bacterial viruses enable their host to acquire antibiotic resistance genes from neighbouring cells. // Nat. Commun. 2016. 7:13333

Hankin E.H.L. Action bactericide des Eaux de la Jumna et du Gange sur le vibron du cholera. // Annales de l'Institut Pasteur. 1896. 10:511

Heller K., Braun V. Polymannose O-antigens of Escherichia coli, the binding sites for the reversible adsorption of bacteriophage T5+ via the L-shaped tail fibers. // J. Virol. 1982. 41:222–27

Heller K.J. Identification of the phage gene for host receptor specificity by analyzing hybrid phages of T5 and BF23. // Virology. 1984. 139:11–21

Heller K.J. Molecular interaction between bacteriophage and the gram-negative cell envelope. // Arch. Microbiol. 1992. 158:235–48

Hill C., Miller L.A., Klaenhammer T.R. Nucleotide sequence and distribution of the pTR2030 resistance determinant (hsp) which aborts bacteriophage infection in lactococci. // Appl. Environ. Microbiol. 1990. 56(7):2255-8

Hooton S.P., Connerton I.F. Campylobacter jejuni acquire new host-derived CRISPR spacers when in association with bacteriophages harboring a CRISPR-like Cas4 protein. // Front. Microbiol. 2015. 5:744

Horvath P., Romero D.A., Coute-Monvoisin A.C. Richards M., Deveau H., Moineau S., Boyaval P., Fremaux C., Barrangou. R. Diversity, activity, and evolution of CRISPR loci in Streptococcus thermophiles. // J. Bacteriol. 2008. 190, 1401–141

Høyland-Kroghsbo N.M., Maerkedahl R.B., Svenningsen S.L. A quorum-sensing induced bacteriophage defense mechanism. // MBio. 2016. 4(1):e00362-12

Høyland-Kroghsbo N.M., Paczkowskia J., Mukherjee S., Broniewskib J., Westra E.R., Bondy-Denomy J., Bassler B.L. Quorum sensing controls the Pseudomonas aeruginosa CRISPR-Cas adaptive immune system. // Proc. Natl. Acad. Sci. USA. 2017. 114(1):131-135

Hunt J.F., van der Vies S.M., Henry L., Deisenhofer J. Structural adaptations in the specialized bacteriophage T4 co-chaperonin Gp31 expand the size of the Anfinsen cage. // *Cell*. 1997. 90(2):361-71.

Hyman P., Abedon S.T. Practical methods for determining phage growth parameters. // *Methods Mol. Biol.* 2009. 501:175–202.

Ishii T., Yamaguchi Y., Yanagida M. Binding of the structural protein Soc to the head shell of bacteriophage T4. // *J. Mol. Biol.* 1978. 120, 533-544

Ishino Y., Shinagawa H., Makino K., Amemura M., Nakata A. Nucleotide sequence of the iap gene, responsible for alkaline phosphatase isozyme conversion in *Escherichia coli*, and identification of the gene product. // *Journal of bacteriology*. 1987 169(12):5429–33.

Jansen R., van Embden J. D. A., Gaastra W., Schouls L. M. Identification of a novel family of sequence repeats among prokaryotes. // *OMICS*. 2002. 6(1):23–33.

Jore M.M., Lundgren M., van Duijn E., Bultema J.B., Westra E.R., Waghmare S.P., Wiedenheft B., Pul U., Wurm R., Wagner R., Beijer M.R. Structural basis for CRISPR RNA-guided DNA recognition by Cascade. // *Nat. Struct. Mol. Biol.* 2011. 18:529-536

Jurczak-Kurek A., Gąsior T., Nejman-Faleńczyk B., Bloch S., Dydecka A., Topka G., Necel A., Jakubowska-Deredas M., Narajczyk M., Richert M., Mieszkowska A., Wróbel B., Węgrzyn G., Węgrzyn A. Biodiversity of bacteriophages: morphological and biological properties of a large group of phages isolated from urban sewage. // *Sci Rep*. 2016. 6:34338

Karlsson F., Malmberg-Hager A.C., Albrekt A.S., Borrebaeck C.A. Genome-wide comparison of phage M13-infected vs. uninfected Escherichia coli. // *Can. J. Microbiol.* 2005. 51:29–35.

Keller B., Bickle T.A. The nucleotide sequence of gene 21 of bacteriophage T4 coding for the prohead protease. // *Gene.* 1986. 49(2):245-51

Kemp P., Gupta M., Molineux I.J. Bacteriophage T7 DNA ejection into cells is initiated by an enzyme-like mechanism. // *Mol. Microbiol.* 2004. 53, 1251–1265

Kick B., Hensler S., Praetorius F., Dietz H., Weuster-Botz D. Specific growth rate and multiplicity of infection affect high-cell-density fermentation with bacteriophage M13 for ssDNA production. // *Biotechnol. Bioeng.* 2017. 114(4):777-784

Kiino D.R., Licudine R., Wilt K., Yang D.H., Rothman-Denes L.B. A cytoplasmic protein, NfrC, is required for bacteriophage N4 adsorption. // *J. Bacteriol.* 1993. 175(21):7074-80

Koonin E.V., Makarova K.S., Zhang F. Diversity, classification and evolution of CRISPR-Cas systems. // *Curr. Opin. Microbiol.* 2017. 37:67-78.

Kostyuchenko V.A., Chipman P.R., Leiman P.G., Arisaka F., Mesyanzhinov V.V., Rossmann M.G. The tail structure of bacteriophage T4 and its mechanism of contraction. // *Nat. Struct. Mol. Biol.* 2005. 12(9):810-3.

Krupovic M., Dutilh B.E., Adriaenssens E.M., Wittmann J., Vogensen F.K., Sullivan M.B., Rumnieks J., Prangishvili D., Lavigne R., Kropinski A.M., Klumpp J., Gillis A., Enault F., Edwards R.A., Duffy S., Clokie M.R., Barylski J., Ackermann H.W., Kuhn

J.H. Taxonomy of prokaryotic viruses: update from the ICTV bacterial and archaeal viruses subcommittee. // *Arch. Virol.* 2016. 161(4):1095-9

Krzywinski M., Schein J., Birol I., Connors J., Gascoyne R., Horsman D., Jones S.J., Marra M.A. Circos: an information aesthetic for comparative genomics. // *Genome Res.* 2009. 19:1639–1645

Kulczyk A.W., Richardson C.C. The Replication System of Bacteriophage T7. // *Enzymes.* 2016. 39:89-136

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

Kutter E. Phage host range and efficiency of plating. // *Methods Mol. Biol.* 2009. 501:141-9

Kutter E., Sulakvelidze A. *Bacteriophages Biology and application* // CRC Press. 2004

Labrie S.J., Samson J.E., Moineau S. Bacteriophage resistance mechanisms. // *Nat.Rev.Microbiol.* 2010. 8:317–27

Landsmann J., Kroger M., Hobom G. The rex region of bacteriophage λ : two genes under three-way control. // *Gene.* 1982. 20:11–24 Parma D.H., Snyder M., Sobolevski S., Nawroz M., Brody E., Gold L. The Rex system of bacteriophage λ : tolerance and altruistic cell death. // *Genes Dev.* 1992. 6:497–510

Lanni Y.T. First-step-transfer deoxyribonucleic acid of bacteriophage T5. // *Bacteriol. Rev.* 1968. 32, 227–242 Davison J. Pre-early functions of bacteriophage T5 and its relatives. // *Bacteriophage.* 2015. 5, e1086500

- Lee S.Y., Rasheed S. A simple procedure for maximum yield of high-quality plasmid DNA. // *Biotechniques*. 1990. 9:676–679
- Leiman P.G., Chipman P.R., Kostyuchenko V.A., Mesyanzhinov V.V., Rossmann, M.G. Three-dimensional rearrangement of proteins in the tail of bacteriophage T4 on infection of its host. // *Cell*. 2004. 118(4):419-29.
- Li M., Wang R., Zhao D., Xiang H. Adaptation of the *Haloarcula hispanica* CRISPR–Cas system to a purified virus strictly requires a priming process. // *Nucleic Acids Res.* 2014. 42:2483–2492.
- Lieb M. The establishment of lysogenicity in *Escherichia coli*. // *J. Bacteriol.* 1953. 65: 642- 51
- Liljas L. Virus assembly. // *Current Opinion in Structural Biology*. 1999. 9:129–134
- Lillestol R., Redder P., Garrett R., Brugger K. A putative viral defense mechanism in archaeal cells. // *Archaea*. 2006. 59e72.
- Lindberg A.A. Bacteriophage receptors. // *Annu. Rev. Microbiol.* 1973. 27:205–41
- Lu M.J., Henning U. Superinfection exclusion by T-even-type coliphages. // *Trends Microbiol.* 1994. 2(4):137-9
- Lu M.J., Henning U. The immunity (imm) gene of *Escherichia coli* bacteriophage T4 // *Viol.* 1989. 63, 3472-3478
- Lundgren M. Exploring the ecological function of CRISPR-Cas virus defense. // *Commun. Integr. Biol.* 2016. 9(5):e1216740
- Luria S.E., Human M.L. A nonhereditary, host-induced variation of bacterial viruses. // *J. Bacteriol.* 1952. 64, 557

Mahony J., McGrath S., Fitzgerald G.F., van Sinderen D. Identification and characterization of lactococcal-prophage-carried superinfection exclusion genes. // *Appl. Environ. Microbiol.* 2008. 74(20):6206-15

Makarova K. S., Grishin N. V., Shabalina S. A., Wolf Y. I., Koonin E. V. 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* 2006. 1:7

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

Makarova K.S., Wolf Y.I., Alkhnbashi O.S., Costa F., Shah S.A., Saunders S.J., Barrangou R., Brouns S.J., Charpentier E., Haft D.H., Horvath P., Moineau S., Mojica F.J., Terns R.M., Terns M.P., White M.F., Yakunin A.F., Garrett R.A., van der Oost J., Backofen R., Koonin E.V. An updated evolutionary classification of CRISPR-Cas systems. // *Nat. Rev. Microbiol.* 2015. 13:722–736.

Makhov A.M., Trus B.L., Conway J.F., Simon M.N., Zurabishvili T.G., Mesyanzhinov V.V., Steven A.C. The short tail-fiber of bacteriophage T4: molecular structure and a mechanism for its conformational transition. // *Virology.* 1993. 194(1):117-27

Mangenot S., Hochrein M., Radler J., Letellier L. Real-time imaging of DNA ejection from single phage particles. // *Curr. Biol.* 2005. 15, 430–435

Maranger R., Bird D.F. Viral abundance in aquatic systems: a comparison between marine and fresh waters. // *Mar. Ecol. Prog. Ser.* 1995. 121: 217–226 Hendrix R.W. Bacteriophages: evolution of the majority. // *Theor. Popul. Biol.* 2002. 61:471–480

Marraffini L.A., Sontheimer E.J. CRISPR interference limits horizontal gene transfer in staphylococci by targeting DNA. // *Science*. 2008. 322(5909):1843-5

Matthews B.W., Remington S.J., Grutter M.G., Anderson W.F. Relation between hen egg white lysozyme and bacteriophage T4 lysozyme: evolutionary implications. // *J.Molelcular Biology*. 1981. 147:545– 558.

McCorquodale D. J., Warner H. R. The Bacteriophages. // Plenum. 1988.1st edn. Vol. 1 439–475

Mesyanzhinov V.V., Leiman P.G., Kostyuchenko V.A., Kurochkina L.P., Miroshnikov K.A., Sykilinda N.N., Shneider M.M. Molecular architecture of bacteriophage T4. // *Biochemistry (Mosc)*. 2004 69(11):1190-202

Miller E.M., Nickoloff J.A. Escherichia coli electro transformation. // *Methods Mol. Biol.* 1995. 47:105

Minot S., Bryson A., Chehoud C. Rapid evolution of the human gut virome. // *Prot. Natl. Acad. Sci. USA* 2013. 110:12450–5.

Mojica F.J.M., Diez-Villasenor C., Garcia-Martinez J., Almendros C. Short motif sequences determine the targets of the prokaryotic CRISPR defence system. // *Microbiology*. 2009. 155; 733–740

Molineux I.J., Panja D. Popping the cork: mechanism of phage genome ejection. // *Nat. Rev. Microbiol.* 2013. 11:194-204

Molineux J. Fifty-three years since Hershey and Chase; much ado about pressure but which pressure is it? // *Virology*. 2006. 344, 221–229

Molineux J., Panja D. Popping the cork: mechanisms of phage genome ejection. // *Nat. Rev. Microbiol.* 2013 11(3):194-204

Morgan M., Anders S., Lawrence M., Aboyoun P., Page's H., Gentleman R. ShortRead: a bioconductor package for input, quality assessment and exploration of high-throughput sequence data. // *Bioinformatics*. 2009. 25:2607–2608

Morley D., Broniewski J.M., Westra E.R., Buckling A., van Houte S. Host diversity limits the evolution of parasite local adaptation. // *Mol. Ecol.* 2017. 26(7):1756-1763

Morozova N., Sabantsev A., Bogdanova E., Fedorova Y., Maikova A., Vedyaykin A., Rodic A., Djordjevic M., Khodorkovskii M., Severinov K. Temporal dynamics of methyltransferase and restriction endonuclease accumulation in individual cells after introducing a restriction-modification system. // *Nucleic Acids Res.* 2016. 44:790–800

Mosig G., Colowick N., Gruidl M.E., Chang A., Harvey A.J. Multiple initiation mechanisms adapt phage T4 DNA replication to physiological changes during T4's development. // *FEMS Microb. Rev.* 1995. 17, 83–98.

Mulepati S., Bailey S. In vitro reconstitution of an Escherichia coli RNA-guided immune system reveals unidirectional, ATP-dependent degradation of DNA target. // *J. Biol. Chem.* 2013. 288:22184–22192

Mulepati S., Bailey S.J. Structural and biochemical analysis of nuclease domain of clustered regularly interspaced short palindromic repeat (CRISPR)-associated protein 3 (Cas3). // *Biol. Chem.* 2011. 286; 31896–31903

Mulepati S., Heroux A., Bailey S. Crystal structure of a CRISPR RNA-guided surveillance complex bound to a ssDNA target. // *Science*. 2014. 345:1479-1484

Musharova O., Klimuk E., Datsenko K.A., Metlitskaya A., Logacheva M., Semenova E., Severinov K., Savitskaya E. Spacer-length DNA intermediates are associated with Cas1 in cells undergoing primed CRISPR adaptation. // *Nucleic Acids Res.* 2017. 45(6):3297-3307

Newton G.J., Daniels C., Burrows L.L., Kropinski A.M., Clarke A.J., Lam J.S. Three-component-mediated serotype conversion in *Pseudomonas aeruginosa* by bacteriophage D3. // *Mol. Microbiol.* 2001. 39:1237–1247

Nguyen H.M., Kang,C. Lysis delay and burst shrinkage of coliphage T7 by deletion of terminator T reversed by deletion of early genes. // *J Virol.* 2014. 88: 2107–2115

Noble E., Spiering M.M., Benkovic S.J. Coordinated DNA Replication by the Bacteriophage T4 Replisome. // *Viruses*. 2015. 7(6):3186-200

Nossal N.G. Protein-protein interactions at a DNA replication fork: Bacteriophage T4 as a model. // *FASEB J.* 1992. 6, 871–878

Nossal N.G., Richardson C.C., Kong D. Role of the bacteriophage T7 and T4 single-stranded DNA-binding proteins in the formation of joint molecules and DNA helicase-catalyzed polar branch migration. // *J. Biol. Chem.* 1997. 272:8380–8387

Nunez J.K., Lee A.S., Engelman A., Doudna J.A. Integrase-mediated spacer acquisition during CRISPR-Cas adaptive immunity. // *Nature*. 2015b. 519; 193–198.

Paez-Espino D., Morovic W., Sun C.L., Thomas B.C., Ueda K., Stahl B. Strong bias in the bacterial CRISPR elements that confer immunity to phage. // *Nature Communications*. 2013. 4:1430

Paez-Espino D., Sharon I., Morovic W., Stahl B., Thomas B.C., Barrangou R., Banfield J.F. CRISPR immunity drives rapid phage genome evolution in *Streptococcus thermophiles*. // *MBio*. 2015. 6(2): e00262-15

Pages H., Aboyoun P., Gentleman R., DebRoy S. Biostrings: String objects representing biological sequences, and matching algorithms. // R package version. 2012. 2.24.1

Panja D. Molineux J. Dynamics of bacteriophage genome ejection in vitro and in vivo. // *Phys. Biol*. 2010. 7, 045006

Papenfort K., Bassler B.L. Quorum sensing signal-response systems in Gram-negative bacteria. // *Nat. Rev. Microbiol*. 2016. 14(9):576–588

Parreira R., Ehrlich S.D., Chopin M.C. Dramatic decay of phage transcripts in lactococcal cells carrying the abortive infection determinant AbiB. // *Mol Microbiol*. 1996. 19:221-230

Patterson A.G., Chang J.T., Taylor C., Fineran P.C. Regulation of the type I-F CRISPR-Cas system by CRP-cAMP and GalM controls spacer acquisition and interference. // *Nucleic Acids Res*. 2015. 43(12):6038–6048.

Pawluk A., Bondy-Denomy J., Cheung V.H.W., Maxwell K.L., Davidson A.R. A new group of phage anti-CRISPR genes inhibits the type I-E CRISPR-Cas system of *Pseudomonas aeruginosa*. // *MBio*. 2014. 5:e00896-14

Perez-Rodriguez R., Haitjema C., Huang Q., Nam K.H., Bernardis S., Ke A., DeLisa M.P. Envelope stress is a trigger of CRISPR RNA-mediated DNA silencing in *Escherichia coli*. // *Mol. Microbiol.* 2011. 79(3):584–599

Peterson J., Garges S., Giovanni M., McInnes P., Wang L., Schloss J.A., Bonazzi V., McEwen J.E., Wetterstrand K.A., Deal C., Baker C.C., di Francesco V., Howcroft T.K., Karp R.W., Lunsford R.D., Wellington C.R., Belachew T., Wright M., Giblin C., David H., Mills M., Salomon R., Mullins C., Akolkar B., Begg L., Davis C., Grandison L., Humble M. The NIH human microbiome project. // *Genome Res.* 2009. 19(12):2317–2323

Pickard D.J. Preparation of bacteriophage lysates and pure DNA. // *Methods Mol. Biol.* 2009. 502:3–9

Prigent M., Leroy M., Confalonieri F., Dutertre M., DuBow M.S. A diversity of bacteriophages forms and genomes can be isolated from the surface sands of Sahara Desert. // *Extremophiles.* 2005. 9:289–296 10

Ptashne M. Two "what if" experiments. // *Cell.* 2004. 116: S71-72

Refardt D., Rainey P.B. Tuning a genetic switch: experimental evolution and natural variation of prophage induction. // *Evolution.* 2010. 64: 1086-97

Richardson A., Landry S. J., Georgopoulos C. The ins and outs of a molecular chaperone machine. // *Trends Biochem. Sci.* 1998. 23, 138-143

Richter C., Dy R.L., McKenzie R.E., Watson B.N., Taylor C., Chang J.T., McNeil M.B., Staals R.H., Fineran P.C. Priming in the Type I-F CRISPR-Cas system triggers strand-

independent spacer acquisition, bi-directionally from the primed protospacer. // *Nucleic Acids Res.* 2014. 42:8516–8526.

Riede I. Receptor specificity of the short tail fibres (gp12) of T-even type *Escherichia coli* phages. // *Mol. Gen. Genet.* 1987. 206:110–15

Roberts R.J., Belfort M., Bestor T., Bhagwat A.S., Bickle T.A., Bitinaite J. A nomenclature for restriction enzymes, DNA methyltransferases, homing endonucleases and their genes. // *Nucleic Acids Res.* 2003. 31, 1805–1812

Rocha E.P., Danchin A., Viari A. Evolutionary role of restriction/modification systems as revealed by comparative genome analysis. // *Genome Res.* 2001. 11:946–58.

Sanguino L., Franqueville L., Vogel T.M. Linking environmental prokaryotic viruses and their host through CRISPRs. // *FEMS Microbiol. Ecol.* 2015. 91

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

Savva C.G., Dewey J.S., Moussa S.H., To K.H., Holzenburg A., Young R. Stable micron-scale holes are a general feature of canonical holins. // *Mol Microbiol.* 2014. 91:57–65.

Sävström C.H., Lisle J., Anesio A.M., Priscu J.C., Laybourn-Parry J. Bacteriophage in polar inland waters. // *Extremophiles.* 2008.12:167–175

Schindelin J., Arganda-Carreras I., Frise E., Kaynig V., Longair M., Pietzsch T., Preibisch S., Rueden C., Saalfeld S., Schmid B., Tinevez J.Y., White D.J., Hartenstein

V., Eliceiri K., Tomancak P., Cardona A. Fiji: an open-source platform for biological-image analysis. // Nat. Methods. 2012 9:676–682.

Schmitt C.K., Kemp P., Molineux J. 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. 1991. 173:6507–14

Schmitt C.K., Molineux J. Expression of gene 1.2 and gene 10 of bacteriophage T7 is lethal to F plasmid-containing *Escherichia coli*. // J. Bacteriol. 1991. 173:1536–43

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

Schrader H.S., Schrader J.O., Walker J.J., Wolf T.A., Nickerson K.W., Kokjohn T.A. Bacteriophage infection and multiplication occur in *Pseudomonas aeruginosa* starved for 5 years. // Can. J. Microbiol. 1997. 43, 1157–1163

Schwartz M. The adsorption of coliphage lambda to its host: effect of variations in the surface density of receptor and in phage-receptor affinity. // J. Mol. Biol. 1976. 103:521–536

Seed K.D., Lazinski D.W., Calderwood S.B., Camilli A. A bacteriophage encodes its own CRISPR/Cas adaptive response to evade host innate immunity. // Nature. 2013. 494(7438):489-91

Sekulovic O., Ospina Bedoya M., Fivian-Hughes A.S., Fairweather N.F., Fortier L.C. The *Clostridium difficile* cell wall protein CwpV confers phase-variable phage resistance. // Mol. Microbiol. 2015. 98(2):329-42

Selle K., Barrangou R. Harnessing CRISPR-Cas systems for bacterial genome editing. // Trends Microbiol. 2015b. 23(4):225–32

Semenova E., Jore M.M., Datsenko K.A., Semanova A., Westra E.R., Wanner B., van der Oost J., Brouns S.J., Severinov K. Interference by clustered regularly interspaced short palindromic repeat (CRISPR) RNA is governed by a seed sequence. // Proc. Natl. Acad. Sci. USA. 2011. 108(25):10098-103

Semenova E., Savitskaya E., Musharova O., Strotskaya A., Vorontsova D., Datsenko K.A., Logacheva M.D., Severinov K. Highly efficient primed spacer acquisition from targets destroyed by the Escherichia coli type I-E CRISPR-Cas interfering complex. // Proc. Natl. Acad. Sci. U.S.A. 2016. 113; 7626–7631

Simmons S.L., Dibartolo G., Deneff V.J., Goltsman D.S., Thelen M.P., Banfield J.F. Population genomic analysis of strain variation in Leptospirillum group II bacteria involved in acid mine drainage formation. // PLoS Biol. 2008. 6(7):e177.

Sinkunas T., Gasiunas G., Fremaux C., Barrangou R., Horvath P., Siksnys V. Cas3 is a single-stranded DNA nuclease and ATP-dependent helicase in the CRISPR/Cas immune system. // EMBO J. 2011. 30; 1335–1342

Sinkunas T., Gasiunas G., Waghmare S.P., Dickman M.J., Barrangou R., Horvath P., Siksnys V. In vitro reconstitution of Cascade-mediated CRISPR immunity in Streptococcus thermophiles. // EMBO J. 2013. 32; 385–394

Skurnik M., Hyytiainen H.J., Happonen L.J., Kiljunen S., Datta N., Mattinen L., Williamson K., Kristo P., Szeliga M., Kalin-Manttari L. Characterization of the genome, Proteome, and structure of Yersiniophage phiR1-37. // Virol. 2012. 86:12625

Smeal S.W., Schmitt M.A., Pereira R.R., Prasad A., Fisk J.D. Simulation of the M13 life cycle II: Investigation of the control mechanisms of M13 infection and establishment of the carrier state. // *Virology*. 2017. 500:275-284

Smith D.E., Tans S.J., Smith S.B., Grimes S., Anderson D.L., Bustamante C. The bacteriophage straight Φ 29 portal motor can package DNA against a large internal force. // *Nature*. 2001. 413, 748–752

Snyder L., McWilliams K. The rex genes of bacteriophage λ can inhibit cell function without phage superinfection. // *Gene*. 1989. 81:17–24

Staden R. A strategy of DNA sequencing employing computer program. // *Nucleic Acids Res.* 1979. 6(7):2601-10

Stewart F.M., Levin B.R. The population biology of bacterial viruses: why be temperate. // *Theoretical Population Biology*. 1984. 26: 93-117

Sumby P., Smith M.C. Genetics of the phage growth limitation (Pgl) system of *Streptomyces coelicolor* A3(2). // *Mol. Microbiol.* 2002. 44:489–500

Sumby P., Smith M.C. Phase variation in the phage growth limitation system of *Streptomyces coelicolor* A3(2). // *J. Bacteriol.* 2003.185:4558–63

Summer E.J., Berry J., Tran T.A., Niu L., Struck D.K., Young R. Rz/Rz1 lysis gene equivalents in phages of Gram-negative hosts. // *J. Mol. Biol.* 2007. 373(5):1098-112.

Sun B., Pandey M., Inman T., Yang Y., Kashlev M., Patel S.S., Wang M.D. T7 replisome directly overcomes DNA damage. // *Nat. Commun.* 2015. 6:10260.

Sun X., Göhler A., Heller K. J., Neve H. The *ltp* gene of temperate *Streptococcus thermophilus* phage TP-J34 encodes a lipoprotein which is expressed during lysogeny. // *Virology*. 2006. 350,146–157

Swarts D.C., Mosterd C., van Passel M.W.J., Brouns S.J.J. CRISPR interference directs strand specific spacer acquisition. // *PLoS ONE*. 2012. 7, e35888

Tan D., Svenningsen S.L., Middelboe M. Quorum sensing determines the choice of antiphage defense strategy in *Vibrio anguillarum*. // *MBio*. 2015. 6(3):e00627

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

Tyndall C., Lehnerr H., Sandmeier U., Kulik E., Bickle T.A. The type IC *hsd* loci of the enterobacteria are flanked by DNA with high homology to the phage P1 genome: implications for the evolution and spread of DNA restriction systems. // *Mol. Microbiol*. 1997. 23(4):729-36.

Tyson G.W., Banfield J.F. Rapidly evolving CRISPRs implicated in acquired resistance of microorganisms to viruses. // *Environ. Microbiol*. 2008. 10:200–7

Vale P.F., Lafforgue G., Gatchitch F., Gardan R., Moineau S., Gandon S. Costs of CRISPR-Cas-mediated resistance in *Streptococcus thermophilus*. // *Proc. Biol. Sci*. 2015. 282(1812):20151270

Van der Oost J., Westra E.R., Jackson R.N., Wiedenheft B. Unravelling the structural and mechanistic basis of CRISPR-Cas systems. // *Nat. Rev. Microbiol*. 2014. 12:479 – 492.

van Houte S., Buckling A., Westra E.R. Evolutionary Ecology of Prokaryotic Immune Mechanisms. // *Microbiol. Mol. Biol. Rev*. 2016. 80(3):745-63.

Vercoe R.B., Chang J.T., Dy R.L., Taylor C., Gristwood T., Clulow J.S., Richter C., Przybilski R., Pitman A.R., Fineran P.C. Cytotoxic chromosomal targeting by CRISPR/Cas systems can reshape bacterial genomes and expel or remodel pathogenicity islands. // *PLoS Genet.* 2013. 9(4):e1003454

Vorontsova D., Datsenko K.A., Medvedeva S., Bondy-Denomy J., Savitskaya E.E., Pougach K., Logacheva M., Wiedenheft B., Davidson A.R., Severinov K. Foreign DNA acquisition by the I-F CRISPR-Cas system requires all components of the interference machinery. // *Nucleic Acids Res.* 2015. 43:10848–10860.

Wayne R., Neilands J.B. Evidence for common binding sites for ferrichrome compounds and bacteriophage ϕ 80 in the cell envelope of *Escherichia coli*. // *J. Bacteriol.* 1975. 121:497–503

Westra E.R., Van Erp P.B., Kunne T., Wong S.P., Staals R.H., Seegers C.L., Bollen S., Jore M.M., Semenova E., Severinov K. CRISPR immunity relies on the consecutive binding and degradation of negatively supercoiled invader DNA by Cascade and Cas3. // *Mol. Cell.* 2012. 46:595–605.

Westra E.R., van Erp P.B.G., Kunne T., Wong S.P., Staals R.H.J., Seegers C.L.C., Bollen S., Jore M.M., Semenova E., Severinov K. CRISPR immunity relies on the consecutive binding and degradation of negatively supercoiled invader DNA by Cascade and Cas3. // *Mol. Cell.* 2012. 46; 595–605

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. Parasite exposure drives

selective evolution of constitutive versus inducible defense. // *Curr. Biol.* 2015. 25:1043–1049

Westra E.R., van Houte S., Oyesiku-Blakemore S., Makin B., Broniewski J.M., Best A. Parasite Exposure Drives Selective Evolution of Constitutive versus Inducible Defense. // *Curr. Biol.* 2015 25(8):1043-9

Whitman W.B., Coleman D.C., Wiebe W.J. Prokaryotes: the unseen majority. // *Proc. Natl. Acad. Sci. USA.* 1998. 95, 6578–6583

Williamson K.E., Radosevich M., Wommack K.E. Abundance and diversity of viruses in six Delaware soils. // *Appl. Environ. Microbiol.* 2005. 71:3119–3125

Xu M., Arulandu A., Struck D.K., Swanson S., Sacchettini J.C., Young R. Disulfide isomerization after membrane release of its SAR domain activates P1 lysozyme. // *Science.* 2005. 307:113–117

Yaung S.J., Esvelt K.M., Church G.M. CRISPR/Cas9-mediated phage resistance is not impeded by the DNA modifications of phage T4. // *PLoS One.* 2014. 9:e98811

York, Miller E.S., Kutter E., Mosig G., Arisaka F., Kunisawa T., Ruger W. Bacteriophage T4 genome. // *Microbiol. Mol. Biol. Rev.* 2003. 67:86–156

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

Young R. Bacteriophage lysis: mechanism and regulation. // *Microbiol. Rev.* 1992 56:430–481

Zagotta M.T., Wilson D.B. Oligomerization of the bacteriophage lambda S protein in the inner membrane of *Escherichia coli*. // *J. Bacteriology.* 1990. 172:912–921.

Zhao H., Sheng G., Wang J., Wang M., Bunkoczi G., Gong W., Wei Z., Wang Y. Crystal structure of the RNA-guided immune surveillance Cascade complex in Escherichia coli.
// Nature. 2014. 515(7525):147-50