SINGLE-CELL STUDIES OF BACTERIOPHAGE LAMBDA AND CRISPR-CAS SYSTEMS

A Dissertation

by

JINGWEN GUAN

Submitted to the Office of Graduate and Professional Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

Chair of Committee,	Lanying Zeng
Committee Members,	Hays Rye
	Ryland F. Young
	Xiuren Zhang
Head of Department,	Dirk Hays

December 2019

Major Subject: Molecular and Environmental Plant Sciences

ABSTRACT

Cellular decision-making is a ubiquitous biological process among all organisms, from simple viruses to complex mammals, occurring when living systems process signals received from the environment and then make appropriate responses for survival. Bacteriophage lambda and its bacterial host *E. coli* have served as a basic paradigmatic system for understanding cellular decision-making. Upon infection by phage λ, an *E. coli* bacterium undergoes either of two alternative life cycles, lysis or lysogeny. Despite extensive efforts to uncover the underlying decision-making mechanisms, several hidden variables have yet been characterized, rendering cell-fate decisions mysterious and unpredictable.

The λ genetic circuit composed of fate-determining genes drives the lysis-lysogeny decision. It has been suggested that the fluctuations in viral gene expression can cause dramatic changes in cell-fate determination. To evaluate the effect of stochastic gene activity on λ decision-making, we quantify the transcriptional level of the λ *cII* gene at the single-cell level, as CII is the master regulator during the decision-making process. We reveal that the average *cII* mRNA levels increase and reach a peak around 10 min after infection, and subsequently drop. Next, by labeling individual phage genomes, we investigate the intracellular organization of phage DNA development in single cells. We observe that infecting phage DNA can organize their own DNA replication and gene expression by assembling separate entities within a single cell. This phage individuality corroborates our previous hypothesis that each phage DNA has the capability of making decisions independently. Furthermore, we provide evidence of heterogeneity in phage subcellular development, most likely as a result of intracellular phage-phage interactions. Additionally, when the effect of the side tail fibers on λ infection is examined, those side tail fibers significantly reduce the ability of λ for successful infection but do not affect the lysis-lysogeny decision-making outcome.

Meanwhile, we visualize viral DNA degradation by a bacterial CRISPR-Cas system at the single-cell level. We quantitatively characterize several factors accounting for effective CRISPR defense and hypothesize that phages can combat the CRISPR system through rapid DNA replication and co-infections. These findings enrich our current understanding of the mechanistic basis of CRISPR-Cas immune systems.

ACKNOWLEDGEMENTS

First and foremost, I would like to express my sincere gratitude to my advisor, Dr. Lanying Zeng, for her patient guidance, encouragement and support throughout the entire course of my graduate career. As a student from the MEPS program, I greatly appreciate that she gave me the same opportunity as other Bio/Bio students to join her lab. I have learned more than I could have imagined over the past five years, especially in the aspect of quantitative data analysis and coding. I also would like to thank my committee members, Dr. Ryland F. Young, Dr. Hays Rye and Dr. Xiuren Zhang, for their insightful suggestions and guidance for my research.

Next, I would like to give my special thanks to the Zeng lab alumni, Dr. Qiuyan Shao and Dr. Jimmy T. Trinh, for their generous help, not only in my scientific research but also in my daily life. Being new to the phage field without any related research background and also coming to this foreign country alone for the first time, I would not have been able to survive the first two years of my Ph.D. program without their help and encouragement. They imparted knowledge and experimental techniques to me, gave me a lot of constructive advice for my research projects and provided as much help as they can whenever I need them in my life. I genuinely hope that both of them will have brilliant and successful life in the future!

I am also grateful to the past and present members of the Zeng lab, Dr. Xu Shi, Kailun Zhang, Laith Harb, Matthew Theodore and Zihao Yu, for their valuable discussions during the lab meetings. I also want to thank to the members of the Young lab,

especially Dr. Karthik Reddy Chamakura and Dr. Jesse Cahill, for kindly sharing strains, reagents, and their helpful suggestions. I am thankful to the undergraduate students, David Ibarra, Anastasia O'Neill and Roberto Burgos, for their efforts and contributions to my experiments. Special thanks go to my mentor during my undergraduate years, Dr. Dong Xiao from the College of Life Sciences at Nankai University in China, who brought me into the fields of molecular biology. He is also the one who taught me every single basic experimental skill at the very beginning of my research career. Those research experiences fully prepared me for pursuing my Ph.D. degree over the past few years.

I greatly appreciate Dr. Mary Bryk and Dr. Dorothy Shippen for giving me the opportunity to be a teaching assistant during the last year of my Ph.D. program. I am also heartily thankful to my friends for their support, companionship and always leaving doors open for me whenever I need help.

Finally, I deeply thank my loving parents and my family for their endless love, patience and understanding in me over the years. I could not have come this far without them.

CONTRIBUTORS AND FUNDING SOURCES

Contributors

This work was supervised by a dissertation committee consisting of Dr. Lanying Zeng, Dr. Ryland F. Young, Dr. Hays Rye and Dr. Xiuren Zhang of the Department of Biochemistry and Biophysics. Dr. Lanying Zeng and Dr. Xiuren Zhang are also in the Molecular and Environmental Plant Sciences (MEPS) program.

In chapter I, Zhicheng Cui of Dr. Junjie Zhang lab at Texas A&M University performed the negative staining of bacteriophage λ and Ur-λ virions and provided the electron micrographs.

In chapter II, Dr. Qiuyan Shao, one of alumni in the Zeng lab, performed part of the smFISH experiments and constructed the analysis framework for the quantification of the *cII* mRNA levels. The construction of the *E. coli* DnaB reporter strain and the charaterization of the subcellular spatial organization of phage DNA was in part conducted by Dr. Jimmy T. Trinh, another alumnus in the Zeng lab. The real-time live-cell infection movies of the engineered *tetO* phage and movie analysis were in part performed by Zihao Yu of the Zeng lab.

In chapter III, the Ur-λ strain without fluorescent labeling was constructed by Dr. Qiuyan Shao. David Ibarra helped with the data analysis for the phage infection movies.

In chapter IV, the transformation of the CRISPR-Cas plasmids into the *E. coli* reporter strain was done by Dr. Xu Shi of the Zeng lab. Roberto Burgos helped with the cell recognition for the live-cell movies. Manuscript writing was in part done by Dr. Lanying Zeng.

All other work conducted for the dissertation was completed independently by the student.

Funding Sources

This work was made possible by Texas A&M Agrilife Research and the National Institutes of Health under Grant Number R01GM107597. This content is solely the responsibility of the authors and do not necessarily represent the official views of Texas A&M AgriLife Research and the National Institutes of Health.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

Page

CHAPTER I

INTRODUCTION AND LITERATURE REVIEW

Bacteriophage Lambda

Bacteriophages, also simply known as phages, are viruses that infect bacteria. They were identified independently as “antagonistic” microbes to bacteria by F. W. Twort and Felix d'Herelle in 1910s [1-3]. Ever since their discovery, they have been extensively used for numerous applications. Pioneering studies on phage genetic basis and structures allow various discoveries in genetic regulatory processes that shared by living organisms, greatly facilitating the development of molecular biology [4, 5]. Given that phages have great potential as antimicrobials, they play an important role in biotechnological fields and have been widely used to control bacterial infections in food industries and to clinically cure bacterial diseases [3, 6, 7].

Bacteriophages are acknowledged as the most abundant organisms ($\sim 10^{31}$ virions) in the biosphere [8, 9]. Among their enormous diversity, phage lambda (λ) is the most comprehensively studied phage. Phage λ was first discovered in 1951 by Esther Lederberg as a byproduct when she studied the genetics of *Escherichia coli* strain K-12, which carried a λ prophage and released λ virions after undergoing treatment with ultraviolet irradiation [10]. Electron microscopy of negatively stained phage λ shows that it has a symmetric and isometric head about 60 nm in diameter and a long non-contractile tail about 150 nm in length (Figure 1.1A) [11-13]. The tail of phage λ is a flexible tube with a conical tail tip (15 nm in length) and a single short tail fiber (23 nm) extending from the center of the tail tip [12, 14]. Notably, there are four long side tail fibers attached at the junction between the tube and the tail tip, as reported for the originally isolated λ variant (Ur-λ) (Figure 1.1B) [15]. These fibers were not found previously because almost all laboratories around the world perform experiments with a mutant λ strain (λ*PaPa*) [16], which is a result of selective breeding for genetic studies during the early days. Each λ particle contains a single molecule of double-stranded DNA with a genomic size of 48,503 bp (48,502 bp for the commonly used, "so called" wild-type strain, λ*PaPa*), packaged in the head. Owing to its modest genome size and the capability of making a decision of whether to kill the bacteria immediately or exist as a quiescent prophage, phage λ has been adequately studied for more than half a century and serves as an important paradigm for various biological developmental processes, such as regulation of gene expression, DNA replication, mechanisms of recombination and cell-fate decision-making, thus making great contributions to molecular genetic research [12, 17-20].

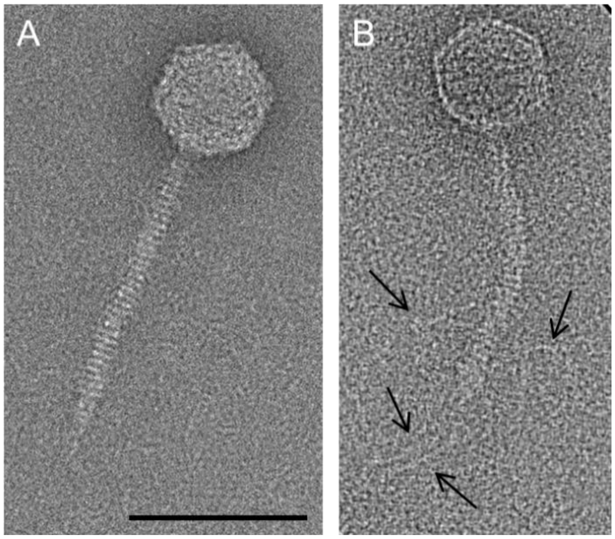

Figure 1.1 Electron micrograph of the phage lambda virions.
(A) λWT or λ*PaPa*. (B) Ur-λ. The side tail fibers are indicated by the black arrows. Phages were purified from a standard CsCl purification procedure and negatively stained with 0.75% uranyl formate. Scale bar represents 100 nm.

Phage Lambda Life Cycle

To initiate an infection, phage λ will first encounter its bacterial host *E. coli* in the environment and then adsorb onto the cell surface through binding to the receptor, followed by phage genome delivery into the cytoplasm. As a temperate phage, phage λ can enter either of two distinct life cycles, lysis or lysogeny (Figure 1.2). During the lytic cycle, λ rapidly replicates its DNA and assembles new phages, resulting in host cell lysis and the subsequent release of about a hundred progeny virions. Alternatively, if the lysogenic cycle is chosen, λ establishes the dormant state by integrating its genome into the host chromosome, existing as a quiescent prophage and replicating passively as a part of the host genome. Although extremely stable, the dormant state can be irreversibly switched back to the lytic state, a process known as prophage induction [18, 21, 22]. The induction is caused by treatments with ultraviolet radiation or chemical agents, such as mitomycin C, which can damage the host DNA to activate the bacterial SOS response, ultimately leading to the cleavage of λ repressor protein, CI [23, 24].

As a virus-host model system for decades, phage λ has garnered much attention for its adsorption dynamics and the subsequent phage DNA translocation. Meanwhile, many efforts have also been made to reveal the genetic regulatory circuits of phage λ, characterize deterministic factors that affect the cell-fate choices and elucidate the underlying mechanism of how phage λ makes the decisions.

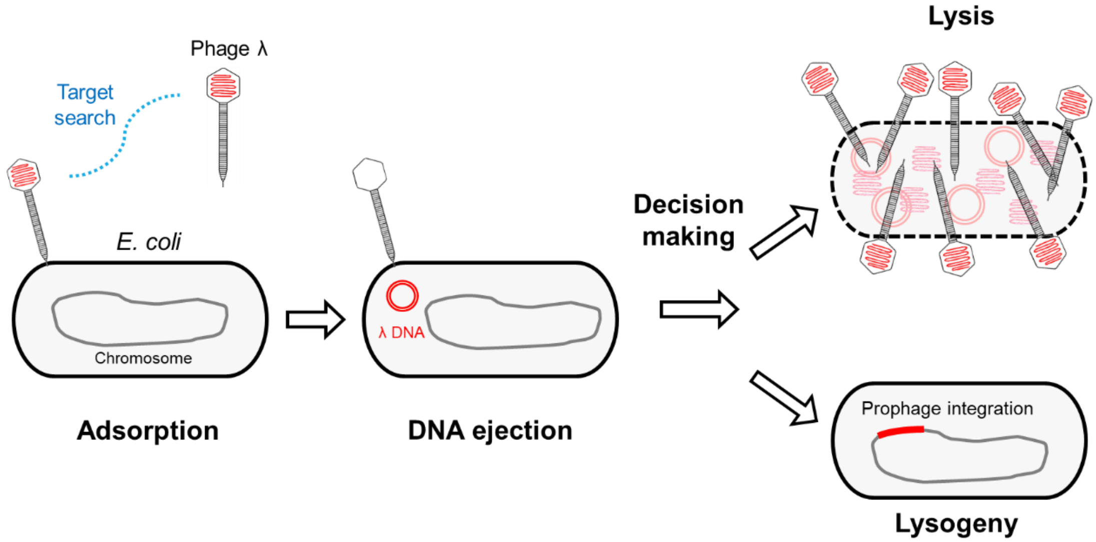

Figure 1.2 Phage lambda life cycle.
Phage λ searches and adsorbs onto the *E. coli* host cell surface, followed by the ejection of phage DNA (red) into the cytoplasm. Subsequently, a decision will be made to choose between lysis and lysogeny. During the lytic development, phage rapidly replicates its DNA, assembles new virions and eventually lyses the host. If the lysogenic cycle is established, the phage genome is integrated into the host chromosome and replicates along with the host.

Adsorption

Phage infection begins with a binding process through a specific interaction between a phage and its receptor on the host cell surface [25]. The receptor for phage λ is LamB (maltoporin), an abundant outer membrane porin, which is utilized by *E. coli* for maltose uptake [26]. Crystal structure shows that LamB is composed of three identical subunits, each of which is formed by an 18-stranded antiparallel β-barrel (Figure 1.3A) [27]. Genetic studies showed that the short central tail fiber of λ, which is encoded by *J* gene, resides at the tail tip and directly interacts with LamB for phage attachment on *E. coli* outer membrane [28]. This interaction between λ gpJ and the receptor LamB determines the host range specificity. When λ gpJ was replaced with the tail fiber of bacteriophage 434, the binding of the engineered phage switched to another outer membrane protein, OmpC, which is typically recognized by phage 434 [29].

In addition to the genetic findings, the target searching process conducted by phage λ has served as a paradigm for studying virus-host interaction system [30, 31]. The virus-host finding process has been proposed as a two-stage, dimensional reduction model [32, 33]. In the first stage, free phages diffuse in a three-dimensional (3D) manner in liquid until they encounter bacteria. The second stage occurs when phages land on the cell and they undergo a two-dimensional (2D) "random walk" along the cell surface until they are captured by LamB receptors, followed by viral DNA ejection. A later study emphasized the interaction between phage λ and the receptors and proposed a kinetic model to describe the phage adsorption as a two-step process with an initial on-and-off reversible virus-host association followed by an irreversible binding for subsequent DNA delivery [32]. This model is based on the measurements that upon incubation of phage λ with *E. coli* strain Ymel, the concentration of free phages in the solution decreases over time. The decrease curve follows a double-exponential function with an initial rapid rate and a later slow decay time. Although these reports provide insights into the behavior of virus-host interaction, the underlying viral targeting finding mechanism, especially in terms of the distribution of the receptors on the cell membrane, was missing due to the limited resolution offered by the traditional genetic and biochemical approaches during the early time.

Recently, with widely applied high-resolution microscopy-based technologies, the previous interaction models based on the assumption that LamB proteins are uniformly distributed on the cell surface was challenged. A study using fluorescently labeled λ tails shows that LamB molecules accumulate in spiral patterns on *E. coli* outer membrane, and the distribution changes rapidly with time [34]. Another study demonstrates this helical organization of LamB receptors by labeling them with quantum dots (QDots) (Figure 1.3B) [35]. Moreover, in the latter study, the dynamics of individual fluorescently labeled phages were monitored in real time at the single-cell level, from when they landed on the cell surface until they arrived at their destination. The phage trajectories also exhibit a non-isotropic, helical motion pattern with a tendency to move along the short axis of the cell, in agreement with the spatial organization of the receptors. Furthermore, when the dwell times of the labeled phage particles were examined, it was found that the phages spent 73.6% of their total trajectory time (~ 1 s) on average in receptor-rich regions. However, in the absence of LamB, the arriving phages immediately fell off from cell surface with a more than 15-fold decreased dwell time compared with that when LamB was present. All of these observations indicate that the viral searching motion on the cell surface, instead of the “random walk”, is predominantly governed by the interaction between the phages and the receptor network [31, 35]. This study not only verified the two-step search model proposed in the earlier work [32], but also incorporated the effect of the phage-receptor interaction into the model, enhancing our understanding of the viral target-searching process.

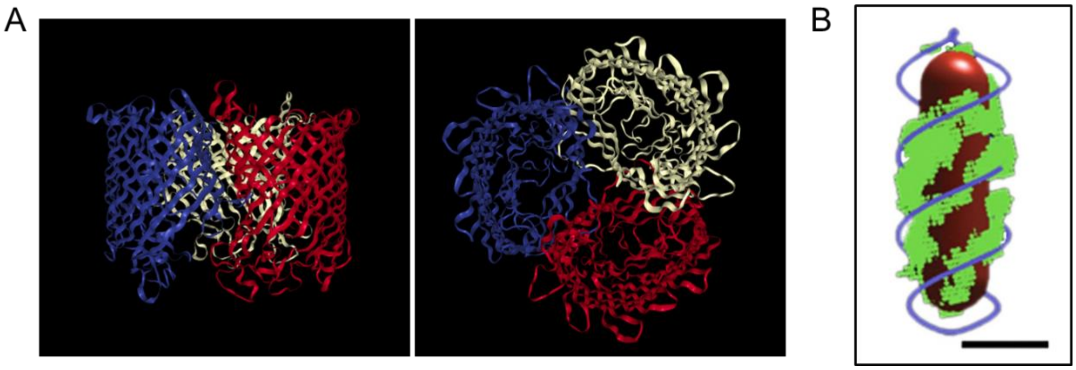

Figure 1.3 Crystal structure of LamB protein and arrangement of LamB network on cell surface.
(A) Crystal structure of LamB adapted from [27]. Left: side view. Right: top view. The three monomers of LamB trimer are shown in different colors (red, blue and yellow). (B) A rendered 3D image obtained by sectioning epifluorescence microscopy showing the distribution pattern of LamB (green) as continuous helices (blue) on cell surface (red). Scale bar represents 2 μm. This image is adapted from [35].

Additionally, by labeling phage λ with either quantum dots or fluorescent protein fused to phage capsid decoration protein gpD, phage λ was found to show a strong preference for adsorption at the cell pole, as well as the mid-cell (future cell pole) region [35-38]. Accordingly, the spatial distribution of LamB also exhibited a higher concentration around the cell pole, further suggesting the role of LamB spatial organization in the λ target searching process. In addition, it is worth noting that this preference for binding to the cell polar positions is not restricted to λ but instead shared by other bacteriophages, such as coliphages T4 and T7, T4-like vibriophage KVP40 and T7-like *Yersinia* phage ϕA1122 [36].

To summarize, the current picture of the phage λ adsorption process describes that a phage first follows a free 3D diffusion movement in liquid until the tail fiber gpJ find its receptor LamB on the host cell surface. The initial reversible interaction between gpJ and LamB slows down the diffusion of the phage and stalls the phage in receptor-rich regions. Subsequently, the phage docks onto a receptor through a strong and irreversible binding. Thereafter, phage DNA ejection is triggered. Quantitative measurements of phage adsorption kinetics in bulk show that the rate of phage adsorption depends on the concentration of magnesium sulfate ($MgSO_4$) in the solution and is strongly affected by temperature, and the higher the temperature, the faster the adsorption [32, 39, 40]. Furthermore, the density of LamB in the *E. coli* outer membrane controlled by whether maltose (an inducer of the LamB expression) or glucose (a repressor of the LamB expression) is presented or not also have a great impact on phage adsorption rate [15, 32, 40].

On the other hand, as mentioned in the previous section, nearly all λ strains currently used around the world are descendants of a mutant strain, λ*PaPa*, which contains a frameshift mutation (a deletion of cytosine) in the *stf* gene, leading to the loss of four side tail fibers [15]. The original variant isolated from the *E. coli* prophage [10], referred to as Ur-λ, was found to make very small plaques on the cell lawn. The small plaque size is perhaps because the side tail fibers extending from virions increase the hydrodynamic drag of phage particles, thus slowing down the phage diffusivity through the agar in phage titration experiments [41]. On the contrary, the large plaques produced by λ*PaPa* make it more suitable for genetics studies of λ, for example, the identification of *cI*, *cII* and *cIII* genes, which are responsible for making decisions between lytic and lysogenic development. Notably, although the side tail fibers seem not to be essential for λ infection, their presence significantly accelerates the rate of adsorption onto the host cell surface by ~7.4 fold [42]. That is because besides the binding to the primary receptor LamB by the tail fiber gpJ, the extra side tail fibers on Ur-λ provide additional interaction with their own receptor, another host outer membrane protein OmpC, thereby contributing to the more efficient adsorption [12, 15]. Previous studies evince that a higher adsorption rate would be beneficial for phage fitness and propagation in nature [42-44]. That is because phages with faster adsorption ability could encounter and hijack a host more rapidly compared with those phages with a lower adsorption rate. Additionally, when T4 tail fiber genes (*gp38* and part of *gp37*) were substituted by part of λ *stf* gene (ORF314) and *tfa* gene (encodes tail fiber assembly protein), the resulting T4-λ hybrid phages had a different host range from T4 phage [45], indicating that the side tail fibers participate in determining the host range specificity as λ central tail fiber. However, to our knowledge, no research has yet been conducted to examine whether the presence of the side tail fibers would affect the phage λ adsorption behaviors and the cell-fate determination at the resolution of individual cells and phages.

DNA Ejection

After locking onto a LamB receptor, mostly at cell poles or mid-cell, phage λ DNA is ejected into the host cytoplasm for subsequent replication and gene expression. However, the mechanism of phage DNA ejection, as well as what triggers phage DNA delivery, remains poorly understood. In addition to LamB, which is required for λ DNA ejection, it has been reported that ManY, an *E. coli* inner membrane protein, facilitates the penetration of λ DNA across the inner membrane [46, 47]. When GFP-ManY fused proteins were provided from a plasmid under the control of the arabinose-inducible promoter *pBAD* in a *manY* mutant *E. coli* strain, GFP-ManY was shown to localize to the cell poles or mid-cell positions where phage DNA ejection preferentially occurs [36]. However, ManY is not required for phage adsorption as the adsorption preference for polar positions was not altered with the deletion of ManY [36].

Early studies on the kinetics of DNA ejection *in vitro* suggest that phages can eject their genomes into LamB-containing liposomes within 1 min [48]. However, in a later report, real-time measurements of λ DNA ejection at the single-molecule resolution show that phage genome translocation can be finished within as little as around 10 s, with the maximal velocity of ~ 60 kbp/s [49]. Recently, *in vivo* experiments at the single-cell/single-phage level have been performed to characterize the kinetics of λ DNA ejection by labeling λ DNA with a fluorescent dye (SYTOX Orange) and measuring the loss of fluorescence intensity from phage capsid and the concomitant increase of signal in the infected cell [50]. In contrast to the *in vitro* studies, the study reveals that it takes a mean of roughly 5 min for DNA ejection to complete. Individual ejection trajectories also exhibit pauses that last for a few minutes, leading to significant cell-to-cell variability in terms of DNA ejection time, ranging from 1 min to 20 min. The difference between *in vitro* and *in vivo* studies indicates that the natural barriers, like the bacterial membrane, would reduce the rate of viral DNA translocation and give rise to an expanded variability of the process.

Furthermore, phage DNA ejection process is known to be affected by phage genome length and some environmental factors, such as temperature and extra-cellular ionic supplies [49, 51-53]. A great majority of λ DNA is packaged very tightly in the capsid, and the pressure caused by DNA-DNA interactions within the head can drive the ejection kinetics [50, 53]. Thus, the length of a phage genome is an important parameter to control the initial DNA ejection rate. *In vitro* measurements showed that longer phage genomes result in faster ejection speed and vice versa [49, 53]. However, it is worth noting that the conclusion about the driving parameter for *in vitro* ejections was contrasted with that drawn from the *in vivo* experiments. The results from *in vivo* experiments suggested that the ejection kinetics was governed by how much DNA has been ejected, instead of the amount of DNA remaining inside the viral capsid for *in vitro* ejections [49, 50]. The inconsistence might be attributed to osmotic pressure inside the host cytoplasm and the presence of bacterial membrane and proteins associated with the ejection process.

Besides, temperature has a strong effect on DNA ejection rate as well [53]. DNA ejection cannot be triggered at temperature as low as 4 °C. However, the triggering occurs immediately at 37 °C, and after a short period of time within the temperature range of 20-30 °C [51]. These findings are of significant importance for understanding the biophysical aspect of phage DNA delivery. Nevertheless, the molecular basis of the ejection process remains enigmatic. In addition, it also remains unknown whether having the side tail fibers, Ur-λ would eject its DNA in a more efficient and synchronized manner by providing stronger attachment of phage particles onto the cell surface.

When it comes to the effect of phage λ DNA entry site on phage post-infection, it appears that infections at the cell pole and mid-cell area show a lower failed infection frequency than those at other regions [37]. Here, the failed infection is likely the result of unsuccessful or incomplete DNA ejection. The higher successful infection frequency at the polar region might be attributed to the localization of ManY to the cell pole, facilitating the translocation of phage DNA through cell inner membrane. On the other hand, it has been pointed out that the entry site does not affect the lysis-lysogeny decision-making outcome [37]. Overall, as DNA ejection is a critical step of λ life cycle, a detailed understanding about its mechanism awaits further investigation.

Phage DNA Intracellular Dynamics

λ DNA is packaged as a linear form in the phage capsid. Once it is released into the host cytoplasm, the linear DNA molecule immediately circularizes through base pairing the complementary single-stranded 5’ extensions of 12 bp at both ends (*cos* sites). The nicked circle is then closed by *E. coli* DNA ligase and eventually becomes a negative supercoiled structure with the help of host DNA gyrase [14, 22]. Early after the circularization, two phage λ-encoded replication proteins, O and P, are expressed from *pR* promoter. The O protein binds to *oriλ*, the origin of replication in λ genome located inside the *O* gene, to initiate the DNA replication; P protein recruits *E. coli* host DnaB helicase and together binds to the O protein at *oriλ*, forming a DnaB-P-O-*oriλ* complex [54-56]. Thus, the bidirectional θ (circle-to-circle) mode of phage DNA replication starts to proceed soon after infection. At this stage, the number of phage DNA molecules doubles every 2-3 min in each infected cell [57]. The bidirectional θ replication proceeds for about 5 to 6 rounds, resulting in the appearance of 50-100 copies of λ DNA in the cytoplasm. Then, at about 15 min after infection, a few of the DNA molecules start to replicate in σ mode, which is also referred to as rolling-circle replication [22]. During the rolling-circle replication, long concatemers of λ DNA, up to ~10 λ genome equivalents in length, will be produced and further cut at *cos* sites, serving as the substrates for encapsidation [14, 22]. It has been proposed that σ replication may be preceded by one round of unidirectional θ replication, due to inefficient transcriptional activation of *oriλ*, caused by depletion of the host DnaA protein [58]. As λ transcription is necessary for the establishment of the bidirectional θ replication in the *in vitro* system and DnaA can stimulate transcriptional activation of *oriλ*, DnaA is thought to be an important factor in triggering the switch from the bidirectional θ to the unidirectional θ mode and then to the σ replication mode [58-60]. Despite some factors responsible for this replication switch and the corresponding models have been characterized and proposed, the mechanism of the regulation of this switch, as well as what triggers the switch, has not been thoroughly elucidated.

With the power of single-molecule visualization techniques, the intracellular phase of λ life cycle has been probed and examined in order to achieve a more comprehensive understanding of phage λ system. A number of fluorescence labeling approaches have been established to label λ genome, allowing for the examination of the genome delivery process and the subsequent intracellular dynamics in real time by fluorescence microscopy. To detect the spatial position where phage DNA ejection occurs *in vivo*, one effect has been made by inserting an array of 64 copies of *lacO* operators into the λ genome and expressing LacI-FP (FP, fluorescent protein) in the infected cells [36, 61]. Following the infection, individual λ genomes carrying the *lacO* array can be visualized as they are bound by LacI-FP. The results show that phage DNA ejection preferentially takes place at cell poles, as does the adsorption of phage particles on the cell surface. Moreover, another group engineered a *parS* sequence into the λ genome, so that λ DNA can be tracked by introducing ParB-FP in host cell during the infection cycle [62]. In this study, phage DNA shows confined local motion close to its entry site, most likely due to λ DNA being anchored to a site on the cell membrane where it enters the cell. Similarly, the visualization of phage λ genome was also achieved by utilizing SeqA-FP binding to methylated DNA, resulting in the labeling of a subset of phage DNA molecules: the initial ejected one and its first replicated copy [38]. The incoming phage DNA molecules exhibit subdiffusive behavior and show two distinct modes of movement, localized motion and motion spanning the whole cell. The motion varies from cell to cell and this heterogeneous phage DNA dynamics might generate different patterns of intracellular phage DNA interactions, thereby impacting the infection outcome [63, 64].

In addition, it appears that there is no correlation between phage DNA dynamics and the cellular decision-making, as phage DNA moves similarly before either the lytic or lysogenic pathway is chosen [38]. However, during the establishment of lysogenic pathway, phage DNA moves locally at confined subcellular areas, while the bacterial *attB* locus, which was fluorescently labeled by either ParB/*parS* or *tetO*/TetR reporter systems, migrates toward the region enriched with λ DNA for prophage integration [62, 65]. This movement of bacterial chromosome seems not to be driven by phage infection, but instead, attributed to DNA segregation driven by *E. coli* DNA replication machinery [62]. Surprisingly, an interesting "lyso-lysis" phenomenon was observed, where phage DNA is integrated into *E. coli attB* locus as does for lysogenization, followed by cell lysis [65]. This suggests that phage DNA molecules can behave differently within the same cell and each phage DNA molecule is able to independently vote for cell fate. In other words, the mechanism of decision making could be zoomed into individual phage DNA level. Taken together, present studies lay the foundation for future investigation on the spatiotemporal dynamics of phage λ genome inside the host cell. Nonetheless, many open questions still remain to be addressed: Does the bacterial chromosome movement influences the intracellular organization of phage λ genomes? How do replicated λ genomes interact with each other to finalize the lysis-lysogeny decision? What effect do the intracellular components, which are produced by both bacterium and phage, have on the phage DNA spatial patterns and the subsequent infection outcome?

Lysis-lysogeny Decision Making

Once phage DNA is circularized after being ejected into the host cytoplasm, λ gene transcription is initiated by the bacterial RNA polymerase binding to the early promoters *pL* and *pR* in opposite directions (Figure 1.4). At this stage, RNA polymerase molecules will stop the transcription at the *tL1* and *tR1* terminators, resulting in the production of only two genes, *N* from *pL* and *cro* from *pR* [17, 21]. N protein is an anti-termination regulator that enables to modify the RNA polymerase molecules so that they can transcribe beyond the *tL1* and *tR1* terminators, allowing the other downstream genes to be transcribed. From the *pL* promoter, the gene *cIII*, whose product helps establish the lysogenic pathway, and the recombination genes are turned on. In the *pR* operon, the first gene transcribed after *cro* is *cII*, known as the central regulator in the lysis-lysogeny decision-making process [12]. Following the transcription of *cII*, phage DNA replication genes *O* and *P*, as well as the late regulatory gene *Q* are further expressed [18].

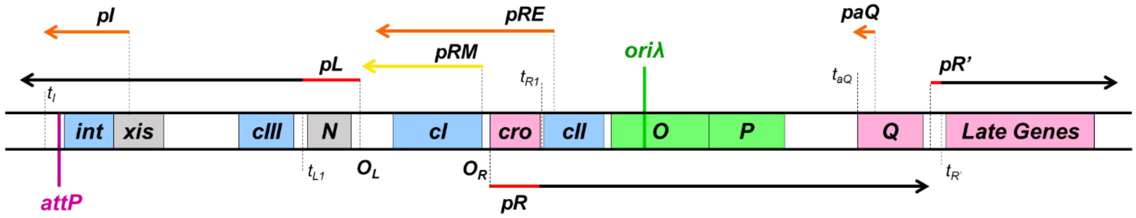

Figure 1.4 Gene map and transcription map of phage λ regulatory region.
Key genes participating in the λ lysis-lysogeny decision-making are shown in their map order as boxes. Genes involved in the lytic cycle are colored in pink, while genes involved in the lysogenic cycle are colored in blue. The horizontal arrows indicate gene transcription with promoter names (p). Red: immediate-early transcripts; black: the early transcripts from *pL* and *pR* and the late transcripts from *pR'*. The CII-activated *pRE*, *pI*, and *paQ* transcripts are indicated with orange arrows. The *pRM* transcript which is activated by CI repressor is shown in yellow. The operators O_L and O_R where CI and Cro bind are shown next to *pL* and *pR*, respectively. Critical transcription terminators (t) are indicated around the genes. Phage λ replication origin site (*oriλ*) and the attachment site (*attP*) are marked in green and magenta, respectively.

Lytic Cycle

During the lytic development, Cro functions as a weak repressor that binds as a dimer to the operators O_L and O_R, each of which is composed of three repressor binding sites (O_{L1}, O_{L2}, O_{L3} and O_{R1}, O_{R2}, O_{R3}, respectively). Each binding site can be occupied by a Cro dimer, thus inhibiting the corresponding transcription [17]. The two operators control the activity of both *pL* and *pR* promoters, as well as the promoter *pRM*, which is responsible for the expression of CI repressor. The suppression of *pR* activity by Cro reduces the CII protein levels, making the infected cells in favor of the lytic cycle [21]. Besides, Cro exhibits a higher binding affinity to O_{R3}, which overlaps with *pRM*, leading to the effective repression of *pRM* to activate *cI* transcription. On the other hand, once the late gene regulator Q is synthesized and reaches a threshold, , it acts as an anti-termination protein, like N, to modify the RNA polymerase and facilitate it to initiate the transcription from the late promoter, *pR'*, which is located on the right next to *Q*. The *pR'* is responsible for the expression of the late genes that encodes the phage lysis and morphogenetic proteins. The lysis genes, *S*, *R* and *Rz*/*Rz1*, are clustered in a cassette immediately downstream of *pR'*, and produce five proteins (the holin S105 and antiholin S107 from gene *S*, the endolysin R, and the two spanin proteins from genes *Rz* and *Rz1*) causing host cell lysis [66]. Following the production of the morphogenetic proteins, i.e., head and tail proteins, new phage virions are assembled and finally released into the environment to infect surrounding cells.

Lysogenic Cycle

The establishment of the lysogenic mode highly depends on the λ repressor, CI. CI blocks the transcription from both *pL* and *pR* promoters through binding to the same operators O_L and O_R as Cro, while activating its own synthetic promoter, *pRM*, which lies immediately adjacent to the *cI* gene [14]. In contrast with Cro, CI dimers preferentially and cooperatively bind to O_{R1} and O_{R2} and O_{L1} and O_{L2} as tetramers to repress *pR* and *pL* transcription respectively. These two sets of tetramers together form a DNA loop between O_L and O_R, thereby enhancing the repression effect for the early promoters while stimulating *cI* itself transcription from *pRM* [17, 18, 21]. Although the repressor CI is crucial for establishing and maintaining the stable lysogenic state, it seems not to be involved in deciding the cell fate. Instead, CII is thought to play a key role in the decision-making process [17, 18, 67]. Once synthesized and accumulated until sufficient enough, CII functions as a transcription factor that activates three promoters, *pRE* transcribing the *cI* gene, *pI* transcribing the *int* gene and *paQ* transcribing an antisense RNA within the *Q* gene [68]. Int integrase introduces single-stranded breaks into host DNA at a specific site, *attB*, allowing phage genome to be integrated into the bacterial chromosome [69, 70]. The Q antisense RNA produced from *paQ* reduces Q protein synthesis, thus preventing cells from entering the lytic cycle [71].

The CII protein is known to be metabolically unstable and can be rapidly degraded by the host protease FtsH, which was originally identified as HflB (high-frequency lysogenization by phage λ) [72-74]. The half-life of the CII protein is critically depending on its initial level. It could be as short as 1.5 min at low initial concentration whereas as long as 22 min at high concentration [75]. Therefore, the cellular concentration of CII which remarkably affects the lysis-lysogeny decision outcome is determined by a balance between the rate of synthesis and the rate of degradation [12]. When the CII level is high, the lysogenic mode is preferred, otherwise the phage defaults to the lytic cycle. Besides, CIII protein functions as an inhibitor of FtsH and protects CII protein from degradation, thereby indirectly promoting the establishment of the lysogenic pathway [67]. Taken together, these regulators, as well as their interactions, constitute the complex phage λ genetic circuitry, giving rise to a commitment for an infected cell whether to lyse or to become a lysogen.

Factors Affecting the Lysis-Lysogeny Decision Making

In addition to the contribution of the λ gene regulatory network, many intracellular and extracellular parameters influencing cell fate choice have been well-characterized. At the population level, it has long been known that the frequency of lysogenization increases with multiplicity of infection (MOI, the number of phages infecting a cell), according to the study conducted by Kourilsky in 1973 [76]. This study also showed that based on the assumption that phage adsorption follows the Poisson distribution, the establishment of lysogeny requires at least two phages to infect a cell. In addition, host cell size or cell volume was correlated with the decision switch [37, 77]. Shorter cells exhibit a higher chance to be lysogenized, whereas longer cells are more likely to lyse. The effect of cell size suggests that instead of the absolute number of infecting phage particles, the viral concentration (given by the ratio of MOI to cell volume), in other words, the concentration of phage-encoded players, determines the decision-making outcome upon phage infection

[77, 78]. Furthermore, the physiological condition of host cell is considered as another important variable affecting cell fate [18, 76, 79]. Cells growing in rich media have higher concentrations of the host protease FtsH, which can degrade CII proteins, and the global regulator RNaseIII, which gives rise to rapid expression of N protein, conjointly resulting in the commitment of the cells to the lytic cycle [80, 81]. However, when cells undergo starvation (e.g., poor carbon condition), the lysogenization is favored as CII is relatively more stable under this condition and the decreased amount of N proteins reduces late gene expression.

During the last few decades, new techniques and powerful tools have been developed to examine the decision-making process at the single-cell/single-phage/single-molecule resolution. The further combination of the observed behaviors of host-phage interaction with quantitative analysis and mathematical modeling makes it possible to uncover the hidden deterministic factors beyond those aforementioned preexsisting variations. Microscopy-based studies have successfully labeled the phage λ capsid with fluorescent proteins through genetic engineering in order to visualize and track individual phage particles during the infection cycle [36, 37, 82, 83]. After coupling with a lysogenic fluorescent indicator, a unanimous voting model was first raised that individual phages are able to independently make decisions between lysis and lysogeny inside the host. Only when all infecting phages vote for cell lysogeny can lead the infected cell to commit to the lysogenic pathway; otherwise, the lytic pathway is favored [37]. Instead of individual phage voting, a ‘partial gene dosage compensation’ scenario provides an alternative explanation for the decision-making based on the same experimental data from [37]. It describes that in the host cell, the viral gene expression from multiple infecting phages may not linearly increase with MOI, but instead, is partially compensated by a degree of MOI$^{1/2}$, thereby influencing the overall cell fate decision-making [84]. Up to this point, it was unclear how individual phages behave and interact with others when they co-infect the same cell. To address this question, a recent work rendered two phages with different fluorescent lytic and lysogenic reporter systems and found that λ phages compete against each other for host resources during cell lysis, while they cooperate during lysogenic development [83]. This finding provides further evidence that each individual phage DNA has the capability of deciding cell fates. Further characterization of hidden influential variables, such as the intracellular phage organization and the interaction between phage DNA molecules at the subcellular level, is still required.

Stochasticity Effect on Cellular Decision Making

Decision-making is a ubiquitous process among all organisms, from simple viruses and bacteria to complex mammals [85]. It occurs when a cell receives signals from the environment, and then makes appropriate responses for optimal fitness [86]. A multicellular organism comprised of a population of cells will take all the decisions made by individual cells to make a response to environment changes, such as protecting itself from harmful stress [86, 87]. At the single-cell level, individual cells can exhibit heterogeneous cell fates even if they are genetically identical and subjected to the same conditions [85, 88, 89]. This heterogeneity has been conjectured to be the result of intrinsic biochemical stochasticity, or noise, in gene regulatory networks [88, 90]. Apparently, gene expression is a stochastic process, as it arises from fluctuations in transcription and translation, reflected by different levels of mRNA and protein, and causes changes in cellular phenotypes [84, 88, 91, 92]. For phage λ, the genetic regulatory circuitry encoded by phage genes could generate the intrinsic stochasticity due to the transcription initiation and regulation, protein degradation and interactions among host factors and multiple phage-encoded players. This intrinsic stochasticity can affect the lysis-lysogeny decision-making outcomes and ultimately render cell-fate choices noisy.

Phage DNA replication is predicted to play an important role in establishing different cell fates, as it provides transcription templates for the expression of cell fate-determining genes, such as *cro, cI* and *cII*, thereby leading to the fluctuations in mRNA and protein production [63]. Phage DNA replication is initiated at the early stage once λ DNA is ejected into host cell [57]. Deficiency in phage DNA replication reduces the lysogenization frequency at the bulk level, implying the importance of DNA replication in the decision-making process [76, 79]. At the single-cell level, in absence of DNA replication, cells infected by a single phage are not be able to commit to either lytic or lysogenic decisions [93]. Moreover, it was recently uncovered that each replicated phage DNA molecule is able to interact with each other and independently vote for either lytic or lysogenic pathway within the same cell [83]. Altogether, by introducing copies of genetic elements and affecting the intracellular viral DNA concentration, DNA replication is capable of altering the overall decision-making outcomes. However, quantitative measurements of phage DNA replication in live-cell infections at the single-cell resolution is currently still incomplete.

To date, many mathematical and theoretical models have been built to elucidate λ decision making in terms of both pre-existing infection parameters and the intrinsic, stochastic genetic network. The theories gained from those studies could be applied on higher biological systems in order to make the response of organisms to environmental signals more predictable. Nevertheless, due to the limitation of resolution, there are still multiple features of cell-fate bifurcation requiring quantitative narrative and the complete story of viral life cycle remains to be investigated.

CRISPR-Cas Immune Systems

Bacteriophages and their host bacteria coexist in nature and are involved in continuous cycles of co-evolution to maintain equilibrium. Phages represent the most abundant biological entities on earth and have been estimated to outnumber bacteria by approximately ten-fold [94]. To preserve bacterial population, bacteria have evolved various defense strategies to circumvent the deadly threat from phage invaders. Those strategies include preventing phage adsorption to host cell surface by modifying the receptors, blocking phage genome delivery by superinfection exclusion (Sie) systems, digesting the invading phage genome by utilizing restriction-modification (R-M) systems to cut foreign DNA at specific recognition sites, and even committing to cell suicide by toxin-antitoxin (TA) systems to abort phage infection and prevent viral propagation in the bacteria population [8, 95, 96].

In the past few decades, a novel adaptive microbial defense system, termed CRISPR-Cas, composed of clustered regularly interspaced short palindromic repeats and CRISPR-associated proteins, has been discovered and generated great interest toward scientists in setting off a wave of research for the molecular mechanisms. The first observation of a CRISPR loci was made in the genome of *Escherichia coli* K12, downstream of the alkaline phosphatase (*iap*) gene [97]. It has been described as a series of short direct repeats of 29 base pairs (bp) interspersed by 32 bp short non-repetitive sequences. In subsequent years, similar structures were found widespread in genomes of various bacteria and archaea [98, 99]. Modern bioinformatic analyses reveal that CRISPR arrays are present in most archaea (87%) and roughly half of sequenced bacterial genomes [100-102]. Since the discovery of CRISPR arrays, a variety of hypotheses with regard to their functions have been raised, including DNA repair, chromosome partitioning and gene regulation [103, 104]. It was not until the year 2005 that three independent groups reported that the non-repeating sequences in CRISPR arrays matched phage genomes and plasmid sequences [105-107]. In conjunction with the findings that *cas* (CRISPR-associated) encoded proteins had putative nuclease and helicase domains [98, 108], CRISPR-Cas started to be considered as a prokaryotic defense system against foreign genetic elements. The function of CRISPR-Cas as an adaptive immune system was first experimentally demonstrated in 2007 by showing that the expansion of the CRISPR arrays, which were acquired following phage challenge, conferred resistance against subsequent infection [109]. Thereafter, numerous evidences confirmed the defense function in different species and further uncovered other novel features of CRISPR-Cas systems, for instance, the role of crRNA guides and the DNA targeting activity [110, 111]. More recently, based on extensive work on the mechanisms of action of CRISPR-Cas systems, researchers have realized the potential of these systems as a powerful tool for eukaryotic genome editing, host-virus evolutional surveys and various industrial applications [103, 112-114].

Diversity and Application

CRISPR-Cas immune systems exhibit remarkable structural and functional diversity due to rapid and constant coevolution with mobile genetic elements (MGEs), such as phages and plasmids [115, 116]. Despite this diversity, all CRISPR-Cas variants that have been discovered so far share some conserved mechanistic features. In general, a functional CRISPR-Cas immune system consists of a CRISPR array and a set of *cas* genes. The type I-E CRISPR-Cas locus is shown as an example in Figure 1.5. The CRISPR array typically comprises an upstream AT-rich leader region, and a series of highly conserved repeat sequences interspaced by variable sequences, known as spacers, derived from invasive virus or plasmids [117-119]. The spacers are the key elements as they function as a bacterium host’s genetic memory of past encounters with invading MGEs. This memory provides protection for the host from being infected by MGEs harboring the complementary sequences [120, 121]. The operon of *cas* genes is always located adjacent to the CRISPR array and encodes the Cas proteins that contain functional domains, such as endonuclease and exonuclease domains, helicase domains and DNA/RNA-binding domains [98, 108]. The CRISPR loci and the *cas* genes vary greatly owing to their heterogeneity in nature, sequence, and architecture. Meanwhile, horizontal gene transfer of the entire CRISPR loci or individual modules among organisms also contributes to the extraordinary diversification of CRISPR-Cas systems, driven by the everlasting virus-host arms race with the invasive MGEs [120, 122].

Classification

Currently, according to the assortment of *cas* genes and the locus architecture, CRISPR-Cas systems have been classified into two classes (class 1 and class 2), each of which is further subdivided into three types (types I, III and IV and types II, V and VI, respectively) and 33 different subtypes (Table 1.1) [102, 123]. The major difference between class 1 and class 2 is the organization of their Cas effector nucleases. Class 1 systems are characterized by multi-subunit surveillance complexes for interference, whereas in class 2 systems, the surveillance and interference are accomplished by a single effector protein. In particular, class 1 systems are the most common and widespread systems in both bacteria and archaea, while class 2 systems are almost completely restricted to bacteria, implying that class 1 might be the ancestry [102, 124]. Moreover, it is worthwhile to note that although the *cas* gene families required for executive functions are greatly variable, *cas1* and *cas2* genes are universally conserved across almost all known CRISPR-Cas systems [115, 124]. Their encoded products, Cas1 and Cas2 proteins are responsible for the genetic recording of infections through the acquisition of spacers from invasive MGEs [125-127]. Furthermore, many organisms are found to harbor more than one type of CRISPR-Cas system, indicating that these systems are compatible and capable of sharing functional components [115, 128].

Table 1.1 Classification of CRISPR-Cas systems.

Class	Type	Signature protein	Effectors	Type of target	Target recognition
Class 1	Type I	Cas3	Cas3, Cascade, crRNA	DNA	PAM
	Type III	Cas10	Cas10, Cmr/Csm, crRNA	DNA and RNA	rPAM
	Type IV	Csf1	unclear	unclear	unclear
Class 2	Type II	Cas9	Cas9, crRNA, tracrRNA	DNA	PAM
	Type V	Cfp1	Cfp1, crRNA, tracrRNA (for some subtypes)	DNA	PAM
	Type VI	C2c2	C2c2, crRNA	RNA	PFS

PAM, protospacer-adjacent motif.
rPAM, RNA PAM [129].
PFS, protospacer-flanking sequence [130].

Class 1 systems encompass the most common and diversified type I and type III, as well as the putative new type IV. Type I systems employ effector complexes, termed Cascade (CRISPR-associated complex for antiviral defense), for foreign DNA recognition, and the signature protein Cas3 to cleave target DNA [110]. Although the Cascade complexes of the type I subtypes share similar overall architectures, each subtype (I-A to I-F and I-U) has defined compositions and distinct features of operon organization [112]. Likewise, type III systems possess Cascade-like complexes (Csm for III-A/D and Cmr for III-B/C) [131]. The signature gene for type III systems is *cas10*, which encodes Cas10 nuclease and serves as the largest subunit of type III effector complexes. Cas10 contains an HD nuclease domain that cleaves foreign ssDNA and two Palm domains, which are responsible for the synthesis of cyclic oligoadenylates (cOAs) [132]. The effector complexes of both type I and type III display similar elaborate architectures, with a backbone comprising paralogous repeat-associated mysterious proteins (RAMPs), such as Cas7 and Cas5 subunits, which contain variants of the RNA recognition motif (RRM) domain [124, 133]. However, in contrast to type I systems, type III immunity requires target transcription and can cleave both DNA and RNA in a co-transcriptional manner [134]. Target degradation is triggered by the binding of the effector complex to nascent mRNA, which is transcribed from foreign DNA, through base paring of CRISPR guide RNA (crRNA) to its complementary transcript. Subsequently, the Csm3 and Cmr4 subunits cleave the transcript (ssRNA) at 6-nt intervals. The HD domain of Cas10 cleaves ssDNA within the transcription bubble. Meanwhile, the synthesized cOAs by Cas10 Palm domains act as second messengers that activate the RNase activity of Csm6 to nonspecifically destroy foreign transcripts. [135-137]. Notably, DNA degradation strictly depends on the transcription of the target sequence for the type III systems. Type IV systems are rather rare and often found on plasmids containing rudimentary CRISPR-Cas loci in the absence of the known effector proteins involved in adaptation (Cas1 and Cas2) and target cleavage (Cas3 or Cas10) [102]. A most recent study revealed that the type IV systems are able to employ crRNA-guided effector complexes [138]. More comprehensive experimental data on type IV-mediated immunity and its functionality are still required.

Class 2 systems have a much simpler and more uniform organization compared with class 1 systems, including the best-characterized type II, much rarer type V and the putative new type VI system. Type II systems require a single effector protein, Cas9, for RNA-guided DNA recognition and cleavage [139, 140]. Cas9 is a large multi-domain endonuclease with two unrelated nuclease domains, HNH and RuvC, which are responsible for the cleavage of the DNA complementary strand (target strand) and the displaced strand (target strand), respectively. The type II CRISPR-Cas locus also encodes a small trans-activating crRNA (tracrRNA) that bears partial complementarity to the repeat portions of crRNA. The tracrRNA is essential for the crRNA maturation process, which is catalyzed by Cas9 and host ribonuclease III [141]. The mature crRNA-tracrRNA duplex is bound and stabilized by Cas9 and then guides Cas9 to the target DNA adjacent to a specific motif, PAM (protospacer adjacent motif) [142, 143]. Consequently, a blunt, double-stranded break (DSB) is introduced by Cas9 at 3 bp upstream of the PAM in the target DNA [140, 144]. In addition to Cas9, all identified type II modules encode *cas1* and *cas2* genes for spacer acquisition [145]. The effectors of type V systems have been identified as Cas12a (Cpf1), Cas12b (C2c1), and Cas12c (C2c3) for the three subtypes V-A, V-B, and V-C, respectively [146, 147]. Cas12 variants all contain only one nuclease domain (RuvC-like) but lack the HNH domain that is present in Cas9. Recent evidence indicate that Cas12a is a single-RNA-guided nuclease that does not require tracrRNA for activity, whereas the maturation of crRNA and target DNA cleavage by Cas12b depends on the presence of both crRNA and tracrRNA [146, 148]. Moreover, unlike Cas9, Cas12a and Cas12b cleave target DNA by generating staggered DSBs with 5- and 7-nt 5’ overhangs, respectively, distal to PAM [148, 149]. Finally, a unique effector protein Cas13 (C2c2) with two HEPN (higher eukaryotes and prokaryotes nucleotide) binding domains, which are frequently found in RNases, is identified as the signature of the type VI systems [146]. This effector is programmed by a single crRNA without the requirement of tracrRNA and the execution is activated by target single-stranded RNA carrying sequences complementary to the crRNA [146, 150]. Remarkably, after activation by the target RNA, Cas13 has been found not only to degrade the target ssRNA, e.g. the genome of the RNA phage MS2, but also to destroy cellular mRNAs nonspecifically, leading to restriction of bacterial growth [150]. It implies that type VI systems appear to prevent viral propagation by causing programmed cell death (PCD) or dormancy induction.

In summary, CRISPR-Cas systems are DNA-encoded and RNA-mediated adaptive immune systems, which provide protection for archaea and bacteria against invasive nucleic acids. Although the current identified CRISPR-Cas variants share similarities in certain structural and functional features, there is substantial variability between the two classes, six types and their subtypes with respect to the gene composition, genomic locus architecture, and the nature and actual sequences of Cas proteins. In terms of the nature of target nucleic acids, type I, II and V cleave DNA, whereas type VI targets RNA. Type III targets both DNA and RNA, while type IV target remains unclear. A major contribution to the enormous diversity of CRISPR-Cas systems is the frequent module shuffling from extensive recombination of adaptation and effector modules between different variants under strong natural selection pressure [124, 133]. Particularly, this diversity reflects their various biological functions and the corresponding potential for applications in scientific, biotechnological and industrial fields.

Application

Most applications of CRISPR-Cas systems have focused on developing programmable genome editing tools in eukaryotes. Type II systems with the unique effector nuclease Cas9 become the best candidate for exploitation in genetic engineering [114, 147, 151]. In the presence of guide RNAs, the cleavage activity of Cas9 could efficiently create indels (insertion/deletion mutations) at specific sites of target DNA. For simplicity, the dual crRNA-tracrRNA has been engineered as a single guide RNA (sgRNA), together with Cas9, constituting a two-component system [140]. Thus, by designing guide sequences, this system can be harnessed for many varieties of genetic manipulations for any gene of interest, including gene knockout and precise editing coupled with cellular DNA repair machinery or homologous recombination [152-155]. Alternatively, by taking advantage of the DNA-binding capacity of Cas9, a catalytically inactive Cas9 variant, dCas9 ("dead"), has been exploited for transcriptional activation or silencing, DNA modification and localization of fluorescent protein labeling [156-158]. Despite the aforementioned advances of CRISPR-Cas9 in genome editing, Cas9 has limitations due to the off-target effects, which can not to be neglected [159]. Much effort has been devoted to detect and minimize off-target cleavage in order to improve the genome-wide specificities of Cas9 nucleases [160, 161]. In addition, some progress has been made on developing methods to deliver CRISPR-Cas9 editing systems to eukaryotic cells and specific tissues for human therapeutics [162, 163].

On the other hand, before the discovery of the defense function, the first application of CRISPR was genome typing [164]. At the very beginning, the developed

CRISPR-based spoligotyping method was used principally for *Mycobacterium tuberculosis* strains, but now has become a standard method for genotyping many other bacterial strains [99, 119]. Furthermore, as CRISPR spacers represent the prior exposure to MGEs, they store a historical and geographical perspective for a particular strain, providing the potential of use for ecological, evolutional and epidemiological research [103]. In addition, the immunity feature of CRISPR opens up possibility for industrial application by protecting industrial bacteria against phage infection. Taken together, owing to their simplicity, efficiency and ease of implementation in a wide variety of organisms, CRISPR-Cas based technologies offer assorted applications in a broad range of fields. Further exploration in CRISPR molecular mechanisms and improvements of CRISPR tools are still in high demand, as they could help expand new applications to other biological research and fields of agriculture and biotechnology, as well as bring potential for therapeutic gene therapy and next-generation genome editing in the near future.

Type I-E CRISPR-Cas system

Among numerous diverse CRISPR-Cas systems, type I systems are the most frequent and widespread systems, accounting for 95% of all known CRISPR systems [115, 165]. In particular, the type I-E CRISPR-Cas system in *Escherichia coli* has been extensively characterized with regard to the structure and mechanism. The type I-E CRISPR-Cas locus in *E. coli* K12 consists of a *cas* operon containing eight CRISPR-associated (*cas*) genes and two CRISPR arrays, CRISPR1 and CRISPR2, located at ~ 62 min on the chromosome (Figure 1.5) [118, 166]. Both arrays are preceded by a conserved

AT-rich leader sequence, which acts as a promoter, and contain the same sequence of palindromic repeats (29 bp) and the intervening 32 or 33 bp spacers [166, 167]. Like other systems, following the invasion of either phages or plasmids, the mechanism of immunity involves three distinct stages: adaptation, expression and interference (Figure 1.6) [104, 112, 124]. During the adaptation stage, short fragments derived from foreign DNA (known as protospacer) are captured, processed and incorporated as new spacers into the host CRISPR array, serving as a genetic memory of prior invasions. In the second expression stage, a long precursor CRISPR RNA (pre-crRNA) is transcribed from the CRISPR locus by host RNA polymerase, followed by processing into small mature CRISPR RNA units (crRNAs). Each crRNA is then assembled with Cas proteins, forming a surveillance complex. Finally, in the interference stage where the immunity is executed, Cas3 nuclease is recruited by the crRNA-guided surveillance complex toward invasive nucleic acids for cleavage and degradation.

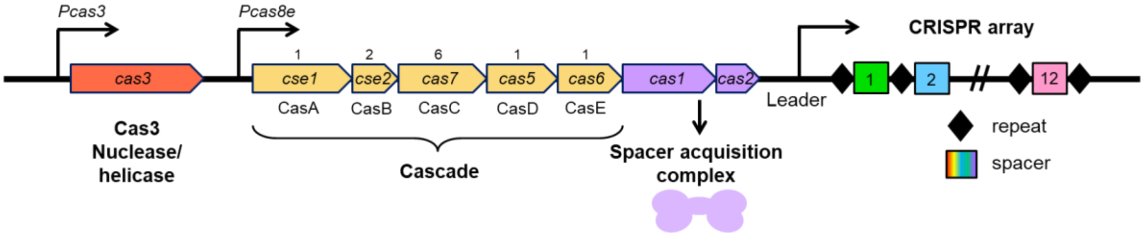

Figure 1.5 Schematic of the type I-E CRISPR-Cas locus in *E. coli* K12.
Five *cas* genes (*A*-*E*) in yellow arrows encode the subunits of the Cascade surveillance complex with their stoichiometry labeled above each gene. The *cas3* gene encodes the signature protein, Cas3, which is responsible for target dsDNA unwinding and cleavage. The gene expression products of *cas1* and *cas2* form the adaptation complex, colored in purple, functioning in the acquisition of new spacers from invading nucleic acids. The shown CRISPR array is CRISPR1, comprising 12 unique spacers indicated as squares with different colors and separated by repeat sequences marked as black diamonds. Different from CRISPR1, CRISPR2 (not shown here) only contains 6 spacers. Promoters are represented as horizontal arrows.

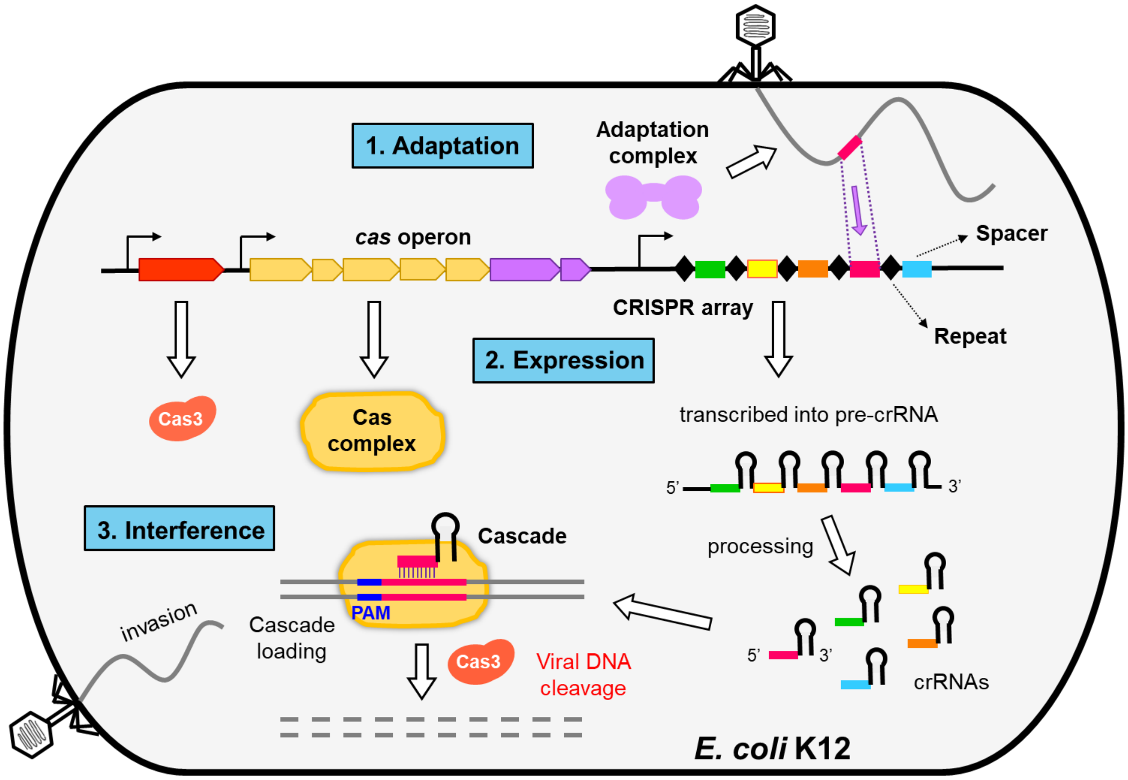

Figure 1.6 Overview of the three stages of the type I-E CRISPR-Cas immune system. The CRISPR-Cas locus consists of the *cas* operon with a set of *cas* genes and CRISPR arrays, which are shaped and colored corresponding to Figure 1.5. During the adaptation stage, short DNA fragments (magenta) derived from viral DNA or other MGEs are captured by the adaptation complex (purple) and integrated into the CRISPR array. In the expression stage, the CRISPR array is transcribed into a long precursor CRISPR RNA (pre-crRNA), which is further processed into small mature crRNAs by Cas6 endonucleases. Meanwhile, *cas* genes encoding diverse Cas proteins (dark yellow) are transcribed and translated, followed by assembling into a Cascade surveillance complex with one crRNA. During the last interference stage, upon infection by viruses or plasmids, crRNA guides the surveillance complex toward the invading DNA (grey) and scans sequence complementary to the spacer sequence (magenta). After recognizing the PAM motif (blue) and base pairing with the target DNA, Cas3 nuclease (red) is recruited by the Cascade complex to catalyze degradation of the target nucleic acid. Consequently, the host is protected and survives.

Adaptation

Adaptation, also referred to as spacer acquisition, is a process in which new spacer sequences are captured and integrated into the CRISPR locus to immunize host cells. This process is accomplished in multiple steps: protospacer selection, prespacer generation and spacer integration [112, 168, 169]. Two different modes have been described for type I systems: naïve and primed adaptation [125, 126, 170]. Naïve adaptation occurs when the host obtains spacers from an invader that has not been encountered previously, whereas primed acquisition relies on a pre-existing spacer that partially matches the invading MGEs, thus stimulating the uptake of new spacers from the same invader [169, 170]. Two proteins, Cas1 and Cas2, play key roles in both modes of adaptation. Two Cas1 dimers are bridged by one Cas2 dimer, forming a ~150 kDa, stable complex in which Cas1 is catalytic and Cas2 serves a structural role (a simplified cartoon is shown in Figure 1.5) [171]. In the *E. coli* type I-E CRISPR-Cas system, Cas1 and Cas2 are the only required Cas proteins for naïve spacer acquisition [125]. However, primed spacer acquisition also requires the involvement of the interference machinery, including both the effector complex Cascade and the Cas3 nuclease [126]. Primed adaptation is much more efficient than naïve adaptation and it allows the host to rapidly accumulate a broader range of resistance against diverse phage variants.

To initiate adaptation, a large pool with short foreign DNA fragments must be first generated in the host cell, serving as the source of new spacers. A study on naïve acquisition in *E. coli* reveals that the bacterial RecBCD repair complex continuously degrades exogenous DNA, which replicates more frequently and contains less Chi sites compared with the host chromosome, leading to the formation of substrates for spacer capture and integration by the adaptation complex [172]. For primed adaptation, the generation of the DNA fragment pool is coupled with the interference machinery. Currently, there are two models explaining this process. Upon loading of Cascade on target DNA, one model proposed that Cas3 unwinds and degrades the target DNA, producing 30 - 100 nt ssDNA fragments as precursors for new spacers [173, 174]. As shown in the *in vitro* studies, instead of ssDNA, dsDNA is the preferred substrate for the Cas1-Cas2 complex [173]. Therefore, the ssDNA fragments generated by Cas3 are most likely to re-anneal before being captured by the adaptation complex. Alternatively, direct visualization of Cascade and Cas3 in real time at the single-molecule level provides evidence for another model: following the target recognition, Cascade surveillance complex recruits nuclease inactive Cas3 and Cas1-Cas2 complex. Then this large Cas3/Cas1-Cas2 complex translocates bi-directionally along the target DNA and directly excises double-stranded prespacers for subsequent processing [175-177]. Notably, primed spacer acquisition occurs in a strand-biased manner, where the newly acquired spacers are biased to match the same strand as the original priming spacer [126, 178].

A question may be raised here: How are the prespacers selected by the adaptation machinery? The selection is based on the recognition of the short protospacer adjacent motif (PAM), which is only present in the foreign nucleic acids and used by host CRISPR-Cas systems for discrimination between self and non-self [143]. For the type I-E system, Cas1 and Cas2 are sufficient for PAM recognition. The formed Cas1-Cas2 complex scans target DNA sequences for potential PAMs, displaying a higher binding affinity for the prespacer with a canonical PAM than that with a noncanonical PAM [179, 180].

Following the generation of the DNA fragment pool, the Cas1-Cas2 complex binds to a prespacer containing a central 23-bp double-stranded DNA with splayed ends [127, 179]. Cas1 cleaves the 3’ single-stranded regions, one of which bears a PAM, exposing a 3’ OH group on each overhang. When the Cas1-Cas2 complex recognizes the CRISPR array, the free 3’ OH groups carry out two consecutive nucleophilic attacks on the phosphodiester bond on the opposite sides of the first repeat, with the first attack on the 5’ end of the repeat bottom strand and the second one targeting the repeat-spacer boundary [181]. Consequently, new spacers are integrated by the Cas1-Cas2 complex with the correct PAM orientation at the leader-proximal end of the CRISPR array [168, 182]. This polarization of spacer incorporation driven by the integration host factor (IHF) is a universal feature of all CRISPR-Cas systems, enabling the host to optimize the immunity response to the latest invader [182, 183].

Expression

In response to the invasion of viral or plasmid DNA, the type I-E CRISPR-Cas system needs to be expressed, including the production of Cas proteins and the transcription of the CRISPR array into a precursor CRISPR RNA (pre-crRNA), which is further processed into mature crRNAs (Figure 1.6). The transcription of the two CRISPR arrays (CRISPR1 and CRISPR2) is directed by their own promoters located within the leader sequences [167, 184]. Genes encoding the subunits of Cascade complex, as well as Cas1 and Cas2, comprise an operon, which is controlled by a single promoter P*cas8e*, while the nuclease Cas3 is transcribed from its own promoter, P*cas3* [167, 169, 184]. The *E. coli* type I-E system is regulated by several factors, including the histone-like nucleoid-structuring (H-NS) protein [167], cAMP receptor protein (CRP) [185], high-temperature protein G [186] and the LysR-type transcription factor LeuO [187]. H-NS, a global transcription repressor, is well known to strongly silence the expression of *cas* genes and the transcription of the CRISPR array, even though all the components are still functional. Only when the engineered host harbors the plasmids that can overexpress those components or *hns* is removed from the host, could phage resistance be observed [110, 187]. Altogether, despite constitutive transcription, the endogenous expression of the type I-E system is insufficient to provide protection for host against phage infection. It still remains unclear how and when the system is triggered to function as a defense mechanism.

The CRISPR array is first transcribed by host polymerase into a long precursor (pre-crRNA) that often contains stable secondary structures (hairpins) within the palindromic repeats. Cas6e endoribonuclease, one of the subunits of Cascade, specifically recognizes the stem-loops and cleaves the pre-crRNA downstream of the stem-loops within the repeat sequences, yielding mature crRNAs [188-190]. A mature crRNA is 61 nucleotide (nt) long and comprises a 32-nt spacer sequence, flanked by 8-nt repeat-derived sequence at 5' handle and 21-nt sequence forming a hairpin with a tetranucleotide loop at the 3' terminus (Figure 1.7A) [191]. After cleavage, Cas6e remains tightly bound to crRNA, allowing other subunits to assemble along the crRNA to form the Cascade surveillance complex. Type I-E Cascade has a molecular weight of 405 kDa and encompasses an uneven stoichiometry of five different Cas proteins (Cse1_1, Cse2_2, Cas7_6, Cas5e$_1$, and Cas6e$_1$) and a crRNA [191]. The overall morphology of Cascade has a helical backbone of six copies of Cas7 with a crRNA lying in its concave groove, forming a seahorse-like architecture [192]. The crRNA is capped at its 5' handle by Cas5e and 3' end by Cas6e. Two small Cse2 subunits interact with Cas7 backbone to help stabilize the crRNA and target DNA. Lastly, Cse1 large subunit is loosely associated to the Cas5 end of the complex.

The crRNA guides Cascade to scan foreign dsDNA for complementary target sequences through specific recognition of a PAM. As mentioned in the previous section, the PAM serves as an authentication and discrimination mechanism, allowing for distinction between host and foreign DNA in nearly all known CRISPR-Cas systems, except the type III system [124, 193]. In *E. coli*, a trinucleotide PAM is located immediately upstream of the target sequence (positions -3 to -1) (Figure 1.7A) and can be recognized by Cse1 subunit in the target strand of foreign DNA [194]. The interaction between Cse1 and PAM promotes Cascade binding to the target but destabilizes the dsDNA duplex at the PAM-proximal region, enabling DNA unwinding and the subsequent protospacer-crRNA base pairing. The first 8 nt in the PAM-proximal region of crRNA (to be more accurate: positions 1-5 and 7-8), termed seed sequence, is critical for high-affinity binding of target DNA by Cascade, as it offers a minimum length requirement for a stable RNA-DNA hybrid (Figure 1.7A) [195-197]. Mutations in seed sequence substantially reduce the binding affinity of Cascade and allow invasive phage to escape immunity [195]. Base pairing between the seed region and the target strand displaces the non-target strand which is then bound by two Cse2 to form an R-loop structure (Figure 1.7A). Subsequently, the full R-loop formation triggers conformational changes in Cascade, licensing Cas3 to bind [191, 198, 199].

Interference

The conformational change of Cascade triggers the recruitment of Cas3 nuclease/helicase. Cas3 comprises an N-terminal histidine-aspartate (HD) domain and a C-terminal superfamily 2 (SF2) helicase domain. The HD domain displays metal-dependent endonuclease activity against ssDNA, while the helicase domain exhibits ATP- and magnesium-dependent unwinding activity in the 3’ to 5’ direction [165, 200]. Single-particle cryoEM of Cas3-bound dsDNA-Cascade reveals that Cse1 subunit of Cascade colocalizes with Cas3 at the PAM-proximal end of target DNA, suggesting that Cse1 plays a critical role in Cas3 recruitment [201]. Moreover, upon the formation of the R-loop structure, the exposed non-target ssDNA by Cascade provides a platform for Cas3 to bind and perform the initial nick on the non-target strand downstream of the PAM by the HD nuclease domain [202]. The nicking activity is stimulated by the presence, rather than hydrolysis, of ATP. The nicked DNA provides Cas3 with the substrate for DNA unwinding by its helicase domain. Subsequently, Cas3 translocates on the non-target strand in the 3’ to 5’ direction and progressively unwinds the dsDNA, followed by degrading the non-target strand in an ATP-dependent manner (Figure 1.7B) [202, 203]. After cleavage of the non-target strand, Cas3 helicase activity may facilitate the displacing of the Cascade-crRNA complex from the invasive DNA, leaving the opposite ssDNA available for Cas3 cleavage [200]. As a result, the entire target DNA is destroyed and the host is protected from phage or plasmid infection.

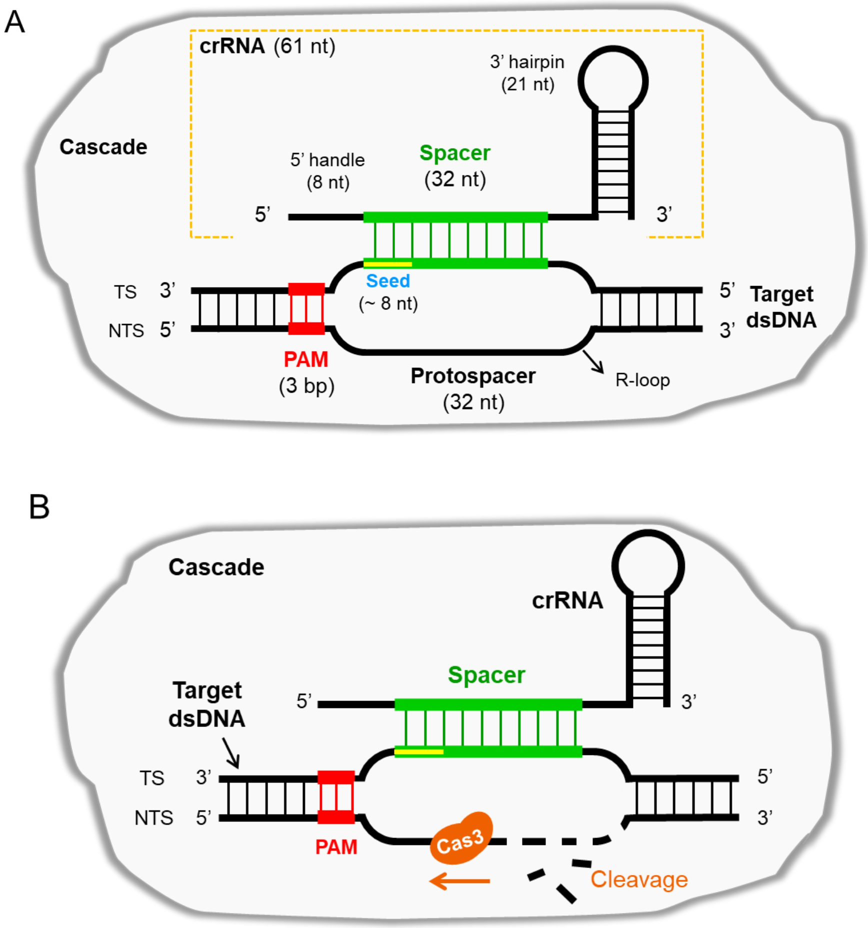

Figure 1.7 Schematic of target DNA recognition by Cascade and Cas3-mediated target degradation during the interference stage.
(A) Target DNA recognition. The crRNA guides Cascade complex (shown as the grey polygon) to search for complementary sequence by first recognizing the PAM sequence (red) on target DNA. PAM recognition destabilizes the target dsDNA and facilitates strand invasion initiated at the seed region (yellow) which is adjacent to the PAM. The subsequent base pairing between the crRNA spacer (green) and the protospacer region (green) on the target strand of the foreign DNA will result in the formation of an R-loop structure, triggering the recruitment of the endonuclease Cas3. TS, target strand. NTS, non-target strand. (B) Target DNA degradation. Cas3 (orange) first nicks the non-target strand, generating substrate for Cas3 to further unwind the dsDNA by its helicase activity. Following that, Cas3 processively cleaves the newly generated ssDNA in a 3' to 5' direction along the non-target strand as indicated by the horizontal arrow, leading to the degradation of the target dsDNA.

Factors affecting the interference efficiency of CRISPR-Cas systems

A number of studies have identified many parameters that can influence the efficiency of CRISPR-Cas interference. First, PAM sequence is known as a strong player, as single-nucleotide mutations in the PAM can significantly inhibit CRISPR immunity, leading to the escape of invaders [142, 195]. High-throughput analyses using plasmid-based target libraries suggest that type I-E Cascade can recognize a broad variety of PAM sequences, but with varying degrees of efficacy [204-206]. By examining the CRISPR immunity response through either plasmid loss assay or phage challenge assay, four canonical PAMs (5'-AAG, AGG, ATG, GAG-3') identified in the earlier studies have been proved to provide effective interference against target DNA, most likely due to the high-affinity binding of Cascade to target DNA which is conferred by those PAMs [203, 205, 207, 208]. Besides, additional sets of PAM sequences have been reported later and categorized together with the canonical ones into three groups based on their ability to induce CRISPR defense: strong interference, intermediate, and those unable to cause interference [205, 209]. Remarkably, a fully ineffective PAM (5'-CCG-3') which is a perfect match to the end of repeat sequence completely abolishes Cascade binding, providing protection for the native CRISPR array from self-cleavage [208].

Second, mismatches in the seed sequence of the protospacer also have a great negative impact on CRISPR interference efficiency. Although seed mutations can be tolerated to some degree, the mismatch substantially reduces the binding affinity of Cascade as does the mutated PAMs, thus disturbing the recruitment and/or cleavage activity of Cas3 [195, 196]. Third, the CRISPR defending ability relies on the sequences of crRNA spacer as well. Even sharing a same PAM, different crRNAs show variable interference degrees according to the result of a plaque assay [205]. Moreover, if a protospacer harbors mismatches, the R-loop-zipping process can be stalled. Therefore, the surveillance complex has to overcome the noncomplementary energy barrier, which may cause declined efficiency of degradation [210]. Furthermore, guanine-cytosine (GC) content in spacers has also been shown to affect interference efficiency with an optimal GC content (~ 62.5%) for execution activity [205]. Last but not least, the abundance of interference machinery is of great importance in effective CRISPR action. Due to the presence of H-NS and other repressors in host, the native expression level of CRISPR components is insufficient to provide phage resistance [184]. Previous research has demonstrated that overexpression of Cascade could allow CRISPR system to overcome the stringent PAM requirements for targeting [110, 194]. Most recently, a study accomplishes the direct visualization and quantification of Cascade complexes in the native type I-E CRISPR-Cas system *in vivo* [211]. It discovered an exponential correlation between Cascade copy number and CRISPR interference levels, where 20 Cascade complexes are required to obtain 50% protection. Additionally, this study found that the transcription of target DNA and CRISPR arrays can influence the interference levels through affecting the integrity of Cascade in host cells. It is worth noting that the effect of those factors on CRISPR defensive response is not restricted to the type I-E system but also holds true for other DNA-targeting CRISPR systems like the type II system with a single effector Cas9. Discovery and examination of other hidden determinants is still in high demand.

Overall, given the fact that CRISPR-Cas systems have become a premier tool for genome editing in a wide range of applications, a more careful and thorough characterization of how multiple factors affect the efficiency of the CRISPR interference is critical for ensuring the accuracy of gene manipulation. As the ultimate goal is to achieve efficient and precise genome editing, this will help increase the specificity of effector nucleases, realize quantitative prediction of off-target cleavage and enable a more rational design for gene modification in biotechnological research and therapeutic applications.

Anti-CRISPR Defense

As a countermeasure, phages have developed various strategies to circumvent bacterial CRISPR-Cas defense machinery, thereby creating a constant evolutionary battle between bacteria and phages. Those strategies include mutation of phage genome sequence to avoid CRISPR-Cas targeting, encoding phage-derived CRISPR-Cas systems to directly hijack bacterial CRISPR systems, and producing anti-CRISPR (Acr) proteins to inactivate host CRISPR function [95].

Phages can escape CRISPR-Cas immunity through a single nucleotide mutation or deletion in the PAM or the seed sequence that is close to PAM in the protospacer region [142, 195]. This is because that both PAM and a perfect complementarity between crRNA and target DNA sequences are required for successful CRISPR interference. Although CRISPR-Cas systems are generally carried by bacteria and archaea, it has been discovered that some phages, such as *Vibrio cholerae* O1 and ICP phages, encode fully functional CRISPR-Cas systems that target and inactivate host antiviral defense system, allowing the completion of viral infection cycle [212, 213]. In 2013, a total of five distinct anti-CRISPR (AcrF1-5) proteins that inhibit the activity of bacterial type I-F CRISPR-Cas system were first discovered in *Pseudomonas aeruginosa* phages [214]. Four Acr protein families (AcrE1-4) that inactivate *P. aeruginosa* type I-E systems were identified in a follow-up study [215]. To date, by use of a combination of bioinformatics and biochemical and genetic screening, there have been 40 distinct families of anti-CRISPR proteins uncovered in several phage species against type I, type II and type V CRISPR-Cas systems [216].

Besides their small size, typically 50-150 amino acids, these anti-CRISPR proteins share no common features in sequence or structural motifs and are not related to any proteins of known function [216-218]. Nevertheless, they are frequently found to have a highly conserved gene, referred to as anti-CRISPR-associated gene (*aca*), immediately downstream of anti-CRISPR genes. The identified *aca* genes (*aca1* and *aca2*) encode predicted helix-turn-helix transcriptional regulators that seem to control the expression of the anti-CRISPR genes for optimal activity [218]. Recent studies reveal that anti-CRISPR proteins can inhibit CRISPR-Cas system via a variety of mechanisms. But all with characterized mechanisms so far act during the interference stages [216]. A majority of the Acr proteins with elucidated mechanisms (8 out of 11) have been shown to block target DNA binding to the Cascade effector complex and Cas9 for type I and type II systems, respectively. However, the means of the DNA binding inhibition vary greatly between those anti-CRISPR proteins. For example, AcrF1, AcrF2 and AcrF10 interact with different subunits of the Cascade complex to constrain its ability to bind to target DNA [219-221]. By contrast, three anti-CRISPR proteins have been determined to prevent target DNA cleavage. AcrIE1 and AcrIF3 directly interact with Cas3 nuclease to inhibit its recruitment to Cascade [221, 222]. AcrIIC1 binds to Cas9 HNH endonuclease domain to obstruct DNA cleavage [223]. The discovery of anti-CRISPR provides additional regulators of CRISPR-Cas activity and would undoubtedly be exploited for genome engineering and many other biotechnological applications.

Last but not least, two recent studies report that phages encoding anti-CRISPR proteins (Acr-phages) cooperate in order to overcome bacterial CRISPR-Cas immunity [224, 225]. The initial infections by Acr-phages are unsuccessful, nonetheless, the production of Acr proteins by these "sacrificial" phages generate immunosuppressed cells, enabling subsequent phage infection to succeed. This observation suggests that individual phage can benefit the whole viral population by weakening the anti-viral systems in bacterial host to avoid extinction. It provides fresh insights into the ongoing evolutionary arms race between phages and bacteria. Further investigations to uncover novel anti-CRISPR protein families and characterize mechanisms of anti-CRISPR activity remain to be done.

Thesis Overview

In this dissertation, two major research topics will be investigated:

1. Multiple factors affect bacteriophage lambda infection process and the lysis-lysogeny decision making.

2. Visualization of phage DNA degradation by a type I-E CRISPR-Cas system *in vivo*.

In the following chapters, several fluorescent labeling systems have been developed and applied for studying these two topics list above at the single-cell/single-phage/single-molecule level in both live cells and chemically fixed cells.

In chapter II, a single-molecule quantification of gene expression technology is adapted and optimized to probe the spatiotemporal expression pattern of λ *cII* gene, the key player involved in the cellular decision making, during phage infection. Moreover, different fluorescent reporter systems are designed to allow monitoring the dynamics of both phage DNA molecules and *E. coli* chromosome in real time during phage intracellular development. The spatial features of phage DNA replication and the interaction between phage DNA and *E. coli* DNA are discovered. Furthermore, a single-molecule phage DNA labeling technique in fixed cells is established in order to confirm the observations in live-cell images and characterize the importance of phage DNA replication in the decision making.

In chapter III, the effect of the side tail fibers of the real wild type phage lambda strain (Ur-λ) on phage infection and post-infection development is focused on. Utilizing a lysis-lysogeny fluorescence reporter system, the role of those side tail fibers played in Ur-λ infection cycle as well as the lysis-lysogeny decision making outcome is characterized at the single-cell level.

In chapter IV, by taking advantage of the *E. coli*-lambda infection system as well as the phage particle and phage DNA labeling techniques, in real-time phage DNA degradation by an artificial type I-E CRISPR-Cas system is visualized *in vivo* at the single-cell resolution. In order to obtain a better understanding on how CRISPR-Cas machinery breaks down invading DNA, some parameters accounting for the effect of cell-to-cell variability on CRISPR activity are also examined and characterized.

In chapter V, conclusions collected from those work are summarized and discussed.

CHAPTER II

SPATIOTEMPORAL EXPRESSION PATTERN OF PHAGE LAMBDA CII GENE AND PHAGE DNA SUBCELLULAR LOCALIZATION DURING DECISION MAKING[*]

Introduction

Decision-making is a ubiquitous process among all organisms, from simple viruses to complex mammals [85]. A multicellular organism comprised of a population of cells will take all the decisions made by individual cells to make an appropriate response to environment signals, such as protecting itself from harmful stress [86, 87]. At the single-cell level, individual cells can exhibit heterogeneous cell fates even when they are genetically identical and subjected to the same conditions [85, 88, 89]. This heterogeneity or “noisy” phenotype has been conjectured to be the result of intrinsic stochasticity in gene regulatory networks [88, 90].

Bacteriophage λ, one of the most comprehensively studied paradigms, has served as an important model for studying basic genetic regulatory processes, such as the bistable genetic switch and cell-fate decision [14, 17, 18]. Upon infection of the bacterial host *E. coli,* phage λ can make a decision to enter either of two distinct life cycles, lysis or lysogeny. In the lytic pathway, λ rapidly produces new phages, resulting in host cell lysis

---

[*] Part of this chapter is reprinted with permission from “Coupling of DNA Replication and Negative Feedback Controls Gene Expression for Cell-Fate Decisions” by Qiuyan Shao, Michael G. Cortes, Jimmy T. Trinh, Jingwen Guan, Gábor Balázsi, and Lanying Zeng, *iScience* **2018** 6, 1-12, Copyright 2018 by The Author(s).

to release about a hundred progeny virions. During the lysogenic cycle, λ establishes a dormant state by integrating its genome into the host chromosome, and existing as a quiescent prophage. The lysis-lysogeny decision making has been considered as the result of the interactions between regulators within the λ genetic circuitry.

Phage λ genetic network has been well characterized over decades. Many genes are involved in determining the lysis-lysogeny decision making outcome. Among them, the product of λ *cII* gene is known as the key regulator during the decision making process [17, 21, 91]. The *cII* gene is transcribed under the promoter *pR* early after phage DNA ejection. As a transcriptional activator, the CII protein turns on the transcription from three promoters, *pRE, pI* and *paQ*, which overall function to inhibit lytic pathway and establish of the lysogenic pathway [71, 226]. The *pI* promoter enables the expression of λ integrase, Int, which facilitates the integration of phage DNA into the host chromosome [227, 228]. CII reduces the *Q* gene product via the *paQ* promoter, which transcribes antisense RNA within the *Q* gene [18, 68]. Q acts to antiterminate the transcription of a single operon, initiated at *pR'*, which harbors both the lysis and morphogenesis genes [18, 71]. Therefore, the reduced Q level by CII will inhibit the late gene expression. CII activates the promoter *pRE* that is responsible for the transcription of λ repressor CI, allowing CI to establish lysogeny. CI is able to maintain a stable lysogenic state by turning on the transcription of itself from promoter *pRM*, meanwhile turning off the transcription of all other phage genes [17, 71]. In order to trigger the lysogenic switch by activating its three promoters, it is critical for CII protein to accumulate until it reaches and exceeds a threshold level [17, 21].

λ lysis-lysogeny decision-making process was studied at the population level in the 1970s, uncovering that the lysogenization frequency increases with the number of infecting phages [76]. A study at the single-cell resolution proposed a unanimous voting model that individual phages are able to independently make decisions between lysis and lysogeny inside the host, and only when all infecting phages vote for the lysogenic commitment can lead the cell to go to the lysogenic pathway [37]. Based on this scenario, it is predicted that phage gene expression level increases linearly with MOI (multiplicity of infection). Instead of individual phage voting, a following study provided an alternative scenario that attributed the decision-making outcome to partial gene dosage compensation [84]. In this study, when the lysogenization frequency was plotted against phage genome concentration (defined as the number of infecting phages divided by cell volume) using the same experimental data from [37], the results suggested that in the host cell, viral gene expression from multiple infecting phages is partially compensated by a degree of $\text{MOI}^{1/2}$, thereby influencing the ultimate cell fate. Nevertheless, quantitative measurements of certain phage-encoded molecules during the infection cycle are still missing.

More details have emerged from recent studies on the mechanism of λ decision-making using high-resolution techniques [38, 65, 83]. It was uncovered that co-infecting phages compete against each other during lysis and cooperate during lysogenization [83]. By labeling *E. coli* chromosome and the initial injected phage DNA to allow for the tracking of post-infection development *in vivo*, our previous work observed that phage can integrate its DNA into the host committed to lysis [229]. These studies suggest that the cellular decision-making might occur at the phage DNA level, where each phage DNA is able to make an individual decision. Nevertheless, the observations of the intracellular dynamics of phage λ DNA were limited to the initial phage DNA molecules [38, 65], and much less is known of how the subsequent replicated viral genomes behave in the cytoplasm. Moreover, the effect of the intracellular phage organization and the interaction between phage DNA molecules at the subcellular level on cell-fate determination awaits further investigation.

In this work, to characterize the overall expression of the fate-determining genes, we chose the *cII* gene whose product is the master viral regulator to begin with. Following phage infection, we quantified the *cII* transcription levels over time at the single-cell level by RNA single-molecule fluorescence *in situ* hybridization (smFISH) [230]. We revealed the intracellular distribution of *cII* mRNAs and discovered that they colocalized with their DNA templates. In parallel, we developed new fluorescent reporter systems to label each individual phage genome and examine the dynamics of the whole population of phage DNA after infection at the single-cell level in real time. In coupled with the labeling of *E. coli* helicase DnaB, an essential DNA replication resource [231], we revealed how phages manage their resources after infection. Moreover, we exploited both live-cell and fixed-cell strategies to simultaneously label the *E. coli* chromosome and phage DNA and investigate the interaction between phage DNA and *E. coli* DNA during the post-infection development. The results suggested that individual phage DNA has the capacity to establish separate subcellular compartments within single cells. The *E. coli* nucleoid acts as physical barriers to segregate those phage-derived compartments, allowing co-infecting phages to develop as individuals to contribute to the final cell fate decision.

Materials and Methods

Bacterial strains, phages, plasmids, and primers

Bacterial strains, phages, plasmids, and primers used in this study are listed in Table 2.1.

To construct the phage λ DNA reporter strains, we began with the modification of λ *D* gene adapted from the approach described in [83]. Briefly, we thermal induced either a wild type λ lysogen or a λ*cII68* (deficiency in lysogenization) lysogen harboring plasmids pBR322-*λD-mNeongreen/mTurquoise2/mKate2-E* for recombination. The resulting lysate containing the phage of interest was titered on MG1655 with the plasmid pACYC177-*PLate*D* for screening. The plaques were checked under a fluorescence dissecting microscope to screen out the corresponding fluorescent phages, followed by amplification and lysogenization on MG1655. Subsequently, the newly obtained lysogens were transferred with the recombination plasmids bearing an array of 24 or 48 copies of *tetO* or *lacO* repeats, a chloramphenicol resistance gene cassette, and homologous sequences probing the *bor* region on λ genome. After a thermal induction of these lysogens, the lysate containing the phage of interest was lysogenized into MG1655 and selected for chloramphenicol resistance. The resulting lysogen strains were further confirmed by PCR for successful insertion of the operator array and being single integration prophages. The successful lysogen strains were transformed with the plasmid pACYC177-*PLate*D* in order to generate stable fluorescent gpD-mosaic phages upon phage purification.

Prior to construct the bacterial *attB* reporter strain, a plasmid bearing a DNA array containing 96 copies of *lacO* sequence and a kanamycin resistance gene which are flanked by homologous region at ~1.6 kb upstream of the *attB* site was first generated. The plasmid was then digested with EcoRI and SalI restriction enzymes to obtain the linear target DNA fragment, followed by engineering into MG1655 using lambda Red recombination system [232]. The resulting cells were then screened for kanamycin resistance. Finally, the recombinant strains were confirmed by PCR for the existence and appropriate length of the array and transferred with the plasmid pACYC177-*pFtski-tetR-mCherry-lacI-eyfp* to obtain the reporter strain allowing the visualization of both *E. coli* chromosome and phage DNA simultaneously.

Table 2.1 Bacterial strains, plasmids, phages, and primers used in this work.

Bacterial strains, phages, plasmids, and primers		
Strain Name	Relevant Genotype/Comments	Source/Reference
Bacterial strains		
MG1655	*sup*^{0}, Wild type *E. coli*	Lab collection
BW14894	*ΔlacIZYA*, *Δphn33-30*	[233]
TK310	*ΔcyaA*, *ΔcpdA*, *ΔlacY*	[234]
LZ204	MG1655, *dam*$^-$, *seqA-ecfp*, *Cm*^{R}	[38]
LZ2001	MG1655 [pZS24 *pFtski-tetR-mNeongreen*]	This work
LZ1677	MG1655 [pZS24 *pFtski-tetR-mNeongreen lacI-mKO2*]	This work
LZ1643	MG1655, *attB-96×lacO Kan*^{R} [pACYC177 *pFtski-tetR-mCherry lacI-eyfp*]	This work
Phage strains		
λLZ613	λ*cI*$_{857}$ *bor::Kan*^{R}, λWT	[38]
λLZ1575	λ*D-mNeongreen cI*$_{857}$ *bor:: 24×tetO Cm*^{R}	This work
λLZ1653	λ*D-mTurquoise2 cI*$_{857}$ *bor::48×tetO Cm*^{R}	This work
λLZ1678	λ*D-mTurquoise2 cI*$_{857}$ *cII*$_{68}$ *bor::48×tetO Cm*^{R}	This work
λLZ1681	λ*D-mKate2 cI*$_{857}$ *cII*$_{68}$ *bor::48×lacO Cm*^{R}	This work
Plasmids		
pCP20	Recombines flanking FRT sites to remove inserts, *Amp*^{R}	Lab collection
pKD46	Inserts target DNA fragments through homologous recombination, *Amp*^{R}	Lab collection
pACYC177 *PLate*D*	gpD under the control of λ late promoter, *Amp*^{R}	Lab collection
pZS24 *pFtski-tetR-mNeongreen*	tetR-mNeongreen fusion under weak constitutive promoter *pFtsKi*, *Kan*^{R}	This work
pZS24 *pFtski-tetR-mNeongreen lacI-mKO2*	tetR-mNeongreen fusion and LacI-mKO2 fusion under weak constitutive promoter *pFtsKi*, *Kan*^{R}	This work
pACYC177 *pFtski-tetR-mCherry lacI-eyfp*	tetR-mCherry fusion and LacI-EYFP fusion under weak constitutive promoter *pFtsKi*, *Amp*^{R}	This work
pUC19 *bor::48×tetO Cm*^{R}	contains homologous regions for recombination of a *tetO* array into phage λ genome, *Amp*^{R} *Cm*^{R}	This work
pUC19 *bor::48×lacO Cm*^{R}	contains homologous regions for recombination of a *lacO* array into phage λ genome, *Amp*^{R} *Cm*^{R}	This work
pUC19 *attB-96×lacO Kan*^{R}	contains a linear DNA fragment with homologous regions and a *lacO* array for Red recombination upstream of *E. coli attB*, *Amp*^{R} *Kan*^{R}	This work
pBR322 *D-mTurquoise2-E*	provides translational fusion *D-mTurquoise2*, *Amp*^{R}	[83]
pBR322 *D-mKate2-E*	provides translational fusion *D-mKate2*, *Amp*^{R}	Lab collection
Primers		
cII-forward	5'-GCAGATCAGCAGGTGGAAGA	This work
cII-reverse	5'-AATCGAGCCATGTCGTCGTC	This work
ihfB-forward	5'-ACCACGTACCGGACGTAATC	This work
ihfB-reverse	5'-ATCGCGCAGTTCTTTACCAG	This work
λ probe-forward	5'-AGAATCGACCATTTCTGCCATCACC	This work
λ probe-reverse	5'-GAGATTTCCGCTTTTGTGCTGGTTG	This work
attB probe-forward	5'-GCGAGATTATCAAGGCGGGTA	This work
attB probe-reverse	5'-CGAATGGGATGAGCGCGATA	This work

RNA single-molecule fluorescence in situ hybridization (RNA smFISH)

To study the expression pattern of λ *cII* gene, 48 DNA oligonucleotides spanning the coding sequence of *cII*, *O* and part of *P* gene were designed and synthesized from Biosearch Technologies. The sequences can be found in Table 2.2. The probes pre-modified with a 3'-end amine group were labeled with Cy5 (GE Healthcare Life Sciences, #PA15000), following the procedures described in [230]. Briefly, 7.5 μl of each of the oligo solutions (100 μM) was pooled and added 40 μl of 1 M sodium bicarbonate (pH 8.5). To generate the Cy5 dye solution, 1 mg of Cy5 was dissolved in 2.5 μl of DMSO and 25 μl of 0.1 M sodium bicarbonate (pH 9.0). Then the oligo solution and the dye solution were mixed thoroughly and incubated in the dark overnight at 37 °C. The next day, 47 μl of 3 M sodium acetate (pH 5.2) was added into the overnight solution, followed by adding 1180 μl of 100% ethanol and incubation -80 °C for 3 hr to precipitate the oligos. The oligos were then spun down and washed twice by dissolving the pellet in 45 μl of DEPC-treated water with addition of 5 μl of 3 M sodium acetate (pH 5.2) and 125 μl of 100% ethanol. Finally, the probes were resuspended in 250 μl of 1× TE resulting in a 10× probe stock solution. The probe solution was diluted with 1× TE by 10 fold to make the 1× probe solution for FISH experiments usage. The labeling efficiency of the probes was measured using a NanoDrop and calculated to be 97.6% beyond the requirement of >90%.

To detect *cII* mRNA level after infection, *E. coli* strain MG1655 was used as the host for infection. The overnight culture MG1655 was diluted 1:1000 into fresh LBMM (LB supplemented with 0.2% maltose and 10 mM $MgSO_4$) and grown at 37 °C with shaking at 265 rpm until OD_{600} ~0.4. The cells were then collected by centrifugation at 4

°C, 2000×g for 15 min and resuspended in one-tenth volume of pre-chilled LBMM. For each sample, 0.75 ml of the concentrated cells were used for infection. An appropriate amount of phages was added to reach requested APIs (0.2, 0.5, 1, 2, 3, 4, and 5) and mixed well with cells. For the negative control, the same volume of SM buffer (phage buffer, 100 mM NaCl, 10 mM $MgSO_4$, 0.01% gelatin, 50 mM Tris-Cl, pH 7.5) was added to the sample. The infection mixture was incubated on ice for 30 min to allow phage adsorption, then transferred to a 35 °C water bath for 5 min to trigger phage DNA ejection. Thereafter, the mixed culture was aliquoted as 750 µl per sample which was transferred into 7 ml of pre-warmed LBGM (LB supplemented with 0.2% glucose and 10 mM $MgSO_4$), followed by incubation in a 30 °C water bath with mild shaking at 225 rpm until certain time points. At the specific time point, the sample was poured into a 15 ml tube with 860 µl of 37% formaldehyde (final concentration 3.7%) for quick fixation, and incubated for 30 min at room temperature using a nutator.

After 30 min fixation, each sample was washed three times with 1 ml of ice-cold 1× PBS to remove excess formaldehyde. The cells were permeabilized in 70% ethanol for 2 hr at room temperature and then spun down at 600×g for 7 min. The cell pellet was resuspended in 1 ml of wash solution (40% (wt/vol) formamide, 2× SSC) and incubated at room temperature with rotating for 5 min and spun down again to remove the wash solution. For hybridization, the cells were resuspended thoroughly in 25 µl of hybridization solution (40% (wt/vol) formamide, 2× SSC, 10% (wt/vol) dextran sulfate, 1 mg/ml *E. coli* tRNA, 2 mM ribonucleoside-vanadyl complex and 0.2 mg/ml BSA) with *cII* probes at a final concentration of 1 µM (2.06 µl of 1× solution). Samples were incubated in a 30 °C water bath for overnight. The next day, 10 μl of cells were washed three times by incubation in 200 μl of wash solution for 30 min in a 30 °C water bath and centrifugation at 650×g for 3.5 min. During the third repeat of washing, 10 μg/ml DAPI was added in the wash solution to stain the *E. coli* and phage DNA. Finally, cells were resuspended in 2× SSC and 1 μl was placed on a coverslip, topped with a piece of 1.5% 1×PBS agarose pad and another coverslip for imaging immediately under the microscope.

Table 2.2 Sequences of the probes for detecting phage λ *cII* mRNA.

1	5’-CGTTTGTTTGCACGAACCAT	25	5’-CCACAGAAAGGTCGTTTTCT
2	5’-TCTCGATTCGTAGAGCCTCG	26	5’-TGAATTGCAGCATCCGGTTT
3	5’-GCGATTTTGTTAAGCAACGC	27	5’-ATGTCAAACATCCACTCTGC
4	5’-TGTCTTCTCAGTTCCAAGCA	28	5’-TGATGGTGCGATAGTCTTCA
5	5’-GCTGATCTGCGACTTATCAA	29	5’-CATCAGGCGGATATCGTTAG
6	5’-AGAACTTTGGAATCCAGTCC	30	5’-TTACCGGACCAGAAGTTGTC
7	5’-CCCATTCAAGAACAGCAAGC	31	5’-TCCACTTATCGCGGAGTTTG
8	5’-AATCGAGCCATGTCGTCGTC	32	5’-TTTGGTTTGCTGGCTGTCAC
9	5’-AATCGCAGCAACTTGTCGCG	33	5’-ATAGATCCACCCCGTAAATC
10	5’-CCGGGCGTTTTTTATTGGTG	34	5’-TCTGCTCACGGTCAAAGTTA
11	5’-GATTTGTTCAGAACGCTCGG	35	5’-CTTTTCGTCGTACTGTTCCG
12	5’-AATGACCTCAGAACTCCATC	36	5’-GAACACACCGTTGATGATCT
13	5’-TGACTCCTGTTGATAGATCC	37	5’-TTCGTTCTGGTCACGGTTAG
14	5’-ATCGAGATCTGCCACATTAC	38	5’-TTTTCCCGAAAAGCCAGAAC
15	5’-TTGATAGTCTGGCGTAACCA	39	5’-CGTTAACCTGTTCCATCGTG
16	5’-GAATAAGCCTCAAGCAGCAT	40	5’-AGAAATGGTCGATTCTGCCG
17	5’-AACTGTCGCTTGGTCAGATC	41	5’-ATATCAACCAGCTCGCTGAC
18	5’-CAGAATGGCAAGCAGCACTT	42	5’-CTTCCGGCAATACTCGTAAA
19	5’-ATCGGTGATTCTGTCCATTG	43	5’-AGTAGTGCGCGTTTGATTTC
20	5’-TTGCACCGTTTGACAGGTAA	44	5’-CTGATACAGGTTGGTAACCA
21	5’-GACGAGTTCTAACTTGGCTT	45	5’-GTAATTCCGCATCAGTAAGC
22	5’-TTTTGAGGGATGCACCATTC	46	5’-CTCACCACGGTTAATTCTCG
23	5’-CTCGTTTTAGGGGATTTTCC	47	5’-GTGCACGATTTAGAGGTCTA
24	5’-ATTCGCCAGAATTCTCTGAC	48	5’-CATACACTTGCTCCTTTCAG

DNA single-molecule fluorescence in situ hybridization (DNA smFISH)

The procedure is adapted from the protocols described in [235]. Briefly, ~3 kb DNA fragments (39.9-42.9 kb region on phage λ genome and 804932-807929 on MG1655 genome) were amplified by PCR and then treated with a PromoFluor-500/640 Nick Translation Labeling Kit to generate DNA-dye fragments ranging from 100-500 bp. The probes were purified by an illustra MicroSpin Columns Kit (GE Healthcare Life Sciences) to remove unincorporated fluorescent nucleotides and finally resuspended in 1× TE. The labeling efficiency was measured using a NanoDrop and calculated to be ~5%.

To perform DNA FISH on infection samples, cells were grown and infected as described in the section of **RNA smFISH**. Samples were collected at given time points, followed by cell fixation in 3.7% formaldehyde solution for 30 min at room temperature using a nutator. Cells were washed three times with 1 ml of ice-cold 1× PBS to remove excess formaldehyde and resuspended in 1 ml of GTE buffer (50 mM glucose, 10 Mm EDTA, 20 mM Tris-Cl, pH 7.5). The cells were permeabilized with 0.2 μg/ml lysozyme for 5 min at room temperature, followed by washing with 1 ml of GTE for three times. The cell pellet was resuspended in 100 μl of GTE to reach an optimal cell density. Subsequently, 10 μl of cells was spread on a pre-coated poly-L-lysine slide for cell attachment and allow to dry completely. Samples then went through a gradient dehydration with 70%, 90% and 100% ethanol and were dried before hybridization. For hybridization, 160 ng of the labeled probes were added into 10 μl of hybridization solution (50% (wt/vol) formamide, 2× SSC, 10% (wt/vol) dextran sulfate, 50 mM $NaPO_4$, pH 7), denatured at 75 °C for 5 min and placed on ice. The denatured probes were then applied onto the cell attachment area on the slides, overlaid with a coverslip and sealed with nail polish. The slides were then placed at 80-85 °C for 5 min and transferred to 0 °C for 3 min. Following that, the samples were placed in a pre-warmed chamber and incubated overnight in the dark at 37 °C to complete hybridization. The next day, after removing nail polish and the coverslips, cells were washed twice by incubation in wash solution (50% (wt/vol) formamide, 2× SSC) for 20 min in a 37 °C water bath. Samples were further washed with a series of SSC solution (1×, 2×, 4×), each for 5 min. 0.5 mg/ml DAPI was added along with the 4× SSC to stain the *E. coli* and phage DNA. Finally, 10 μl of 2× SSC was dropped on the cell area and the slides were covered with coverslips for imaging under the microscope.

Quantitative RT-PCR (qRT-PCR)

Cell culture and phage infection were set up following the same procedure as for smFISH. At certain time points, infection samples were immediately poured into 5 ml ice-cold methanol for fixation. Samples were then spun down at 4000×g for 4 minutes at 4 °C. Cell pellet was resuspended in 1 ml of Bacteria RNA protection Reagent (Qiagen, 76506), followed by incubation at room temperature for 5 minutes. Subsequently, cells were spun down at 4500×g for 10 minutes at 4 °C. After discarding the supernatant, the cells were kept at -20 °C until all samples were collected and ready for total RNA extraction. RNA extraction was performed using the RNeasy Mini Kit (Qiagen, 74104), followed by DNA digestion with the TURBO DNA-free kit (Ambion, AM1907) according to the manufacturer's instructions. Then 400 ng of total RNA was used as the template for reverse transcription by the High Capacity RNA-to-cDNA kit (Applied Biosystems, 4387406). The synthesized cDNA was then quantified using the SYBR Green PCR master mix in a CFX96 real-time PCR thermocycler (Bio-Rad). The thermal cycling program was as follows: an initial denaturation at 95.0 °C for 30 s; 39 cycles of denaturation at 95.0 °C for 5 s, annealing at 60.0 °C for 30 s; and a final melting curve from 65.0 °C to 95.0°C for 5 min by 0.5 °C increment. The *E. coli ihfB* was used as a reference gene to provide the calibration for *cII* mRNA quantification since its transcriptional level does not change significantly with API and time. All samples within one experiment were run on the same 96-well plate to avoid between-run variations. Gene expression quantification was performed by the CFX Manager software (Bio-Rad).

Single-cell Infection Assay

Single-cell infection was performed as previously described [37, 38, 83], using M9 minimal medium as the growth medium in order to obtain optimal fluorescence signals. 1 ml of host MG1655 cells was grown in M9 minimal medium (11.3 g/L M9 salts, 1 mM $MgSO_4$, 0.5 µg/ml thiamine HCl, 0.1% casamino acids, 100 µM $CaCl_2$) supplemented with 0.4% maltose (M9M) with appropriate antibiotics for overnight. The overnight culture was diluted 1:100 into 5 ml of fresh M9M medium and grown until the optical density (OD_{600}) reached 0.3 at 37°C. The cells were harvested and concentrated by 10 fold by centrifugation at 2000×g for 2 minutes at room temperature, and then resuspended in ice-cold M9M. 10 µl of phages were mixed with 10 µl of the resuspended cells to reach an API of 1~2, followed by an incubation on ice for 30 min and another 5 min incubation at 35°C water bath to trigger phage DNA ejection. The phage-cell mixture was then diluted into M9 and 1 µl of the diluted sample was placed onto a 1.5% agarose pad of M9M (~1 mm thick) resting on a small coverslip until visibly dry (~1 min), topped with another large coverslip. The sample was moved to the fluorescence microscope for time-lapse imaging, where the time = 0 is set to the first time-lapse image taken, which is typically 15~20 min after cells are treated at 35°C.

Microscopy and imaging

Imaging was performed on an inverted epifluorescence microscope (Ti-E, Nikon, Tokyo, Japan) within a cage incubator (InVivo Scientific, St. Louis, MO) set at 30°C. Images were acquired using a 100× objective (Plan Fluo, NA 1.40, oil immersion) with standard filter sets using a mercury lamp as the light source (X-Cite 200DC, Excelitas Technologies) and a cooled EMCCD camera (iXon3 897, Andor, Belfast, United Kingdom).

For a typical real-time live-cell movie, 16 stages were selected where cells were well separated and a series of 5 z-stacks images with spacing of 300 nm in the CFP/Far-red channel using 200 ms exposure time were taken in the first frame in order to visualize all infecting phages surrounding cells. The time-lapse movie was subsequently taken every 5 minutes for a total length of 4 hours through the phase-contrast (100 ms exposure), mTurquoise2 (100 ms exposure, Nikon 96361), mKate2 (100 ms exposure, custom filter, Chroma 49310), mneonGreen (200 ms exposure, custom filter, Chroma 49308), mKO_2 (400 ms exposure, custom filter, Chroma 49309), YFP channel (100 ms exposure, Nikon 96363), mCherry (200 ms exposure, Nikon 96365) channels. To image RNA smFISH samples, a series of 11 z-stacks images with 200-nm spacing in the Cy5 channel for *cII* mRNA (200 ms exposure, Nikon 96366). To image DNA smFISH samples, a series of 5 z-stacks images with 300-nm spacing in the YFP channel (300 ms exposure, Nikon 96363) for phage DNA and Cy5 channel (400 ms exposure, Nikon 96366) for *E. coli attB*. For all smFISH images, one image was taken at the focal plan for the phase contrast (100 ms exposure) and DAPI channel (30 ms exposure, DAPI filter, Nikon 96310).

Image processing and data analysis

To quantify λ *cII* mRNA level in smFISH, images were first exported and then analyzed using the cell recognition program Schnitzcell (gift from Michael Elowitz, California Institute of Technology) which recognizes and segments individual cells. The total cell fluorescence intensity was calculated with homemade script in Matlab by summing the intensities within the cell boundaries and subtracting the background fluorescence in the same images for calibration. The average intensity of each cell was calculated by dividing the total intensity by the cell area. At 0 min after infection, cells with *cII* signals typically displayed as a single focus, possibly representing a single mRNA or a few mRNA molecules clustering together. The total fluorescence intensities of those cells and the cells from the negative samples were fitted into multi-Gaussian functions and the differences between them were designated as the intensity for a single *cII* mRNA. In the later time point samples, the *cII* mRNA copy number in each cell was calculated following the equation, $(T\text{-}m \times S)/A$, where T is the total intensity of the cell, m is the median of the average intensity of the cells from the negative sample without phage infection, S is the cell area, and A is the intensity of a single mRNA calculated as described above. To study the intracellular localization of phage DNA and mRNA, images were processed using MicrobeTracker [236] first to outline cells. Thereafter, fluorescent spots were recognized automatically using SpotFinderZ, followed by manually correction using SpotFinderM.

Results

Quantification of cII mRNA expression at the single-cell level by smFISH

To quantify *cII* mRNA expression levels upon infection of *E. coli* MG1655, a set of 48 oligonucleotide probes, each 20 bases long, was designed and labeled with Cy5 fluorescent dye molecules to target the transcript of λ *cII*. Because the short length of the *cII* gene (294 bp) is not sufficient to meet the minimum requirement of the number of FISH probes (typically 48), the *O* gene and part of the *P* gene, which are located within the same transcript as *cII*, had also been included in the probe design. We first preformed experiments with λWT phages at a low API of 0.2. As the probability of a cell being infected by a certain number of phages follows a Poisson distribution [37], 18.13% of the total cells will be infected. Consequently, at an API of 0.2, the estimated percentage of MOI=1 infections is 90.29%. In other words, the majority of infected cells were infected by only one phage. As shown in Figure 2.1A, the infected cells exhibited strong fluorescence signals corresponding to the *cII* mRNA molecules, whereas fluorescence signals were not detected in the cells without phage infection in the negative sample. At 0 min when samples were fixed immediately after phage infection, the infected cells displayed small fluorescent spots, which might represent one single *cII* mRNA molecule or a few mRNAs. At later time points, the *cII* mRNA appeared as clusters and became larger and brighter inside the cells, suggesting *cII* mRNA levels increased over time. We measured the total cell fluorescence intensity after performing a calibration by subtracting the intensity in the negative sample and background area to estimate the average *cII* expression level over time. In parallel, the whole population level of *cII* mRNA expression was performed by RT-qPCR. The results were in agreement with the findings obtained by smFISH. The overall trend of the *cII* transcription over time was that upon phage infection, *cII* mRNA level first increased until it reached a peak at around at around 6 -12 min and subsequently decreased (Figure 2.1B). It indicates that after a phage genome entering the host cell, *cII* expresses and its mRNAs accumulate until a peak is reached. After that, the decreased levels of *cII* mRNA suggest that the cell-fate decision has been made and the *cII* stops being transcribed likely due to the repression of the *pR* promoter by either CI or Cro [17, 18].

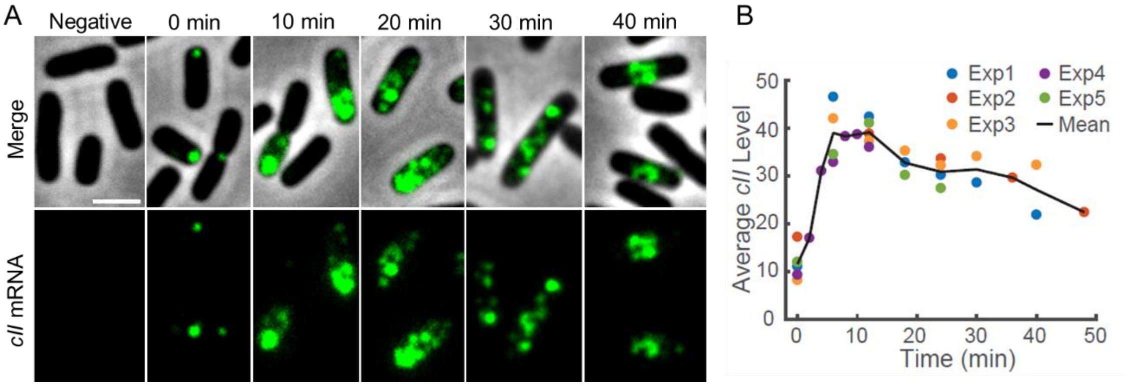

Figure 2.1 λ *cII* mRNA expression over time at the single-cell level.
(A) Representative cells showing *cII* mRNA signals by RNA smFISH. Negative: cells without phage infection. None of the cells show *cII* signal. Cells with λWT infection at API = 0.2 were fixed at different time points. At 0 min, cells show distinct foci, representing one single *cII* mRNA or a few mRNA molecules aggregating together. At later time points, *cII* mRNAs appear as clusters instead of punctate foci. Scale bar, 2 µm. (B) Average *cII* mRNA levels over time upon infection. Data from 5 experiments (dots) and the mean (black line) are shown. Only infected cells with *cII* signals were included in the calculation. *cII* expression reaches a peak at around 6 -12 min after infection and subsequently decreases.

The location preference of cII mRNA in infected cells

By labeling phage λ genome, there is evidence that the ejected phage DNA preferentially appears at cells poles [36]. However, the location of phage mRNA in the cytoplasm has not been uncovered yet. To examine the subcellular localization of *cII* mRNA, we measured the position of *cII* signals at 0 min when *cII* mRNA level was low and the fluorescence signals exhibited as distinct foci in the cells (Figure 2.2A). Similar to phage DNA, the *cII* mRNAs localized predominantly at the middle and the poles of the infected cells (Figure 2.2B). It would be considered as the result of the preference of phages attaching to cell poles and mid-cell positions for infection as described in the literatures [36, 37]. After phage DNA is injected into the cells, transcription of λ genes is immediately initiated by host RNA polymerase from the two early promoters *pL* and *pR* [12, 14]. λ *cII* is the second gene that is transcribed in the *pR* operon. Theoretically, *cII* is transcribed around the region where its DNA template stays. Therefore, when the infected cells were collected at 0 min, the localization of the first one or a few *cII* mRNA molecules reflected the position where the phage genome ejection occurred. To rule out the possibility that the smFISH technique will artificially affect the location preference of a gene of interest, a control experiment was performed using an *E. coli* strain TK310 to target the expression of the gene *lacZ*, which should not have any location preferences for the transcription in the cytoplasm [230]. The results showed that the distribution of *lacZ* mRNAs exhibited no location preference as expected, indicating that the technique does not alter the location of the gene expression (Figure 2.2C).

We next sought to confirm whether λ *cII* mRNA is transcribed around its DNA template. We took advantage of our previously established SeqA-DNA binding reporter system to label phage DNA [38]. The λWT phages were prepared to be fully methylated and were used to infect a methylation-deficient *E. coli* host strain (*dam*$^-$) and constitutively expresses the fluorescent fusion protein SeqA-ECFP. SeqA protein has a high binding-affinity to fully methylated and hemimethylated DNA. It binds specifically to GATC sequences, which are methylated on the adenine residues [237]. There are 116 GATC sites distributed across the λ genome. Accordingly, the injected phage DNA can be visualized as a CFP fluorescent focus under the fluorescence microscope. In conjunction with the smFISH method targeting *cII* mRNA, phage DNA and its expressed mRNA can be detected in the infected cells concurrently. Since *cII* mRNA accumulated and spread all over the cells at later time points, here we fixed the infection mixture at 5 min after triggering the DNA injection. The result showed that SeqA-ECFP formed punctate foci, marking the injected phage DNAs (Figure 2.2D). Among the cells with fluorescent signals (4315 in total), 15.4% of them displayed both phage DNA (CFP) and *cII* mRNA (Cy5) signals, while 35% and 49.6% showed only CFP and Cy5 signals, respectively. The cells with only DNA signals indicated that *cII* transcription had not occurred yet in those cells. The cells showing only mRNA signals suggested that some SeqA-CFP proteins bound to phage DNA might not be able to survive through certain steps of the smFISH procedure, i.e., several washing steps before and after hybridization with probes. When we calculated the minimum distance between λ DNA and *cII* mRNA in the cells with both signals (15.4%, 666 out of 4315 cells), 91.4% (609 out of 666 cells) of them exhibited good colocalization in which the minimum distances were less than 0.5 µm, which is a reliable indicator for colocalization [65] (Figure 2.2E). Overall, the colocalization between phage λ DNA and its expressed mRNA was visualized for the first time inside cells, and the results demonstrate that the transcription of phage genes in the cytoplasm always takes place near their templates.

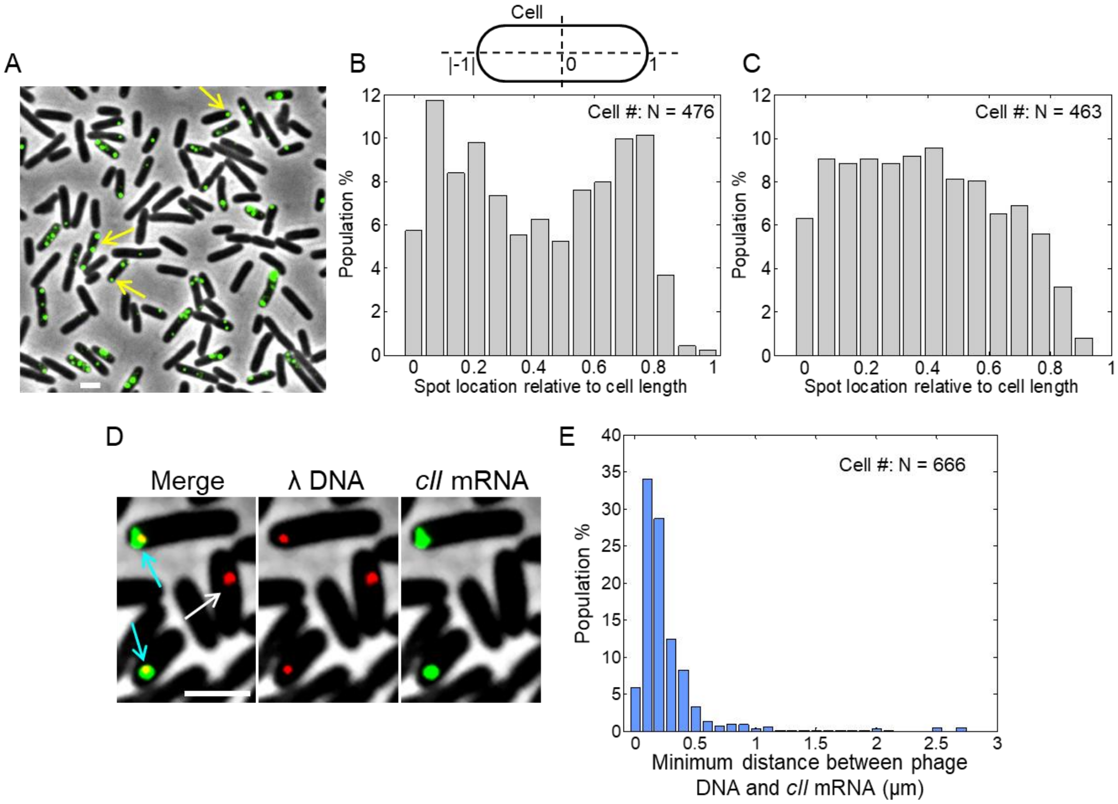

Figure 2.2 *cII* mRNA localization and correlation with λ genome in the cytoplasm. (A) A smFISH image showing *cII* mRNA (green signals) intracellular localization at 0 min. Yellow arrows point to the representative cells where *cII* mRNAs prefer to localize at the mid-cell region and cell poles. (B) Distribution of *cII* mRNA location along the cell major axis, where "0" represents the mid-cell area and "1" represents the cell poles. A schematic of the cell coordinates is shown on the top. (C) Distribution of *lacZ* mRNA locations in *E. coli* TK310. The *lacZ* mRNA does not exhibit any location preferences inside cells. (D) Representative cells showing the colocalization between λ genomes (red dots) and the *cII* transcripts (green dots), pointed by cyan arrows. Infected cells were fixed at 5 min after infection. An example cell pointed by the white arrow shows only phage DNA without mRNA expression. (E) Distribution of the minimum distance between phage DNA and *cII* mRNA. The majority of the distances are less than 0.5 μm, indicating the colocalization. Scale bars, 2 μm.

Visualization of the intracellular phage DNA replication and development

Given the fact that the SeqA-FP reporter system only allows for accessing the original infecting phage genome and its first replicated copy, it remains unknown the behavior of subsequent replicated DNA molecules. To obtain a full picture of the spatiotemporal dynamics of an injected λ genome over the entire infection cycle, we constructed two new phage DNA fluorescence labeling systems, *tetO*/TetR-FP and *lacO*/LacI-FP, following the approach in [61], where an array of 48×*tetO* or 48×*lacO* sequences were engineered into the *bor* region on λ genome (Figure 2.3A). The phage DNA can be detected when TetR-FP or LacI-FP fusion proteins are expressed from a plasmid and then bind to the operator array on the phage gnome to form fluorescent foci. Hence, all offspring phage DNA molecules can be tracked. In this work, we inserted the *tetR-mNeongreen* and/or *lacI-mKO2* under the constitutive promoter *FtsKi* into a low copy number vector, pZS24, which bears the *pSC101* origin (10-12 copies per genome) [238], to ensure that the background signal is low enough to visualize phage DNA foci. Moreover, by substituting the λ *D* gene, which expresses phage capsid decoration protein (gpD), with either *D-mTurquoise2* or *D-mKate2* translational fusion, the infecting phage λ particles and the development of the lytic life cycle can be visualized during the infection.

With the new phage DNA reporter systems, we next performed time-lapse movies to characterize the dynamics of phage DNA replication at the single-cell level. Following the infection of the reporter strains, λ DNA first appeared as a small focus (green or orange) at the entry point (Figure 2.3B and C). The DNA fluorescence foci grew and gradually formed as a cluster, corresponding to λ DNA replication. In the lytic cells, the phage DNA fluorescent clusters accumulated and expanded along the cell major axis over time. At the late stage, the cyan or far-red fluorescent protein, gpD-mTurquoise2 or gpD-mKate2, appeared within the clusters, signifying synthesis of the λ capsid proteins and the assembly of phage progeny, and eventually the cells lysed (Figure 2.3B and C). For those cells that committed to the lysogenic pathway, λ DNA replicated from a small focus to a bright cluster during the early infection (0-30 min), similar to the lytic cells. However, we observed that the formed DNA clusters barely grew after 30 min and exhibited both localized motion and active motion across the cell (Figure 2.3D). Following the cell division, a single DNA focus can be observed in all daughter cells, marking the integration of the phage DNA into the host chromosome. The initially formed DNA cluster was passed to one of the daughter cells and stayed until the end of the movies (240 min). The limited size of phage DNA clusters during the early infection time window suggests that it does not seem to require many phage DNAs for a cell to commit to either lytic or lysogenic decision. Moreover, once the lysogenic pathway is established, the expression of the *O* and *P* genes, which are required for DNA replication, is inhibited by CI repressor, leading to the limited amount of DNA in the lysogenic cells during the late infection period.

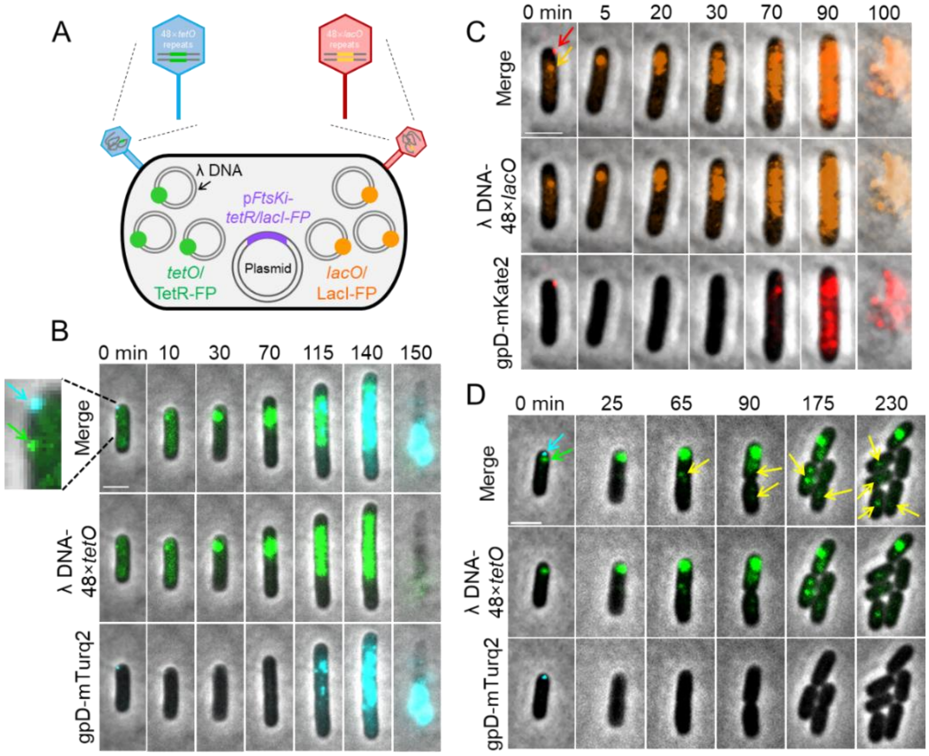

Figure 2.3 Detection of phage λ DNA development upon infection in real time at the single-cell/single-molecule level.

(A) A schematic description of phage λ DNA labeling with *tetO*/TetR-FP and *lacO*/LacI-FP reporter systems. λ genome is inserted by an array of 48 *tetO*/*lacO* repeats and the TetR-mNeongreen (green) and LacI-mKO2 (orange) fusion proteins are expressed under a constitutive promoter *Ftski* in a plasmid. Phage decoration capsid protein gpD is fused by either mTurquoise2 (blue) or mKate2 (red). (B) and (C) Representative lytic cells infected by a *tetO* phage (B) or a *lacO* (C) phage, respectively. Cyan and red arrows point to the labeled phages attached on the cell surface. Phage DNA appears as a small distinct dot pointed by the green and orange arrows at 0 min. Over time, the DNA signal expands and forms a large cluster filling most of the cell, indicating phage DNA replication. At later time points, gpD signals (cyan and red) arise within the DNA cluster, corresponding to phage progeny assembly. Eventually, both cells lyse. (D) A representative lysogenic cell infected by a *tetO* phage. During the early infection stage (0-30 min), phage DNA forms a cluster, similar to that in a lytic cell. After 30 min, in contrary to the continuously growing DNA signals as does in a lytic cell, the DNA cluster stops expanding, suggesting that lysogenic cells have limited phage DNA replication. A small, separate green focus appears at 65 min signifies the event of phage DNA integration into the *E. coli* chromosome. Following cell division, single green dots can be found in each daughter cell representing the integrated phage DNA. Scale bars, 2 µm.

In particular, when multiple phages carrying the same reporter system co-infected a cell through distinct locations, we observed that each ejected phage DNA could form its own cluster, which remained separated from others in the cytoplasm during the post-infection development (Figure 2.4A). To further test whether those separated phage DNA clusters can collect host resources for their own usage, we translationally fused the *E. coli* DnaB with mTurquoise2 by replacing the native *dnaB* gene with *dnaB-mTurquoise2* on the *E. coli* chromosome. DnaB is a replicative helicase that is required by both *E. coli* and phage λ for DNA replication [54, 231]. We also introduced a plasmid constitutively expressing TetR-mCherry fusion proteins into the host with the engineered *dnaB* gene and infected this reporter strain with a D-mNeongreen phage strain carrying an array of a 24× *tetO* tandem array. We observed that when multiple phage DNA clusters formed in different areas in a cell, DnaB foci typically colocalized with those phage DNA clusters over time (Figure 2.4B). It indicates that after entering host cells, each phage DNA is able to recruit host resources, like DnaB, to spatially assemble its own compartment at the subcellular level.

In addition, we noticed that the size of phage DNA clusters differed even within the same cell. This implies that the amount of phage DNA generated from each compartment varied, which is likely the result of the interaction between phage DNAs inside cells. Moreover, our mRNA smFISH images showed that early after infection (~10 min), the *cII* transcripts existed in separate space within single cells (Figure 2.4C). In conjunction with our previous findings that at the early infection stage, the expression of phage genes occurs near phage genome (Figure 2.2E), the phage DNA subcellular organization might play a significant role in the lysis-lysogeny decision making. A single phage DNA is capable of organizing and maintaining its own compartment for replication and gene expression, and different infecting phages could develop as individuals within a single cell and commit to different decisions. Next, we infected a single cell with two phages carrying different DNA reporter systems, *tetO*/TetR-FP and *lacO*/LacI-FP. We observed that one phage DNA rapidly replicated to fill the most cell space, while the other exhibited constrained replication by remaining as a small cluster throughout the infection cycle (Figure 2.4D and E). Eventually, the phage with a larger DNA zone produced its corresponding lytic reporter and took over the cell. This observation provides direct evidence that individual phage DNA develops as its own and the competition between two phage DNAs occurs during DNA replication, resulting in one phage dominating the other, most likely through occupying most host resources [83]. However, it is worth mentioning that this predominantly happened when the two phages ejected their genomes at distant locations on the cell surface (63.27%, 31 out of 49 infected cells at MOI = 2). In the cases when the genomes of two infecting phages entered the cells in close proximity, they generally mixed within the same subcellular area and replicated together, followed by the appearance of both lytic reporter signals (85.53%, 65 out of 76 infected cells at MOI = 2) (Figure 2.4F). It suggests that for a co-infection event, phage genome entry site has a great impact on the subcellular organization of different phages, as well as their interactions during the post-infection development.

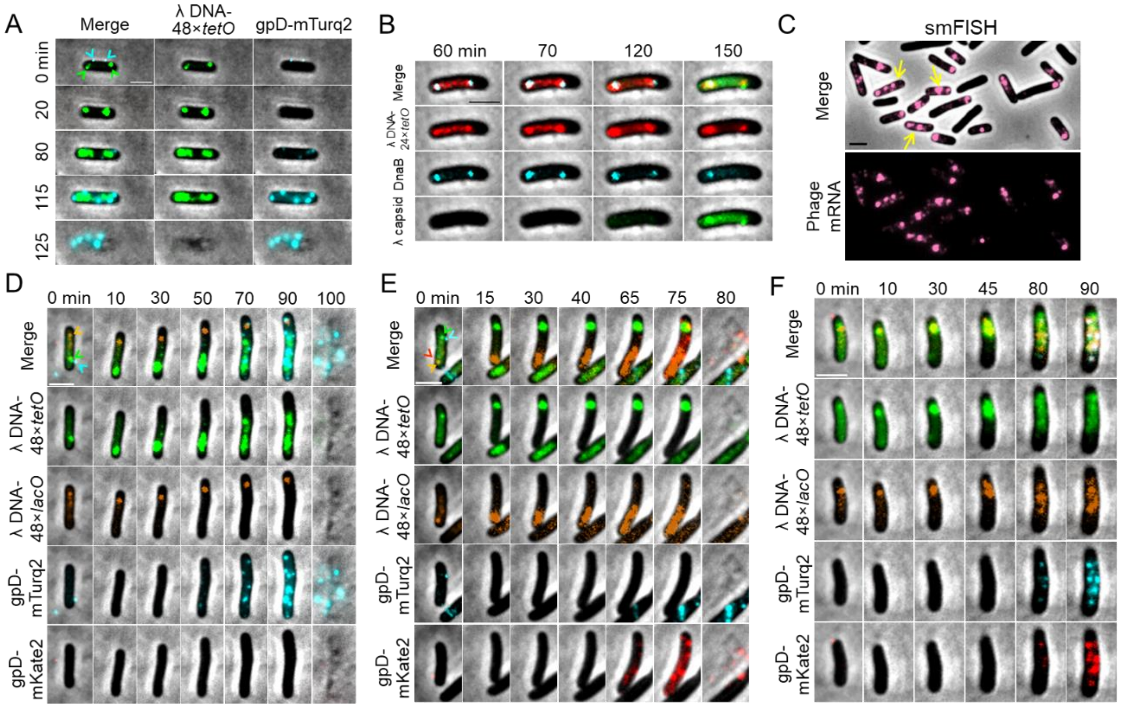

Figure 2.4 Phage DNA can organize individual development by assembling separate compartments in a single cell during infection.
(A) A representative cell infected by two phages labeled with *tetO*/TetR-FP system. The two ejected phage DNAs establish their own subcellular compartments which remain separated throughout the lytic development. The lytic reporter signals emerge from both phage DNA clusters, indicating that virion assembly occurs within both compartments. Cyan and green carets point at phage particles adsorbed to the cell and the injected phage DNA, respectively. (B) A representative cell with two subcellular phage-derived compartments (red). Each compartment recruits host resource DnaB (cyan) to its proximity for its own development. During the late stage, phage capsid gpD-mNeongreen (green) accumulates within the different compartments. (C) RNA smFISH targeting λ *cII* transcripts. *cII* mRNAs form separate clusters in single cells, marked by yellow arrows. (D) and (E) Representative lytic cells co-infected by a *tetO* phage and a *lacO* phage displaying phage competition within a single cell via inhibition of the viral DNA replication. A *tetO* phage dominates a *lacO* phage is shown in (D) and a *lacO* phage dominates a *tetO* phage is shown in (E). In (D), the cell lacks the *lacO* phage which is supposed to attach on the cell surface, meaning that it fell off after ejecting its DNA (orange dot) into the cell. (F) A representative lytic cell showing that the co-infecting *tetO* and *lacO* phage DNAs develop as one unit. At 0 min, the two phages eject their genomes in close proximity. Their replicated DNAs mix together over the entire infection cycle. Finally, both viral progenies are produced indicated by the appearance of the lytic reporters (cyan and red). Scale bars, 2 µm.

The spatial organization of phage DNA is correlated with the presence of E. coli chromosome

We hypothesized that phage DNA clusters remaining separated is due to the presence of the *E. coli* chromosome. It is well known that *E. coli* has a single circular, highly compacted chromosome, occupying a large volume of the cytoplasm [239]. It might behave as a barrier to physically segregate replicating phage DNA. To this end, we inserted an array of 96×*lacO* repeats at ~1.6 kb upstream of the *E.coli attB* site and transferred a plasmid expressing LacI-YFP fusion into the above engineered *E. coli* strain (Figure 2.5A). The labeling of the *attB* site also makes it possible to capture the integration event during lysogenic development for later studies. We next infected the engineered *E. coli* strain with our *tetO* phages (genotype: $\lambda D\text{-}mTurquoise2\ cI_{857}\ bor{::}48{\times}tetO\ Cm^{R}$) to perform single-cell time-lapse movies. As expected, we observed that *E. coli attB* avoided overlapping with phage DNAs during phage DNA replication (Figure 2.5B-E). Collectively, the interaction between phage DNA and bacterial DNA in lytic cells displayed four distinct behaviors: 1) a phage DNA cluster expanded from one side of the cell and pushed the *E. coli* DNA toward the other cell pole (Figure 2.5B), 2) a phage DNA cluster expanded from the middle of the cell and pushed the *E. coli* DNA toward both directions (Figure 2.5C), 3) two phage DNA clusters expanded from the cells poles and confined the *E. coli* DNA within the midcell region (Figure 2.5D), and 4) *E. coli* DNA acted as physical barriers to separate the replicated phage DNAs into multiple clusters (Figure 2.5E). These observations confirm our hypothesis that *E. coli* chromosome participates in establishing and maintaining phage DNA subcellular localization.

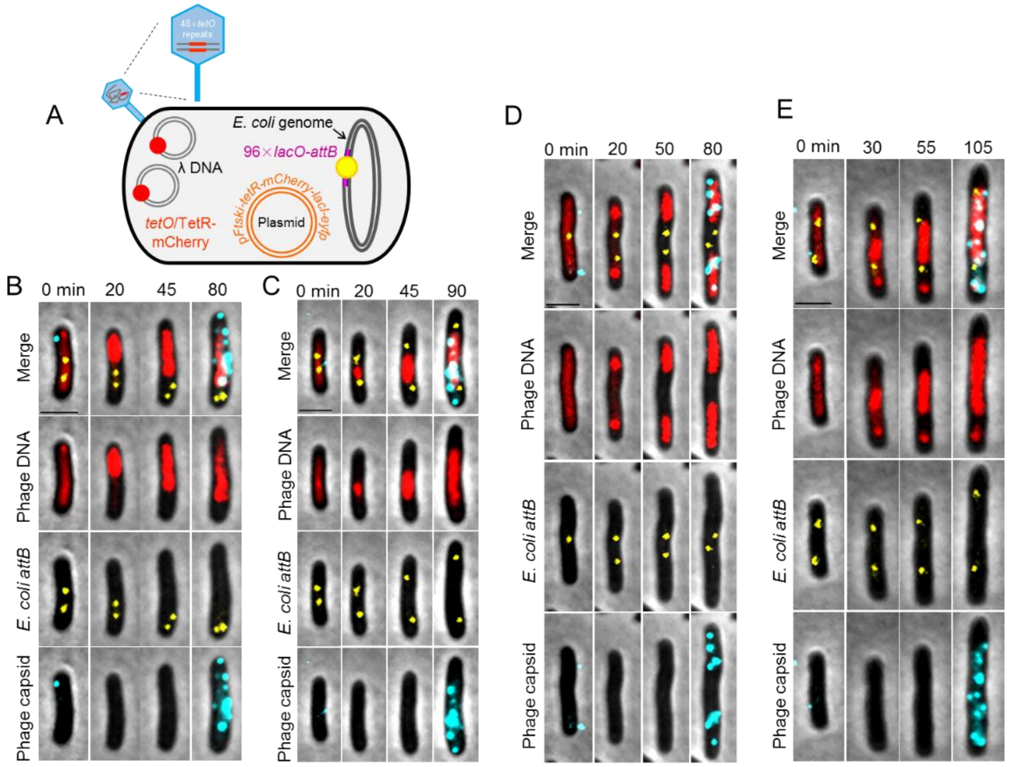

Figure 2.5 Phage DNA intracellular compartments are physically separated by *E. coli* chromosome.

(A) A schematic of *E. coli* DNA and phage λ DNA labeling. A 96× *lacO* array is inserted upstream of the *attB* locus on *E. coli* genome. Phage λ carries a 48× *tetO* array and the gpD-mTurquoise2 lytic reporter as described earlier. A plasmid expressing both TetR-mCherry and LacI-EYFP under the promoter *Ftski* is introduced into the engineered *E. coli* host in order to label bacterial and phage DNA concurrently. (B-E) Representative lytic cells showing four distinct behaviors of the interaction between bacterial *attB* (yellow dots) and phage genomes (red clusters) as described in the text. Scale bars, 2 µm.

Detection of phage DNA in fixed cells by single-molecule DNA FISH

To rule out the possibility that our fluorescent repressor operator systems (*tetO*/TetR-FP and *lacO*/LacI-FP) artificially caused viral genomes to stick together and influenced the biological behaviors, we adapted the DNA single-molecule fluorescence *in situ* hybridization (DNA smFISH) technique to characterize the spatial distribution of phage DNA replication in fixed cells upon infection. We designed fluorescent probes against a 3 kb region (39.9 ~ 42.9 kb) of phage λWT genome and fixed the cell-phage mixtures at different time points after infection (0, 10, 20, 30, 40, 50, and 60 min), followed by the hybridization with the probes. At 0 min, phage DNA appeared as single foci, primarily at the cell poles and mid-cell region. Over time, phage DNA signals increased significantly and formed large clusters similar to those observed in the live-cell movies with the repressor-operator reporter systems (Figure 2.6). At 50-60 min, some infected cells lysed, leaving only phage DNAs sitting on the slides (pointed by red arrows). The observations demonstrate that our phage DNA reporter using the genetic modified operator-binding repressor system does not alter the subcellular behaviors and localization of phage DNA.

In corresponding to the live-cell studies where the *E. coli attB* site was labeled by a *lacO*/LacI-FP system, we further designed probes against the region covering the *attB* site and performed dual-color DNA FISH experiments to target both phage DNA and *E. coli attB* loci. Meanwhile, we stained the fixed, hybridized cells with DAPI to locate the entire *E. coli* chromosome. It is important to emphasize that DAPI theoretically stains both phage DNA and *E. coli* DNA with no preference. However, as the genome size of λ (48.5

kbp) is far smaller than *E. coli* DNA (4.6 Mbp), a single or a few phage DNA molecules are not detectable by DAPI during the early infection period. For this reason, phage DNA and *E. coli* DNA can be distinguished by FISH signals and DAPI staining, respectively. At early time points (i.e., 10 min), phage DNA clusters preferred to stay in the *E. coli* nucleoid-free areas (Figure 2.7A), supporting the live-cell observations that phage DNA and bacterial DNA were not colocalized in the cytoplasm. At later time points, viral and bacterial DNA signals overlapped in the DAPI channel due to the increased amount of phage DNA as a result of DNA replication. Therefore, we moved to examine the location of the *attB* locus, which was labeled by FISH probes, to represent where the *E. coli* DNA stays. We found that *attB* circumvented the phage DNA clusters (Figure 2.7A), consistent with the live-cell data (Figure 2.5B-E). Additionally, our previous *cII* mRNA smFISH samples were also treated with DAPI to stain the *E. coli* chromosomes. From there, we observed that during the early infection stage, similar to phage DNA, λ *cII* mRNA clusters colocalized in the subcellular space where *E. coli* DNA did not reside (Figure 2.7B), signifying that gene expression occurs within the phage-derived compartments, which are partitioned by *E. coli* chromosome. Altogether, the FISH data verified the reliability of the live-cell results that the initial infecting phage DNAs have the ability to develop and form individual compartments, which are physically separated by *E. coli* chromosomes in the cytoplasm. On the other hand, the established DNA FISH technique can be utilized to quantify phage DNA copy number during the infection for further investigation of the importance of phage DNA replication in cellular decision making at the single-cell level.

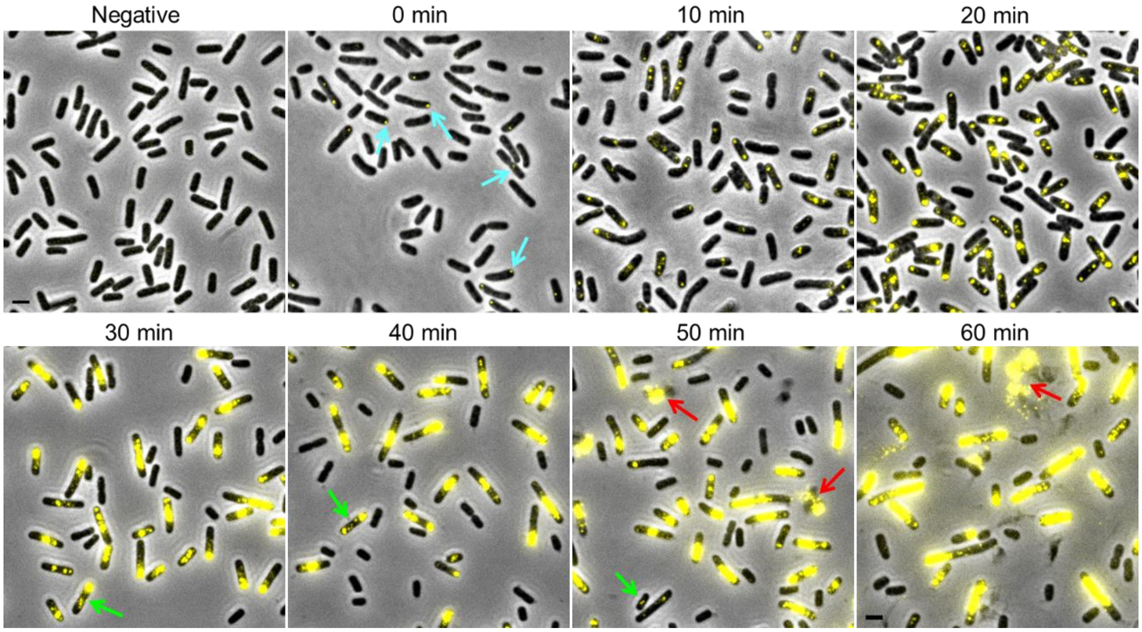

Figure 2II.6 Examination of phage DNA intracellular localization by DNA smFISH. Representative images showing phage λ DNA replication over time by DNA smFISH. Negative: cells without phage infection. None of the cells show phage DNA signal. Cells with λWT infection at API = 1 were fixed at different time points. At 0 min, cells show distinct foci in yellow, representing one or a few copies of phage DNA. Those foci are frequently found at the mid-cell areas and cell poles, pointed by cyan arrows. Over time, as phage DNA replicates, the foci become large clusters, similar to the live-cell images. At the late time points (50-60 min), the clusters fill most cell space and some lysed cells can be observed (pointed by red arrows). Green arrows point to the potential lysogenic cells. Scale bar, 2 µm.

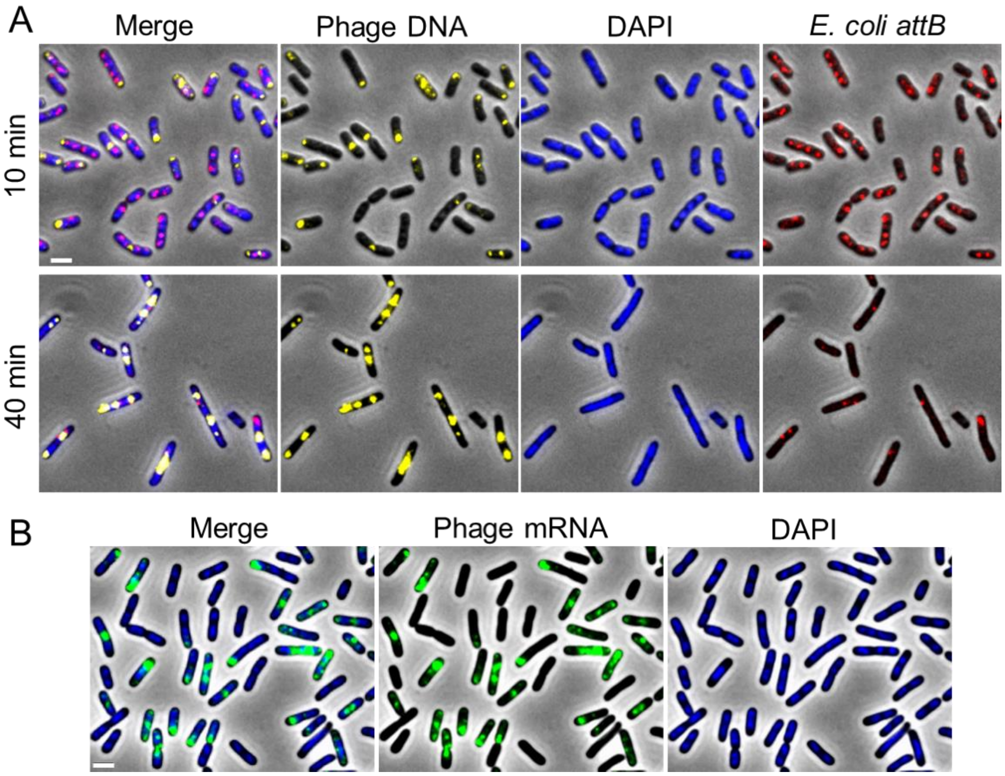

Figure II.7 Phage DNA replication and gene transcription do not colocalize with *E. coli* DNA during the post-infection development.

(A) Representative cells targeting phage DNA (yellow) and *E. coli attB* (red) at an early time point (10 min) and a late time point (40 min) after infection. *E. coli attB* appear as bright dots in each cell. Cells were stained with DAPI, allowing for detection of the entire *E. coli* chromosome, as well as phage DNA. Early after infection, phage DNA clusters are organized in *E. coli* nucleoid-free areas. During the late stage, replicated phage DNAs stay away from *E. coli attB* in the cells. (B) Representative cells from RNA smFISH experiments showing that λ *cII* mRNA (green) prefers *E. coli* nucleoid-free regions as does phage DNA. Scale bars, 2 μm.

Discussion

The lysis-lysogeny decision making of bacteriophage λ has served as an important paradigm for a binary choice between alternative cell fates. The cell-fate choice is driven by λ's genetic regulatory circuit, as well as the infecting parameters, such as environmental conditions and the number of infecting phages (MOI) [18, 71, 76]. Several mathematical models have also been built in order to interpret the decision making process [78, 240]. Recently, a stochasticity scenario has been proposed in which random fluctuations in expression of cell-fate determining genes at the individual-cell level have a great impact on decision outcome [241]. However, quantitative single-cell measurements on the expression of the genetic players at both transcriptional level and translational level during λ infection have not been achieved yet.

In this work, we measured λ *cII* transcription over time upon infection of *E. coli* at the single-cell level. The peak of the overall trend of the *cII* expression within the decision time window (around 10 min) might represent a certain threshold that CII protein need to reach in order to activate the expression of the *cI* gene for lysogeny to be established, as proposed in previous studies [18, 71] Moreover, the *cII* mRNA expression pattern also represents the transcription of the *pR* promoter, which is negatively controlled by Cro and λ repressor CI after they are produced. In addition, phage DNA replication is believed to contribute to the *cII* or *pR* transcription, because it provides more templates for RNA polymerase to bind with, leading to the accumulation of *cII* mRNAs and the transcripts of its negative regulator gene *cro*. However, the increased level of Cro can counteract the effect of increasing template number on *cII* expression. Therefore, *pR* transcription and phage DNA replication might be constantly affecting each other during the decision is being made. Further measurements on the transcriptional level of *cro* and the changes of phage DNA copy number are required for investigating the correlation between phage gene expression and DNA replication.

In addition to phage DNA replication, an alternative source to provide additional DNA templates for transcription is to increase the number of co-infecting phages. It was documented that when multiple viruses infect the same host, nonlinear gene regulatory dynamics can lead to qualitative changes in steady-state gene expression even with a small change in the rate of transcription, thereby ultimately affecting the cell fate decision [78]. Future experiments to characterize the effect of MOI on the expression of the fate-determining genes remain to be done. Besides, it is worth mentioning that our results came from the measurements of the transcription of phage genes and the conclusions were drawn based on the assumption that the translation of phage proteins is proportional to the total amount of their respective transcripts. Since phage-encoded proteins are the factors taking effect in the cell-fate selection, quantification of the protein concentration of those regulators using techniques, like immunofluorescence, is of great importance to characterize the role of genetic regulatory network plays in λ cellular decision making.

Beyond the genetic regulatory network, previous studies point out that during co-infection, individual phages are able to "vote" for the choice of whether to lyse or lysogenize the cell [37]. Here, by examining the spatial distribution of phage and *E. coli* DNA, we provided direct evidence of how multiple phages with identical genetic materials develop as individuals in the limited space of a single cell. When multiple phages co-infected a single cell, their injected DNAs can individually establish separate subcellular compartments, allowing for space-restricted DNA replication and viral gene expression, which sustains heterogeneous viral development in single cells. These phage DNA compartments are physically separated by the *E. coli* chromosome and might have different selections for a cell fate. As a consequence, the overall cell-fate decision is expected to be made by multiple voting from separate subcellular spaces in a single cell. Nonetheless, it is worthwhile to point out that the separate spatial organization of phage DNA does not mean that their intracellular development is completely independent of each other. On the contrary, we observed that DNA replication of one infecting phage remarkably inhibit that of the other, supporting the previous prediction that the competition among co-infecting phages occurs during the DNA replication [83]. Once again, we have been focusing on the phage DNA and mRNA levels but have neglected the spatial distribution of the fate-determining proteins, such as CII, CI and Cro, which directly control the viral gene expression and affect the cell-fate choice. Detection of those proteins at the single-cell level is essential to uncover what degree the regulators can be shared between different phage compartments in the subcellular environment and how the key players interplay to determine the decision-making outcome.

In summary, we investigated phage λ decision-making from the aspects of both phage gene expression pattern and the subcellular spatial organization of phage DNA and mRNA at the single-cell level. Phage gene expression is a consequence of the interaction between multiple factors involved in the complex genetic circuit. The variation of the expression level is affected by the degree of the regulation of those factors and the external environmental fluctuation. The subcellular organization of phage biomolecules generates the heterogeneous development within single cells, which further contributes to noisiness of decision-making at the population level. Nevertheless, there remain unresolved questions of how individual phage DNA senses the local environmental signals to make appropriate response for development and how phage DNA individualities interact in the subcellular space to together make an optimal decision for a cell. Understanding the intracellular spatial structures may yield new insights into the cause of cell-to-cell heterogeneity and make the lysis-lysogeny decision outcome more predictable.

CHAPTER III

THE ROLE OF SIDE TAIL FIBERS DURING THE INFECTION CYCLE OF PHAGE LAMBDA[*]

Introduction

As the most abundant organisms in the world, bacteriophages (simply phages) are viruses that infect bacteria. Phage λ was discovered in 1951 by Esther Lederberg [10]. Ever since its discovery, phage λ has become one of the most comprehensively studied phage systems and serves as paradigms of different biological processes, such as regulation of gene expression, mechanisms of recombination, and cell-fate decision-making [12, 14, 17, 18]. To start an infection, phage λ first carries out the interaction with the host cell through its tail fiber binding to the receptor on the host cell. Subsequently, λ DNA is ejected into the cell, followed by a decision to enter either of two distinct pathways, lysis or lysogeny. In the lytic pathway, λ replicates its DNA and produces new phages, resulting in cell lysis to release about a hundred progeny virions. Alternatively, in the lysogenic pathway, λ establishes the dormant state by integrating its genome into the host chromosome, and existing as a quiescent prophage [17, 21].

The phage λ used in most laboratories around the world, which has been thought of as wild-type λ (λWT) is actually not the original strain that was isolated from a prophage

---

[*] Part of this chapter is reprinted with permission from "The Role of Side Tail Fibers during the Infection Cycle of Phage Lambda" by Jingwen Guan, David Ibarra, and Lanying Zeng, *Virology* **2019** 18, 527: 57-63, Copyright 2019 by Elsevier B.V.

in *E. coli* strain K-12 in 1951 [10]. It is a mutant strain derived from a cross between a λ strain in use in Pasadena and another strain in use in Paris, and later referred to as lambda *PaPa* (λ*PaPa*) [15]. The real wild type strain is referred to as Ur-lambda (Ur-λ). Compared with λWT (or λ*PaPa*) which only has a short tail fiber at the tail tip, Ur-λ virions have additional four long, thin side tail fibers that extend from the side of the tail tip [15, 19]. The absence of the side tail fibers from λWT is the result of a frameshift mutation (a deletion of cytosine) in the side tail fiber (*stf*) gene in λWT genome sequence [15]. Without those side tail fibers, λWT produces larger plaques on cell lawn, which makes them more suitable for genetics studies in the early days [19, 41]. Although the side tail fibers are not essential for λ plaque formation, they greatly accelerate the rate of adsorption onto the host cell surface by ~7.4 fold [42]. That is because in addition to the binding of the λ receptor, LamB, on the outer membrane of *E. coli* by the tail fiber gpJ protein, the extra side tail fibers on Ur-λ can bind to another host outer membrane protein OmpC [12, 15]. Moreover, the side tail fibers presumably slow down the diffusion of Ur-λ through the top agar layer, resulting in the smaller plaque size [41]. However, how the side tail fibers affect phage lambda infection remains largely unknown.

In this study, we characterize the infection cycle of Ur-λ at the single-cell level using fluorescent reporter systems [83]. By following the phage infection, we reveal the differences between Ur-λ and λWT such as the lysogenic response to the number of infecting phages and the failed infection frequency.

Materials and Methods

Bacterial strains, plasmids and phages

All bacterial strains, plasmids and phages used are listed in Table 3.1.

E. coli wild-type strain MG1655 was used as the host strain for bulk lysogenization assay and normal phage movies. *E. coli* strain LZ1386 (MG1655 *seqA-mKO2* Cm^R *Δdam::*Kan^R) used for tracking phage DNA ejection was obtained from our recent work [83]. *E. coli* strain defective in OmpC expression (JW2203, *ΔompC768::kan*) was obtained from the Keio collection (CGSC#9781) [242]. *E. coli* BW25113 was used as a control strain to determine the effect of OmpC on phage infection. Wild-type phage λ (λ_{LZ613}, λcI_{857} *bor::*Kan^R) and wild-type phage λ reporter strain (λ_{LZ1367}, λ*D-mTurquoise2* cI_{857}*-mKO2 bor::*Cm^R) were constructed previously [38, 83]. Ur-λ (λ_{LZ610}, λcI_{857} *stf+ bor::*Kan^R) and Ur-λ reporter strain (λ_{LZ1636}, λ*D-mTurquoise2 stf+* cI_{857}*-mKO2 bor::*Kan^R) were produced by restoring the frameshift mutation present in the side tail fiber gene (*stf*) back into the genome of λWT and λWT reporter strain respectively through site-directed mutagenesis and recombination as described in [42]. Plasmids pZE1-*ΔJ-*Cm^R*-stf* and pZE1-*J-stf+* were received as gifts from Ryland Young, Texas A&M University. All phages were produced through the heat-induction of lysogens, followed by a standard CsCl purification procedure following the protocol described in [243].

Table 3.1 Bacterial strains, phages, and plasmids used in this work.

Bacterial strains, plasmids, and phages		
Strain Name	Relevant Genotype/Comments	Source/Reference
Bacterial strains		
MG1655	*sup*^{0}, Wild type *E. coli*	Lab collection
BW25113	The parent strain of JW2203	Lab collection
JW2203	*ΔompC768::kan*^{R} (CGSC#9781)	[242]
LZ1386	MG1655 *seqA-mKO2 Δdam::Kan*^{R} *Cm*^{R}	[83]
Phage strains		
λ_{LZ613}	λ*cI*$_{857}$ *bor::Kan*^{R}, referred to as λWT or λ*PaPa*	[38]
λ_{LZ610}	λ*cI*$_{857}$ *stf+ bor::Kan*^{R}, referred to as Ur-λ	This work
λ_{LZ1367}	λ*D-mTurquoise2 cI*$_{857}$*-mKO2 bor::Cm*^{R}, referred to as λWT-reporter	[83]
λ_{LZ1636}	λ*D-mTurquoise2 stf+ cI*$_{857}$*-mKO2 bor::Kan*^{R}, referred to as Ur-λ-reporter	This work
Plasmids		
pZE1-*ΔJ-Cm*^{R}*-stf*	Contains part of *J* and *Cm*^{R} replacing the region between *J* and *orf314*, *Cm*^{R} *Amp*^{R}	[42]
pZE1-*J-stf+*	Contains part of *J* and functional *stf*, *Amp*^{R}	[42]

Bulk adsorption assay

The procedure for determination of λ phages adsorption rate in bulk is adopted from Moldovan et al. [32]. Briefly, *E. coli* strain MG1655 was cultured from an overnight in LBMM (LB + 0.2% maltose + 10 mM $MgSO_4$) at 37 °C, 265 rpm, until OD_{600} of 0.4. The cells were concentrated through the centrifugation (1000×g for 10 min at 4 °C) by 10 fold in pre-chilled fresh LBMM, and 10 μl of the concentrated cell culture was aliquoted in microcentrifuge tubes. 10 μl of phages were then added to the cell culture aliquots to reach an API of 1 at specific time intervals. The cell-phage infection mixture was incubated either on ice or at 37 °C in a water bath for adsorption, and the reactions was terminated simultaneously for all samples by a 50 fold dilution using ice-cold SM buffer (100 mM NaCl, 10 mM $MgSO_4$, 0.01% gelatin, and 50 mM Tris-Cl, pH 7.5). The samples were centrifuged at 16,000×g for 1.5 min to sediment the phage-adsorbed bacteria. The concentration of unabsorbed phages remaining in the supernatants was determined as Pt. The phage titer in the supernatant of the initial sample (0 min) was defined as Pt_0. The adsorption rate was calculated as $(Pt_0 - Pt)/Pt_0 \times 100\%$.

Bulk lysogenization assay

As described in previous work [37, 83], host *E. coli* MG1655 cells were diluted 1:100 from an overnight culture and grown in LBMM at 37 °C, 265 rpm, until an OD_{600} of ~0.4. The cells were collected by centrifugation (1000×g for 10 min at 4 °C) and concentrated 10× in pre-chilled fresh LBMM. Phages were diluted from original stocks to reach a maximal API of ~10, and then proceeded into a 2-fold series of dilution until 2^{-8}. 20 μl of the cell suspension was infected with 20 μl of phages at different concentrations by an incubation on ice for 30 min, followed by another incubation at 35 °C for 5 min to trigger phage DNA ejection. Subsequently, 10 µl of each infection mixture was then added into 1 ml of pre-warmed LBGM (LB + 0.2% glucose + 10 mM $MgSO_4$), and incubated with shaking at 265 rpm at 30 °C for 45 min. The samples were then properly diluted with 1×PBS and spread onto LB + Kan or LB + Cm plates to allow 100-200 lysogens to grow at 30 °C for overnight.

Single-cell infection assay

1 ml of host cell MG1655 was grown in M9 minimal medium (11.3 g/L M9 salts, 1 mM $MgSO_4$, 0.5 µg/mL thiamine HCl, 0.1% casamino acids, 100 µM $CaCl_2$) supplemented with 0.4% maltose (M9M) at 37 °C for overnight. The overnight culture was subsequently diluted 1:100 into 5 ml of M9M and grown at 37 °C with 265 rpm shaking until OD_{600} of ~0.3. 1 ml of cells were then concentrated by centrifugation at 2000×g for 2 minutes at room temperature, and resuspended in ice-cold M9M to OD_{600} of ~3. 10 µl of phage stock was mixed with 10 µl of the resuspended cells to reach an appropriate API, followed by incubation on ice for 30 min and an additional 5 min incubation at 35 °C water bath to trigger phage DNA ejection. The phage-cell mixture was then diluted by 10 fold into 50 µl of M9M at room temperature. 1 µl of the diluted mixture was placed onto a 1.5% agarose pad of M9M (~1 mm thick) until visibly dry (~1 min). A coverslip (No.1, Fisher Scientific) was gently laid over the mixture and the sample was imaged under the fluorescence microscope at 30 °C within a cage incubator (InVivo Scientific, St. Louis, MO). For the phage DNA reporter movies, the same protocol was performed but with the reporter host strain (MG1655 *seqA-mKO2* Cm^R $\Delta dam::Kan^R$).

Fluorescence microscopy and imaging

Microscopy was performed on an inverted epifluorescence microscope (Ti-E, Nikon, Tokyo, Japan) using a 100×objective (Plan Fluo, NA 1.40, oil immersion) with a 2.5×TV relay lens and standard filter sets. Images were acquired using a cooled EMCCD camera (iXon3 897, Andor, Belfast, United Kingdom). Acquisition was performed using Nikon Elements software. Typically, 16 stages with well-separated cells were selected for each movie. In order to visualize all infecting phages surrounding the cells, a series of 7 z-stack (vertical) images at a spacing of 300 nm were taken for the first frame through the blue (mTurquoise2/CFP, 200 ms exposure) channel. During the time-lapse movie, the sample was imaged every 5 min at the focal plane for 4 hours through the channel of the phase contrast (100 ms exposure, for cell recognition), blue (mTurquoise2/CFP, 40 ms exposure, for phage lytic reporter), and yellow (mKO2, 100 ms exposure, for phage lysogenic reporter).

For the phage DNA reporter movies, the first frame of the sample was imaged with 7 z-stacks at a spacing of 300 nm under both blue (mTurquoise2/CFP, 300 ms exposure) and yellow (mKO2, 200 ms exposure) channels to localize infecting phages surrounding the cells and detect phage DNA foci inside the cells. The time-lapse movies were taken at a time interval of 2 min, until the phage DNA foci were not detectable (1 hours) due to photobleaching. The images were acquired in phase contrast (100 ms exposure, for cell recognition), blue (mTurquoise2/CFP, 100 ms exposure, for phage lytic reporter) and yellow (mKO2, 200 ms exposure, with a series of 7 z-axis images at a spacing of 300 nm, for phage DNA detection) channels.

Data analysis

Movie images were analyzed using the cell recognition program Schnitzcell (gift of Michael Elowitz, California Institute of Technology) in the phase-contrast channel. All data analysis was performed in Matlab (The MathWorks, Natick, MA) using our homemade script.

Results

The side tail fibers increase the adsorption of Ur-λ

In order to characterize the infection of Ur-λ at the single-cell level, we took advantage of our previously established fluorescent reporter system to visualize infecting phage particles and report the lysis-lysogeny pathways under the fluorescence microscope [83]. Briefly, fluorescent protein gene *mTurquoise2* was translationally fused to λ*D* gene, encoding λ capsid decoration protein, which allows for the visualization of the infecting phages and reporting the lytic pathway. Meanwhile, the lysogenic pathway is reported by fluorescent protein mKO2 which was transcriptionally fused after λ*cI* gene, whose product is required for lysogenic establishment and maintenance (Figure 3.1A). These two modifications result in the λWT reporter strain (λ_{LZ1502}, λ*D-mTurquoise2* cI_{857}*-mKO2 bor::Kan*^{R}). Subsequently, through the genetic recombination based on the method in [42], we restored the functional λ*stf* gene back into the genome of the above λWT reporter strain and sequenced the entire *stf* gene and confirmed the insertion of an additional cytosine at positions 20833 to 20835 relative to λWT genome sequence. Eventually, an Ur-λ phage with the same dual-color reporter system was constructed: λ_{LZ1636}, λ*D-mTurquoise2 stf+*

cI$_{857}$*-mKO2 bor::Kan*^{R}. Thereafter, we performed single-cell infection experiments with both Ur-λ and λWT reporter phages.

We first compared the abilities of Ur-λ and λWT to adsorb onto the host cells. As reported in the literature [15] and our bulk measurements [39], Ur-λ has an increased adsorption rate compared with λWT (Figure 3.2). To examine the adsorption at the single-cell level, we followed our standard single-cell infection protocols [37, 83]. Briefly, *E. coli* host cells were mixed with Ur-λ or λWT phages, followed by an incubation on ice for 30 min to allow phages adsorb onto host cells and an additional incubation at 35°C for 5 min to trigger phage DNA ejection. The infection mixture was then transferred onto an agarose pad of medium for subsequent imaging. From the images, by calculating the ratio of the number of fluorescent phages attached on the cell surface to the number of all phages shown in the frame, we were able to quantify the adsorption extents of both phages. 84.0 ± 2.2% (6746 attached phages out of 8034 total phages in 11 experiments) of Ur-λ phages were adsorbed on *E. coli* host cells, significantly more than 77.3 ± 2.6% (6365 out of 8239 phages in 9 experiments) of λWT phages (Mann-Whitney *U*-test, $p < 0.05$) (Figure 3.1B). This confirms that the side tail fibers help Ur-λ to encounter the host for infection.

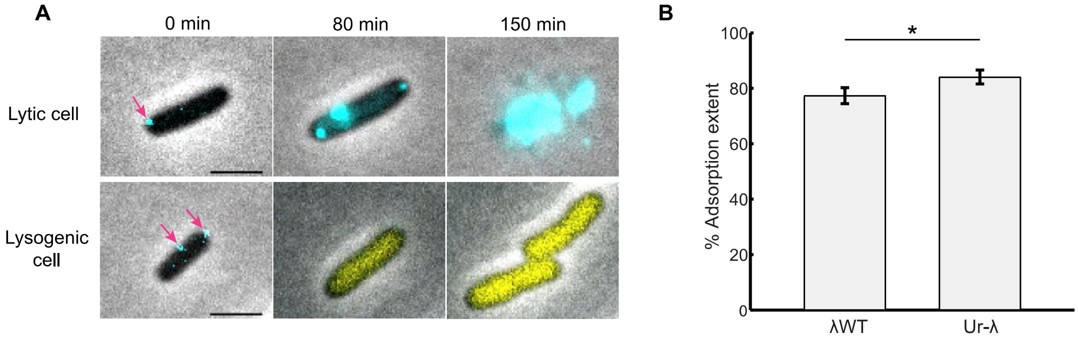

Figure 3.1 Visualization of phage λ decision-making and the adsorption assay in single cells.

(A) Top: a representative lytic cell was infected by one λ phage at 0 min (blue dot, pointed by magenta arrow), and subsequently gpD-mTurquoise2 fluorescence (blue) developed over time, followed by cell lysis at 150 min. Bottom: a representative lysogenic cell was infected by two λ phages at 0 min (blue dots, pointed by magenta arrows), and subsequently *cI*-mKO2 fluorescence (yellow) developed over time, followed by cell growth and cell division. Scale bar denotes 2 µm. (B) Ur-λ has a higher adsorption extent (84.0 ± 2.2%) than λWT (77.3 ± 2.6%) at the single-cell level. *: $p < 0.05$. Error bars denote standard error of the mean (SEM).

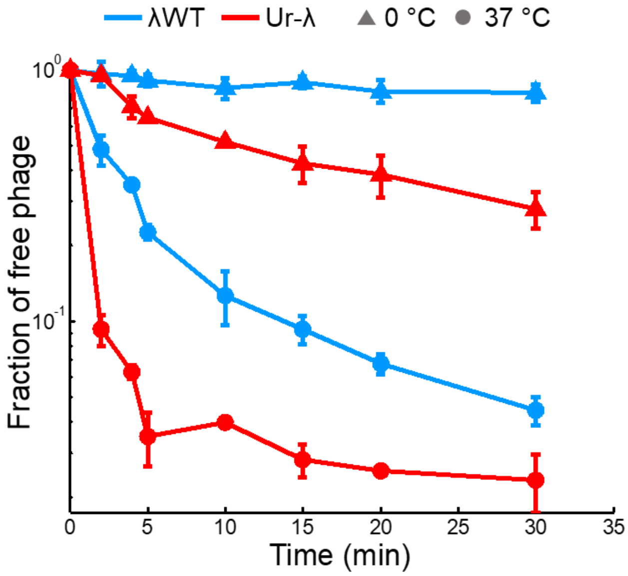

Figure 3.2 Phage adsorption assay of λWT and Ur-λ at the bulk level.
After mixing λWT or Ur-λ phages with *E. coli* host cells, the normalized free phage concentration present in solution is measured and plotted as a function of time for two temperatures, 0 °C (triangles) and 37 °C (circles). Ur-λ (red) adsorbs to host cells substantially faster than λWT (blue). The data are collected from three independent experiments and the mean are shown. Error bars denote standard error of the mean (SEM).

Ur-λ has a higher chance of failure in the infection

We next sought to determine whether the adsorbed phage can successfully infect the cell. A failed infection event is marked when a cell fails in establishing either lytic or lysogenic pathway even with phages attached (Figure 3.3A). It was reported that phage lambda fails around 20-25% under the standard infection condition [37, 51]. With the side tail fibers, we expect that Ur-λ may have a reduced failed infection frequency as those side tail fibers could possibly stabilize the adsorption and facilitate phage DNA ejection process. However, to our surprise, Ur-λ exhibited a much higher failed infection frequency (42.0 ± 1.6%, 261 out of 606 cells in 11 experiments) than λWT (29.2 ± 1.8%, 173 out of 593 cells in 9 experiments) at MOI = 1 (multiplicity of infection, the number of infecting phages for each cell) ($p < 0.001$) (Figure 3.3B). This phenomenon holds true for higher MOIs (Figure 3.3C).

Phage λ tail fiber gpJ directly interacts with the receptor LamB for infection [244]. To rule out the possibility that our Ur-λ phage possesses a deficient tail fiber gpJ due to any mutations in the sequence of gene *J*, leading to the higher failed infection frequency, we sequenced the *J* gene of the unlabeled and fluorescently labeled λWT and Ur-λ phages and confirmed that the sequences of the *J* gene in all four phages are identical to the *J* gene sequence reported in Sanger et al. [245] (RefSeq accession no. NC_001416.1).

The failed infection could be due to unsuccessful or incomplete ejection of phage DNA into the cytoplasm (failed ejection), or inability to finish either the lytic or lysogenic pathway with a successful DNA ejection (failed establishment). Our previous work showed that the former case is the dominant cause for the failed infection [38]. Here, to compare Ur-λ with λWT, we utilized our established SeqA-DNA binding reporter system to visualize ejected phage DNA inside the cell [38]. Briefly, we prepared E. coli host with dam- mutation and seqA-mKO2 fusion on its genome [83]. Phage λ DNA is known to be partially methylated when phages are propagated in wild-type *E. coli* cells [246]. Once a successful ejection of phage DNA happens, the uniformly distributed SeqA-mKO2 protein will bind to the ejected DNA forming a fluorescent focus. The successful ejection is indicated by the formation of SeqA foci, and when the cell is followed by lytic fluorescence signal (lytic cell) or *cI*-mKO2 signal followed by cell division (lysogenic cell), this marks as successful infection. In contrast, the failed ejection was indicated when SeqA-mKO2 protein remains as uniform distribution inside the cell with fluorescent phage(s) attached on the cell surface. At MOI = 1, 38.4 ± 3.0% (135 out of 352 cells in 10 experiments) of Ur-λ phages failed to eject their DNAs into the cells, whereas λWT phages remained a lower frequency of 24.7 ± 1.3% (96 out of 388 cells in 9 experiments) ($p < 0.01$) (Figure 3.3B). These numbers are comparable to the failed infection frequencies measured above, which confirms that the higher failed infection frequency of Ur-λ is mainly caused by failed DNA ejection rather than failed establishment of cell fate. This leads to our hypothesis that the side tail fibers of Ur-λ might be involved in the process of phage DNA ejection.

We then speculated whether our single-cell infection protocol artificially increased the failed infection frequency of Ur-λ. The current picture of phage lambda infection is that phage first follows a free 3D diffusion in solution. When it lands on the cell, the 3D diffusion will transit into a 2D motion on the cell surface. Finally, the phage's tail fiber gpJ irreversibly binds to a LamB receptor, preferentially at cell poles or midcell (future pole) positions for DNA ejection [35, 36]. This was achieved by a relaxed liquid condition of phage lambda tail fiber gpJ to find the LamB receptor on the cell surface. However, in our standard protocol for the single-cell studies, phages and cells were first mixed in a tube and incubated on ice for 30 min, followed by another 5 min incubation at 35°C to trigger phage DNA ejection. After that, the phage-cell mixture was transferred onto a thin 1.5% agarose pad and covered by a coverslip for subsequent microscopy imaging. Consequently, it is possible that under our experimental conditions, Ur-λ has not found a LamB receptor at its preferential positions on the cell to accomplish DNA ejection before the phage-cell mixture was transferred onto the agarose pad. In addition, those Ur-λ phages which have already arrived at a LamB receptor site, the tail fiber gpJ might not be able to perfectly interact with LamB due to the agarose pad, leading to the failed DNA ejection. In order to ensure that phages have sufficient time to search for the receptor sites and accomplish DNA ejection, we performed the same single-cell experiments but extended the triggering time for phage DNA ejection at 35°C from 5 min to 10 min and 20 min. Ur-λ still failed more frequently than λWT, with 41.4 ± 3.4% (48 out of 116 cells) versus 29.8 ± 4.0% (17 out of 57 cells) and 44.2 ± 1.9% (58 out of 130 cells) versus 29.2 ± 3.0% (28 out of 96 cells) for 10 min and 20 min incubation at MOI = 1, respectively (Figure 3.3B). This indicates that our standard single-cell experimental procedure does not artificially give rise to the higher failed infection frequency of Ur-λ.

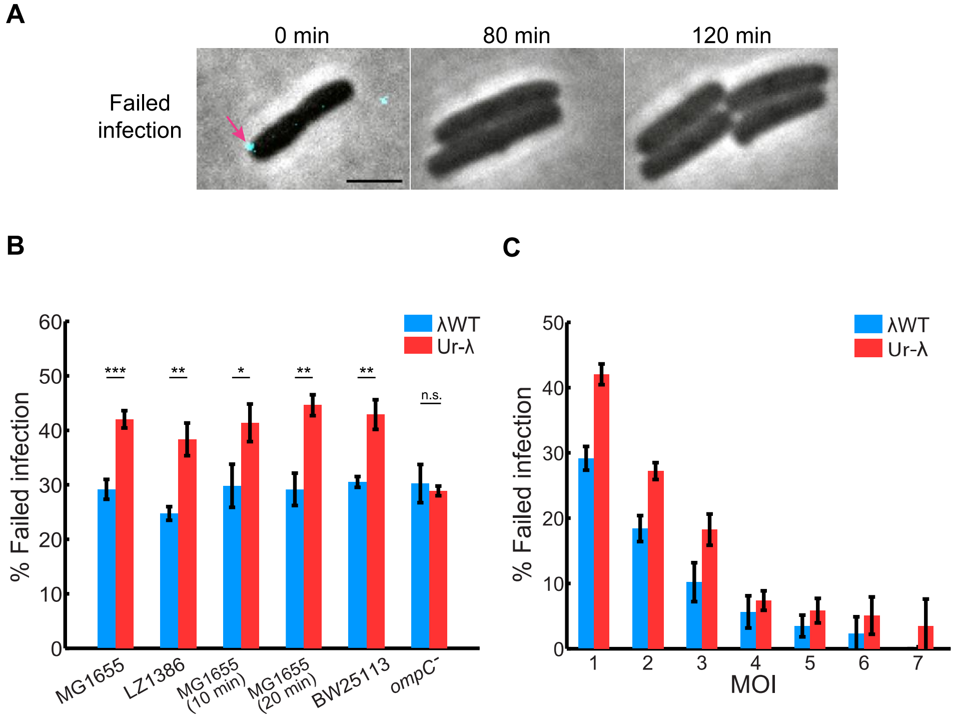

Figure 3.3 The side tail fibers of Ur-λ contribute to failure in the infection.
(A) A representative cell showing the failed infection event by one phage λ (blue dot, pointed by magenta arrow at 0 min). The cell did not show either gpD-mTurquoise2 or *cI*-mKO2 fluorescence over time and divides like uninfected cells, indicating that the phage failed to establish either the lytic or lysogenic development. Scale bar denotes 2 µm. (B) The failed infection frequencies of Ur-λ and λWT at MOI = 1 in different *E. coli* strains. Ur-λ shows a higher frequency of failure in ejecting phage DNA than λWT in MG1655, LZ1386, MG1655 with 10 min treatment, MG1655 with 20 min treatment, and BW25113. In ompC- strain, Ur-λ has a similar failure frequency as λWT. *: $p < 0.05$, **: $p < 0.01$, ***: $p < 0.001$, and n.s.: not significant. λWT: blue, Ur-λ: red. (C) The failed infection frequency of λWT and Ur-λ as a function of MOI in MG1655. For both phages, the failed infection frequency decreases with MOI. Ur-λ (red) fails more frequently than λWT (blue) at different MOIs. In all plots, error bars denote standard error of the mean (SEM).

The interaction between the side tail fibers and their receptors contributes to the failed infection

We next asked how the presence of side tail fibers causes the high frequency of failed infection for Ur-λ. It has been reported that the side tail fibers of phage λ can recognize an outer membrane protein OmpC on *E. coli* [41, 45]. Once phage λ attaches on *E. coli* host, its tail fiber searches for LamB receptors on the cell surface to initiate DNA ejection. Given that OmpC proteins are located at random locations and tend to move freely on *E. coli* outer membrane [247, 248], it is possible that by interacting with OmpC receptor, the phage side tail fibers indirectly impede the binding of phage tail fiber gpJ to LamB, which has also been shown to be mobile [34], for DNA ejection. In order to test this hypothesis, we used an *E. coli ompC*$^-$ strain (JW2203, *ΔompC768::kan*^{R}) to determine the effect of OmpC on phage failed infection frequency. We also used *E. coli* BW25113, the parent strain of JW2203, as a control. Through the same single-cell experiments outlined in the previous section, we found that the failed infection frequency of Ur-λ decreased significantly from (42.9 ± 2.7%, 142 out of 331 cells at MOI = 1 in 4 experiments) in BW25113 host to 28.9 ± 0.9% (63 out of 218 cells at MOI = 1 in 3 experiments) in *ompC*$^-$ host ($p < 0.01$), while λWT kept a similar failed infection frequency in BW25113 to that in *ompC*$^-$ host (30.5 ± 1.0%, 127 out of 416 cells at MOI = 1 in 4 experiments, versus 30.2 ± 3.5%, 75 out of 248 cells at MOI = 1 in 3 experiments, $p = 0.62$) (Figure 3.3B). The data confirm the hypothesis that the interaction between the side tail fibers and their receptors on the host surface is involved in interfering with the Ur-λ infection process. When this interaction is removed, Ur-λ is able to restore the higher frequency of successful infection.

The side tail fibers make a stronger binding Ur-λ on the cell surface

In addition to the failed infection, we examined dark infection, indicated by successful infection with cells ending up with lytic or lysogenic fates even without any fluorescent phage particles detectable on the cell surface (Figure 3.4A). It is possibly a result of phages shearing off from the host cells after ejecting their DNA, likely caused by the performance of the experiments [37]. Or some of the dark infection cells could be the daughter cells dividing from the initial infected cells before imaging [37]. The dark infection frequency is calculated as the ratio of the number of dark infections to the total cell number of MOI = 1 plus the number of dark infections, based on the assumption that all dark infection cells are mostly infected by only one phage. We expect Ur-λ having a reduced dark infection frequency, because besides the binding by the tail fiber to the main receptor LamB, its side tail fibers provide an additional binding to OmpC proteins on the host cell surface, resulting in a tighter adsorption. Our single-cell studies showed that Ur-λ indeed had a lower dark infection frequency (8.3 ± 1.4%, 66 out of 728 cells at MOI = 1 in 11 experiments) compared with 15.0 ± 2.3% (105 out of 593 cells at MOI = 1 in 9 experiments) of λWT ($p < 0.05$) when infecting MG1655 (Figure 3.4B). On the other hand, upon infection of *ompC*$^-$ host, Ur-λ exhibited an indistinguishable dark infection frequency (15.2 ± 2.8%, 39 out of 218 cells at MOI = 1 in 3 experiments), compared with λWT (15.1 ± 1.8%, 44 out of 248 cells at MOI = 1 in 3 experiments) ($p = 0.49$) (Figure 3.4B). As a control, Ur-λ still had a lower dark infection frequency than λWT in BW25113 (9.1 ±

0.6%, 33 out of 331 cells at MOI = 1 in 4 experiments, versus 12.6 ± 2.2%, 60 out of 416 cells at MOI = 1 in 4 experiments, $p < 0.05$, Figure 3.4B). These results suggest that the side tail fibers make a stronger adsorption of Ur-λ on cell surface through binding to OmpC, which leads to less dark infection.

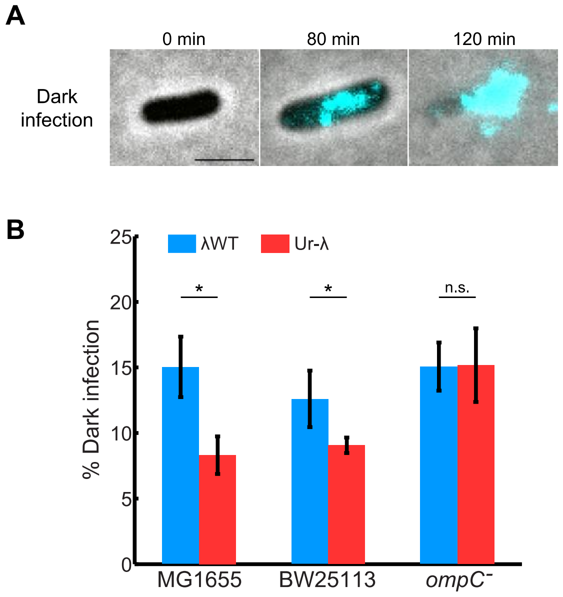

Figure 3.4 The side tail fibers enhance the adsorption of Ur-λ on the host cell. (A) A representative cell showing the dark infection event. No fluorescent phage particle was detected on the cell surface. However, the cell showed lytic reporter signal over time, followed by cell lysis at 120 min. Scale bar denotes 2 µm. (B) The dark infection frequencies of Ur-λ are lower than those of λWT at MOI = 1 in wild-type *E. coli* MG1655 and BW25113. In *ompC*$^-$ strain, Ur-λ shows a similar dark infection frequency compared with λWT. *: $p < 0.05$, and n.s.: not significant. λWT: blue, Ur-λ: red. Error bars denote standard error of the mean (SEM).

The effect of the side tail fibers on Ur-λ post-infection

We first characterized the effect of the side tail fibers on Ur-λ DNA ejection time, using the aforementioned phage DNA labeling system. For phage λWT, it was reported that the DNA ejection is completed with a mean of 5 min ranging from 1 min to 20 min with great cell-to-cell variability [50]. Our previous work using SeqA-DNA reporter system also showed that approximately 95% of phage DNAs appeared inside the cell within 5 min following phage infection [38, 249]. Here, by examining the appearance time of fluorescently labeled phage DNA inside the cell, we did not observe a significantly different distribution between Ur-λ and λWT (Figure 3.5A). 79.7 ± 5.1% (118 out of 148 cells in 10 experiments) of Ur-λ DNA fluorescence foci showed up within 6 min, comparable to 85.2 ± 4.0% (195 out of 229 cells in 9 experiments) for λWT ($p = 0.35$). Considering the higher failed infection frequency of Ur-λ, the data further imply that the side tail fibers might not be able to promote Ur-λ to eject its DNA in a faster, more successful and synchronized manner.

We next sought to investigate whether Ur-λ exhibits different cellular decision-making behaviors from λWT. We first examined the lysogenization response to MOI in bulk. In the bulk assay, MOI is referred to as API (average phage input), the ratio of phage concentration in pfu (plaque forming units) to cell concentration in cfu (colony forming units) [76]. As expected, Ur-λ increased with API, but exhibited a higher lysogenization frequency than λWT. This higher lysogenization frequency might be resulted from the higher adsorption extent of Ur-λ giving rise to a higher effective API. In addition, Ur-λ followed the same $N \geqslant 2$ Poisson distribution as λWT, indicating that the regulation for lysis-lysogeny decision-making remains the same (Figure 3.5B). This is reasonable since *stf* gene lies in the late operon and is only turned on during the late stage of infection when a lytic decision has been made. Furthermore, the two reporter phage strains lysogenized like their corresponding unlabeled strains (Figure 3.5B), suggesting that the genetic modification for the fluorescent reporter does not affect the lysogenization and adsorption behavior. We then moved to test whether Ur-λ lysogenizes the cell in the same manner as λWT at the single-cell level. We found that the lysogenization probability of Ur-λ increased with MOIs, similar to λWT. At MOI = 1, Ur-λ and λWT exhibited a similar lysogenization frequency since Ur-λ shares the same lysis-lysogeny decision-making circuity as λWT as expected. However, at MOI > 1, Ur-λ phages lysogenize the cell slightly less frequently than λWT (Figure 3.5C). This is probably due to the higher failed infection frequency (42.0%) of Ur-λ than that of λWT (29.2%) resulting in a lower effective MOI of Ur-λ than that of λWT.

Furthermore, we examined another important downstream effect, lysis time. Lysis time is an important phage trait, which often defines phage fitness [42]. It was reported that phage λ with and without the side tail fibers both had a lysis time of 52.3 min in bulk [42]. Here, our single-cell data showed that Ur-λ shared a similar distribution of lysis time to λWT (Figure 3.5D) with the average lysis time of 112 min (898 lytic cells) for Ur-λ and 104 min (895 lytic cells) for λWT respectively. Note that our growth condition is very different from that in [42], where a rich medium LB was used as the growth medium and the cell-phage mixtures were incubated in culture flasks with good aeration [42]. Whereas, in this study, we used M9 minimal medium for cell growth in order to reduce the fluorescence background under the microscope. Moreover, the cell-phage mixtures were placed under an agarose pad without shaking. In addition, the lysis time of λWT in this study agrees well with a previous study under a similar growth condition [83].

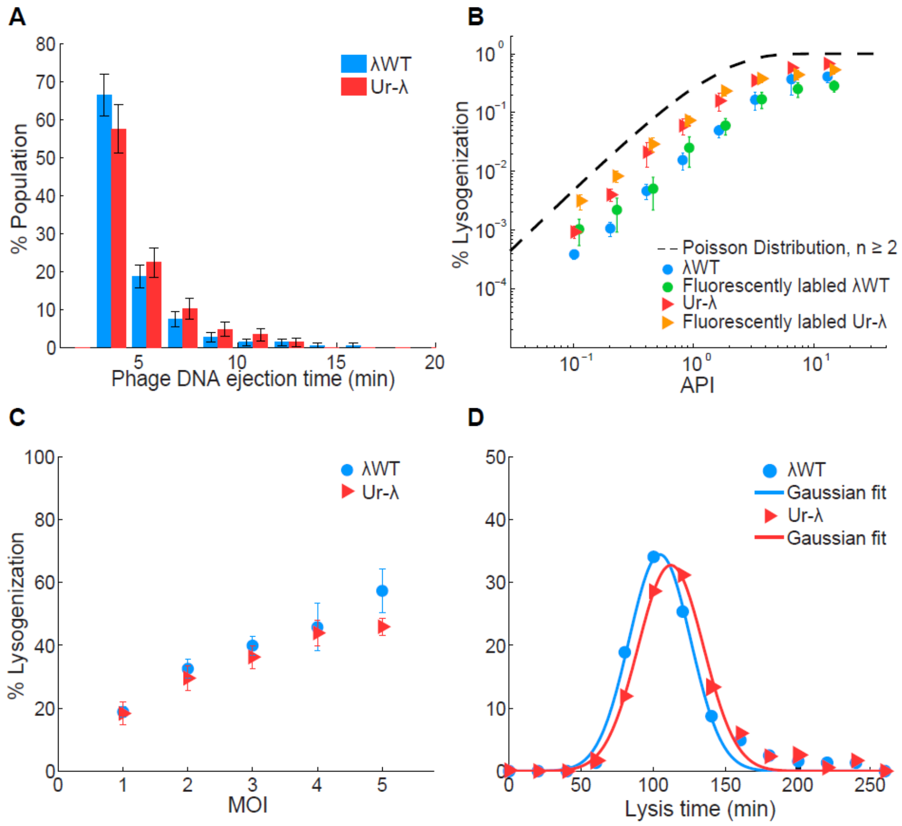

Figure 3.5 The effect of the side tail fibers on Ur-λ post-infection.

(A) Histogram of λWT and Ur-λ phage DNA spot appearance time. 85.2% of λWT DNA fluorescence foci appear within the first 6 min in the movies, while 79.7% of Ur-λ DNA foci appear within the same time frame. λWT: blue, Ur-λ: red. Data are represented as mean ± SE (based on counting error). (B) Lysogenization frequency of λWT and Ur-λ as a function of API in bulk. All phage strains follow the theoretical prediction of Poisson distribution of n $\geqslant 2$ (black dashed line). Ur-λ (red triangles) has higher lysogenization frequencies than λWT (blue circles) at different APIs. Fluorescent Ur-λ (orange triangles) and fluorescent λWT (green circles) lysogenize similarly as their corresponding unlabeled phage strains. Error bars denote standard error of the mean (SEM). (C) Lysogenization frequency of λWT and Ur-λ as a function of MOI at the single-cell level. The lysogenization frequencies of both phages increase with MOI. At MOI = 1, Ur-λ has a similar lysogenization frequency (18.3 ± 3.5%) to that of λWT (18.8 ± 1.8%). At MOI > 1, Ur-λ exhibits slightly lower lysogenization frequencies. Error bars denote standard error of the mean (SEM). λWT: blue circles, Ur-λ: red triangles. (D) Distribution of lysis time for λWT and Ur-λ lytic cells. Ur-λ (red triangles) and λWT (blue circles) show similar distributions of lysis time, with an average lysis time of 112 min and 104 min respectively. Experimental data were well fitted to a Gaussian function (lines).

Discussion

With the advancement of high-resolution microscopy, researchers can examine biological systems at unprecedented levels. It turned out that phage lambda has been discovered to have quite surprising behaviors than previously thought [37, 39, 50, 83, 93]. For example, lambda ejects its DNA into its host cell with great cell-cell variability [50], which can be due to the mobility of the encapsidated phage DNA [52]. The timing of phage lambda DNA ejection greatly affects the downstream lysis-lysogeny decision making and development of the phage [39, 83]. We would expect that the four side tail fibers of Ur-λ facilitate the DNA ejection process due to their extra binding on the cell surface. However, we found Ur-λ exhibited more frequent failed DNA ejection. The DNA ejection happens when phage lambda tail fiber gpJ interacts LamB receptor. So, it seems that the binding of side tail fibers to OmpC receptor disturbs the optimal interaction between gpJ and LamB. This raised the question why the real wild type lambda even needs the four side tail fibers. We know that with the four side tail fibers, Ur-λ can adsorb to the host at a much faster rate and higher extent. Therefore, Ur-λ can have more efficient infection even when each individual phage might fail more frequently compared to λWT. This is probably why Ur-λ would still be a phage with higher fitness [42].

CHAPTER IV

VISUALIZATION OF PHAGE DNA DEGRADATION BY A TYPE I CRISPR-CAS SYSTEM AT THE SINGLE-CELL LEVEL[*]

Introduction

CRISPR-Cas system is a widespread adaptive immune system in prokaryotes including almost half of bacteria and most archaea [124]. Short fragments of foreign DNA, or spacers, are integrated into Clustered, Regularly Interspaced Short Palindromic Repeat (CRISPR) loci on the host genome. The prokaryotes then apply the acquired spacers, which serve as memory elements, to target and cleave invading nucleic acids to defend against viruses and plasmids [104, 119, 124, 170, 250]. In the CRISPR loci, the spacers are separated by host-derived, short, repeated sequences, or repeats. Other important components of CRISPR system include CRISPR-associated (*cas*) genes and a leader sequence, which are found near the CRISPR region. There are multiple kinds of *cas* genes associated with a specific CRISPR [119, 251]. During CRISPR action, the repeat-spacer array is first transcribed into a single transcript, which is further processed by Cas proteins into small RNAs (CRISPR RNA, or crRNA) [104, 165, 252]. The mature crRNA and Cas proteins then assemble to form multicomponent CRISPR ribonucleoprotein (crRNP)

---

[*] Part of this chapter is reprinted with permission from “Visualization of Phage DNA Degradation by a Type I CRISPR-Cas System at the Single-cell Level” by Jingwen Guan, Xu Shi, Roberto Burgos, and Lanying Zeng, *Quantitative Biology* **2017** 5(1), 67-75, Copyright 2017 by Higher Education Press and Springer-Verlag GmbH 2017.

complexes. The crRNP scans invading DNA by base pairing and on successful recognition, and then the target is eventually degraded by Cas nucleases [200, 251].

According to the recent classification of CRISPR-Cas systems, there are two classes with fundamental differences in the organization of the effector module [113]. Class 1 systems include types I, III and IV, and class 2 systems include types II, V, and VI [124, 170]. Much of our understanding of type I system is gained from the studies of type I-E system of *E. coli* K12 [202, 251]. Recently, research has significantly advanced our knowledge about the less understood processes such as the acquisition of new spacers, for example, how the new spacers are produced and integrated into the CRISPR arrays [170, 173, 183, 253]. In addition, the mysteries of the structures and mechanisms of the target recognition have been largely uncovered recently [175, 201, 203, 250, 254, 255]. In the *E. coli* type I-E system, there is a surveillance complex known as Cascade (CRISPR-associated complex for antiviral defense), a 405-kD complex of short crRNAs and five Cas proteins [110]. Cascade binds to target DNA sequences and recruits Cas3 enzyme (a trans-acting nuclease-helicase, the signature Cas protein for type I system) to unwind and degrade the bound foreign DNA [175, 201, 202]. Recently, *in vitro* studies showed that linear DNA target is degraded in a unidirectional manner, and the degradation rate of negatively supercoiled DNA is 4.5 fold faster than the linear one [202]. However, direct demonstration of DNA cleavage *in vivo* has not been provided yet.

Here, we performed *in vivo* experiments with the type I-E CRISPR system of *E. coli* at the single-cell level. An artificial, plasmid-based CRISPR system was introduced into *E. coli* to target bacteriophage lambda [110], where we monitored CRISPR action by infecting cells with our phage strain. With our reporter system, we were able to track *in vivo* lambda DNA degradation over time under the fluorescence microscope. This work provides insights on how CRISPR breaks down invading DNA.

Materials and Methods

Bacterial strains, plasmids and phages

All bacterial strains, plasmids and phages are listed in Table 4.1.

The CRISPR/control strains (LZ1437/LZ1436) are BA16 (or MG1655 *seqA-yfp* dam^-) [256] harboring three plasmids pWUR397A (*cas3*, Amp^R) [257], pWUR400 (*casABCDE*, Str^R) and pWUR478 (λ phage targeting spacers, Cm^R) / pWUR477 (control spacers, Cm^R) [110].

Fully methylated phage λ_{LZ760} was produced by inducing the lysogen MG1655 (λ*D-mTurqoise2* cI_{857} *bor::*Kan^R) [pBR322-pLate-*D*] harboring a Dam methylase overproducing plasmid pZA32-*dam* (with 1 mM IPTG induction) [65]. The crude lysate was then purified from a standard CsCl purification procedure following the protocol described in [243].

Table 4.1 The bacteria, plasmids, and phages used in this study.

Bacterial strains, plasmids, and phages		
Strain Name	Relevant Genotype	Source/Reference
Bacterial strains		
BA16	MG1655, *dam*$^{-}$, *seqA-yfp*, *Cm*^{R}	[256]
LZ1436	BA16 [pWUR397A, pWUR400, pWUR477], *Amp*^{R}, *Str*^{R}, *Cm*^{R}	This work
LZ1437	BA16 [pWUR397A, pWUR400, pWUR478], *Amp*^{R}, *Str*^{R}, *Cm*^{R}	This work
Phage strains		
λ_{LZ760}	Fully methylated, gpD-mosaic, λ*D-mTurquoise2 cI*$_{857}$ *bor::Kan*^{R}	[229]
Plasmids		
pWUR397A	*cas3* in pRSF-1b	[257]
pWUR400	*casA-casB-casC-casD-casE* in pCDF-1b	[110]
pWUR477	Non-targeting CRISPR/ spacers from *E. coli* K12 with no homology to phage lambda in pACYCDuet-1	[110]
pWUR478	Template CRISPR/ template strand of λ genes *J*, *O*, *R*, and *E* in pACYCDuet-1	[110]

Bulk lysogenization assay

To measure the lysogenization frequency, we followed the protocol as described in [32]. Briefly, 2 ml of the host CRISPR/control strains LZ1437/LZ1436 was grown in LBMM (LB + 0.2% maltose + 10 mM MgSO$_4$) with appropriate antibiotics for overnight and subsequently diluted 1:100 into 12 ml of LBMM and grown at 37 °C. 0.2% L-arabinose and 0.1 mM IPTG were added into the culture to induce CRISPR plasmids for about 1.5 hours. When grown to OD$_{600}$ ~ 0.4, cells were centrifuged (1000×g for 10 min at 4 °C), concentrated 10× and resuspended to OD$_{600}$ ~ 4 in pre-chilled LBMM with 0.2% L-arabinose and 0.1 mM IPTG. 20 µl of the resuspended cells were then infected with 20 µl of phages at different concentrations by incubation for 30 min on ice. The samples were then transferred to 35 °C water bath for 5 min to allow for phage DNA ejection, followed by 100-fold dilution into pre-warmed LBGM (LB + 0.2% glucose + 10 mM $MgSO_4$) with 0.2% L-arabinose and 0.1 mM IPTG and incubation with shaking at 265 rpm at 30 °C for 45 min. The samples were then properly diluted with ice-cold 1× PBS and plated on LB + Kan plates to allow around 100 colonies to grow at 30 °C.

Fluorescence microscopy and imaging

An overnight culture of the host CRISPR/control strains LZ1437/LZ1436 was diluted 1:100 in M9 minimal medium (11.3 g/L M9 salts, 1 mM $MgSO_4$, 0.5 μg/mL thiamine HCl, 0.1% casamino acids, 100 μM $CaCl_2$) supplemented with 0.4% maltose (M9M) at 37 °C. 0.2% L-Arabinose and 0.1 mM IPTG were added into the culture to induce CRISPR plasmids for about 1.5 hours. Cells were grown to OD_{600} ~ 0.3, and then concentrated and resuspended in ice-cold M9M with 0.2% L-arabinose and 0.1 mM IPTG to OD_{600} ~ 3. λ_{LZ760} phages were added to reach a desired API (10 μl of cells + 10 μl of phages), followed by incubation on ice for 30 min and an additional 5 min incubation at 35 °C to trigger phage DNA ejection. The phage-cell mixture was diluted 1:10 in M9 with 0.2% L-arabinose and 0.1 mM IPTG, and 1 μl of the diluted phage-cell mixture was placed on a thin 1.5% agarose slab of M9M with 0.2% L-arabinose and 0.1 mM IPTG (~ 1 mm thick). After 1 min, a coverslip (No.1.5, Fisher Scientific) was gently laid over the mixture and the sample was imaged under the fluorescence microscope at 30°C with a cage incubator (InVivo Scientific, St. Louis, MO). Microscopy was performed on an inverted epifluorescence microscope (Ti-E, Nikon, Tokyo, Japan) using a 100× objective (Plan Fluo, NA 1.40, oil immersion) with a 2.5x TV relay lens and standard filter sets. Images were acquired using a cooled EMCCD camera (iXon3 897, Andor, Belfast, United Kingdom). Acquisition was performed using Nikon Elements software.

To localize all phages surrounding the cells, a series of 5 z-axis (vertical) images at a spacing of 400 nm were taken through the mTurquoise2/CFP channel using 200 ms exposure. To obtain more data in each time-lapse movie, cells were imaged at multiple stage positions (typically 16) in each experiment. During the time-lapse movie, the sample was imaged in phase contrast (100 ms exposure, for cell recognition), YFP (200 ms exposure, for phage DNA detection inside the cell, with a series of 5 z-axis images at a spacing of 400 nm in order to localize phage DNA foci), mTurquoise2/CFP (200 ms exposure, for phage lytic reporter) channels. The time-lapse movies were taken at a time interval of 5 min, until the cell fate was observable (2 hours).

Data analysis

All data analysis was performed in Matlab (The MathWorks, Natick, MA). Images were processed using MicrobeTracker [236]. Briefly, cells were first outlined by the phase-contrast channel using MicrobeTracker, after which spots were recognized first automatically using SpotFinderZ, then manually corrected using SpotFinderM [236]. Cell lineage tracking and fluorescent intensity were processed and calculated using our home-made Matlab script.

Results

Phage lysogenization is significantly reduced with the CRISPR system

For phage lambda, it was found that the *E. coli* CRISPR system protects against lambda lysogenization, where the lysogenization frequency is 100-fold lower in the presence of the CRISPR compared to without [257]. To test the efficiency of the artificial CRISPR system in our *E. coli* phage DNA reporter host cell BA16 (or MG1655 *seqA-yfp dam*$^-$) [256], we utilized the same CRISPR plasmids: pWUR397A (*cas3*) [257], pWUR400 (*casABCDE*) and pWUR478 (λ phage targeting spacers) / pWUR477 (control spacers) [110]. The CRISPR plasmid pWUR478 contains four anti-lambda spacers with homology to lambda genes *J*, *O*, *R*, and *E*. The control plasmid pWUR477 contains spacers without homology to lambda genome. To examine the function of this CRISPR system against the phage lysogenization, we performed experiments with the CRISPR or control plasmids at two different APIs (average phage input, i.e. average number of phages per cell) through a bulk lysogenization assay. Following the function of the CRISPR-Cas system, foreign phage DNA should be degraded and be unable to integrate into the host cell genome, resulting in reduced efficiency of lysogenization. As shown in Figure 4.1A, the lysogenization frequency of our host cells with the CRISPR or control plasmids is very low, within the order of 10^{-7} - 10^{-3} (the colony forming units, or cfu, and plaque forming units, or pfu, counts together with API and the calculated lysogenization frequencies are listed in Table 4.2). To determine the efficiency of this CRISPR-Cas system, we normalized the lysogenization frequency to that of the cells containing the control plasmid within two categories, i.e., low API (1 - 6) and high API (15 - 65). The lysogenization of the CRISPR targeting plasmid is only about 2% - 7% as efficient compared to the control (Figure 4.1B). This reduced lysogenization frequency indicates that the artificial CRISPR-Cas system works at the bulk level. However, the 2% - 7% efficiency in lysogenization is slightly higher than the 1% of previously reported *hns* mutant strain harboring the same

CRISPR plasmids at API of 1 [257]. This might be mainly because our host strain MG1655 does not harbor a gene encoding T7 RNA polymerase in its genome like the host strain BL21-AI used in the original system [110] to overexpress the Cas proteins and crRNAs from the plasmids, leading to the limited concentrations of the CRISPR components inside cells. Besides, the relative lower interference efficiency is also possibly due to our host strain MG1655 containing the *hns* product, heat-stable nucleoid structuring (H-NS) protein, a global transcriptional repressor in *E. coli*, or other host differences in our phage DNA reporter strain, for example, our cells with the *dam*$^-$ allele lack host DNA methylation, thereby abrogating the regulation of *E. coli* replication initiation by SeqA [167, 184, 237, 258].

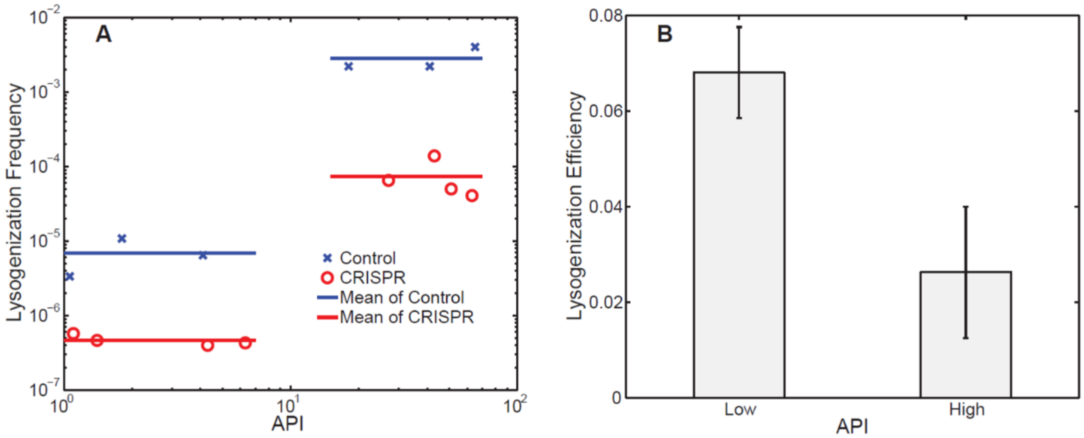

Figure 4.1 CRISPR system reduces lysogenization efficiency.
(A) The lysogenization frequencies of CRISPR versus control system for our phage DNA reporter strain at a range of APIs (low API of 1 - 6 and high API of 15 - 65). The control spacer: blue cross marker with blue lines as the mean for low and high APIs; the CRISPR spacer: red circle with red lines as the mean of the low and high APIs. (B) The lysogenization efficiency of the CRISPR system, defined as the percentage of average lysogenization frequency of the CRISPR system relative to that of the control system for low and high APIs. Error bars represent standard deviation (S.D.).

Table 4.2 The counts of colony and plaque forming units, and the calculated API and lysogenization frequency in the lysogenization experiments.

Original Phage Stock						
	Plaque counts with the dilution factor of 10^{-6}				**pfu (mean±SD)**	
Exp #1	233		241		2.46±0.15 × 10^{10}	
Exp #2	235		278			
Exp #3	250		240			
	Original Cell (Dilution factor/duplicate cell counts/cfu, mean±SD)		**Lysogen (Dilution factor/duplicate cell counts/cfu, mean±SD)**		**API**	**Lysogenization Frequency**
Control spacer strain						
Exp #1	10^{-7}		10^{-1}		1.06	3.36 × 10^{-6}
	62	54	186	203		
	5.80±0.40 × 10^{8}		1.95±0.85 × 10^{3}			
Exp #2	10^{-7}		10^{-2}		1.72	1.08 × 10^{-5}
	159	126	140	168		
	1.43±0.17 × 10^{9}		1.54±0.14 × 10^{4}			
Exp #3	10^{-7}		10^{-2}		3.97	6.42 × 10^{-6}
	58	66	32	48		
	6.20±0.4 × 10^{8}		4.00±0.08 × 10^{4}			
Exp #4	10^{-7}		10^{-4}		17.2	2.21 × 10^{-3}
	159	126	310	323		
	1.43±0.17 × 10^{9}		3.17±0.65 × 10^{6}			
Exp #5	10^{-7}		10^{-4}		39.7	2.19 × 10^{-3}
	58	66	132	140		
	6.20±0.4 × 10^{8}		1.36±0.04 × 10^{6}			
Exp #6	10^{-7}		10^{-4}		64.7	4.00 × 10^{-3}
	40	36	167	136		
	3.80±0.20 × 10^{8}		1.52±0.16 × 10^{6}			
CRISPR spacer strain						
Exp #1	10^{-7}		10^{0}		0.98	5.82 × 10^{-7}
	65	60	302	415		
	6.25±0.25 × 10^{8}		3.64±0.52 × 10^{2}			
Exp #2	10^{-7}		10^{0}		1.46	4.66× 10^{-7}
	30	52	193	188		
	4.1±1.1 × 10^{8}		1.91±0.25 × 10^{2}			
Exp #1	10^{-7}		10^{-1}		4.10	3.50 × 10^{-7}
	58	62	21	-		
	6.00±0.20 × 10^{8}		2.1 × 10^{2}			
Exp #2	10^{-7}		10^{-1}		6.07	4.32 × 10^{-7}
	44	37	18	17		
	4.05±0.35 × 10^{9}		1.75±0.50 × 10^{4}			
Exp #3	10^{-7}		10^{-3}		27.0	6.48 × 10^{-5}
	88	94	62	56		
	9.10±0.30 × 10^{8}		5.90±0.30 × 10^{4}			

Table 4.2 Continued

	Original Cell (Dilution factor/duplicate cell counts/cfu, mean±SD)		**Lysogen (Dilution factor/duplicate cell counts/cfu, mean±SD)**		**API**	**Lysogenization Frequency**
CRISPR spacer strain						
Exp #4	10^{-7}		10^{-3}		41.0	1.38×10^{-4}
	58	62	94	72		
	$6.00 \pm 0.20 \times 10^{8}$		$8.30 \pm 1.10 \times 10^{4}$			
Exp #5	10^{-7}		10^{-2}		49.2	5.00×10^{-5}
	53	47	256	243		
	$5.00 \pm 0.30 \times 10^{8}$		$2.50 \pm 0.65 \times 10^{4}$			
Exp #6	10^{-7}		10^{-2}		60.7	4.07×10^{-5}
	44	37	201	128		
	$4.05 \pm 0.35 \times 10^{8}$		$1.65 \pm 0.37 \times 10^{4}$			

Visualizing CRISPR function in single cells

To visualize the phage DNA degradation process, we employed a reporter system for visualizing infecting phage particles on the cell surface, and phage DNA inside the cell [38]. The infecting phage λ_{LZ760} is fluorescently labeled by the co-expression of gpD-mTurquoise2 and wild type gpD (λ capsid decoration protein) on the phage capsid, and the packaged phage DNA is prepared to be fully methylated. The host CRISPR/control strains (LZ1437/LZ1436) constitutively express a fluorescent SeqA fusion, SeqA-YFP, and the host DNA is not methylated owing to a *dam*$^{-}$ mutation (methylation deficient) [256]. SeqA specifically binds to fully methylated and hemi-methylated DNA, so the fully methylated DNA of phage λ_{LZ760} forms a bright YFP focus once ejected into the cell. This reporter system allows the tracking of each initial phage DNA and its first, hemi-methylated, replicated copy. However, the replicated DNAs are not always necessarily visualized as two separate fluorescent spots since the replicated DNAs can stick together and spatially overlap, or stay not far away from each other, thus appearing as one spot during imaging under the microscope.

To determine how efficiently CRISPR functions, we compared the cell fates after phage infection in both control movies (i.e., cells containing three plasmids: pWUR397A (*cas3*), pWUR400 (*casABCDE*) and pWUR477 (control spacers)) and CRISPR movies (i.e., cells containing three plasmids: pWUR397A (*cas3*), pWUR400 (*casABCDE*) and pWUR478 (λ phage targeting spacers)). A successful phage infection is indicated by the appearance of a YFP focus during imaging. In the control movies (592 cells in 7 movies), 398 cells (67%) start with one YFP focus staying as one focus throughout the movies (120 minutes) (classified as group 1), 186 cells (31%) splitting into two foci (group 2), and 8 cells (2%) splitting into 3 or more probably due to recombination (group 3). In the CRISPR movies (866 cells in 12 movies), there are 647 cells (75%) in group 1, 206 cells (24%) in group 2, and 13 cells (1%) in group 3. For the simplicity of spot tracking, we only analyzed cells with only one YFP focus throughout the movies (cells in group 1). Two typical cell fates are observed: lytic (86%, 341 out of 398 cells in the control movies; 66%, 425 out of 647 cells in the CRISPR movies) and CRISPR cells (26%, 167 out of 647 cells in the CRISPR movies). There are also rare cases of lysogeny or non-growing cells in both types of movies. The lytic cells show the accumulation of the fluorescent lytic reporter, gpD-mTurquoise2 protein inside the cell. The CRISPR cells are defined as those that have phage DNA with much shorter SeqA-YFP spot lifetime and do not lyse, but instead show normal cell growth. Figure 4.2 shows the snapshots of representative cells over time. At 0 min, phage DNA of a fluorescently labeled phage (appears as a cyan focus on the cell surface) has been ejected into the cell forming a YFP focus (appears as a yellow focus). The fluorescence intensity of the phage DNA decreases over time. The DNA foci then disappear at 110 min (the lytic cell in the control movie, Figure 4.2A), 120 min (the lytic cell in the CRISPR movie, Figure 4.2B) and 45 min (the CRISPR cell in the CRISPR movie, Figure 4.2C). The lytic cells eventually lyse, releasing fluorescent phage particles outside the cell while the CRISPR cell divides and grows normally like the uninfected cell. As our movies only last for 2 hours, some lytic cells do not lyse by the end of the movies, but the lytic reporter (gpD-mTurquoise2, appeared as cyan color) accumulates greatly indicating their lytic cell fate.

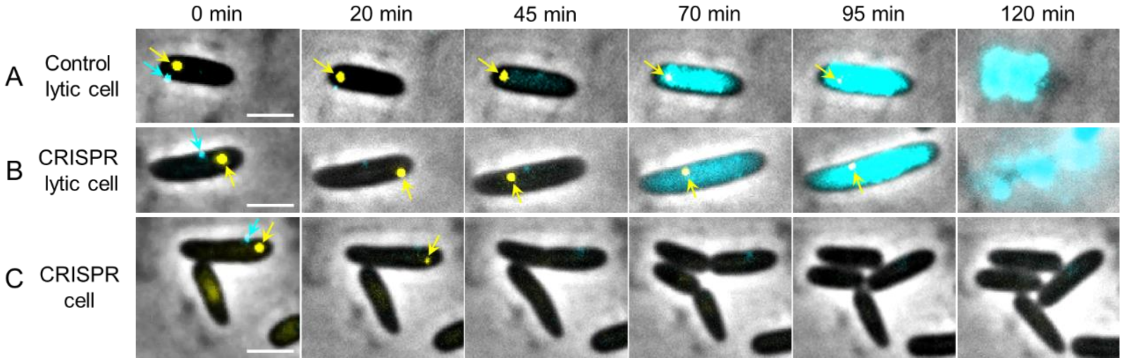

Figure 4.2 CRISPR system apparently degrades labeled phage DNA in single cells. The DNA of the fluorescently labeled infecting phage (cyan spot, pointed by cyan arrows) is ejected into the host *E. coli* cell forming a distinct fluorescent spot (yellow spot, pointed by yellow arrows) at 0 min. The intensity of phage DNA (yellow spot) decreases over time and finally disappears. (A) A representative lytic cell in a control movie. The yellow spot indicating the ejected phage DNA disappears at around 110 min likely due to the viral genome packaging which could strip off SeqA-YFP proteins from the phage DNA. At 120 min, the cell lyses. (B) A representative lytic cell in a CRISPR movie. The phage DNA disappears around 120 min or persists until cell lysis. At 120 min, the cell lyses. (C) A representative CRISPR cell (top) and an uninfected cell (bottom) in a CRISPR movie. The phage DNA in the CRISPR cell disappears at around 45 min, which is much earlier than that in the lytic cell in (A) and (B). At 70 min, the cell divides, similar to that of the uninfected cell.

Factors associated with effective CRISPR degradation

To determine and quantify the degradation rate of DNA by CRISPR in each cell, we plot the spot intensity over time. The spot intensity in the cell population mostly maximizes within 5 minutes from the beginning of the movie indicating that phage ejection is fairly synchronized among cells (Figure 4.3). The small variation in the timing of maximum intensity is likely due to the variability of DNA ejection time and the time delay from setting up our microscopy movies [38, 50]. In addition, the average maximum DNA focus intensities in lytic cells are similar to those in CRISPR cells in CRISPR movies, therefore all cells have similar levels of SeqA-YFP binding prior to CRISPR functioning independent of cell fate (Figure 4.4). To simplify the analysis and eliminate the variance of the maximum intensities among cells, we normalized spot intensities at each time point by the maximum intensity of each trace. We find that the spot intensity in CRISPR cells decreases much faster than that in lytic cells, as shown in Figure 4.5A, which indicates that CRISPR is functioning to degrade labeled phage DNA. The DNA foci become undetectable by 43.9 ± 1.5 minutes to provide an estimated degradation rate (Figure 4.5B). Lytic cells also show fluorescence signal loss, but this is probably due to photobleaching, which provides a control for the effect of photobleaching under our experimental conditions. The spots in lytic cells usually disappear much later than those in CRISPR cells or the disappearance is not observed over the entire course of the movies (120 minutes) (Figure 4.5C). The spot disappearance in lytic cells is likely due to the phage DNA packaging, which would strip SeqA-YFP proteins off the phage DNA. As shown in Figure 4.5D, the spot disappearance time is correlated with cell lysis time with a correlation coefficient of 0.91, p-value of 0.01. When we compare the lytic cells in CRISPR movies with those in control movies, their DNA fluorescence decreases similarly indicating that CRISPR spacers can fail in targeting the invading DNA, similarly to what happens with the control spacers (Figure 4.6). Under our experimental conditions, the efficiency of CRISPR in targeting and degrading the phage DNA is around 26%, 167 CRISPR-active cells out of 647 cells with one focus throughout the movies.

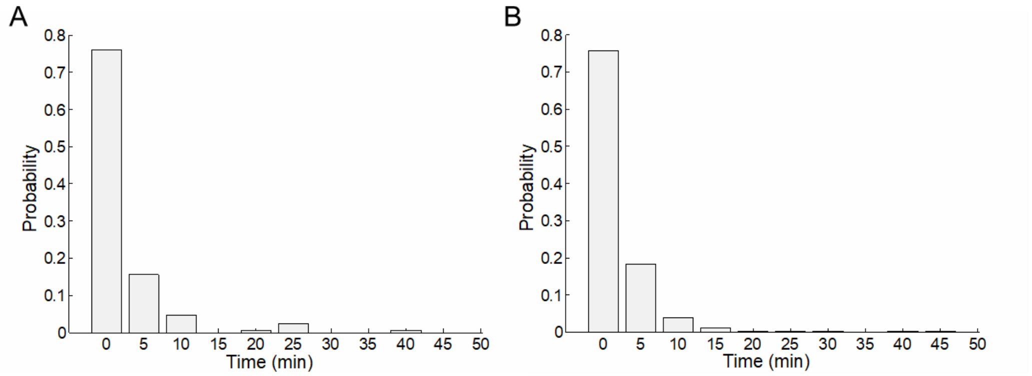

Figure 4.3 Histograms of the appearance time of SeqA-YFP foci corresponding to the phage DNA ejected into the cytoplasm with the maximum fluorescence intensity. The maximum spot intensity mostly appears within the first 5 minutes of the time-lapse movies indicating phage DNA ejection with the CRISPR cells of 91.6% (A) and lytic cells of 93.9% (B).

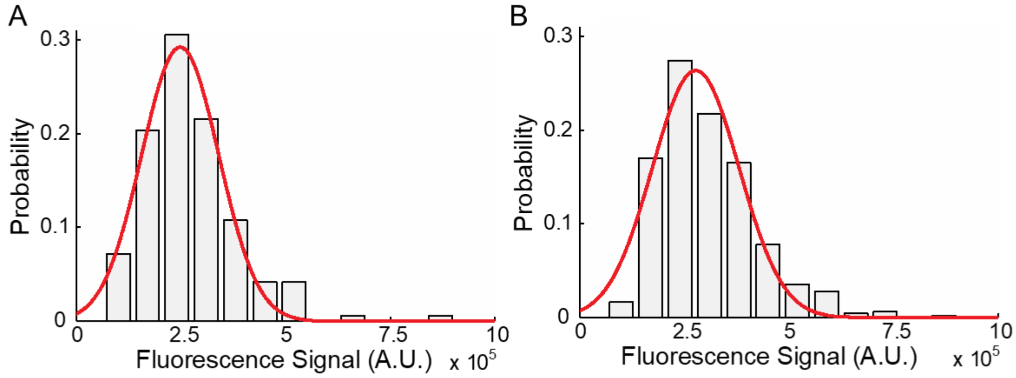

Figure 4.4 Histograms of initial spot intensity/maximum intensity of phage DNA inside CRISPR cells and lytic cells.

(A) CRISPR cells. (B) Lytic cells. There is little difference in the average maximum intensity of phage DNA between CRISPR cells ($(2.5 \pm 0.9) \times 10^5$) and lytic cells ($(2.8 \pm 1.0) \times 10^5$) in CRISPR movies. Experimental data (bars) were well fitted to a Gaussian function (red lines).

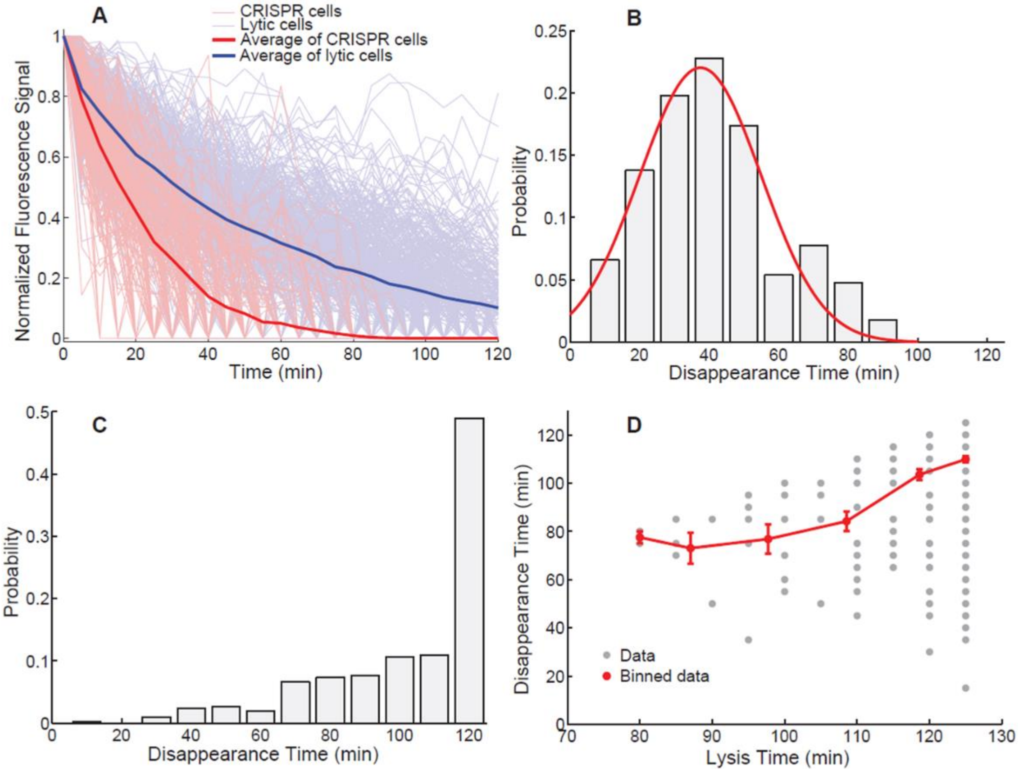

Figure 4.5 Phage DNA intensity decreases faster in CRISPR-active cells.
(A) The fluorescence intensity of phage DNA spots is plotted over time for CRISPR cells and lytic cells in the CRISPR movies. The phage DNA intensities of the CRISPR cells (red line, N = 167) decrease much faster than those of the lytic cells (blue line, N = 423), indicating CRISPR is actively functioning to degrade the invading phage DNA. The averages are shown as the thick lines. The signal of each phage DNA trace is defined as the spot fluorescence intensity at each time point normalized to the maximum intensity of each trace. (B) The histogram of phage DNA spot disappearance time corresponding to the degradation time for CRISPR cells is well fitted to a Gaussian distribution (red line). The time to totally degrade phage DNA is 43.9 ± 1.5 min on average. (C) The histogram of phage DNA spot disappearance time accounting for the photobleaching and/or phage DNA packaging into the phage head. Around 50% of the lytic cells still have the phage DNA spot at the end of the movies (120 min). (D) The phage DNA spot disappearance time is correlated with cell lysis time with a correlation coefficient of 0.91, p-value of 0.01. Error bars denote standard error of the mean (S.E.M.).

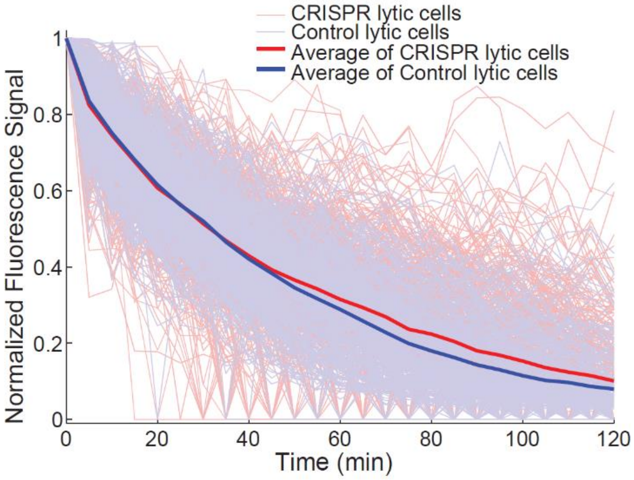

Figure 4.6 The fluorescence intensity of phage DNA spots in lytic cells over time. The phage DNA intensities decrease over time similarly for lytic cells in CRISPR movies (red line, N = 423) and control movies (blue line, N = 331). The decrease is probably just due to photobleaching, and the final disappearance may correspond to phage DNA packaging into the phage head. The averages are shown as the thick lines.

We next examined various factors associated with the time of CRISPR spot disappearance corresponding to the degradation rate of invading phage DNA. Considering the initial maximum spot intensity, due to time variation of taking movies, both between experiments and different stages within the same movie (up to a few minutes), the maximum intensity captured for each cell trace may correspond to a partially degraded phage DNA. If this were the case, the maximum intensity would be expected to have a positive correlation with spot disappearance time. Indeed, when we plot spot maximum intensity against disappearance time, we find the correlation coefficient to be 0.97 with p-value of 0.03 (Figure 4.7A). Next, we examined the effect of initial cell size, as cell size (or volume, or length as proxy) influences protein concentration, here the Cas proteins. Different concentrations of Cas proteins may degrade phage DNA at different rates. We measured the length of the infected cells and found that the CRISPR cells and the lytic cells shared a similar distribution of cell size (Figure 4.8). Then when we correlated the phage DNA spot disappearance time with the cell size, it seems that the phage DNA degradation rate does not change with cell size (correlation coefficient of -0.55, with a large p-value of 0.45) as shown in Figure 4.7B. This indicates that the expressed Cas proteins are sufficiently concentrated to degrade phage DNA. Finally, we examined how phage DNA location inside the cell affects CRISPR action. Previous work has showed that phage lambda tends to adsorb to the cell pole and eject its DNA [35-37]. Although the exact mechanism underlying this polar preference has not been uncovered, cell pole is a special location where many proteins associated with the infection cycle, like ManYZ and FtsH accumulate [36]. We then ask whether the degradation of phage DNA inside the cell has any correlation with the initial phage DNA location. After the phage is ejected inside the cell, our recent work has shown that some phage DNAs move locally while others move over the whole cell [38]. When we correlate the phage DNA location with the efficiency of CRISPR degradation, we find that there appears to be no difference in CRISPR efficiency, whether the initial phage location is at polar/mid-cell (future pole) or non-polar positions (27 ± 4% at pole/mid-cell versus 25 ± 2% at non-pole, $p = 0.02$) (Figure 4.7C). Additionally, the timing of phage DNA degradation (spot disappearance time) is unaffected by phage DNA location (42.1 ± 2.4 min at pole/mid-cell and 44.7 ± 1.8 min at non-pole, $p = 0.02$) (Figure 4.7D). This indicates that the Cas proteins are uniformly distributed in the cytoplasm and can effectively target invading DNA regardless of where the DNA is located.

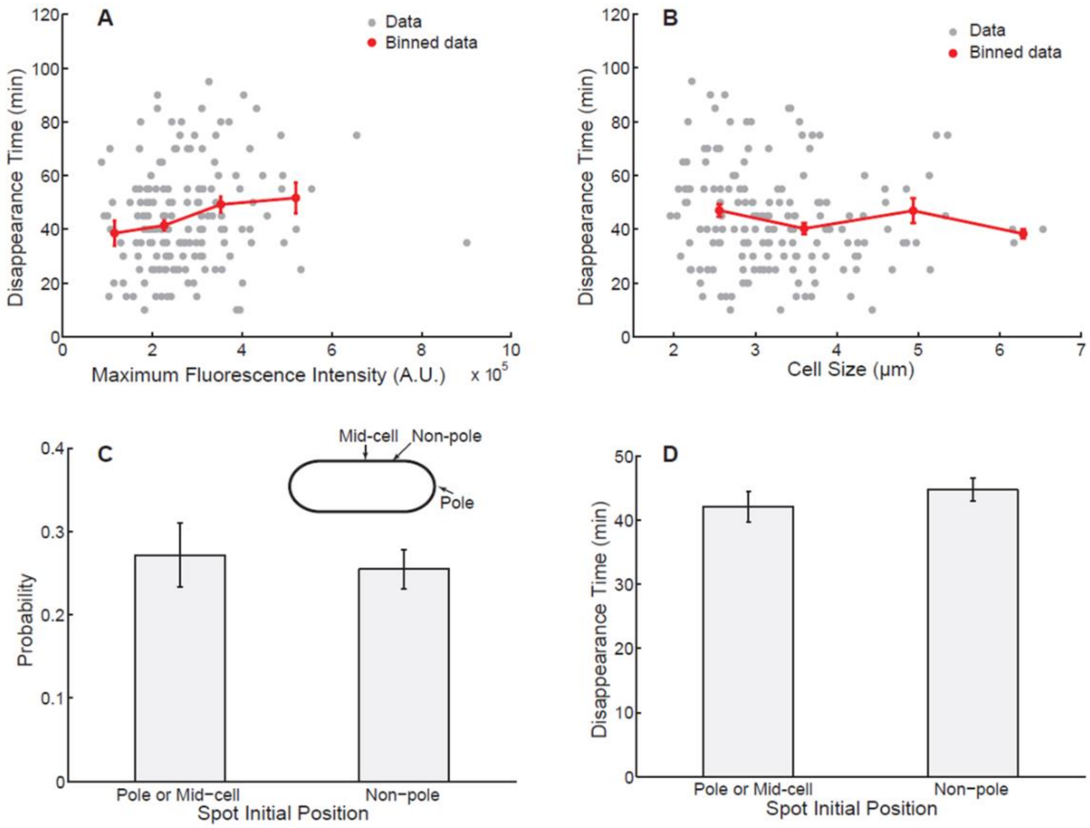

Figure 4.7 Phage DNA degradation correlates with spot intensity, not with cell size or initial DNA location.

(A) The complete phage DNA degradation or spot disappearance time positively correlates with the maximum intensity of the spot at the beginning of the movie with a correlation coefficient of 0.97, p-value of 0.03. The binned data were obtained with a bin interval of 1.5×10^5 A.U. (B) The spot disappearance time does not change with the initial cell size with a correlation coefficient of -0.55, p-value of 0.45. The binned data were obtained with a bin interval of 1 μm. (C) The efficiency of CRISPR is very similar for the initial invading phage DNA at polar/mid-cell (0.27 ± 0.04) or non-polar (0.25 ± 0.02) positions with a p-value of 0.02. The diagram of the cell is shown on the top right. (D) The spot disappearance time does not seem to correlate with the initial phage DNA location showing similar disappearance time for polar/mid-cell (42.1 ± 2.4 min) or non-polar (44.7 ± 1.8 min) cell location with a p-value of 0.02. Error bar represents S.E.M.

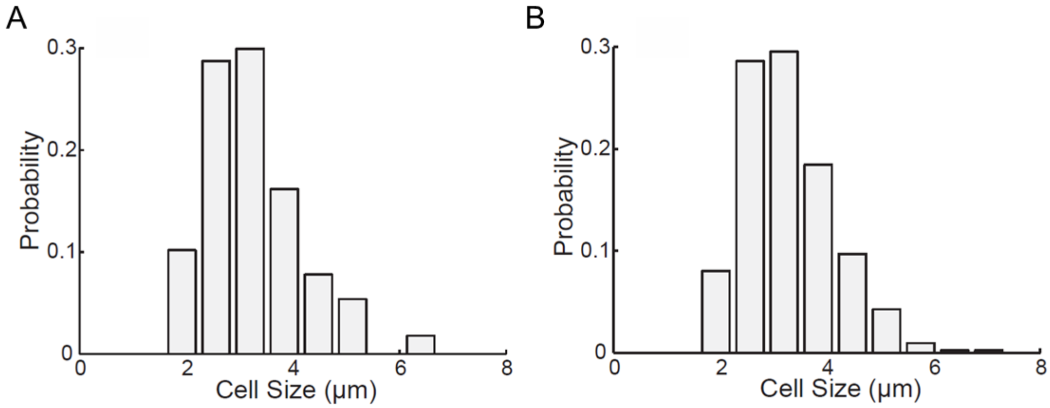

Figure 4.8 The distribution of cell size for CRISPR cells and lytic cells.
The cell length, which is served as a proxy for cell size, of CRISPR cells (A) and lytic (B) cells shows similar distributions. The average cell lengths are 3.27 ± 0.07 μm and 3.25 ± 0.07 μm for the CRISPR cells and the lytic cells, respectively.

Phage escapes bacterial CRISPR defense through rapid DNA replication

Interestingly and surprisingly, when we compared the phage DNA disappearance time of the lytic cells in the CRISPR movies with those in the control movies, we observed that a considerable population (~40%) of the lytic cells with CRISPR targeting plasmids lost injected phage DNA foci during the early infection period (Figure 4.9). Given the fact that our phage DNA labeling approach only allows for visualization of the initial phage DNA and its first replicated copy and constricts us from accessing the whole population of phage DNA copies as they replicate over time, the observation implies that the CRISPR system might only degrade the ejected phage DNA but fail to destroy the subsequent replicated copies. Accordingly, we came up with a hypothesis that there might be a competition between the CRISPR function and the phage DNA replication. When a single phage DNA is ejected into host, if phage DNA replication happens before CRISPR system executes its function, or phage DNA replication rate is faster than the rate of CRISPR degrading foreign DNA, CRISPR system could only destroy a few DNA molecules but allow escape of other copies, resulting in failed bacterial defense and cell lysis. On the contrary, if the DNA clearance occurs as soon as an invader enters the host as a single copy, thus depriving the foreign DNA of the opportunity to replicate and take over the cell, the host cell can survive. Notably, the CRISPR function rate presumably should be correlated to the cellular concentration of the surveillance complex and the Cas3 effector nuclease. Therefore, further characterization of the competition between bacterial CRISPR activity and phage DNA replication in terms of cell-to-cell variability in the CRISPR expression level is required.

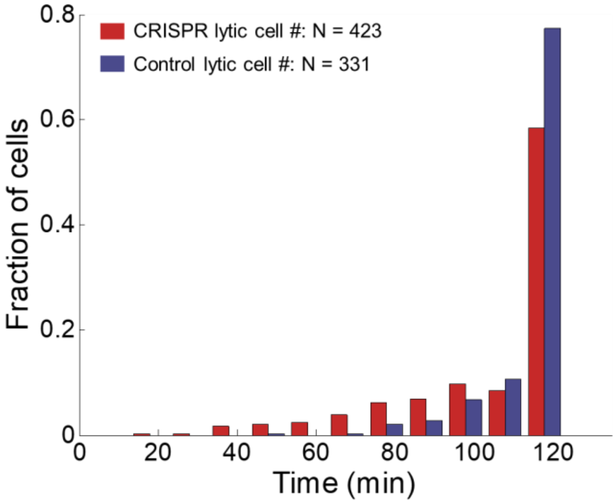

Figure 4.9 The distribution of the disappearance time of phage DNA spots in lytic cells in CRISPR movies and control movies.

There are ~60% and ~80% of the lytic cells where phage DNA spots disappear at or after the end of the time-lapse movies (120 min) in CRISPR movies (red bar) and control movies (blue bar), respectively. Phage DNA signals in some lytic cells in CRISPR movies become undetectable as early as 10-50 min, comparable to the average DNA degradation time (44 min) in the cells where CRISPR system functions successfully.

Discussion

In this study, we used our recently developed fluorescence technique to visualize phage DNA inside the cell upon infection by phage lambda to explore single-cell CRISPR function. With an artificial type I-E CRISPR system expressed in the host, the CRISPR system is able to degrade the invading phage DNA effectively with an efficiency of around 26% at the single-cell level. Consistent with the literature, the CRISPR system decreases the lysogenization frequency of the phage to be around 2-7% at the bulk level, which shows the effectiveness of CRISPR system attacking the invading phage DNA during the early infection stage to prevent the establishment of the lysogenic cycle. Remarkably, both 26% and 2-7% of the efficiency are much lower than the interference efficiency reported for the same artificial CRISPR-Cas system, which conferred $>10^7$-fold resistance against λ infection in plaque assays [110]. We attributed these lower efficiencies to the insufficient expression of CRISPR components in the host cells. In the previously described system, the *casABCDE*, *cas3*, and CRISPR arrays, which are under the control of a T7 promoter, can be overexpressed when the inducer IPTG is added and when the T7 RNA polymerase gene engineered on the *E. coli* chromosome is fully induced by the addition of arabinose. However, the *E. coli* host we used in this study does not carry a T7 RNA polymerase gene in its genome. Therefore, the expression level of the CRISPR system under our experimental condition completely relied on the leakage of the system, resulting in low concentrations of Cas proteins and the transcribed crRNAs in the host, which would attenuate the effectiveness of the CRISPR defense.

To our surprise, the fluorescent DNA foci in CRISPR cells do not divide into several smaller foci upon degradation, since the targeting plasmid pWUR478 contains four anti-lambda spacers with homology to lambda genes *J*, *O*, *R*, and *E*, which are fairly far apart from each other. If each spacer is targeted, the expected result would be that the phage genome is broken into multiple pieces, resulting in a few small fluorescent foci. However, our time-lapse movies show that the initial fluorescent DNA focus always stays as one before it finally disappears. Therefore, it is possible that only one anti-lambda spacer can target the phage DNA effectively one time with a synergistic effect of individual spacers and the Cas proteins degrade the DNA from that site thereafter, which may be due to the different sensitivities of the phage lambda DNA to the anti-lambda spacers [110]. But we cannot rule out the possibility that our SeqA-YFP system holds the cut DNA together somehow such that there is only one visible spot.

Under our experimental conditions under the fluorescence microscope, the lambda DNA takes about 44 minutes to be degraded. As a comparison, in *in vitro* experiments, a 5 kb negatively supercoiled DNA and linear DNA can be degraded within a few minutes (~ 1 min and 5 min) [202]. In addition to the insufficient expression of the CRISPR system as mentioned above, the much longer lambda DNA, 48.5 kb compared to this 5 kb DNA, almost 10-fold longer, might be the other cause why it takes almost 10-fold or more time to be digested. Moreover, another contributor of different degrading time is the different experimental conditions of the *in vitro* experiments versus our *in vivo* experiments under the fluorescence microscope. The average fluorescence decrease in the lytic cells in the control and CRISPR movies (Figure 4.6) is very similar, probably representing the photobleaching effect. The cell-to-cell variability of all the individual cells is probably due to the fluorescently labeled phage DNA moving in and out of focus, stochastic photobleaching of fluorescent molecules on the attached phage DNA and some unknown cell-to-cell variability. In addition to the stochastic photobleaching effect, the cell-to-cell variability of the CRISPR cells might also come from how easily the invading lambda DNA, probably the supercoiled form, is exposed for the Cas proteins to dock and degrade.

From our single-cell analysis, the time to fully degrade the invading phage DNA is dependent on the initial fluorescent intensity of the phage DNA at the beginning of the movie. The lower maximum intensity of a spot in the cell might correspond to a partially degraded phage DNA due to the time to set up a movie, which is consistent with a fast initiation of CRISPR action. However, the phage DNA degradation does not depend on the initial cell size or phage DNA location. It is reasonable that Cas proteins are well mixed inside the cell, so they can degrade the invading phage DNA with no preference of the location.

Last but not least, we observed that some cells eventually entered the lytic cycle even though the infecting phage DNAs apparently were degraded by the CRISPR system. It is well known that phages develop several strategies to evade the degradation of their DNA by bacterial CRISPR systems [95]. Since our phage λ strain does not carry mutations in the genome or encode any anti-CRISPR proteins to inactivate bacterial CRISPR function, we speculate that phages have the capability of fighting against bacterial CRISPR system through their rapid DNA replication outpacing CRISPR action inside cells. To examine whether phage DNA replication is a hidden antagonist against CRISPR interference within the time scale of phage infection at the single-cell level, the labeling and visualization of all phage genomes need to be achieved. Moreover, it is important to emphasize that under our experimental condition, the CRISPR components remained low concentrations, which might allow some phage DNA to escape from the degradation. Previous work has demonstrated that endogenous levels of *E. coli* CRISPR-Cas expression are not sufficient to provide resistance for phage infection, due to the presence of negative regulators [184]. Most likely, only at low CRISPR expression levels, can phage DNA replication be able to compete with CRISPR function. These findings open many questions, such as in the natural environments, how CRISPR-Cas systems are regulated in response to viral infection, what factors trigger the CRISPR activity to act as a defense system, and whether viruses can evolve to accelerate their genome replication rate in order to outpace CRISPR-Cas immunity. Addressing these questions will greatly expand our understanding of the evolutionary arms race between bacteria and their invaders.

CHAPTER V

SUMMARY AND CONCLUSIONS

Bacteriophage lambda (λ) has long served as an important paradigm for studying a variety of biological processes. With the advance of technology, the research on λ has moved onto single-cell studies, and further zoomed into the subcellular level through precisely labeling individual phages and individual bacterial and viral biomolecules. In this work, by taking advantage of the single-virus/single-molecule labeling technology, we developed various fluorescence reporter systems to label *E. coli* and λ genomes, as well as phage gene transcripts, in both live cells and chemically fixed cells. These reporter systems open up the possibility to characterize the spatiotemporal pattern of the expression of certain viral genes, to investigate the interaction between host genome and infecting phage DNA, to explore the effect of phage morphology on phage infection and final decision-making, and to visualize how bacterial CRISPR-Cas system breaks down invasive viral nucleic acid in real time at the single-cell level. The discoveries gained from this work provide direct evidence for previous hypotheses relating to λ cellular decision-making and reveal interesting features of the λ infection process as well as CRISPR-Cas immune systems.

Phage λ gene expression pattern and spatial organization in single cells

Phage λ lysis-lysogeny decision-making is complicated in terms of both the intrinsic genetic regulatory circuitry and the dependence on extrinsic infection parameters. Cell-to-cell variability in decision-making outcomes has been attributed to stochastic gene expression in the λ genetic regulatory network. In this work, we launched an investigation into the expression profile of phage-derived fate-determining regulators during the decision-making process at the single-cell level. We started with the λ *cII* gene, whose product is considered as the central player in establishing lysis-lysogeny decision making. Utilizing RNA-smFISH technique, we quantified the transcriptional change of λ *cII* over time upon phage infection at the single-cell level. We revealed that *cII* transcription is initiated soon after a phage genome is injected into the cytoplasm. Over time, *cII* expression increases and peaks around 10 min and subsequently goes down. This work lays the foundation for future quantitative measurements of other fate-determining regulators at both the transcriptional and translational levels. Further characterization of the λ gene expression profile and how the interplay between different regulators affects lytic-lysogenic switches would complement the quantitative description of bacteriophage lysis-lysogeny decision making.

Next, to fulfill the curiosity of how replicated phage genomes behave and localize in the subcellular environment during the development of phage post-infection, we exploited two distinct fluorescence reporter systems, *tetO*/TetR-FP and *lacO*/LacI-FP, to label individual phage genomes. In combination with the previously established lytic reporters, we followed phage intracellular development and found that phage DNA formed clusters as it replicated inside cells. When multiple phages co-infected the same cell, their DNAs are capable of maintaining individuality by assembling their own compartments for DNA replication and gene expression in different spaces in single cells. This phenomenon holds true for singe-phage infection in the cases where phage DNA exhibits active motions in the cytoplasm once being ejected. In conjunction with the labeling of *E. coli* chromosome (*attB* locus) by a *lacO*/LacI-FP reporter scheme, it turns out that *E. coli* DNA participates in separating individual phage-derived compartments inside cells. In parallel, similar phenotypes were observed in fixed infected cells using FISH when bacterial and phage genomes were labeled with fluorescent probes, supplementing the finding in the live-cell imaging. Remarkably, the amounts of phage DNA and mRNA in those phage-derived compartments differ from each other, most likely due to the competition between phage DNAs for host resources as well as some other factors, which may help interpret the origin of stochasticity in viral gene expression. All the evidence support our previous hypothesis that each infecting phage or phage DNA has the ability to vote for the decision of whether to lyse or lysogenize a cell. That is possibly because individual phage DNA can organize its own development by creating a compartment in certain area of a cell. Different DNA replication rate and gene expression levels arise from those separate compartments will result in distinct cell-fate choices within a single cell and overall contribute to the final decision. These single-cell observations uncover the subcellular structure formation and both phage-phage, phage-host intracellular interaction behaviors during the λ post-infection development. The findings provide hints into the unexplored importance of the spatial structure in affecting viral gene expression and the ultimate decision-making outcome. Future investigation of the mechanism of those interaction behaviors, particularly in terms of spatial distribution patterns of cell-fate-determining regulators and their correlations with phage-derived subcellular compartments, will expand the mechanistic understanding of cellular decision-making.

In recent years, phages started to be used as antibacterial agents to treat bacterial infections in plants, domestic animals, humans, and biocontrol in food industries. For clinical applications, temperate phages, like phage λ, are less attractive, because the establishment of lysogeny can cause the lysogenic cell to become immune to superinfection by the same phage strain and some prophages encode genes that can alter the phenotype of the host cell [259]. However, due to the highly abundance of temperate phages in nature and a wide range of host bacteria they can infect, researchers have begun to explore the potential of temperate phages for therapeutic purposes [260]. Therefore, a comprehensive and quantitative understanding of the post-infection development of phage λ, which has been serving as a basic paradigm, at the subcellular level will be of great importance to provide mechanistic basis for phage therapy.

The side tail fibers of λ give rise to more frequent failure in phage infection

Bacteriophage λ has served as an important model for molecular biology and different cellular processes over the past few decades. However, it was not until 1992 that the λ strain used in most laboratories around the world, thought of as λ wild type, was discovered to carry a frameshift mutation in the *stf* gene, which encodes four side tail fibers that were actually present in the originally isolated λ strain in 1951. The original strain was later on known as Ur-λ. In order to characterize the effect of the side tail fibers on phage infection, we utilized a dual-color fluorescent reporter system to label individual phage particles and report the two alternative life cycles, lysis and lysogenization, upon infection at the single-cell level. As expected, Ur-λ exhibits a stronger attachment onto the cell with the help of those side tail fibers via providing additional binding to the cell membrane. However, to our surprise, the side tail fibers interfere with the phage DNA ejection process, most likely through the binding to their receptors, OmpC, on the host cell outer membrane, leading to a higher failed infection frequency compared with λWT. Moreover, when the development of post-infection was followed, the side tail fibers do not alter the lysis time or affect the lysis-lysogeny decision-making outcome. Consequently, it seems that the presence of the side tail fibers does not give Ur-λ more advantages in infecting host cells at the single-cell level. Nevertheless, they are apparently of great significance in nature. The side tail fibers participate in determining the host range specificity as one of roles of the λ central tail fiber. It is also well known that the side tail fibers significantly accelerate the adsorption rate of phage λ onto bacterial host. This higher adsorption rate can shorten the optimal lysis time and may compensate for the less successful infection frequency for individual phage and ultimately cause a larger bacterial population rapidly being infected in either the natural environment or clinical applications of phage therapy. Taken together, having the side tail fibers would still endows Ur-λ with a higher fitness from the evolutionary point of view.

Visualization of viral DNA degradation by a CRISPR-Cas system *in vivo*

Although CRISPR-Cas systems, which are a widespread prokaryotic defense system against invasive viruses and plasmids, have been extensively studied over the past few decades, a direct visualization of CRISPR-Cas degrading the invading DNA in real time had not yet been achieved at the single-cell/single-virus/single-molecule level. To this end, we took advantage of our well-established fluorescence reporter system to label individual phage lambda particles and their packaged genomes, allowing for the tracking of a phage DNA when it enters the *E. coli* reporter host with a plasmid-based CRISPR-Cas type I-E system via fluorescence microscopy. We found that at the bulk level, the lysogenization frequency of the cells harboring active CRISPR plasmids targeting λ genome is significantly lower compared to the cells with non-targeting CRISPR plasmids (served as the negative control). At the single-cell level, infecting phages failed to establish either the lytic or lysogenic life cycles in the host cells with active CRISPR function, represented by the disappearance of phage DNA foci in the cytoplasm. Those cells exhibited normal growth similar to uninfected cells, where the efficiency of the successful interference is around 26% under the experimental condition that the cellular concentration of CRISPR components maintained at the basal level. During the course of the time-lapse movies, the average fluorescence intensity of invasive phage DNA foci in cells with successful CRISPR interference decays more rapidly compared to that in the control cells. Phage DNA is degraded by around 44 minutes on average, indicated by the disappearance of phage DNA spots. Moreover, when several parameters accounting for the effect of cell-to-cell variability on the time of phage DNA spot disappearance were characterized, we found that the degradation of phage DNA appears to be independent of cell size or the phage DNA ejection site on the cell surface.

Most interestingly, we observed that in some lytic cells, phage DNA disappeared very early with the disappearance time comparable to that of CRISPR functioning cells. It suggests that the CRISPR system in those cells has successfully degraded a subset of phage DNA molecules but unfortunately allows some phage DNAs to escape and develop, possibly due to insufficient recognition by the Cascade complex or depletion of Cas proteins. Accordingly, we hypothesize that the overlooked viral DNA replication might be a potential anti-CRISPR mechanism in the everlasting battle between bacteria and viruses, which is currently our subject of exploration. Moreover, if the viral DNA replication could compete with CRISPR activities by rapidly increasing the cellular phage DNA concentration, we anticipate that phages might be able to take over the host through multiple infections (MOI > 1). Further investigations of how co-infecting phages cooperate in competing with bacterial CRISPR-Cas activities in space and time at the single-cell/single-phage level remains to be done. The gained findings will advance our mechanistic understanding of the continuous arms race between bacteria and their predators under natural conditions. Additionally, a thorough description of the molecular basis of CRISPR-Cas systems in living organisms at the single-cell level will benefit diverse applications of CRISPR-based technologies in genome editing and clinical gene therapy, since genetic manipulation by CRISPR-Cas doubtlessly requires single-cell accuracy.

REFERENCES

1. Summers, W.C. (2006). Phage and the early development of molecular biology. In The Bacteriophages, R. Calendar, ed.

2. Summers, W.C. (2004). Bacteriophage Research: Early History. In Bacteriophages: Biology and Applications, A.S. Elizabeth Kutter, ed.

3. Marks, T., and Sharp, R. (2000). Bacteriophages and biotechnology: a review. J Chem Technol Biot *75*, 6-17.

4. Calendar, R. (2006). The bacteriophages, 2nd Edition, (Oxford ; New York: Oxford University Press).

5. Sharp, R. (2001). Bacteriophages: biology and history. J Chem Technol Biot, 667-672.

6. Carl R. Merril, D.S., Sankar Adhya (2006). Phage Therapy. In The Bacteriophages, R. Calendar, ed.

7. Pires, D.P., Cleto, S., Sillankorva, S., Azeredo, J., and Lu, T.K. (2016). Genetically Engineered Phages: a Review of Advances over the Last Decade. Microbiol Mol Biol Rev *80*, 523-543.

8. Labrie, S.J., Samson, J.E., and Moineau, S. (2010). Bacteriophage resistance mechanisms. Nat Rev Microbiol *8*, 317-327.

9. Grose, J.H., and Casjens, S.R. (2014). Understanding the enormous diversity of bacteriophages: the tailed phages that infect the bacterial family Enterobacteriaceae. Virology *468-470*, 421-443.

10. Lederberg, E.M. (1951). Lysogenicity in E-Coli K-12. Genetics *36*, 560-560.

11. Kaiser, A.D. (1966). On the internal structure of bacteriophage lambda. J Gen Physiol *49*, 171-178.

12. Roger W. Hendrix, S.C. (2006). Bacteriophage λ and its genetic neighborhood. In The Bacteriophages, 2nd Edition Edition, R.L. Calendar, ed. (Oxford, UK: Oxford University Press), pp. 409-447.

13. Bayer, M.E., and Bocharov, A.F. (1973). The capsid structure of bacteriophage lambda. Virology *54*, 465-475.

14. Hendrix, R.W. (1983). Lambda II, (Cold Spring Harbor, N.Y.: Cold Spring Harbor Laboratory).

15. Hendrix, R.W., and Duda, R.L. (1992). Bacteriophage lambda PaPa: not the mother of all lambda phages. Science *258*, 1145-1148.

16. Dove, W.F. (1969). Strains of phage lambda in current use. Virology *38*, 349-351.

17. Ptashne, M. (1986). A genetic switch : gene control and phage [lambda], (Cambridge, Mass. Palo Alto, Calif.: Cell Press; Blackwell Scientific Publications).

18. Oppenheim, A.B., Kobiler, O., Stavans, J., Court, D.L., and Adhya, S. (2005). Switches in bacteriophage lambda development. Annu Rev Genet *39*, 409-429.

19. Casjens, S.R., and Hendrix, R.W. (2015). Bacteriophage lambda: Early pioneer and still relevant. Virology *479-480*, 310-330.

20. Golding, I. (2016). Single-Cell Studies of Phage lambda: Hidden Treasures Under Occam's Rug. Annu Rev Virol *3*, 453-472.

21. Court, D.L., Oppenheim, A.B., and Adhya, S.L. (2007). A new look at bacteriophage lambda genetic networks. J Bacteriol *189*, 298-304.

22. Taylor, K., and Wegrzyn, G. (1995). Replication of coliphage lambda DNA. FEMS Microbiol Rev *17*, 109-119.

23. Rokney, A., Kobiler, O., Amir, A., Court, D.L., Stavans, J., Adhya, S., and Oppenheim, A.B. (2008). Host responses influence on the induction of lambda prophage. Mol Microbiol *68*, 29-36.

24. Little, J.W., and Michalowski, C.B. (2010). Stability and instability in the lysogenic state of phage lambda. J Bacteriol *192*, 6064-6076.

25. Marsh, M., and Helenius, A. (2006). Virus entry: open sesame. Cell *124*, 729-740.

26. Randall-Hazelbauer, L., and Schwartz, M. (1973). Isolation of the bacteriophage lambda receptor from Escherichia coli. J Bacteriol *116*, 1436-1446.

27. Schirmer, T., Keller, T.A., Wang, Y.F., and Rosenbusch, J.P. (1995). Structural basis for sugar translocation through maltoporin channels at 3.1 A resolution. Science *267*, 512-514.

28. Wang, J., Hofnung, M., and Charbit, A. (2000). The C-terminal portion of the tail fiber protein of bacteriophage lambda is responsible for binding to LamB, its receptor at the surface of Escherichia coli K-12. J Bacteriol *182*, 508-512.

29. Fuerst, C.R., and Bingham, H. (1978). Genetic and physiological characterization of the J gene of bacteriophage lambda. Virology *87*, 437-458.

30. Storms, Z.J., and Sauvageau, D. (2015). Modeling tailed bacteriophage adsorption: Insight into mechanisms. Virology *485*, 355-362.

31. Chatterjee, S., and Rothenberg, E. (2012). Interaction of bacteriophage l with its E. coli receptor, LamB. Viruses *4*, 3162-3178.

32. Moldovan, R., Chapman-McQuiston, E., and Wu, X.L. (2007). On kinetics of phage adsorption. Biophys J *93*, 303-315.

33. Adam, G., and M. Delbruck (1968). Reduction of dimensionality in biological diffusion processes. . In Structural Chemistry and Molecular Biology., A.R.a.N. Davidson, ed. (San Francisco, NC, USA: W.H. Freeman & Company).

34. Gibbs, K.A., Isaac, D.D., Xu, J., Hendrix, R.W., Silhavy, T.J., and Theriot, J.A. (2004). Complex spatial distribution and dynamics of an abundant Escherichia coli outer membrane protein, LamB. Mol Microbiol *53*, 1771-1783.

35. Rothenberg, E., Sepulveda, L.A., Skinner, S.O., Zeng, L., Selvin, P.R., and Golding, I. (2011). Single-virus tracking reveals a spatial receptor-dependent search mechanism. Biophys J *100*, 2875-2882.

36. Edgar, R., Rokney, A., Feeney, M., Semsey, S., Kessel, M., Goldberg, M.B., Adhya, S., and Oppenheim, A.B. (2008). Bacteriophage infection is targeted to cellular poles. Mol Microbiol *68*, 1107-1116.

37. Zeng, L., Skinner, S.O., Zong, C., Sippy, J., Feiss, M., and Golding, I. (2010). Decision making at a subcellular level determines the outcome of bacteriophage infection. Cell *141*, 682-691.

38. Shao, Q., Hawkins, A., and Zeng, L. (2015). Phage DNA dynamics in cells with different fates. Biophys J *108*, 2048-2060.

39. Cortes, M.G., Trinh, J.T., Zeng, L., and Balazsi, G. (2017). Late-Arriving Signals Contribute Less to Cell-Fate Decisions. Biophys J *113*, 2110-2120.

40. Schwartz, M. (1976). 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 *103*, 521-536.

41. Gallet, R., Kannoly, S., and Wang, I.N. (2011). Effects of bacteriophage traits on plaque formation. BMC Microbiol *11*, 181.

42. Shao, Y., and Wang, I.N. (2008). Bacteriophage adsorption rate and optimal lysis time. Genetics *180*, 471-482.

43. Wang, I.N. (2006). Lysis timing and bacteriophage fitness. Genetics *172*, 17-26.

44. Gallet, R., Shao, Y., and Wang, I.N. (2009). High adsorption rate is detrimental to bacteriophage fitness in a biofilm-like environment. BMC Evol Biol *9*, 241.

45. Montag, D., Schwarz, H., and Henning, U. (1989). A component of the side tail fiber of Escherichia coli bacteriophage lambda can functionally replace the receptor-recognizing part of a long tail fiber protein of the unrelated bacteriophage T4. J Bacteriol *171*, 4378-4384.

46. Scandella, D., and Arber, W. (1974). An Escherichia coli mutant which inhibits the injection of phage lambda DNA. Virology *58*, 504-513.

47. Esquinas-Rychen, M., and Erni, B. (2001). Facilitation of bacteriophage lambda DNA injection by inner membrane proteins of the bacterial phosphoenol-pyruvate: carbohydrate phosphotransferase system (PTS). J Mol Microbiol Biotechnol *3*, 361-370.

48. Novick, S.L., and Baldeschwieler, J.D. (1988). Fluorescence measurement of the kinetics of DNA injection by bacteriophage lambda into liposomes. Biochemistry *27*, 7919-7924.

49. Grayson, P., Han, L., Winther, T., and Phillips, R. (2007). Real-time observations of single bacteriophage lambda DNA ejections in vitro. Proc Natl Acad Sci U S A *104*, 14652-14657.

50. Van Valen, D., Wu, D., Chen, Y.J., Tuson, H., Wiggins, P., and Phillips, R. (2012). A single-molecule Hershey-Chase experiment. Curr Biol *22*, 1339-1343.

51. Mackay, D.J., and Bode, V.C. (1976). Events in lambda injection between phage adsorption and DNA entry. Virology *72*, 154-166.

52. Evilevitch, A. (2018). The mobility of packaged phage genome controls ejection dynamics. Elife *7*.

53. Lof, D., Schillen, K., Jonsson, B., and Evilevitch, A. (2007). Forces controlling the rate of DNA ejection from phage lambda. J Mol Biol *368*, 55-65.

54. Zylicz, M., Gorska, I., Taylor, K., and Georgopoulos, C. (1984). Bacteriophage lambda replication proteins: formation of a mixed oligomer and binding to the origin of lambda DNA. Mol Gen Genet *196*, 401-406.

55. Hayes, S., Erker, C., Horbay, M.A., Marciniuk, K., Wang, W., and Hayes, C. (2013). Phage Lambda P protein: trans-activation, inhibition phenotypes and their suppression. Viruses *5*, 619-653.

56. Wegrzyn, G., and Wegrzyn, A. (2005). Genetic switches during bacteriophage lambda development. Prog Nucleic Acid Res Mol Biol *79*, 1-48.

57. Better, M., and Freifelder, D. (1983). Studies on the replication of Escherichia coli phage lambda DNA. I. The kinetics of DNA replication and requirements for the generation of rolling circles. Virology *126*, 168-182.

58. Baranska, S., Gabig, M., Wegrzyn, A., Konopa, G., Herman-Antosiewicz, A., Hernandez, P., Schvartzman, J.B., Helinski, D.R., and Wegrzyn, G. (2001). Regulation of the switch from early to late bacteriophage lambda DNA replication. Microbiology *147*, 535-547.

59. Wegrzyn, A., and Wegrzyn, G. (1995). Transcriptional activation of ori lambda regulates lambda plasmid replication in amino acid-starved Escherichia coli cells. Biochem Biophys Res Commun *214*, 978-984.

60. Wegrzyn, G., Wegrzyn, A., Konieczny, I., Bielawski, K., Konopa, G., Obuchowski, M., Helinski, D.R., and Taylor, K. (1995). Involvement of the host initiator function dnaA in the replication of coliphage lambda. Genetics *139*, 1469-1481.

61. Lau, I.F., Filipe, S.R., Soballe, B., Okstad, O.A., Barre, F.X., and Sherratt, D.J. (2003). Spatial and temporal organization of replicating Escherichia coli chromosomes. Mol Microbiol *49*, 731-743.

62. Tal, A., Arbel-Goren, R., Costantino, N., Court, D.L., and Stavans, J. (2014). Location of the unique integration site on an Escherichia coli chromosome by bacteriophage lambda DNA in vivo. Proc Natl Acad Sci U S A *111*, 7308-7312.

63. Shao, Q., Trinh, J.T., and Zeng, L. (2019). High-resolution studies of lysis-lysogeny decision-making in bacteriophage lambda. J Biol Chem *294*, 3343-3349.

64. Golding, I. (2018). Infection by bacteriophage lambda: an evolving paradigm for cellular individuality. Curr Opin Microbiol *43*, 9-13.

65. Shao, Q., Trinh, J.T., McIntosh, C.S., Christenson, B., Balazsi, G., and Zeng, L. (2017). Lysis-lysogeny coexistence: prophage integration during lytic development. Microbiologyopen *6*.

66. Young, R. (2013). Phage lysis: do we have the hole story yet? Curr Opin Microbiol *16*, 790-797.

67. Hoyt, M.A., Knight, D.M., Das, A., Miller, H.I., and Echols, H. (1982). Control of phage lambda development by stability and synthesis of cII protein: role of the viral cIII and host hflA, himA and himD genes. Cell *31*, 565-573.

68. Hoopes, B.C., and McClure, W.R. (1985). A cII-dependent promoter is located within the Q gene of bacteriophage lambda. Proc Natl Acad Sci U S A *82*, 3134-3138.

69. Craig, N.L., and Nash, H.A. (1983). The mechanism of phage lambda site-specific recombination: site-specific breakage of DNA by Int topoisomerase. Cell *35*, 795-803.

70. Kotewicz, M., Chung, S., Takeda, Y., and Echols, H. (1977). Characterization of the integration protein of bacteriophage lambda as a site-specific DNA-binding protein. Proc Natl Acad Sci U S A *74*, 1511-1515.

71. Kobiler, O., Rokney, A., Friedman, N., Court, D.L., Stavans, J., and Oppenheim, A.B. (2005). Quantitative kinetic analysis of the bacteriophage lambda genetic network. Proc Natl Acad Sci U S A *102*, 4470-4475.

72. Cheng, H.H., Muhlrad, P.J., Hoyt, M.A., and Echols, H. (1988). Cleavage of the cII protein of phage lambda by purified HflA protease: control of the switch between lysis and lysogeny. Proc Natl Acad Sci U S A *85*, 7882-7886.

73. Shotland, Y., Koby, S., Teff, D., Mansur, N., Oren, D.A., Tatematsu, K., Tomoyasu, T., Kessel, M., Bukau, B., Ogura, T., et al. (1997). Proteolysis of the phage lambda CII regulatory protein by FtsH (HflB) of Escherichia coli. Mol Microbiol *24*, 1303-1310.

74. Shotland, Y., Shifrin, A., Ziv, T., Teff, D., Koby, S., Kobiler, O., and Oppenheim, A.B. (2000). Proteolysis of bacteriophage lambda CII by Escherichia coli FtsH (HflB). J Bacteriol *182*, 3111-3116.

75. Rattray, A., Altuvia, S., Mahajna, G., Oppenheim, A.B., and Gottesman, M. (1984). Control of bacteriophage lambda CII activity by bacteriophage and host functions. J Bacteriol *159*, 238-242.

76. Kourilsky, P. (1973). Lysogenization by bacteriophage lambda. I. Multiple infection and the lysogenic response. Mol Gen Genet *122*, 183-195.

77. St-Pierre, F., and Endy, D. (2008). Determination of cell fate selection during phage lambda infection. Proc Natl Acad Sci U S A *105*, 20705-20710.

78. Weitz, J.S., Mileyko, Y., Joh, R.I., and Voit, E.O. (2008). Collective decision making in bacterial viruses. Biophys J *95*, 2673-2680.

79. Kourilsky, P. (1974). Lysogenization by bacteriophage lambda. II. Identification of genes involved in the multiplicity dependent processes. Biochimie *56*, 1511-1516.

80. Wilson, H.R., Yu, D., Peters, H.K., 3rd, Zhou, J.G., and Court, D.L. (2002). The global regulator RNase III modulates translation repression by the transcription elongation factor N. EMBO J *21*, 4154-4161.

81. Slominska, M., Neubauer, P., and Wegrzyn, G. (1999). Regulation of bacteriophage lambda development by guanosine 5'-diphosphate-3'-diphosphate. Virology *262*, 431-441.

82. Alvarez, L.J., Thomen, P., Makushok, T., and Chatenay, D. (2007). Propagation of fluorescent viruses in growing plaques. Biotechnol Bioeng *96*, 615-621.

83. Trinh, J.T., Szekely, T., Shao, Q., Balazsi, G., and Zeng, L. (2017). Cell fate decisions emerge as phages cooperate or compete inside their host. Nat Commun *8*, 14341.

84. Joh, R.I., and Weitz, J.S. (2011). To lyse or not to lyse: transient-mediated stochastic fate determination in cells infected by bacteriophages. PLoS Comput Biol *7*, e1002006.

85. Balazsi, G., van Oudenaarden, A., and Collins, J.J. (2011). Cellular decision making and biological noise: from microbes to mammals. Cell *144*, 910-925.

86. Perkins, T.J., and Swain, P.S. (2009). Strategies for cellular decision-making. Mol Syst Biol *5*, 326.

87. Kuchina, A., Espinar, L., Garcia-Ojalvo, J., and Suel, G.M. (2011). Reversible and noisy progression towards a commitment point enables adaptable and reliable cellular decision-making. PLoS Comput Biol *7*, e1002273.

88. Elowitz, M.B., Levine, A.J., Siggia, E.D., and Swain, P.S. (2002). Stochastic gene expression in a single cell. Science *297*, 1183-1186.

89. Koseska, A., Zaikin, A., Kurths, J., and Garcia-Ojalvo, J. (2009). Timing cellular decision making under noise via cell-cell communication. PLoS One *4*, e4872.

90. Raj, A., and van Oudenaarden, A. (2008). Nature, nurture, or chance: stochastic gene expression and its consequences. Cell *135*, 216-226.

91. McAdams, H.H., and Arkin, A. (1997). Stochastic mechanisms in gene expression. Proc Natl Acad Sci U S A *94*, 814-819.

92. Cai, L., Friedman, N., and Xie, X.S. (2006). Stochastic protein expression in individual cells at the single molecule level. Nature *440*, 358-362.

93. Shao, Q., Cortes, M.G., Trinh, J.T., Guan, J., Balazsi, G., and Zeng, L. (2018). Coupling of DNA Replication and Negative Feedback Controls Gene Expression for Cell-Fate Decisions. iScience *6*, 1-12.

94. Brussow, H., and Hendrix, R.W. (2002). Phage genomics: small is beautiful. Cell *108*, 13-16.

95. Samson, J.E., Magadan, A.H., Sabri, M., and Moineau, S. (2013). Revenge of the phages: defeating bacterial defences. Nat Rev Microbiol *11*, 675-687.

96. Gasiunas, G., Sinkunas, T., and Siksnys, V. (2014). Molecular mechanisms of CRISPR-mediated microbial immunity. Cell Mol Life Sci *71*, 449-465.

97. Ishino, Y., Shinagawa, H., Makino, K., Amemura, M., and Nakata, A. (1987). Nucleotide sequence of the iap gene, responsible for alkaline phosphatase isozyme conversion in Escherichia coli, and identification of the gene product. J Bacteriol *169*, 5429-5433.

98. Jansen, R., Embden, J.D., Gaastra, W., and Schouls, L.M. (2002). Identification of genes that are associated with DNA repeats in prokaryotes. Mol Microbiol *43*, 1565-1575.

99. Sorek, R., Kunin, V., and Hugenholtz, P. (2008). CRISPR--a widespread system that provides acquired resistance against phages in bacteria and archaea. Nat Rev Microbiol *6*, 181-186.

100. Grissa, I., Vergnaud, G., and Pourcel, C. (2007). The CRISPRdb database and tools to display CRISPRs and to generate dictionaries of spacers and repeats. BMC Bioinformatics *8*, 172.

101. Rousseau, C., Gonnet, M., Le Romancer, M., and Nicolas, J. (2009). CRISPI: a CRISPR interactive database. Bioinformatics *25*, 3317-3318.

102. 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., et al. (2015). An updated evolutionary classification of CRISPR-Cas systems. Nat Rev Microbiol *13*, 722-736.

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

104. Bhaya, D., Davison, M., and Barrangou, R. (2011). CRISPR-Cas systems in bacteria and archaea: versatile small RNAs for adaptive defense and regulation. Annu Rev Genet *45*, 273-297.

105. Bolotin, A., Quinquis, B., Sorokin, A., and Ehrlich, S.D. (2005). Clustered regularly interspaced short palindrome repeats (CRISPRs) have spacers of extrachromosomal origin. Microbiology *151*, 2551-2561.

106. Mojica, F.J., Diez-Villasenor, C., Garcia-Martinez, J., and Soria, E. (2005). Intervening sequences of regularly spaced prokaryotic repeats derive from foreign genetic elements. J Mol Evol *60*, 174-182.

107. Pourcel, C., Salvignol, G., and Vergnaud, G. (2005). CRISPR elements in Yersinia pestis acquire new repeats by preferential uptake of bacteriophage DNA, and provide additional tools for evolutionary studies. Microbiology *151*, 653-663.

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

109. Barrangou, R., Fremaux, C., Deveau, H., Richards, M., Boyaval, P., Moineau, S., Romero, D.A., and Horvath, P. (2007). CRISPR provides acquired resistance against viruses in prokaryotes. Science *315*, 1709-1712.

110. 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., and van der Oost, J. (2008). Small CRISPR RNAs guide antiviral defense in prokaryotes. Science *321*, 960-964.

111. Marraffini, L.A., and Sontheimer, E.J. (2008). CRISPR interference limits horizontal gene transfer in staphylococci by targeting DNA. Science *322*, 1843-1845.

112. Hille, F., Richter, H., Wong, S.P., Bratovic, M., Ressel, S., and Charpentier, E. (2018). The Biology of CRISPR-Cas: Backward and Forward. Cell *172*, 1239-1259.

113. Barrangou, R. (2015). Diversity of CRISPR-Cas immune systems and molecular machines. Genome Biol *16*, 247.

114. Doudna, J.A., and Charpentier, E. (2014). Genome editing. The new frontier of genome engineering with CRISPR-Cas9. Science *346*, 1258096.

115. Makarova, K.S., Haft, D.H., Barrangou, R., Brouns, S.J., Charpentier, E., Horvath, P., Moineau, S., Mojica, F.J., Wolf, Y.I., Yakunin, A.F., et al. (2011). Evolution and classification of the CRISPR-Cas systems. Nat Rev Microbiol *9*, 467-477.

116. Takeuchi, N., Wolf, Y.I., Makarova, K.S., and Koonin, E.V. (2012). Nature and intensity of selection pressure on CRISPR-associated genes. J Bacteriol *194*, 1216-1225.

117. Horvath, P., Romero, D.A., Coute-Monvoisin, A.C., Richards, M., Deveau, H., Moineau, S., Boyaval, P., Fremaux, C., and Barrangou, R. (2008). Diversity, activity, and evolution of CRISPR loci in Streptococcus thermophilus. J Bacteriol *190*, 1401-1412.

118. Diez-Villasenor, C., Almendros, C., Garcia-Martinez, J., and Mojica, F.J. (2010). Diversity of CRISPR loci in Escherichia coli. Microbiology *156*, 1351-1361.

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

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

121. Westra, E.R., Buckling, A., and Fineran, P.C. (2014). CRISPR-Cas systems: beyond adaptive immunity. Nat Rev Microbiol *12*, 317-326.

122. Godde, J.S., and Bickerton, A. (2006). The repetitive DNA elements called CRISPRs and their associated genes: evidence of horizontal transfer among prokaryotes. J Mol Evol *62*, 718-729.

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

124. Mohanraju, P., Makarova, K.S., Zetsche, B., Zhang, F., Koonin, E.V., and van der Oost, J. (2016). Diverse evolutionary roots and mechanistic variations of the CRISPR-Cas systems. Science *353*, aad5147.

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

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

127. Nunez, J.K., Harrington, L.B., Kranzusch, P.J., Engelman, A.N., and Doudna, J.A. (2015). Foreign DNA capture during CRISPR-Cas adaptive immunity. Nature *527*, 535-538.

128. Wiedenheft, B., Sternberg, S.H., and Doudna, J.A. (2012). RNA-guided genetic silencing systems in bacteria and archaea. Nature *482*, 331-338.

129. Elmore, J.R., Sheppard, N.F., Ramia, N., Deighan, T., Li, H., Terns, R.M., and Terns, M.P. (2016). Bipartite recognition of target RNAs activates DNA cleavage by the Type III-B CRISPR-Cas system. Genes Dev *30*, 447-459.

130. Leenay, R.T., and Beisel, C.L. (2017). Deciphering, Communicating, and Engineering the CRISPR PAM. J Mol Biol *429*, 177-191.

131. Dorsey, B.W., Huang, L., and Mondragon, A. (2019). Structural organization of a Type III-A CRISPR effector subcomplex determined by X-ray crystallography and cryo-EM. Nucleic Acids Res *47*, 3765-3783.

132. Kazlauskiene, M., Kostiuk, G., Venclovas, C., Tamulaitis, G., and Siksnys, V. (2017). A cyclic oligonucleotide signaling pathway in type III CRISPR-Cas systems. Science *357*, 605-609.

133. Makarova, K.S., Wolf, Y.I., and Koonin, E.V. (2018). Classification and Nomenclature of CRISPR-Cas Systems: Where from Here? CRISPR J *1*, 325-336.

134. Samai, P., Pyenson, N., Jiang, W., Goldberg, G.W., Hatoum-Aslan, A., and Marraffini, L.A. (2015). Co-transcriptional DNA and RNA Cleavage during Type III CRISPR-Cas Immunity. Cell *161*, 1164-1174.

135. Taylor, D.W., Zhu, Y., Staals, R.H., Kornfeld, J.E., Shinkai, A., van der Oost, J., Nogales, E., and Doudna, J.A. (2015). Structural biology. Structures of the CRISPR-Cmr complex reveal mode of RNA target positioning. Science *348*, 581-585.

136. You, L., Ma, J., Wang, J., Artamonova, D., Wang, M., Liu, L., Xiang, H., Severinov, K., Zhang, X., and Wang, Y. (2019). Structure Studies of the CRISPR-Csm Complex Reveal Mechanism of Co-transcriptional Interference. Cell *176*, 239-253 e216.

137. Mogila, I., Kazlauskiene, M., Valinskyte, S., Tamulaitiene, G., Tamulaitis, G., and Siksnys, V. (2019). Genetic Dissection of the Type III-A CRISPR-Cas System Csm Complex Reveals Roles of Individual Subunits. Cell Rep *26*, 2753-2765 e2754.

138. Ozcan, A., Pausch, P., Linden, A., Wulf, A., Schuhle, K., Heider, J., Urlaub, H., Heimerl, T., Bange, G., and Randau, L. (2019). Type IV CRISPR RNA processing and effector complex formation in Aromatoleum aromaticum. Nat Microbiol *4*, 89-96.

139. Gasiunas, G., Barrangou, R., Horvath, P., and Siksnys, V. (2012). Cas9-crRNA ribonucleoprotein complex mediates specific DNA cleavage for adaptive immunity in bacteria. Proc Natl Acad Sci U S A *109*, E2579-2586.

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

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

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

143. Mojica, F.J., Diez-Villasenor, C., Garcia-Martinez, J., and Almendros, C. (2009). Short motif sequences determine the targets of the prokaryotic CRISPR defence system. Microbiology *155*, 733-740.

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

145. Chylinski, K., Makarova, K.S., Charpentier, E., and Koonin, E.V. (2014). Classification and evolution of type II CRISPR-Cas systems. Nucleic Acids Res *42*, 6091-6105.

146. Shmakov, S., Abudayyeh, O.O., Makarova, K.S., Wolf, Y.I., Gootenberg, J.S., Semenova, E., Minakhin, L., Joung, J., Konermann, S., Severinov, K., et al. (2015). Discovery and Functional Characterization of Diverse Class 2 CRISPR-Cas Systems. Mol Cell *60*, 385-397.

147. Shmakov, S., Smargon, A., Scott, D., Cox, D., Pyzocha, N., Yan, W., Abudayyeh, O.O., Gootenberg, J.S., Makarova, K.S., Wolf, Y.I., et al. (2017). Diversity and evolution of class 2 CRISPR-Cas systems. Nat Rev Microbiol *15*, 169-182.

148. Zetsche, B., Gootenberg, J.S., Abudayyeh, O.O., Slaymaker, I.M., Makarova, K.S., Essletzbichler, P., Volz, S.E., Joung, J., van der Oost, J., Regev, A., et al. (2015). Cpf1 is a single RNA-guided endonuclease of a class 2 CRISPR-Cas system. Cell *163*, 759-771.

149. Yang, H., Gao, P., Rajashankar, K.R., and Patel, D.J. (2016). PAM-Dependent Target DNA Recognition and Cleavage by C2c1 CRISPR-Cas Endonuclease. Cell *167*, 1814-1828 e1812.

150. Abudayyeh, O.O., Gootenberg, J.S., Konermann, S., Joung, J., Slaymaker, I.M., Cox, D.B., Shmakov, S., Makarova, K.S., Semenova, E., Minakhin, L., et al. (2016). C2c2 is a single-component programmable RNA-guided RNA-targeting CRISPR effector. Science *353*, aaf5573.

151. Knott, G.J., and Doudna, J.A. (2018). CRISPR-Cas guides the future of genetic engineering. Science *361*, 866-869.

152. Cho, S.W., Kim, S., Kim, J.M., and Kim, J.S. (2013). Targeted genome engineering in human cells with the Cas9 RNA-guided endonuclease. Nat Biotechnol *31*, 230-232.

153. Jiang, W., Bikard, D., Cox, D., Zhang, F., and Marraffini, L.A. (2013). RNA-guided editing of bacterial genomes using CRISPR-Cas systems. Nat Biotechnol *31*, 233-239.

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

155. Mali, P., Yang, L., Esvelt, K.M., Aach, J., Guell, M., DiCarlo, J.E., Norville, J.E., and Church, G.M. (2013). RNA-guided human genome engineering via Cas9. Science *339*, 823-826.

156. Bikard, D., Jiang, W., Samai, P., Hochschild, A., Zhang, F., and Marraffini, L.A. (2013). Programmable repression and activation of bacterial gene expression using an engineered CRISPR-Cas system. Nucleic Acids Res *41*, 7429-7437.

157. Chen, B., Gilbert, L.A., Cimini, B.A., Schnitzbauer, J., Zhang, W., Li, G.W., Park, J., Blackburn, E.H., Weissman, J.S., Qi, L.S., et al. (2013). Dynamic imaging of genomic loci in living human cells by an optimized CRISPR/Cas system. Cell *155*, 1479-1491.

158. Gilbert, L.A., Larson, M.H., Morsut, L., Liu, Z., Brar, G.A., Torres, S.E., Stern-Ginossar, N., Brandman, O., Whitehead, E.H., Doudna, J.A., et al. (2013). CRISPR-mediated modular RNA-guided regulation of transcription in eukaryotes. Cell *154*, 442-451.

159. Fu, Y., Foden, J.A., Khayter, C., Maeder, M.L., Reyon, D., Joung, J.K., and Sander, J.D. (2013). High-frequency off-target mutagenesis induced by CRISPR-Cas nucleases in human cells. Nat Biotechnol *31*, 822-826.

160. Tsai, S.Q., and Joung, J.K. (2016). Defining and improving the genome-wide specificities of CRISPR-Cas9 nucleases. Nat Rev Genet *17*, 300-312.

161. Tsai, S.Q., Zheng, Z., Nguyen, N.T., Liebers, M., Topkar, V.V., Thapar, V., Wyvekens, N., Khayter, C., Iafrate, A.J., Le, L.P., et al. (2015). GUIDE-seq enables genome-wide profiling of off-target cleavage by CRISPR-Cas nucleases. Nat Biotechnol *33*, 187-197.

162. D'Astolfo, D.S., Pagliero, R.J., Pras, A., Karthaus, W.R., Clevers, H., Prasad, V., Lebbink, R.J., Rehmann, H., and Geijsen, N. (2015). Efficient intracellular delivery of native proteins. Cell *161*, 674-690.

163. Ramakrishna, S., Kwaku Dad, A.B., Beloor, J., Gopalappa, R., Lee, S.K., and Kim, H. (2014). Gene disruption by cell-penetrating peptide-mediated delivery of Cas9 protein and guide RNA. Genome Res *24*, 1020-1027.

164. Kamerbeek, J., Schouls, L., Kolk, A., van Agterveld, M., van Soolingen, D., Kuijper, S., Bunschoten, A., Molhuizen, H., Shaw, R., Goyal, M., et al. (1997). Simultaneous detection and strain differentiation of Mycobacterium tuberculosis for diagnosis and epidemiology. J Clin Microbiol *35*, 907-914.

165. Huo, Y., Nam, K.H., Ding, F., Lee, H., Wu, L., Xiao, Y., Farchione, M.D., Jr., Zhou, S., Rajashankar, K., Kurinov, I., et al. (2014). Structures of CRISPR Cas3 offer mechanistic insights into Cascade-activated DNA unwinding and degradation. Nat Struct Mol Biol *21*, 771-777.

166. Touchon, M., and Rocha, E.P. (2010). The small, slow and specialized CRISPR and anti-CRISPR of Escherichia and Salmonella. PLoS One *5*, e11126.

167. Pul, U., Wurm, R., Arslan, Z., Geissen, R., Hofmann, N., and Wagner, R. (2010). Identification and characterization of E. coli CRISPR-cas promoters and their silencing by H-NS. Mol Microbiol *75*, 1495-1512.

168. Marraffini, L.A. (2015). CRISPR-Cas immunity in prokaryotes. Nature *526*, 55-61.

169. Xue, C., and Sashital, D.G. (2019). Mechanisms of Type I-E and I-F CRISPR-Cas Systems in Enterobacteriaceae. EcoSal Plus *8*.

170. Sternberg, S.H., Richter, H., Charpentier, E., and Qimron, U. (2016). Adaptation in CRISPR-Cas Systems. Mol Cell *61*, 797-808.

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

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

173. Kunne, T., Kieper, S.N., Bannenberg, J.W., Vogel, A.I., Miellet, W.R., Klein, M., Depken, M., Suarez-Diez, M., and Brouns, S.J. (2016). Cas3-Derived Target DNA Degradation Fragments Fuel Primed CRISPR Adaptation. Mol Cell *63*, 852-864.

174. Semenova, E., Savitskaya, E., Musharova, O., Strotskaya, A., Vorontsova, D., Datsenko, K.A., Logacheva, M.D., and Severinov, K. (2016). 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 *113*, 7626-7631.

175. Redding, S., Sternberg, S.H., Marshall, M., Gibb, B., Bhat, P., Guegler, C.K., Wiedenheft, B., Doudna, J.A., and Greene, E.C. (2015). Surveillance and Processing of Foreign DNA by the Escherichia coli CRISPR-Cas System. Cell *163,* 854-865.

176. Dillard, K.E., Brown, M.W., Johnson, N.V., Xiao, Y., Dolan, A., Hernandez, E., Dahlhauser, S.D., Kim, Y., Myler, L.R., Anslyn, E.V., et al. (2018). Assembly and Translocation of a CRISPR-Cas Primed Acquisition Complex. Cell *175*, 934-946 e915.

177. Xue, C., Zhu, Y., Zhang, X., Shin, Y.K., and Sashital, D.G. (2017). Real-Time Observation of Target Search by the CRISPR Surveillance Complex Cascade. Cell Rep *21*, 3717-3727.

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

179. Wang, J., Li, J., Zhao, H., Sheng, G., Wang, M., Yin, M., and Wang, Y. (2015). Structural and Mechanistic Basis of PAM-Dependent Spacer Acquisition in CRISPR-Cas Systems. Cell *163,* 840-853.

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

181. Nunez, J.K., Lee, A.S., Engelman, A., and Doudna, J.A. (2015). Integrase-mediated spacer acquisition during CRISPR-Cas adaptive immunity. Nature *519*, 193-198.

182. Nunez, J.K., Bai, L., Harrington, L.B., Hinder, T.L., and Doudna, J.A. (2016). CRISPR Immunological Memory Requires a Host Factor for Specificity. Mol Cell *62*, 824-833.

183. McGinn, J., and Marraffini, L.A. (2016). CRISPR-Cas Systems Optimize Their Immune Response by Specifying the Site of Spacer Integration. Mol Cell *64*, 616-623.

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

185. Yang, C.D., Chen, Y.H., Huang, H.Y., Huang, H.D., and Tseng, C.P. (2014). CRP represses the CRISPR/Cas system in Escherichia coli: evidence that endogenous CRISPR spacers impede phage P1 replication. Mol Microbiol *92*, 1072-1091.

186. Yosef, I., Goren, M.G., Kiro, R., Edgar, R., and Qimron, U. (2011). High-temperature protein G is essential for activity of the Escherichia coli clustered regularly interspaced short palindromic repeats (CRISPR)/Cas system. Proc Natl Acad Sci U S A *108*, 20136-20141.

187. Westra, E.R., Pul, U., Heidrich, N., Jore, M.M., Lundgren, M., Stratmann, T., Wurm, R., Raine, A., Mescher, M., Van Heereveld, L., et al. (2010). H-NS-mediated repression of CRISPR-based immunity in Escherichia coli K12 can be relieved by the transcription activator LeuO. Mol Microbiol *77*, 1380-1393.

188. Hochstrasser, M.L., and Doudna, J.A. (2015). Cutting it close: CRISPR-associated endoribonuclease structure and function. Trends Biochem Sci *40*, 58-66.

189. Haurwitz, R.E., Jinek, M., Wiedenheft, B., Zhou, K., and Doudna, J.A. (2010). Sequence- and structure-specific RNA processing by a CRISPR endonuclease. Science *329*, 1355-1358.

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

191. 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., et al. (2011). Structural basis for CRISPR RNA-guided DNA recognition by Cascade. Nat Struct Mol Biol *18*, 529-536.

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

193. Pyenson, N.C., Gayvert, K., Varble, A., Elemento, O., and Marraffini, L.A. (2017). Broad Targeting Specificity during Bacterial Type III CRISPR-Cas Immunity Constrains Viral Escape. Cell Host Microbe *22*, 343-353 e343.

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

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

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

197. Ivancic-Bace, I., Al Howard, J., and Bolt, E.L. (2012). Tuning in to interference: R-loops and cascade complexes in CRISPR immunity. J Mol Biol *422*, 607-616.

198. Hayes, R.P., Xiao, Y., Ding, F., van Erp, P.B., Rajashankar, K., Bailey, S., Wiedenheft, B., and Ke, A. (2016). Structural basis for promiscuous PAM recognition in type I-E Cascade from E. coli. Nature *530*, 499-503.

199. Xiao, Y., Luo, M., Hayes, R.P., Kim, J., Ng, S., Ding, F., Liao, M., and Ke, A. (2017). Structure Basis for Directional R-loop Formation and Substrate Handover Mechanisms in Type I CRISPR-Cas System. Cell *170*, 48-60 e11.

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

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

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

203. 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., et al. (2012). CRISPR immunity relies on the consecutive binding and degradation of negatively supercoiled invader DNA by Cascade and Cas3. Mol Cell *46*, 595-605.

204. Leenay, R.T., Maksimchuk, K.R., Slotkowski, R.A., Agrawal, R.N., Gomaa, A.A., Briner, A.E., Barrangou, R., and Beisel, C.L. (2016). Identifying and Visualizing Functional PAM Diversity across CRISPR-Cas Systems. Mol Cell *62*, 137-147.

205. Fu, B.X., Wainberg, M., Kundaje, A., and Fire, A.Z. (2017). High-Throughput Characterization of Cascade type I-E CRISPR Guide Efficacy Reveals Unexpected PAM Diversity and Target Sequence Preferences. Genetics *206*, 1727-1738.

206. Fineran, P.C., Gerritzen, M.J., Suarez-Diez, M., Kunne, T., Boekhorst, J., van Hijum, S.A., Staals, R.H., and Brouns, S.J. (2014). Degenerate target sites mediate rapid primed CRISPR adaptation. Proc Natl Acad Sci U S A *111*, E1629-1638.

207. Caliando, B.J., and Voigt, C.A. (2015). Targeted DNA degradation using a CRISPR device stably carried in the host genome. Nat Commun *6*, 6989.

208. Westra, E.R., Semenova, E., Datsenko, K.A., Jackson, R.N., Wiedenheft, B., Severinov, K., and Brouns, S.J. (2013). Type I-E CRISPR-cas systems discriminate target from non-target DNA through base pairing-independent PAM recognition. PLoS Genet *9*, e1003742.

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

210. Rutkauskas, M., Sinkunas, T., Songailiene, I., Tikhomirova, M.S., Siksnys, V., and Seidel, R. (2015). Directional R-Loop Formation by the CRISPR-Cas Surveillance Complex Cascade Provides Efficient Off-Target Site Rejection. Cell Rep *10*, 1534-1543.

211. Jochem N.A. Vink, K.J.A.M., Marnix Vlot, Rebecca E. McKenzie, Cristóbal Almendros, Boris Estrada Bonilla, Daan J.W. Brocken, Johannes Hohlbein, Stan J.J.Brouns (2019). Direct visualization of native CRISPR target search in live bacteria reveals Cascade DNA surveillance mechanism. bioRxiv.

212. Seed, K.D., Lazinski, D.W., Calderwood, S.B., and Camilli, A. (2013). A bacteriophage encodes its own CRISPR/Cas adaptive response to evade host innate immunity. Nature *494*, 489-491.

213. Naser, I.B., Hoque, M.M., Nahid, M.A., Tareq, T.M., Rocky, M.K., and Faruque, S.M. (2017). Analysis of the CRISPR-Cas system in bacteriophages active on epidemic strains of Vibrio cholerae in Bangladesh. Sci Rep *7*, 14880.

214. Bondy-Denomy, J., Pawluk, A., Maxwell, K.L., and Davidson, A.R. (2013). Bacteriophage genes that inactivate the CRISPR/Cas bacterial immune system. Nature *493*, 429-432.

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

216. Hwang, S., and Maxwell, K.L. (2019). Meet the Anti-CRISPRs: Widespread Protein Inhibitors of CRISPR-Cas Systems. CRISPR J *2*, 23-30.

217. Maxwell, K.L. (2017). The Anti-CRISPR Story: A Battle for Survival. Mol Cell *68*, 8-14.

218. Pawluk, A., Davidson, A.R., and Maxwell, K.L. (2018). Anti-CRISPR: discovery, mechanism and function. Nat Rev Microbiol *16*, 12-17.

219. Chowdhury, S., Carter, J., Rollins, M.F., Golden, S.M., Jackson, R.N., Hoffmann, C., Nosaka, L., Bondy-Denomy, J., Maxwell, K.L., Davidson, A.R., et al. (2017). Structure Reveals Mechanisms of Viral Suppressors that Intercept a CRISPR RNA-Guided Surveillance Complex. Cell *169*, 47-57 e11.

220. Guo, T.W., Bartesaghi, A., Yang, H., Falconieri, V., Rao, P., Merk, A., Eng, E.T., Raczkowski, A.M., Fox, T., Earl, L.A., et al. (2017). Cryo-EM Structures Reveal Mechanism and Inhibition of DNA Targeting by a CRISPR-Cas Surveillance Complex. Cell *171*, 414-426 e412.

221. Bondy-Denomy, J., Garcia, B., Strum, S., Du, M., Rollins, M.F., Hidalgo-Reyes, Y., Wiedenheft, B., Maxwell, K.L., and Davidson, A.R. (2015). Multiple mechanisms for CRISPR-Cas inhibition by anti-CRISPR proteins. Nature *526*, 136-139.

222. Pawluk, A., Shah, M., Mejdani, M., Calmettes, C., Moraes, T.F., Davidson, A.R., and Maxwell, K.L. (2017). Disabling a Type I-E CRISPR-Cas Nuclease with a Bacteriophage-Encoded Anti-CRISPR Protein. MBio *8*.

223. Harrington, L.B., Doxzen, K.W., Ma, E., Liu, J.J., Knott, G.J., Edraki, A., Garcia, B., Amrani, N., Chen, J.S., Cofsky, J.C., et al. (2017). A Broad-Spectrum Inhibitor of CRISPR-Cas9. Cell *170*, 1224-1233 e1215.

224. Borges, A.L., Zhang, J.Y., Rollins, M.F., Osuna, B.A., Wiedenheft, B., and Bondy-Denomy, J. (2018). Bacteriophage Cooperation Suppresses CRISPR-Cas3 and Cas9 Immunity. Cell *174*, 917-925 e910.

225. Landsberger, M., Gandon, S., Meaden, S., Rollie, C., Chevallereau, A., Chabas, H., Buckling, A., Westra, E.R., and van Houte, S. (2018). Anti-CRISPR Phages Cooperate to Overcome CRISPR-Cas Immunity. Cell *174*, 908-916 e912.

226. Ho, Y., Lewis, M., and Rosenberg, M. (1982). Purification and properties of a transcriptional activator. The cII protein of phage lambda. J Biol Chem *257*, 9128-9134.

227. Nash, H.A. (1981). Integration and excision of bacteriophage lambda: the mechanism of conservation site specific recombination. Annu Rev Genet *15*, 143-167.

228. Nash, H.A., and Robertson, C.A. (1981). Purification and properties of the Escherichia coli protein factor required for lambda integrative recombination. J Biol Chem *256*, 9246-9253.

229. Shao, Q., Trinh, J.T., McIntosh, C.S., Christenson, B., Balazsi, G., and Zeng, L. (2016). Lysis-lysogeny coexistence: prophage integration during lytic development. Microbiologyopen.

230. Skinner, S.O., Sepulveda, L.A., Xu, H., and Golding, I. (2013). Measuring mRNA copy number in individual Escherichia coli cells using single-molecule fluorescent in situ hybridization. Nat Protoc *8*, 1100-1113.

231. LeBowitz, J.H., and McMacken, R. (1986). The Escherichia coli dnaB replication protein is a DNA helicase. J Biol Chem *261*, 4738-4748.

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

233. Yakovleva, G.M., Kim, S.K., and Wanner, B.L. (1998). Phosphate-independent expression of the carbon-phosphorus lyase activity of Escherichia coli. Appl Microbiol Biotechnol *49*, 573-578.

234. Kuhlman, T., Zhang, Z., Saier, M.H., Jr., and Hwa, T. (2007). Combinatorial transcriptional control of the lactose operon of Escherichia coli. Proc Natl Acad Sci U S A *104*, 6043-6048.

235. Visser, B.J., Joshi, M.C., and Bates, D. (2017). Multilocus Imaging of the E. coli Chromosome by Fluorescent In Situ Hybridization. Methods Mol Biol *1624*, 213-226.

236. Sliusarenko, O., Heinritz, J., Emonet, T., and Jacobs-Wagner, C. (2011). High-throughput, subpixel precision analysis of bacterial morphogenesis and intracellular spatio-temporal dynamics. Mol Microbiol *80*, 612-627.

237. Slater, S., Wold, S., Lu, M., Boye, E., Skarstad, K., and Kleckner, N. (1995). E. coli SeqA protein binds oriC in two different methyl-modulated reactions appropriate to its roles in DNA replication initiation and origin sequestration. Cell *82*, 927-936.

238. Lutz, R., and Bujard, H. (1997). Independent and tight regulation of transcriptional units in Escherichia coli via the LacR/O, the TetR/O and AraC/I1-I2 regulatory elements. Nucleic Acids Res *25*, 1203-1210.

239. Krogh, T.J., Moller-Jensen, J., and Kaleta, C. (2018). Impact of Chromosomal Architecture on the Function and Evolution of Bacterial Genomes. Front Microbiol *9*, 2019.

240. Bednarz, M., Halliday, J.A., Herman, C., and Golding, I. (2014). Revisiting bistability in the lysis/lysogeny circuit of bacteriophage lambda. PLoS One *9*, e100876.

241. Arkin, A., Ross, J., and McAdams, H.H. (1998). Stochastic kinetic analysis of developmental pathway bifurcation in phage lambda-infected Escherichia coli cells. Genetics *149*, 1633-1648.

242. Baba, T., Ara, T., Hasegawa, M., Takai, Y., Okumura, Y., Baba, M., Datsenko, K.A., Tomita, M., Wanner, B.L., and Mori, H. (2006). Construction of Escherichia coli K-12 in-frame, single-gene knockout mutants: the Keio collection. Mol Syst Biol *2*, 2006 0008.

243. Zeng, L., and Golding, I. (2011). Following cell-fate in E. coli after infection by phage lambda. J Vis Exp, e3363.

244. Wang, J., Michel, V., Hofnung, M., and Charbit, A. (1998). Cloning of the J gene of bacteriophage lambda, expression and solubilization of the J protein: first in vitro studies on the interactions between J and LamB, its cell surface receptor. Res Microbiol *149*, 611-624.

245. Sanger, F., Coulson, A.R., Hong, G.F., Hill, D.F., and Petersen, G.B. (1982). Nucleotide sequence of bacteriophage lambda DNA. J Mol Biol *162*, 729-773.

246. Szyf, M., Avraham-Haetzni, K., Reifman, A., Shlomai, J., Kaplan, F., Oppenheim, A., and Razin, A. (1984). DNA methylation pattern is determined by the intracellular level of the methylase. Proc Natl Acad Sci U S A *81*, 3278-3282.

247. Begg, K.J. (1978). Cell surface growth in Escherichia coli: distribution of matrix protein. J Bacteriol *135*, 307-310.

248. Begg, K.J., and Donachie, W.D. (1984). Concentration of a major outer membrane protein at the cell poles in Escherichia coli. J Gen Microbiol *130*, 2339-2346.

249. Guan, J., Shi, X., Burgos, R., and Zeng, L. (2017). Visualization of phage DNA degradation by a type I CRISPR-Cas system at the single-cell level. Quant Biol *5*, 67-75.

250. van Erp, P.B., Jackson, R.N., Carter, J., Golden, S.M., Bailey, S., and Wiedenheft, B. (2015). Mechanism of CRISPR-RNA guided recognition of DNA targets in Escherichia coli. Nucleic Acids Res *43*, 8381-8391.

251. van der Oost, J., Westra, E.R., Jackson, R.N., and Wiedenheft, B. (2014). Unravelling the structural and mechanistic basis of CRISPR-Cas systems. Nat Rev Microbiol *12*, 479-492.

252. Hatoum-Aslan, A., Maniv, I., and Marraffini, L.A. (2011). Mature clustered, regularly interspaced, short palindromic repeats RNA (crRNA) length is measured by a ruler mechanism anchored at the precursor processing site. Proc Natl Acad Sci U S A *108*, 21218-21222.

253. Amitai, G., and Sorek, R. (2016). CRISPR-Cas adaptation: insights into the mechanism of action. Nat Rev Microbiol *14*, 67-76.

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

255. Mulepati, S., Heroux, A., and Bailey, S. (2014). Structural biology. Crystal structure of a CRISPR RNA-guided surveillance complex bound to a ssDNA target. Science *345*, 1479-1484.

256. Babic, A., Lindner, A.B., Vulic, M., Stewart, E.J., and Radman, M. (2008). Direct visualization of horizontal gene transfer. Science *319*, 1533-1536.

257. Edgar, R., and Qimron, U. (2010). The Escherichia coli CRISPR system protects from lambda lysogenization, lysogens, and prophage induction. J Bacteriol *192*, 6291-6294.

258. Lu, M., Campbell, J.L., Boye, E., and Kleckner, N. (1994). SeqA: a negative modulator of replication initiation in E. coli. Cell *77*, 413-426.

259. Gill, J.J., and Hyman, P. (2010). Phage choice, isolation, and preparation for phage therapy. Curr Pharm Biotechnol *11*, 2-14.

260. Monteiro, R., Pires, D.P., Costa, A.R., and Azeredo, J. (2019). Phage Therapy: Going Temperate? Trends Microbiol *27*, 368-378.