DETERMINANTS THAT INFLUENCE DNA UPTAKE INTO CRISPR LOCI OF *PYROCOCCUS FURIOSUS*

by

ELIZABETH ANN WATTS

(Under the Direction of Michael P. Terns)

ABSTRACT

The constant battle between viruses and host organisms provides evolutionary pressure for both to evolve powerful defense and counter-defense mechanisms to one another. Many prokaryotes defend themselves from viruses and other mobile genetic elements (MGEs) such as plasmids, using adaptive immune systems called CRISPR-Cas (Clustered Regularly Interspaced Short Palindromic Repeats-CRISPR Associated) systems. These CRISPR-Cas systems function by integrating short fragments of DNA sequences (called spacers) from the MGEs into host cell CRISPR genomic arrays to provide a heritable record of the captured MGE sequences (a process called adaptation). Subsequent CRISPR array expression and RNA processing leads to the production of small CRISPR RNAs (crRNAs) that guide recognition and Cas nuclease-mediated destruction of invasive MGE nucleic acids to prevent further MGE infection. While integration of invasive MGE DNA fragments into a host CRISPR array is normally an exceptionally rare event, we observed that introduction of the first natural MGE (the pT33.3 conjugative plasmid) tested for our model organism, the hyperthermophilic archaeon *Pyrococcus furiosus*, stimulated a strikingly robust spacer acquisition response

whereby the majority of cells captured DNA fragments specifically against this pT33.3 plasmid into the host CRISPR arrays. Our investigation revealed that the observed "hyper-adaptation" response was mediated by a specialized pathway known as "primed" adaptation" that relied upon the presence of a partially matching, naturally acquired spacer in one of the CRISPR arrays. In other work described in this thesis, we obtained key mechanistic insight into the important question for how new spacers become integrated in a directional manner at the leader end repeat rather than internal repeats within CRISPR arrays. Here, through a combination of micrococcal nuclease DNA protection assays, *in vivo* and *in vitro* adaptation studies and high throughput sequencing, we discovered that the archaeal histones of *Pyrococcus furiosus* play a major role in the guidance of new spacers to the first repeat of CRISPR arrays. The work in this dissertation highlights the unique response against the first natural MGE for *Pyrococcus furiosus* and reveals a novel role of archaeal histone proteins in shaping integration of new spacer DNA in a polarized manner at CRISPR arrays.

INDEX WORDS: CRISPR; Cas; Adaptation; Conjugative; Histones; spacer integration; mobile genetic elements; *Pyrococcus furiosus*

DETERMINANTS THAT INFLUENCE DNA UPTAKE INTO CRISPR LOCI OF *PYROCOCCUS FURIOSUS*

by

ELIZABETH ANN WATTS

B.S., Eastern Kentucky University, 2012

A Dissertation Submitted to the Graduate Faculty of The University of Georgia in Partial Fulfillment of the Requirements for the Degree

DOCTOR OF PHILOSOPHY

ATHENS, GEORGIA

2022

DETERMINANTS THAT INFLUENCE DNA UPTAKE INTO CRISPR LOCI OF *PYROCOCCUS FURIOSUS*

by

ELIZABETH ANN WATTS

Major Professor:	Michael P. Terns
Committee:	Bill Lanzilotta
	Janet Westpheling
	Michael McEachern

Electronic Version Approved:

Ron Walcott
Vice Provost for Graduate Education and Dean of the Graduate School
The University of Georgia
May 2022

DEDICATION

This dissertation is dedicated to my fiancée, Dustin Williams. Without you, I honestly would not have been able to make it through this journey. Thank you for your constant support and love, and making sure that I took the time to take care of myself emotionally. I can't thank you enough for everything that you do, you are an amazing person and I'm so glad that I found someone like you to spend the rest of my life with. Let's chase those dreams!

I would also like to dedicate this to my dad. Thank you for always encouraging me to go my own way. Your hard work ethic has always motivated me to become the best version of myself, and I'm so honored to be your daughter.

ACKNOWLEDGEMENTS

First and foremost, I would like to thank Dr. Terns for all his support throughout the years. Thank you for the tough love at times and pushing me to strive for excellence. I would also like to thank my committee members, Dr. Janet Westpheling, Dr. Bill Lanzilotta, and Dr. Michael McEachern (and previous members Dr. Rebecca Terns and Dr. Claiborne Glover) for always having insightful questions and suggestions during our meetings, and your support for my future.

Our lab has always been composed of great colleagues who are always willing to help and are supportive of each other. A huge thank you to Clare Cooper (along with Derek and the "snack pack") for always being there for support, and your sincere friendship. Thank you to all the Terns lab members that I have had the pleasure of knowing throughout the years, which include Yunzhou Wei, Masami Shiimori, Julie Grainy, Kawanda Foster, Xinfu Zhang, Walter Woodside, Jenny Kim, Sandra Garrett, Justin Mclean, Ryan Catchpole, Cécile Philippe, Katie Johnson, Chris Noble-Molnar, Conor Pittman, Raven Tucker, Ela Mitchell, and Arianna Hagins. I would also like to thank Landon Clark, for helping me to develop as a mentor, and for your dedication to our work. You always brought a positive attitude and made some of the hardships we have went through a little easier to deal with.

And finally, thank you to Dylan, Harley, and Tootie (our cats). You were always there to cuddle at the end of the day, I will always be thankful for your emotional support.

TABLE OF CONTENTS

Page

CHAPTER 1

INTRODUCTION AND LITERATURE REVIEW

The CRISPR-Cas (Clustered Regularly Interspaced Short Palindromic Repeat-CRISPR associated) system is an adaptive prokaryotic immune system that targets and silences invading DNA from viruses and other mobile genetic elements. There are many diverse CRISPR-Cas systems, consisting of three main steps: adaptation, crRNA (CRISPR RNA) biogenesis, and invader silencing. This dissertation seeks to study the least understood part of this system, adaptation.

Our model organism, *Pyrococcus furiosus*, contains three CRISPR-Cas systems as well as seven CRISPR arrays. There are 200 spacers within these arrays, however, the origins of these spacers are unknown. Therefore, we have had to use standard shuttle vectors to study adaptation in our systems. Excitingly, an archaeal conjugative plasmid was recently discovered to transfer between many Thermococcales. We were able to utilize this plasmid in our system as well and study the response of the CRISPR-Cas system against this natural plasmid. It was found to cause a hyperadaptation effect, and we ultimately found that our system was undergoing a specialized pathway known as primed adaptation against this invader.

It was found that in addition to Cas proteins, there are also "non-Cas" proteins that are involved in some of the steps in adaptation. IHF (Integration Host Factor), for example, was found to be essential in the polarization of newly integrated spacers in the

genome of *Escherichia Coli* (*E. Coli*) [1]. This leads us to assume that other non-Cas factors may be involved in the various steps of adaptation in *Pyrococcus furiosus*. In fact, we have found that archaeal histones influence polarization of new spacers into CRISPR arrays.

In this dissertation, we seek to explore determinants that affect integration at CRISPR arrays in *Pyrococcus furiosus*.

CRISPR-Cas Systems

CRISPR-Cas systems are adaptive immune systems of prokaryotic organisms [2-4] (Figure 1.1). These systems allow for the acquisition of short DNA fragments from invading viruses or plasmids, and integrates the nucleolytically processed fragments into the host genome [2, 4-9]. The entire CRISPR locus is then transcribed and the resultant primary transcript is processed by Cas proteins to generate a mature crRNA [10-14]. The crRNAs then associate with specific Cas proteins to assemble into an interference complex, and to guide the complex to the target sequence [5, 6, 15-22] (Figure 1.1).

Our current model includes several conceptual steps for adaptation. The first step is the nuclease excision of protospacers (DNA derived from invading elements) from foreign DNA sources [23]. After a "pool" of protospacers have been created, Cas adaptation associated proteins then identify the protospacers by recognition of a PAM (Protospacer Adjacent Motif) and bind the protospacers [24-29]. These protospacers are then "trimmed" down to an appropriate size, resulting in the loss of the PAM sequence. These protospacers are then integrated into the genome of the host prokaryote in a

polarized manner into a CRISPR array consisting of many repeats and spacers [2, 4-9]. Once integrated into the array, these protospacers are now referred to as spacers.

In adaptation, a protospacer (the fragment of DNA from the invader) is targeted adjacent to a PAM sequence [24-29]. PAM sequences can be from three to five nucleotides in length, and are required for selection of the protospacer during the adaptation step, as well as a method to aid in to selection of the foreign DNA rather the host genome destruction during the downstream silencing stage [21, 27]. It has been found that newly acquired spacers were acquired from regions corresponding to active transposons, CRISPR loci, ribosomal RNA gens, rolling circle origins of replication in our system (*Pyrococcus furiosus*), which combined with data from other systems, suggest that free DNA termini and PAMS are important for spacer acquisition events [30] [23, 31-33]. The protospacers are attained by the universal Cas1 and Cas2 integrase complex, which are known to allow for integration into the CRISPR locus [34-36]. *In vivo*, spacer integration occurs primarily at the repeat that is most proximal to the leader, resulting in the duplication of an additional repeat as well [1, 36-38]. After integration, this protospacer is then referred to as a spacer within the CRISPR array. This CRISPR array is composed of several components: 1) the leader, which has elements that allow for transcription and integration of new spacers, 2) repeats, which are direct repeats of conserved nucleotides throughout the CRISPR array, and 3) the spacers, which are on average 30-45 base-pairs in length (but are typically of a very narrow size distribution in a given CRISPR array) and are mostly derived from invading MGEs (Figure 1.2). These

spacers are incorporated into the CRISPR array in a polarized fashion, therefore the spacer in closest proximity of the leader is typically the more recent invader [39].

There are multiple CRISPR-Cas systems based on diverse characteristics of each system. There are two main classes, six types, and over 30 subtypes [40]. The three major types of these systems consist of type I, type II, and type III. They are divided based on their distinctive proteins involved in interference. type I is associated with Cas3, type II Cas9, and type III Cas10 [41-45]. Both type I and type III systems are found in *Pyrococcus furiosus* [25]. These types of systems are further divided up into subtypes, as these CRISPR-Cas systems are so diverse [40-45]. The genes associated with the CRISPR-Cas systems are usually in close proximity to each other, and consist of the interference, crRNA biogenesis, and adaptation cassettes, although some of the genes involved in the CRISPR-Cas system are located elsewhere in the genome [46, 47].

Pyrococcus furiosus and its CRISPR-Cas systems

Pyrococcus furiosus is a hyperthermophilic archaeon organism that thrives optimally at 100˚C. This species belongs to the order of Thermococcales, and lives in hydrothermal vents of the deep sea [46]. The strain of *Pyrococcus furiosus* that we are using as wildtype in our studies is the COM1 strain, which is naturally competent, and allows for easy genetic manipulation of the genome [63-65].

There are seven active CRISPR loci within the *Pyrococcus furiosus* genome [66]. All seven loci produce crRNAs and allow for efficient targeting of invading DNA or RNA [66-68]. These loci are of different lengths depending on how many spacers they

encode, with the longest being CRISPR 1 (51 spacers) and the shortest being CRISPR 8 (11 spacers) [69] (Figure 1.3).

Pyrococcus furiosus has three invader silencing systems, type I-A (Csa), type I-B (Cst), and type III-B (Cmr) [25, 67]. Type I-A and type I-B systems target invading DNA within the cell [25, 45, 67]. This is true with both the type I and type II systems. The type III-B system, Cmr, targets invading RNA and DNA [19, 68, 70]. Having the ability to target both RNA and DNA invaders has its evolutionary advantages, and makes our model organism more relevant, as we can use both DNA and RNA targets.

Adaptation in *Pyrococcus furiosus*

There are two universally conserved Cas proteins, Cas1 and Cas2. These two proteins have been shown to be necessary and sufficient in catalyzing integration of a protospacer into the CRISPR array [34, 35, 45, 71, 72]. In the absence of Cas1 or Cas2, integration does not occur [1, 34, 35]. In our specific system (*Pyrococcus furiosus*), there are a total of four known Cas proteins that play a role in adaptation: Cas1, Cas2, Cas4-1, and Cas4-2 [73]. The roles of Cas4-1 and Cas4-2 have been found to be involved in several steps of adaptation. These include trimming of the protospacer, identification of the PAM sequence, and correct orientation of new spacers being integrated into CRISPR arrays [47, 58, 74]. While these proteins do play major roles in adaptation, we have found that the presence of Cas1 and Cas2 is enough for integration of new spacers to take place, both *in vivo* and *in vitro* [47, 75]. However, there are still questions about how new spacers are integrated into the CRISPR array in a polarized manner *in vivo*. In my dissertation, I identify another protein involved in adaptation, a trans-acting (non-Cas)

factor, archaeal histones, whose typical housekeeping functions are not related to the CRISPR-Cas system.

Primed Adaptation

The initial protospacer acquisition and integration event during adaptation is referred to as naïve adaptation [5, 48]. This is the initial step, in which the CRISPR-Cas system acquires a spacer from the invading DNA and integrates it into the CRISPR locus [34-36] (Figure 1.1). Upon re-infection of the same invader, primed adaptation can take place.

Primed adaptation occurs when the CRISPR locus has a spacer perfectly (although if perfectly matching, the interference complex may degrade the invader too fast to allow for the acquisition of new spacers) or partially matching the invader (invaders are known to have mutations occur in the PAM or protospacer sequence that allow for escape from the CRISPR-Cas system)[49, 50]. The interference RNA protein complex takes crRNA to target and bind to the complementary sequence on the invader DNA. The DNA then forms a R loop, in which the crRNA within the interference complex is bound to the complementary DNA, and the other strand of DNA forms a "loop" when the other strand of DNA is displaced [51]. Cas3 (the nuclease of the interference complex in type I systems) then begins to degrade the displaced DNA that it is bound to, providing substrates for the adaptation complex to take and integrate into the

CRISPR loci [52, 53]. This allows for the new, perfect matching spacers to target the DNA and allow for efficient interference to take place.

Primed adaptation has been found to occur in several systems, including Type I-B in *Pyrococcus furiosus*, type I-B in *Haloarcula hispanica*, type I-C in *Legionella pneumophila*, the type I-E in *E. coli*, type I-E in *Thermobifida fusca*, type I-F in *Pectobacterium atrosepticum*, as well as a type II-A system [33, 49, 50, 54-59]. There has been several studies done to understand the requirements behind primed adaptation, and it has been found that the major nuclease of the CRISPR Cas system, Cas3 in type I systems, the immune effector complex, and the adaptation machinery of that system are essential for primed adaptation to take place [33, 49, 58, 60]. This connection between the adaptation and interference pathways in primed adaptation allows for the adaptation machinery to efficiently acquire new spacers from Cas3 degradation products. During this Cas3 degradation, the effector complex, Cas1 and Cas2, and Cas3 are found to associate with one another and one model proposes that all of these components move along the DNA from the initial protospacer site until a PAM is encountered and a new spacer is acquired. Another model suggests that the Cas1,Cas2, and Cas3 form their own complex and translocate along the DNA together until a new PAM is found [33, 52, 59, 61]. This degradation has been found to occur on both the non-target and target strands, although it observed primarily on the non-target strand in type I-E systems [52, 62]. The ever-standing fight between invaders and host organisms continues to play out, with this primed adaptation pathway evolving to help target and degrade invaders that have

mutations within the protospacer (targeted) region so that the CRISPR-Cas system can still fight effectively against this invader.

Spacer integration in the CRISPR array

The proper integration of new spacers at the CRISPR array is essential for successful downstream interference to take place. This requires the trimming of spacer sequences to the correct size, integration at the known repeat borders to successfully duplicate the repeat between spacers, and correct orientation of new spacers into the array (dependent on PAM motifs).

The Cas1-Cas2 integrase complex is found in the majority of CRISPR-Cas systems and integrates new spacers into the CRISPR array [40, 43, 45]. Cas1 has been found to have nuclease activity in *E. coli* on both single and double-stranded DNA, and RNA [76]. The nuclease active site of Cas1 was found to be metal-dependent [77]. Cas2 was also found to have nuclease activity against double-stranded DNA in *B. halodurans* [78]. However, the main function of Cas2 in the integrase complex has been demonstrated to be a scaffold for the dimers of Cas1 proteins, and its active site is not essential for integration of new spacers [1, 35, 79, 80].

Integration of new spacers occurs at the borders of the repeat in the CRISPR array. The Cas1-Cas2 integrase complex brings the new spacer (prespacer) to the CRISPR array, and the 3' hydroxyl ends of the spacer undergo two nucleophilic attacks at both borders of the repeat (the leader side of the repeat, and the spacer side of the repeat) [79, 81-83]. It has been demonstrated some Type I and Type II systems that the leader side of the repeat is attacked first, although in *P. furiosus* it has been found to be

promiscuous [35, 75, 81]. Whichever border of the repeat is targeted by the nucleophilic attack of the 3' hydroxyl of the spacer first, this is referred to as a half-site intermediate. When the other border of the repeat is subsequently attacked by the other 3' hydroxyl of the spacer, it effectively becomes a full site integration product. This causes the strands of the repeat to dissociate, and allows for the spacer to insert between the two strands of the repeat, and DNA polymerase (found to be essential in adaptation) fill in the gaps of the single stranded repeats, leading to the formation of a new repeat-spacer unit in the CRISPR array [35, 82, 84].

CRISPR Leaders

The leader is a key element of every CRISPR locus. It determines where the integration of new spacers occur *in vivo* [1, 34, 37, 38]. The leader contains the promoter region that allows for the binding of the transcription factors, which then leads to the transcription of the CRISPR array [46]. This leads to the biogenesis of crRNAs. It was found that the spacers in closer proximity to the leader were more abundant as crRNAs in the cell than those spacers more distant from the leader [66]. These "surveillance" RNAs are probably more abundant the closer to the leader due to the recent infection from the invader and are more likely to provide immunity than the older spacers. The polarization of the spacer was found to be important to the survival of the cell, as it provides a more robust response against the most recent invader [39].

The leaders vary in size, and can be as small as 47 nucleotides in bacteria, and up to 500 nucleotides in archaea [37]. It has been found, in *E. coli*, that the 40-60 nucleotides of the (total leader examined was 90 base-pairs) leader, upstream of the first repeat, is

essential for the integration of new spacers [37]. When the leader of *Sulfolobus* was deleted from the -47 to the -70 positions of the (total leader examined was 230 base-pairs) leader, in relation to the first repeat, it decreased the specificity of spacer integration at the repeat most proximal to the leader and instead integrated at other repeats of the CRISPR array, and also lowered adaptation overall [37].

The leader sequence across prokaryotes has a generally low level of sequence conservation, especially among bacterial species [37]. This makes it hard to define the leader, as low conservation occurs even across similar species of archaea and bacteria [37]. The hyperthermophilic archaea have higher levels of conservation amongst related species, with the explanation being that due the extreme environment that the archaea are more isolated and there is less biodiversity among them [37]. In euryarchaeal Pyrococcales, there are high levels of conservation among the *Pyrococcus* genus in terms of their leaders [46] (Figure 1.4). There are small differences between the leaders, but the BRE/TATA elements are conserved throughout [46], as well as the nucleotide directly adjacent to the first direct repeat [46] (Figure 1.4). It is important to note that the leaders in these archaeal species are upwards of 500 base-pairs long, and just the repeat proximal region of the leader in the *Pyrococcus* genus is shown in Figure 1.4. Specifically in *Pyrococcus furiosus*, the leaders are around 450 base-pairs in size and have many conserved regions throughout, some examples including the transcriptional start site, the BRE/TATA elements, and the initial two nucleotides that are adjacent to the first repeat (Figure 1.4) [66, 85].

Archaeal Mobile Genetic Elements

Although there are no currently known natural invaders for *Pyrococcus furiosus*, there is literature describing MGEs for a variety of archaea. There are many viral invaders known, including those specific to archaea or ones that are more cosmopolitan [86]. Currently all archaeal viruses have DNA genomes (single-stranded, double-stranded, linear, circular), though some RNA viruses were detected through metagenomic approaches (however, their host is uknown) [87-89].

Host response against these viral invaders is important to the survival of the host and the virus, and it was found that host-virus interactions varied significantly among archaeal viruses. For example, the SSV1 (Sulfolobus spindle-shaped virus 1) infection, only a few of the host genes were differentially expressed [86, 90]. In contrast, SSV2 elicited a more robust host response, which included upregulation of the CRIPSR loci and corresponding *cas* genes [86, 91]. This was also the case with SIRV2 (*Sulfolobus islandicus* rod-shaped virus 2) and STSV2 (Sulfolobus tengchongensis spindle-shaped virus 2) infections, both inducing transcriptional activation of genes involved in antiviral defense which included CRISPR-Cas systems as well as toxin-antitoxin systems [86, 92, 93].

The Sulfolobaceae family are the only members of archaea known to have conjugative plasmids in published literature. There are three major regions of genes found on archaeal conjugative plasmids; an origin of replication, genes that are involved with plasmid replication and an integrase, and a region that contains genes involved in conjugation [94-96]. In this region with genes pertaining to conjugation, there are genes that have been found to have homology to the ATPase VirB4 and the coupling protein

VirD4 of bacterial type IV secretion systems (T4SSs) [94, 97-99]. There is also a protein thought to be involved with the formation of the translocation pore for DNA transfer, composed of 10-12 transmembrane helices [94, 96]. In the currently published work, there are a lack of homologues of conjugative relaxases or other DNA transfer and replication proteins, with the exception of VirB4 and VirD4 [94, 96]. This is odd in that archaeal conjugative plasmids are self-transmissible, and these components that are unknown in archaea play essential roles in bacterial conjugation systems.

While the mechanisms behind the act of conjugation in archaea is still a mystery, we do know that cell to cell contact is important. There is an OriT (origin of transfer) present on the conjugative plasmid, which is nicked by a relaxase, which then protects the exposed end of DNA, and that is guided to the Type IV-like secretion system by the coupling protein VirD4 to exit the donor cell and enter the recipient cell [96, 100-105].

In my dissertation, I will be describing experiments using a recently discovered archaeal conjugative plasmid from *Thermococcus 33.3*. This is one of the first examples of an archaeal conjugative plasmid outside of the Sulfolobaceae family. This allows for us to study a natural MGE in *Pyrococcus furiosus*, and the response of the CRISPR-Cas system against this element.

Trans-acting factors involved in adaptation

Known Cas proteins have been studied and characterized to elucidate their role in adaptation. Cas1 and Cas2 form an integrase complex to catalyze integration into the leader end of the CRISPR array [35, 36]. Additional Cas proteins were found to be

essential for adaptation in other systems, for example, Cas9 in type II-A systems [56, 106].

It has also been found that other factors not involved with the CRISPR-Cas system played a role in adaptation as well. Integration Host Factor (IHF) was found to play an essential role in adaptation in *E. coli* [1, 80, 107, 108]. IHF binds and bends the DNA of the leader to allow for proper DNA topology to allow for the Cas1 and Cas2 integrase complex to bind and catalyze integration at the repeat most proximal to the leader [1, 108]. Specifically, IHF bends the leader by about 120°, causing the Cas1-Cas2 integration complex to be brought closer to upstream leader motifs [80, 109]. While IHF bends the DNA of the leader, it has also been found to have direct interactions with Cas1 as well [80]. IHF was found to be necessary for polarized integration at the repeat adjacent to the leader *in vitro* and *in vivo* [1]. The published literature has not yet identified other mechanisms that promote leader-proximal integration of spacers as well in Type I systems. In my dissertation, I do demonstrate that the archaeal histones A and B in *Pyrococcus furiosus* have been found to influence integration at the leader adjacent repeat at CRISPR arrays (Chapter 3).

DNA repair and genome stability proteins have been demonstrated to play roles in both naïve and primed adaptation [23, 84]. The RecBCD helicase and nuclease complex was found to be a key player in the acquisition of new spacers due to its processing of double stranded DNA breaks [23]. RecBCD helps repair double stranded DNA breaks by going to the exposed end of the break, and unwinding the DNA and processing the DNA until it reaches a Chi (Crossover Hotspot Instigator) sequence site that causes it to stall [23]. These Chi sites and replication fork stalls (which frequently occur at sites of double

stranded DNA breaks) were found to be hotspots for protospacer acquisition, and it was hypothesized that Cas1 and Cas2 takes these protospacers from the degradation products from the RecBCD activity [23]. Studies also found that some proteins were important for primed adaptation (RecG helicase and PriA protein in *E. coli*), and others were important for both primed and naïve adaptation (DNA polymerase I in *E. coli*) [84]. Sites of replication fork stalls and double stranded DNA breaks or nicks of the DNA all seem to be sites of active protospacer acquisition [23]. While these specific proteins may not be present in the archaeal species, there are equivalents. For example, NurA (Nuclease of Archaea) and HerA (Helicase of Archaea) of *Pyrococcus furiosus* are involved in the processing of double stranded DNA breaks to allow for homologous recombination to repair the DNA, very much like the RecBCD complex in *E. coli* [110-113]. Hef (helicase associated endonuclease for fork-structured DNA) in *Pyrococcus furiosus* is also a homolog of the RecG helicase, and has been predicted to play a role in resolving stalled replication forks [114, 115]. These trans-acting (non-Cas) protein equivalents present in *Pyrococcus furiosus* are ideal candidates to test for roles in adaptation in our system.

Histones

There are no known homologues for IHF in archaea. There are DNA binding proteins that are involved in manipulating DNA architecture. There are several of these in *Pyrococcus furiosus*, and these include TrmBL2, Alba, and histones A and B. TrmBL2 has been found to form stiff filaments and compete with histones for specific binding sites [116-118]. Alba (Acetylation Lowers Binding Affinity) also forms stiff filaments and cross bridges between DNA, and plays a role in the regulation of genes [118].

Histones have been found to regularly chromatize the genome in archaea, as well as being involved in the expression of certain genes and has been found to be involved with the transformation of DNA in *Thermococcus kodakarensis* [117, 119-121].

Most archaea, with the exception of some crenarchaea, have histones present (Figure 1.5) [122, 123]. They have been found to have similar properties to eukaryotic histones, and crystal structures demonstrate how they interact with the DNA in a very similar fashion to each other (Figure 1.6). Histone folds are highly conserved amongst eukaryotes and archaeal histones (Figure 1.6B, Figure 1.6C). Eukaryotic histones form a histone octamer (histones H2A, H2B, H3, and H4) [124]. Archaeal histones are found to form dimers in solution, and can form either heterodimers or homodimers [119]. Eukaryotic histones have long N terminal tails that can undergo posttranslational modifications and also lead to tighter DNA packaging, whereas most archaeal histones do not have these [123].

A unique feature of archaeal histones includes the fact that they can form these "endless" hypernucleosomes, where they can continuously wrap DNA and assemble into an endless left-handed rod and can wrap DNA from 30-300 bp in length [123, 125]. Eukaryotic histones cannot form these assemblies and wrap around 147 bp of DNA.

We have found in *Pyrococcus furiosus* that micrococcal nuclease assays also revealed 30-180 bp assemblies of DNA protection, similar to that of other studied archaea. Upon closer examination of these regions of protection, we found that DNA protection at the CRISPR leader, revealed two major peaks, each about 60 bp in length. This is the same size bound by two histone dimers. Leaders in *Pyrococcus furiosus* are extremely conserved with each other (Figure 1.4), and one peak was about 100 bp

upstream of the first repeat in the leader sequence and the other spanned the leader repeat junction, both underlying DNA sequences in these regions were preferable for histone binding. This led us to the hypothesis that histones may be involved with the integration of new spacers in the CRISPR array (Chapter 3).

In this dissertation, we will explore how adaptation is influenced by various determinants. Through testing of a natural invader, we discovered a robust adaptation response against it, and found that it was undergoing a specialized primed adaptation pathway against the plasmid. Additionally, we identified archaeal histones as playing an important role in the polarization of new spacers into the CRISPR array, demonstrating a new mechanism of integration in an archaeal system.

References

1. Nunez, J.K., et al., *CRISPR Immunological Memory Requires a Host Factor for Specificity.* Mol Cell, 2016. **62**(6): p. 824-33.
2. Barrangou, R., et al., *CRISPR provides acquired resistance against viruses in prokaryotes.* Science, 2007. **315**(5819): p. 1709-12.
3. Jansen, R., et al., *Identification of genes that are associated with DNA repeats in prokaryotes.* Mol Microbiol, 2002. **43**(6): p. 1565-75.
4. Terns, M.P. and R.M. Terns, *CRISPR-based adaptive immune systems.* Curr Opin Microbiol, 2011. **14**(3): p. 321-7.
5. Abedon, S.T., *Facilitation of CRISPR adaptation.* Bacteriophage, 2011. **1**(3): p. 179-181.
6. Barrangou, R., *CRISPR-Cas systems and RNA-guided interference.* Wiley Interdiscip Rev RNA, 2013. **4**(3): p. 267-78.
7. Barrangou, R., *The roles of CRISPR-Cas systems in adaptive immunity and beyond.* Curr Opin Immunol, 2015. **32**: p. 36-41.
8. Barrangou, R. and L.A. Marraffini, *CRISPR-Cas systems: Prokaryotes upgrade to adaptive immunity.* Mol Cell, 2014. **54**(2): p. 234-44.
9. Terns, R.M. and M.P. Terns, *CRISPR-based technologies: prokaryotic defense weapons repurposed.* Trends Genet, 2014. **30**(3): p. 111-8.
10. Carte, J., et al., *Binding and cleavage of CRISPR RNA by Cas6.* RNA, 2010. **16**(11): p. 2181-8.

11. Carte, J., et al., *Cas6 is an endoribonuclease that generates guide RNAs for invader defense in prokaryotes.* Genes Dev, 2008. **22**(24): p. 3489-96.

12. Deltcheva, E., et al., *CRISPR RNA maturation by trans-encoded small RNA and host factor RNase III.* Nature, 2011. **471**(7340): p. 602-7.

13. Li, M., et al., *Characterization of CRISPR RNA biogenesis and Cas6 cleavage-mediated inhibition of a provirus in the haloarchaeon Haloferax mediterranei.* J Bacteriol, 2013. **195**(4): p. 867-75.

14. Niewoehner, O., M. Jinek, and J.A. Doudna, *Evolution of CRISPR RNA recognition and processing by Cas6 endonucleases.* Nucleic Acids Res, 2014. **42**(2): p. 1341-53.

15. Almendros, C. and F.J. Mojica, *Exploring CRISPR Interference by Transformation with Plasmid Mixtures: Identification of Target Interference Motifs in Escherichia coli.* Methods Mol Biol, 2015. **1311**: p. 161-70.

16. Bhaya, D., M. Davison, and R. Barrangou, *CRISPR-Cas systems in bacteria and archaea: versatile small RNAs for adaptive defense and regulation.* Annu Rev Genet, 2011. **45**: p. 273-97.

17. Bikard, D., et al., *CRISPR interference can prevent natural transformation and virulence acquisition during in vivo bacterial infection.* Cell Host Microbe, 2012. **12**(2): p. 177-86.

18. Elmore, J.R., et al., *Programmable plasmid interference by the CRISPR-Cas system in Thermococcus kodakarensis.* RNA Biol, 2013. **10**(5): p. 828-40.

19. Hale, C.R., et al., *RNA-guided RNA cleavage by a CRISPR RNA-Cas protein complex.* Cell, 2009. **139**(5): p. 945-56.

20. Marraffini, L.A. and E.J. Sontheimer, *CRISPR interference: RNA-directed adaptive immunity in bacteria and archaea.* Nat Rev Genet, 2010. **11**(3): p. 181-90.

21. Marraffini, L.A. and E.J. Sontheimer, *Self versus non-self discrimination during CRISPR RNA-directed immunity.* Nature, 2010. **463**(7280): p. 568-71.

22. Semenova, E., et al., *Interference by clustered regularly interspaced short palindromic repeat (CRISPR) RNA is governed by a seed sequence.* Proc Natl Acad Sci U S A, 2011. **108**(25): p. 10098-103.

23. Levy, A., et al., *CRISPR adaptation biases explain preference for acquisition of foreign DNA.* Nature, 2015. **520**(7548): p. 505-10.

24. Anders, C., et al., *Structural basis of PAM-dependent target DNA recognition by the Cas9 endonuclease.* Nature, 2014.

25. Elmore, J., et al., *DNA targeting by the type I-G and type I-A CRISPR-Cas systems of Pyrococcus furiosus.* Nucleic Acids Res, 2015.

26. Leenay, R.T. and C.L. Beisel, *Deciphering, communicating, and engineering the CRISPR PAM.* J Mol Biol, 2016.

27. Li, M., R. Wang, and H. Xiang, *Haloarcula hispanica CRISPR authenticates PAM of a target sequence to prime discriminative adaptation.* Nucleic Acids Res, 2014. **42**(11): p. 7226-35.

28. Westra, E.R., et al., *Type I-E CRISPR-cas systems discriminate target from non-target DNA through base pairing-independent PAM recognition.* PLoS Genet, 2013. **9**(9): p. e1003742.

29. Wu, D., et al., *Structural basis of stringent PAM recognition by CRISPR-C2c1 in complex with sgRNA.* Cell Res, 2017.

30. Shiimori, M., et al., *Role of free DNA ends and protospacer adjacent motifs for CRISPR DNA uptake in Pyrococcus furiosus.* Nucleic Acids Res, 2017. **45**(19): p. 11281-11294.

31. Yosef, I., M.G. Goren, and U. Qimron, *Proteins and DNA elements essential for the CRISPR adaptation process in Escherichia coli.* Nucleic Acids Res, 2012. **40**(12): p. 5569-76.

32. Stoll, B., et al., *Requirements for a successful defence reaction by the CRISPR-Cas subtype I-B system.* Biochem Soc Trans, 2013. **41**(6): p. 1444-8.

33. Richter, C., et al., *Priming in the Type I-F CRISPR-Cas system triggers strand-independent spacer acquisition, bi-directionally from the primed protospacer.* Nucleic Acids Res, 2014.

34. Moch, C., et al., *DNA binding specificities of Escherichia coli Cas1-Cas2 integrase drive its recruitment at the CRISPR locus.* Nucleic Acids Res, 2016.

35. Nunez, J.K., et al., *Cas1-Cas2 complex formation mediates spacer acquisition during CRISPR-Cas adaptive immunity.* Nat Struct Mol Biol, 2014. **21**(6): p. 528-34.

36. Rollie, C., et al., *Intrinsic sequence specificity of the Cas1 integrase directs new spacer acquisition.* Elife, 2015. **4**.

37. Alkhnbashi, O.S., et al., *Characterizing leader sequences of CRISPR loci.* Bioinformatics, 2016. **32**(17): p. i576-i585.

38. Wei, Y., et al., *Sequences spanning the leader-repeat junction mediate CRISPR adaptation to phage in Streptococcus thermophilus.* Nucleic Acids Res, 2015. **43**(3): p. 1749-58.

39. McGinn, J. and L.A. Marraffini, *CRISPR-Cas Systems Optimize Their Immune Response by Specifying the Site of Spacer Integration.* Mol Cell, 2016.

40. Makarova, K.S., et al., *Evolutionary classification of CRISPR-Cas systems: a burst of class 2 and derived variants.* Nat Rev Microbiol, 2020. **18**(2): p. 67-83.

41. Chylinski, K., et al., *Classification and evolution of type II CRISPR-Cas systems.* Nucleic Acids Res, 2014. **42**(10): p. 6091-105.

42. Haft, D.H., et al., *A guild of 45 CRISPR-associated (Cas) protein families and multiple CRISPR/Cas subtypes exist in prokaryotic genomes.* PLoS Comput Biol, 2005. **1**(6): p. e60.

43. Makarova, K.S., et al., *Evolution and classification of the CRISPR-Cas systems.* Nat Rev Microbiol, 2011. **9**(6): p. 467-77.

44. Makarova, K.S. and E.V. Koonin, *Annotation and Classification of CRISPR-Cas Systems.* Methods Mol Biol, 2015. **1311**: p. 47-75.

45. Makarova, K.S., et al., *An updated evolutionary classification of CRISPR-Cas systems.* Nat Rev Microbiol, 2015. **13**(11): p. 722-36.

46. Norais, C., et al., *Diversity of CRISPR systems in the euryarchaeal Pyrococcales.* RNA Biol, 2013. **10**(5): p. 659-70.

47. Shiimori, M., et al., *Cas4 Nucleases Define the PAM, Length, and Orientation of DNA Fragments Integrated at CRISPR Loci.* Mol Cell, 2018. **70**(5): p. 814-824 e6.

48. Staals, R.H., et al., *Interference-driven spacer acquisition is dominant over naive and primed adaptation in a native CRISPR-Cas system.* Nat Commun, 2016. **7**: p. 12853.

49. Datsenko, K.A., et al., *Molecular memory of prior infections activates the CRISPR/Cas adaptive bacterial immunity system.* Nat Commun, 2012. **3**: p. 945.

50. Swarts, D.C., et al., *CRISPR interference directs strand specific spacer acquisition.* PLoS One, 2012. **7**(4): p. e35888.

51. Ivancic-Bace, I., J.A. Howard, and E.L. Bolt, *Tuning in to interference: R-loops and cascade complexes in CRISPR immunity.* J Mol Biol, 2012. **422**(5): p. 607-16.

52. Kunne, T., et al., *Cas3-Derived Target DNA Degradation Fragments Fuel Primed CRISPR Adaptation.* Mol Cell, 2016. **63**(5): p. 852-64.

53. Ramachandran, A. and S. Bailey, *Memory Upgrade: Insights into Primed Adaptation by CRISPR-Cas Immune Systems.* Mol Cell, 2016. **64**(4): p. 641-642.

54. Nussenzweig, P.M., J. McGinn, and L.A. Marraffini, *Cas9 Cleavage of Viral Genomes Primes the Acquisition of New Immunological Memories.* Cell Host Microbe, 2019. **26**(4): p. 515-526 e6.

55. Fineran, P.C., et al., *Degenerate target sites mediate rapid primed CRISPR adaptation.* Proc Natl Acad Sci U S A, 2014. **111**(16): p. E1629-38.

56. Li, M., et al., *Adaptation of the Haloarcula hispanica CRISPR-Cas system to a purified virus strictly requires a priming process.* Nucleic Acids Res, 2014. **42**(4): p. 2483-92.

57. Rao, C., D. Chin, and A.W. Ensminger, *Priming in a permissive type I-C CRISPR-Cas system reveals distinct dynamics of spacer acquisition and loss.* RNA, 2017. **23**(10): p. 1525-1538.

58. Garrett, S., et al., *Primed CRISPR DNA uptake in Pyrococcus furiosus.* Nucleic Acids Res, 2020.

59. Dillard, K.E., et al., *Assembly and Translocation of a CRISPR-Cas Primed Acquisition Complex.* Cell, 2018. **175**(4): p. 934-946 e15.

60. Vorontsova, D., et al., *Foreign DNA acquisition by the I-F CRISPR-Cas system requires all components of the interference machinery.* Nucleic Acids Res, 2015. **43**(22): p. 10848-60.

61. Redding, S., et al., *Surveillance and Processing of Foreign DNA by the Escherichia coli CRISPR-Cas System.* Cell, 2015. **163**(4): p. 854-65.

62. Semenova, E., et al., *Highly efficient primed spacer acquisition from targets destroyed by the Escherichia coli type I-E CRISPR-Cas interfering complex.* Proc Natl Acad Sci U S A, 2016.

63. Bridger, S.L., et al., *Genome sequencing of a genetically tractable Pyrococcus furiosus strain reveals a highly dynamic genome.* J Bacteriol, 2012. **194**(15): p. 4097-106.

64. Farkas, J., et al., *Recombinogenic properties of Pyrococcus furiosus strain COM1 enable rapid selection of targeted mutants.* Appl Environ Microbiol, 2012. **78**(13): p. 4669-76.

65. Lipscomb, G.L., et al., *Natural competence in the hyperthermophilic archaeon Pyrococcus furiosus facilitates genetic manipulation: construction of markerless*

deletions of genes encoding the two cytoplasmic hydrogenases. Appl Environ Microbiol, 2011. **77**(7): p. 2232-8.

66. Hale, C.R., et al., *Essential features and rational design of CRISPR RNAs that function with the Cas RAMP module complex to cleave RNAs.* Mol Cell, 2012. **45**(3): p. 292-302.

67. Majumdar, S., et al., *Three CRISPR-Cas immune effector complexes coexist in Pyrococcus furiosus.* RNA, 2015.

68. Elmore, J.R., et al., *Bipartite recognition of target RNAs activates DNA cleavage by the Type III-B CRISPR-Cas system.* Genes Dev, 2016. **30**(4): p. 447-59.

69. Terns, R.M. and M.P. Terns, *The RNA- and DNA-targeting CRISPR-Cas immune systems of Pyrococcus furiosus.* Biochem Soc Trans, 2013. **41**(6): p. 1416-21.

70. Hale, C.R., et al., *Target RNA capture and cleavage by the Cmr type III-B CRISPR-Cas effector complex.* Genes Dev, 2014. **28**(21): p. 2432-43.

71. Musharova, O., et al., *Spacer-length DNA intermediates are associated with Cas1 in cells undergoing primed CRISPR adaptation.* Nucleic Acids Res, 2017.

72. Nunez, J.K., et al., *Foreign DNA capture during CRISPR-Cas adaptive immunity.* Nature, 2015.

73. Shiimori, M., et al., *Role of free DNA ends and protospacer adjacent motifs for CRISPR DNA uptake in Pyrococcus furiosus.* Nucleic Acids Res, 2017.

74. Kieper, S.N., et al., *Cas4 Facilitates PAM-Compatible Spacer Selection during CRISPR Adaptation.* Cell Rep, 2018. **22**(13): p. 3377-3384.

75. Grainy, J., et al., *CRISPR repeat sequences and relative spacing specify DNA integration by Pyrococcus furiosus Cas1 and Cas2.* Nucleic Acids Res, 2019.

76. Babu, M., et al., *A dual function of the CRISPR-Cas system in bacterial antivirus immunity and DNA repair.* Mol Microbiol, 2011. **79**(2): p. 484-502.

77. Wiedenheft, B., et al., *Structural basis for DNase activity of a conserved protein implicated in CRISPR-mediated genome defense.* Structure, 2009. **17**(6): p. 904-12.

78. Nam, K.H., et al., *Double-stranded endonuclease activity in Bacillus halodurans clustered regularly interspaced short palindromic repeats (CRISPR)-associated Cas2 protein.* J Biol Chem, 2012. **287**(43): p. 35943-52.

79. Xiao, Y., et al., *How type II CRISPR-Cas establish immunity through Cas1-Cas2-mediated spacer integration.* Nature, 2017. **550**(7674): p. 137-141.

80. Wright, A.V., et al., *Structures of the CRISPR genome integration complex.* Science, 2017. **357**(6356): p. 1113-1118.

81. Kim, J.G., et al., *CRISPR DNA elements controlling site-specific spacer integration and proper repeat length by a Type II CRISPR-Cas system.* Nucleic Acids Res, 2019. **47**(16): p. 8632-8648.

82. Nunez, J.K., et al., *Integrase-mediated spacer acquisition during CRISPR-Cas adaptive immunity.* Nature, 2015. **519**(7542): p. 193-8.

83. *<16-Wright-Protecting genome integrity.pdf>.*

84. Ivancic-Bace, I., et al., *Different genome stability proteins underpin primed and naive adaptation in E. coli CRISPR-Cas immunity.* Nucleic Acids Res, 2015. **43**(22): p. 10821-30.

85. Phok, K., et al., *Identification of CRISPR and riboswitch related RNAs among novel noncoding RNAs of the euryarchaeon Pyrococcus abyssi.* BMC Genomics, 2011. **12**: p. 312.

86. Krupovic, M., et al., *Viruses of archaea: Structural, functional, environmental and evolutionary genomics.* Virus Res, 2018. **244**: p. 181-193.

87. Prangishvili, D., et al., *The enigmatic archaeal virosphere.* Nat Rev Microbiol, 2017. **15**(12): p. 724-739.

88. Bolduc, B., et al., *Viral assemblage composition in Yellowstone acidic hot springs assessed by network analysis.* ISME J, 2015. **9**(10): p. 2162-77.

89. Bolduc, B., et al., *Identification of novel positive-strand RNA viruses by metagenomic analysis of archaea-dominated Yellowstone hot springs.* J Virol, 2012. **86**(10): p. 5562-73.

90. Frols, S., et al., *Elucidating the transcription cycle of the UV-inducible hyperthermophilic archaeal virus SSV1 by DNA microarrays.* Virology, 2007. **365**(1): p. 48-59.

91. Fusco, S., et al., *Transcriptome analysis of Sulfolobus solfataricus infected with two related fuselloviruses reveals novel insights into the regulation of CRISPR-Cas system.* Biochimie, 2015.

92. Quax, T.E., et al., *Massive activation of archaeal defense genes during viral infection.* J Virol, 2013. **87**(15): p. 8419-28.

93. Leon-Sobrino, C., W.P. Kot, and R.A. Garrett, *Transcriptome changes in STSV2-infected Sulfolobus islandicus REY15A undergoing continuous CRISPR spacer acquisition.* Mol Microbiol, 2016. **99**(4): p. 719-28.

94. Greve, B., et al., *Genomic comparison of archaeal conjugative plasmids from Sulfolobus.* Archaea, 2004. **1**(4): p. 231-9.

95. She, Q., B. Shen, and L. Chen, *Archaeal integrases and mechanisms of gene capture.* Biochem Soc Trans, 2004. **32**(Pt 2): p. 222-6.

96. Wagner, A., et al., *Mechanisms of gene flow in archaea.* Nat Rev Microbiol, 2017. **15**(8): p. 492-501.

97. Erauso, G., et al., *Two novel conjugative plasmids from a single strain of Sulfolobus.* Microbiology (Reading), 2006. **152**(Pt 7): p. 1951-1968.

98. Basta, T., et al., *Novel archaeal plasmid pAH1 and its interactions with the lipothrixvirus AFV1.* Mol Microbiol, 2009. **71**(1): p. 23-34.

99. She, Q., et al., *Genetic profile of pNOB8 from Sulfolobus: the first conjugative plasmid from an archaeon.* Extremophiles, 1998. **2**(4): p. 417-25.

100. Guglielmini, J., et al., *Key components of the eight classes of type IV secretion systems involved in bacterial conjugation or protein secretion.* Nucleic Acids Res, 2014. **42**(9): p. 5715-27.

101. Ilangovan, A., S. Connery, and G. Waksman, *Structural biology of the Gram-negative bacterial conjugation systems.* Trends Microbiol, 2015. **23**(5): p. 301-10.

102. Alvarez-Martinez, C.E. and P.J. Christie, *Biological diversity of prokaryotic type IV secretion systems.* Microbiol Mol Biol Rev, 2009. **73**(4): p. 775-808.

103. Cabezon, E., et al., *Towards an integrated model of bacterial conjugation.* FEMS Microbiol Rev, 2015. **39**(1): p. 81-95.

104. Christie, P.J., N. Whitaker, and C. Gonzalez-Rivera, *Mechanism and structure of the bacterial type IV secretion systems.* Biochim Biophys Acta, 2014. **1843**(8): p. 1578-91.

105. Chandran Darbari, V. and G. Waksman, *Structural Biology of Bacterial Type IV Secretion Systems.* Annu Rev Biochem, 2015. **84**: p. 603-29.

106. Wei, Y., R.M. Terns, and M.P. Terns, *Cas9 function and host genome sampling in Type II-A CRISPR-Cas adaptation.* Genes Dev, 2015. **29**(4): p. 356-61.

107. Wei, Y. and M.P. Terns, *CRISPR Outsourcing: Commissioning IHF for Site-Specific Integration of Foreign DNA at the CRISPR Array.* Mol Cell, 2016. **62**(6): p. 803-4.

108. Yoganand, K.N., et al., *Asymmetric positioning of Cas1-2 complex and Integration Host Factor induced DNA bending guide the unidirectional homing of protospacer in CRISPR-Cas type I-E system.* Nucleic Acids Res, 2016.

109. Lee, H. and D.G. Sashital, *Creating memories: molecular mechanisms of CRISPR adaptation.* Trends Biochem Sci, 2022.

110. Blackwood, J.K., et al., *Structural and functional insights into DNA-end processing by the archaeal HerA helicase-NurA nuclease complex.* Nucleic Acids Res, 2012. **40**(7): p. 3183-96.

111. Byrne, R.T., et al., *Molecular architecture of the HerA-NurA DNA double-strand break resection complex.* FEBS Lett, 2014. **588**(24): p. 4637-44.

112. Cheng, K., et al., *Biochemical and Functional Characterization of the NurA-HerA Complex from Deinococcus radiodurans.* J Bacteriol, 2015. **197**(12): p. 2048-61.

113. De Falco, M., et al., *NurA Is Endowed with Endo- and Exonuclease Activities that Are Modulated by HerA: New Insight into Their Role in DNA-End Processing.* PLoS One, 2015. **10**(11): p. e0142345.

114. Komori, K., et al., *Cooperation of the N-terminal Helicase and C-terminal endonuclease activities of Archaeal Hef protein in processing stalled replication forks.* J Biol Chem, 2004. **279**(51): p. 53175-85.

115. Nishino, T., et al., *Crystal structure and functional implications of Pyrococcus furiosus hef helicase domain involved in branched DNA processing.* Structure, 2005. **13**(1): p. 143-53.

116. Efremov, A.K., et al., *Transcriptional Repressor TrmBL2 from Thermococcus kodakarensis Forms Filamentous Nucleoprotein Structures and Competes with Histones for DNA Binding in a Salt- and DNA Supercoiling-dependent Manner.* J Biol Chem, 2015. **290**(25): p. 15770-15784.

117. Maruyama, H., et al., *Histone and TK0471/TrmBL2 form a novel heterogeneous genome architecture in the hyperthermophilic archaeon Thermococcus kodakarensis.* Mol Biol Cell, 2011. **22**(3): p. 386-98.

118. Wierer, S., et al., *TrmBL2 from Pyrococcus furiosus Interacts Both with Double-Stranded and Single-Stranded DNA.* PLoS One, 2016. **11**(5): p. e0156098.

119. Nishida, H. and T. Oshima, *Archaeal histone distribution is associated with archaeal genome base composition.* J Gen Appl Microbiol, 2017. **63**(1): p. 28-35.

120. Sandman, K. and J.N. Reeve, *Archaeal chromatin proteins: different structures but common function?* Curr Opin Microbiol, 2005. **8**(6): p. 656-61.

121. Cubonovaa, L., et al., *An archaeal histone is required for transformation of Thermococcus kodakarensis.* J Bacteriol, 2012. **194**(24): p. 6864-74.

122. Laursen, S.P., S. Bowerman, and K. Luger, *Archaea: The Final Frontier of Chromatin.* J Mol Biol, 2021. **433**(6): p. 166791.

123. Henneman, B., et al., *Structure and function of archaeal histones.* PLoS Genet, 2018. **14**(9): p. e1007582.

124. *<1997-Luger-Crystal structure of the nucleosome.pdf>*.

125. Mattiroli, F., et al., *Structure of histone-based chromatin in Archaea.* Science, 2017. **357**(6351): p. 609-612.

Figure 1.1. The CRISPR-Cas System.

A schematic demonstrating the CRISPR-Cas defense pathway. The initial invasion of a MGE is followed by the first step, adaptation. 1. **Adaptation** involves the acquisition of a spacer from an invading MGE and subsequent integration of that spacer into the CRISPR array. 2. The second step, **crRNA biogenesis**, is the transcription of the CRISPR array and the processing of this transcript into mature crRNAs. 3. The final step**, invader silencing**, is the association of the crRNA with an effector complex which will then guide it to the complementary protospacer to silence the invading MGE. Adapted from Terns et al, 2014 [9].

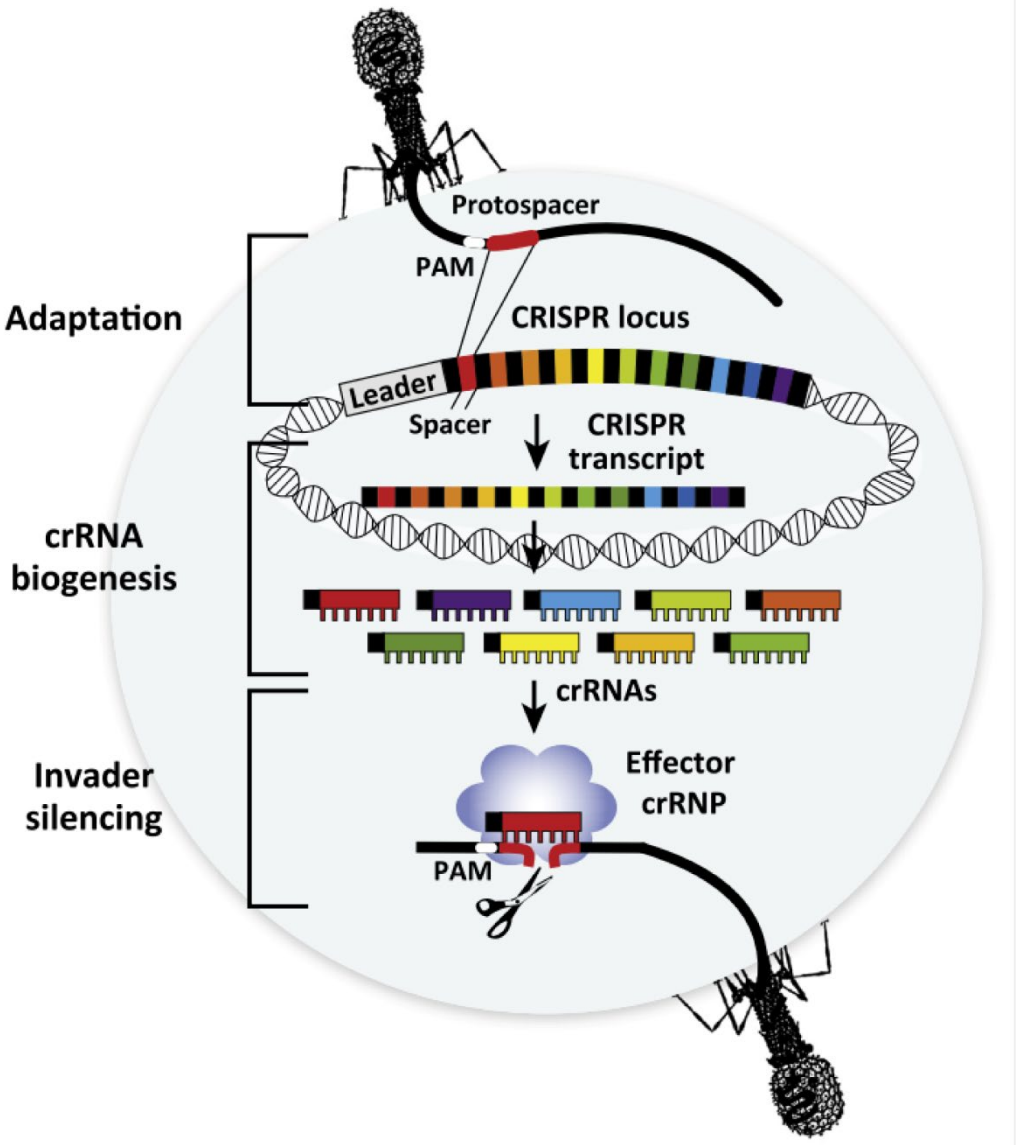

Protospacer
PAM
Adaptation
CRISPR locus
Leader
Spacer
CRISPR
transcript
crRNA
biogenesis
crRNAs
Effector
crRNP
Invader
silencing
PAM

Figure 1.2. The CRISPR array.

A CRISPR array is composed of three major parts. A leader (gray) containing a promoter allowing for transcription of the CRISPR array, repetitive sequences throughout the array (black repeats), and the various spacer sequences between the repeats.


Leader
R
S
R
S
R
S
R
S
R
S
R

Figure 1.3. The CRISPR-Cas system within *Pyrococcus furiosus*.

(A) Distribution of the CRISPR loci and Cas genes throughout the genome of Pyrococcus furiosus (Csa=Type I-A, Cst=Type I-B, Cmr=Type III-B). **(B)** The seven CRISPR loci of *Pyrococcus furiosus* (L=Leader, black boxes=repeats, variety of colors=spacers). The total number of spacers retained in each array are noted on the right. Adapted from Terns and Terns et al, 2013, and Shiimori et al, 2018 [47, 69].

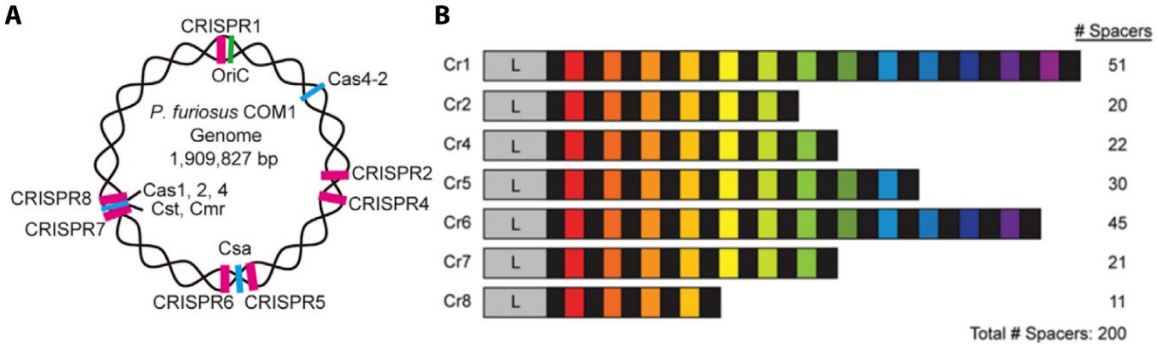

A
CRISPR1
OriC
Cas4-2
P. furiosus COM1
Genome
1,909,827 bp
CRISPR2
CRISPR8
Cas1, 2, 4
Cst, Cmr
CRISPR4
CRISPR7
Csa
CRISPR6
CRISPR5
B
Spacers
Cr1
L
51
Cr2
L
20
Cr4
L
22
Cr5
L
30
Cr6
L
45
Cr7
L
21
Cr8
L
11
Total # Spacers: 200

Figure 1.4. Aligned leader sequences for Pyrococcus species.

Red arrows indicate transcriptional start sites. DR1=Direct Repeat 1. Red letter indicate conserved sequences throughout all noted pyrococcal leader regions. Adapted from Norais et al, 2013 [46].

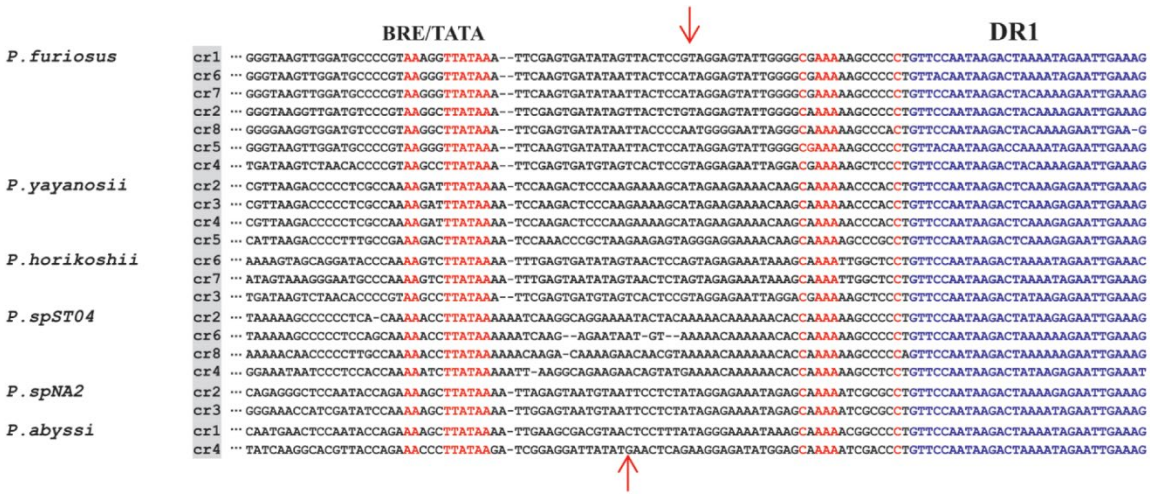

BRE/TATA
DR1
P.furiosus
cr1 ··· GGGTAAGTTGGATGCCCCGTAAAGGTTATAAA--TTCGAGTGATATAGTTACTCCGTAGGAGTATTGGGGCGAAAAAGCCCCCTGTTCCAATAAGACTAAAATAGAATTGAAAG
cr6 ··· GGGTAAGTTGGATGCCCCGTAAGGGTTATAAA--TTCAAGTGATATAATTACTCCATAGGAGTATTGGGGCGAAAAAGCCCCCTGTTACAATAAGACTAAAATAGAATTGAAAG
cr7 ··· GGGTAAGTTGGATGCCCCGTAAGGGTTATAAA--TTCAAGTGATATAATTACTCCATAGGAGTATTGGGGCGAAAAAGCCCCCTGTTCCAATAAGACTACAAAAGAATTGAAAG
cr2 ··· GGGTAAGGTTGATGTCCCGTAAGGCTTATAAA--TTCGAGTGATATAGTTACTCTGTAGGAGTATTGGGGCAAAAAAGCCCCCTGTTCCAATAAGACTACAAAAGAATTGAAAG
cr8 ··· GGGGAAGGTGGATGTCCCGTAAGGCTTATAAA--TTCGAGTGATATAATTACCCCAATGGGGAATTAGGGCAAAAAAGCCCACTGTTCCAATAAGACTACAAAAGAATTGAA-G
cr5 ··· GGGTAAGTTGGATGCCCCGTAAGGGTTATAAA--TTCAAGTGATATAATTACTCCATAGGAGTATTGGGGCGAAAAAGCCCCCTGTTACAATAAGACCAAAATAGAATTGAAAG
cr4 ··· TGATAAGTCTAACACCCCGTAAGCCTTATAAA--TTCGAGTGATGTAGTCACTCCGTAGGAGAATTAGGACGAAAAAGCTCCCTGTTCCAATAAGACTACAAAAGAATTGAAAG
P.yayanosii
cr2 ··· CGTTAAGACCCCCTCGCCAAAAGATTTATAAAA-TCCAAGACTCCCAAGAAAAGCATAGAAGAAAACAAGCAAAAAACCCACCTGTTCCAATAAGACTCAAAGAGAATTGAAAG
cr3 ··· CGTTAAGACCCCCTCGCCAAAAGATTTATAAAA-TCCAAGACTCCCAAGAAAAGCATAGAAGAAAACAAGCAAAAAACCCACCTGTTCCAATAAGACTCAAAGAGAATTGAAAG
cr4 ··· CGTTAAGACCCCCTCGCCAAAAGATTTATAAAA-TCCAAGACTCCCAAGAAAAGCATAGAAGAAAACAAGCAAAAAACCCACCTGTTCCAATAAGACTCAAAGAGAATTGAAAG
cr5 ··· CATTAAGACCCCTTTGCCGAAAGACTTATAAAA-TCCAAACCCGCTAAGAAGAGTAGGGAGGAAAACAAGCAAAAAGCCCGCCTGTTCCAATAAGACTCAAAGAGAATTGAAAG
P.horikoshii
cr6 ··· AAAAGTAGCAGGATACCCAAAAGTCTTATAAAA-TTTGAGTGATATAGTAACTCCAGTAGAGAAATAAAGCAAAATTGGCTCCTGTTCCAATAAGACTAAAATAGAATTGAAAC
cr7 ··· ATAGTAAAGGGAATGCCCAAAAGTCTTATAAAA-TTTGAGTAATATAGTAACTCTAGTAGAGAAATAAAGCAAAATTGGCTCCTGTTCCAATAAGACTAAAATAGAATTGAAAG
cr3 ··· TGATAAGTCTAACACCCCGTAAGCCTTATAAA--TTCGAGTGATGTAGTCACTCCGTAGGAGAATTAGGACGAAAAAGCTCCCTGTTCCAATAAGACTATAAGAGAATTGAAAG
P.spST04
cr2 ··· TAAAAAGCCCCCCTCA-CAAAAACCTTATAAAAAATCAAGGCAGGAAAATACTACAAAAACAAAAAACACCAAAAAAGCCCCCTGTTCCAATAAGACTATAAGAGAATTGAAAG
cr6 ··· TAAAAAGCCCCCTCCAGCAAAAACCTTATAAAAAATCAAG--AGAATAAT-GT--AAAAACAAAAAACACCAAAAAAGCCCCCTGTTCCAATAAGACTAAAAAAGAATTGAAAG
cr8 ··· AAAAACAACCCCCTTGCCAAAAACCTTATAAAAAACAAGA-CAAAAGAACAACGTAAAAACAAAAAACACCAAAAAAGCCCCCAGTTCCAATAAGACTAAAAAAGAATTGAAAG
cr4 ··· GGAAATAATCCCCTCCACCAAAAATCTTATAAAAAATT-AAGGCAGAAGAACAGTATGAAAACAAAAAACACCAAAAAAGCCTCCTGTTCCAATAAGACTATAAGAGAATTGAAAT
P.spNA2
cr2 ··· CAGAGGGCTCCAATACCAGAAAAGCTTATAAAA-TTAGAGTAATGTAATTCCTCTATAGGAGAAATAGAGCAAAAATCGCGCCTGTTCCAATAAGACTAAAAGAGAATTGAAAG
cr3 ··· GGGAAACCATCGATATCCAAAAAGCTTATAAAA-TTGGAGTAATGTAATTCCTCTATAGAGAAAATAGAGCAAAAATCGCGCCTGTTCCAATAAGACTAAAATAGAATTGAAAG
P.abyssi
cr1 ··· CAATGAACTCCAATACCAGAAAAGCTTATAAAA-TTGAAGCGACGTAACTCCTTTATAGGGAAAATAAAGCAAAAACGGCCCCTGTTCCAATAAGACTAAAATAGAATTGAAAG
cr4 ··· TATCAAGGCACGTTACCAGAAACCCTTATAAGA-TCGGAGGATTATATGAACTCAGAAGGAGATATGGAGCAAAAATCGACCCTGTTCCAATAAGACTAAAATAGAATTGAAAG

Figure 1.5. Distribution of histones throughout archaea.

Crystal structure of a histone dimer (medium and light green), and a phylogeny showing the distribution of histones among archaea. The gap in the histone bar indicates those that do not have histones, also indicated by the bold writing. Adapted from Laursen et al, 2020 [122].

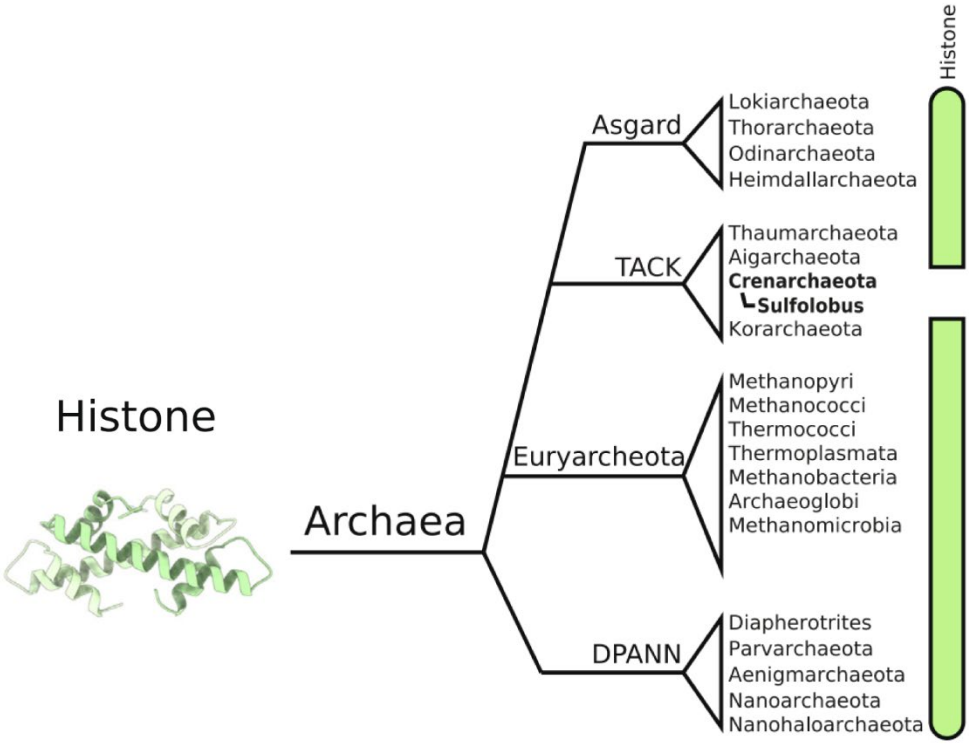
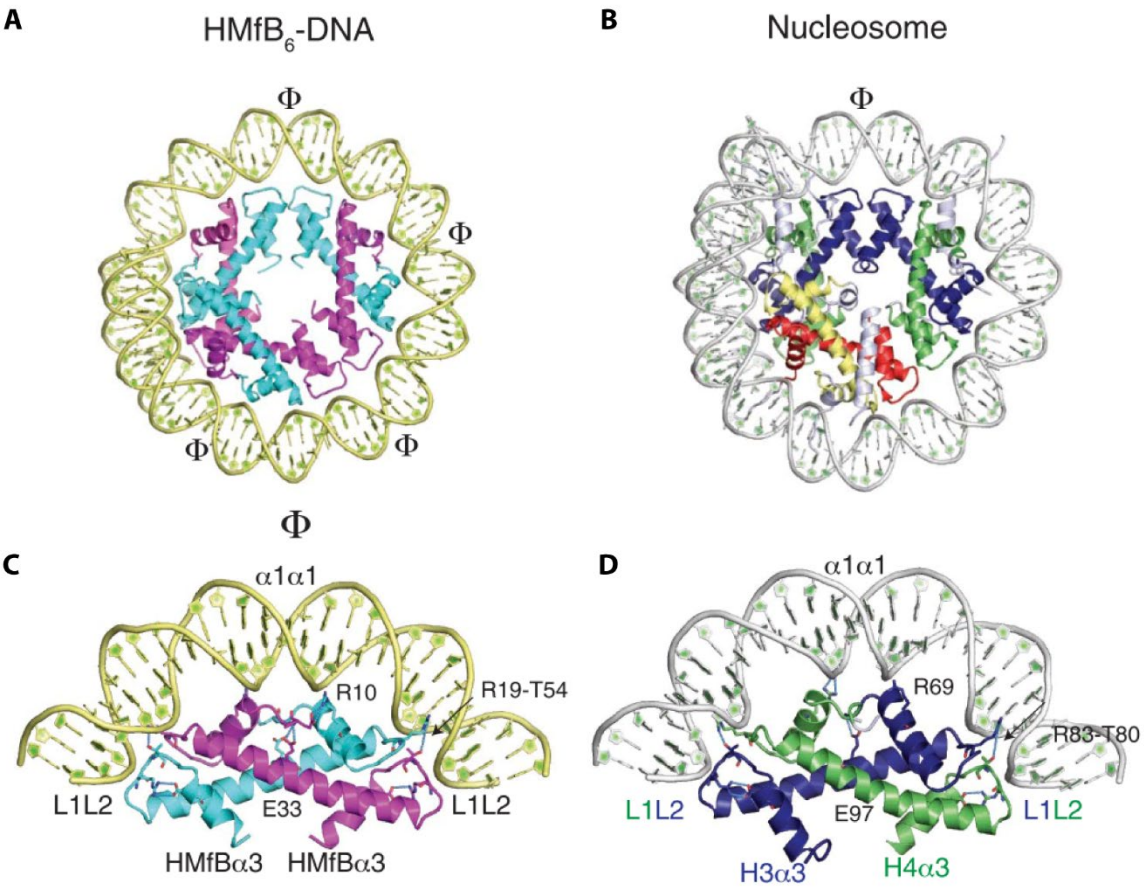

Histone
Histone
Archaea
Asgard
Lokiarchaeota
Thorarchaeota
Odinarchaeota
Heimdallarchaeota
TACK
Thaumarchaeota
Aigarchaeota
Crenarchaeota
Sulfolobus
Korarchaeota
Euryarcheota
Methanopyri
Methanococci
Thermococci
Thermoplasmata
Methanobacteria
Archaeoglobi
Methanomicrobia
DPANN
Diapherotrites
Parvarchaeota
Aenigmarchaeota
Nanoarchaeota
Nanohaloarchaeota

Figure 1.6. Comparison of archaeal and eukaryotic histones.

(A) Three histone homodimers of histone B from *Methanothermus fervidus* bound to 90 bp of DNA (4 Å crystal structure). **(B)** The nucleosome hexasome, created by removing one H2A-H2B heterodimer and the histone tails from the published structure. Φ Indicates axes of symmetry in both protein assemblies. **(C)** Histone folds (HFs) of one of the *Methanothermus fervidus* dimers, **(D)** and the H3-H4 heterodimer from the eukaryotic histone. Adapted from Mattiroli et al, 2018 [125].


A
HMfB6-DNA
B
Nucleosome
Φ
Φ
Φ
Φ
Φ
Φ
C
α1α1
R10
R19-T54
L1L2
E33
L1L2
HMfBα3
HMfBα3
D
α1α1
R69
R83-T80
L1L2
E97
L1L2
H3α3
H4α3

CHAPTER 2

HYPER-STIMULATION OF *PYROCOCCUS FURIOSUS* ADAPTATION BY A SELF-TRANSMISSIBLE PLASMID[1]

[1]Elizabeth A. Watts, Ryan Catchpole, Sandra Garrett, Landon M. Clark, Brenton G. Graveley, Michael P. Terns. *To be submitted to Biochemistry.*

Abstract

Pyrococcus furiosus is a hyperthermophilic archaeon with seven functional CRISPR loci and three effector crRNP complexes (types I-A, I-B, and III-B) that each employ crRNAs derived from all CRISPR arrays. Here, we investigate the CRISPR adaptation response to a newly discovered and self-transmissible plasmid from a *Thermococcus* strain. This large conjugative plasmid is transferred from *Thermococcus* species to *Pyrococcus furiosus* via mating induced by the plasmid. Transconjugate strains of *Pyrococcus furiosus* exhibited elevated levels of spacer integration at all CRISPR loci relative to the strain lacking the conjugative plasmid. High throughput sequence analysis demonstrated that the majority of new spacers were derived from the conjugative plasmid rather than the host genome. The new spacers were preferentially selected from DNA surrounding a particular region of the plasmid and exhibited a bi-directional pattern of strand bias that is a hallmark of primed adaptation by Type I systems. We found that one of the CRISPR arrays of our *Pyrococcus furiosus* lab strain encodes a spacer that closely matches the region of the conjugative plasmid that is targeted for adaptation. Our results indicate that *Pyrococcus furiosus* naturally encountered this conjugative plasmid or a related mobile genetic element in the past and responds to infection with robust primed adaptation.

Introduction

There is a constant battle for survival between prokaryotes and the viruses and other mobile genetic elements (MGEs) that infect them. As such, prokaryotes have evolved ways to defend themselves against parasitization by MGEs, and conversely,

MGEs have evolved mechanisms to evade these defenses. Host defense systems can be broadly classified in two categories: innate defense systems such as restriction-modification and toxin-antitoxin systems which act to non-specifically neutralize invaders; or adaptive defense systems that target specific, remembered invaders. Among prokaryotes, the CRISPR-Cas (Clustered Regularly Interspaced Short Palindromic Repeats-CRISPR associated) system is the only adaptive system identified so far [1-3]. CRISPR-Cas systems are encoded in 42.3% of sequenced bacterial genomes, and upwards of 85.2% of archaeal (96.7% of hyperthermophiles) genomes [4]. There is a diverse range of CRISPR-Cas systems discovered thus far, and these are divided into two classes delineated by the type of interference/effector complex genes they employ. These two classes are further divided into six types and over 30 subtypes [4]. For most CRISPR systems the effector complex genes are adjacent to a CRISPR array on the host genome (with some exceptions, like the Type IV-B system which does not have an array). The CRISPR array consists of a series of repeats with variable spacers, immediately downstream from a leader region which encodes a promoter for transcription of the array.

The immunity conferred by CRISPR-Cas systems is composed of three major steps [5, 6]. Adaptation, the first step, is the process by which DNA from an MGE is acquired and integrated in the CRISPR array, usually at the first (leader-adjacent) repeat. With every new spacer integration, the repeat is duplicated, resulting in a new repeat-spacer insert. In the next step, CRISPR RNA (crRNA) biogenesis, the CRISPR array is transcribed, and the resulting transcript is processed by either a Cas gene (Cas6 in Type I and Type III systems; Cas12/13 in Type V/VI) or other trans-acting factors (For example, RNase III in Type II systems) resulting in cleavage at repeat sequences [3, 7, 8]. These

immature crRNAs are further trimmed at the 3' end, with the final crRNA product referred to as a mature crRNA [8-10]. This mature crRNA then associates with a Cas effector complex to form a crRNA-ribonucleoprotein complex (crRNP). In the final step of immunity, interference, this crRNP binds nucleic acids complementary to the guide crRNA and destroys invaders from which spacers have been acquired [11-14].

In certain cases, the adaptation stage of CRISPR immunity can become somewhat iterative, being stimulated by interference activities employing existing spacers in a process known as primed adaptation [15, 16]. Here, a spacer acquired by naïve adaptation (the initial acquisition of a new spacer) to an invading MGE guides effector complexes to degrade invader genomes. The degradation products are then used as substrates by the adaptation proteins and incorporated as new spacers in the CRISPR array. Importantly, this mechanism can still operate if the original crRNA-target binding is weak (base-pairing mismatches are present), allowing rapid adaptation of systems to mutant MGEs or homologues of previously infecting elements [17-21].

While all steps of the CRISPR-Cas pathway are essential for effective defense, the adaptation stage is what makes CRISPR unique by allowing heritable adaptive immunity against invaders [22, 23]. Without the initial spacer to guide the effector complex to the invader, this whole process would be non-functional. Intriguingly, it remains difficult to identify targets of natural CRISPR spacers - studies have found that only 1-19% (depending on CRISPR-Cas system type/organism) of the spacers have corresponding protospacers (complementary region on the invading DNA from which the spacer was derived) in sequence databases [24]. Of those, 80-90% are against known

viral genomes, including proviruses, and the remainder target genes involved in plasmid conjugation and replication that had been inserted into microbial genomes [24].

The near ubiquity of CRISPR-Cas systems in Archaea suggests they must regularly encounter and defend themselves against numerous MGEs. Despite this, only ~60 viruses and ~60 plasmids have been identified from the *Crenarchaeota* and *Euryarchaeota* phylums [25]. Conjugative plasmids are even rarer in isolated archaeal strains, with all published isolates from a single family of Crenarchaea – the Sulfolobaceae [26]. Archaea within this group have been extensively studied in regards to their CRISPR-Cas systems, especially *Saccharolobus solfataricus* and *Sulfolobus islandicus*, which carry type I-A and III-B systems. These systems are active in interference against both viruses and plasmids, conferring immunity with either perfectly matching protospacers or those with a few mismatches (<4) in the "seed" sequence [27, 28]. However, there has only been a handful of studies investigating conjugative plasmids and the CRISPR-Cas system in archaea. One such study in *Saccharolobus solfataricus* P2 found that adaptation was only stimulated against the pMGB1 conjugative plasmid when co-infected with the SMV1 (Sulfolobus monocaudavirus 1) [29]. Here, spacers were acquired seemingly randomly across the plasmid, with a small bias towards the antisense strand of protein-coding genes. In contrast, in *E. coli* spacers were preferentially acquired from sequences of the plasmid corresponding to the leading or lagging regions relative to the conjugative plasmid [30].

Pyrococcus furiosus (*P. furiosus*) is a hyperthermophilic archaeon isolated from thermal marine sediments [31, 32]. This organism encodes three distinct CRISPR-Cas systems: type I-A, type I-B, and type III-B [33, 34]. These three systems all share one set

of adaptation proteins that include Cas1, Cas2, Cas4-1, and Cas4-2 [35], and this set is responsible for adaptation at all seven CRISPR arrays. Previous studies have found that all three effector complexes can utilize the 200 crRNAs from any of the arrays [36-39]. Therefore, the CRISPR-Cas system within *P. furiosus* is extremely cooperative.

Despite encoding 200 spacers in these CRISPR arrays, surprisingly, none of these spacers match protospacer sequences in any currently known virus or other MGE. This makes it impossible to study the response of the *P. furiosus* CRISPR-Cas system to natural invaders, and to date all studies have relied on small plasmids to simulate conditions of naïve and primed adaptation [35, 40, 41].

Recently, our collaborators discovered a novel conjugative plasmid in an isolate of *Thermococcus 33.3,* known as pT33.3. This is the largest plasmid in Thermococcales, at 103,230 bp in length. While this plasmid encodes around 180 ORFs, their functions remain mostly unknown. Among those to which function can be ascribed are several genes homologous to bacterial conjugative transfer proteins, and an associated origin of transfer (OriT). These features confer the capacity for pT33.3 to readily transfer between many Thermococcales species.

In light of these new data, we sought to study this natural MGE in our *P. furiosus* model system. Here, we demonstrate a striking CRISPR hyper-adaptation phenotype in *P. furiosus* in response to pT33.3, with almost all new spacers arising from the conjugative plasmid. The pattern of spacer uptake is reminiscent of primed adaptation and allowed us to identify a protospacer on pT33.3 with a partial match to a spacer in one *P. furiosus* CRISPR array. We show this spacer is the source of the primed adaptation we observe; both the spacer and the I-B CRISPR-Cas system are necessary for the hyper-

adaptation phenotype. The pT33.3 plasmid thus provides a new tool to study the natural dynamics of adaptive immunity in an archaeal model system.

Results

The presence of pT33.3 leads to hyperadaptation in *Pyrococcus furiosus*

pT33.3 is a naturally occurring conjugative plasmid capable of transfer between Thermococcales species. We hypothesized that a large, natural plasmid such as pT33.3 may cause different CRISPR spacer acquisition or effector activation than our lab-generated shuttle vectors and that the patterns could be more akin to those that occur in the environment. To test this, we transferred marker-encoding variants of pT33.3 from donor strains of *Thermococcus litoralis* (*T. litoralis*) or *Thermococcus kodakarensis* (*T. kodakarensis*) to *P. furiosus* (Pfu) (Figure 2.1A). We found that both donors were readily able to transfer the plasmid to *P. furiosus*, and that the plasmid stably replicated, forming transconjugate colonies on selective media. These transconjugates were inoculated into selective liquid media for genomic DNA isolation.

To observe any changes in adaptation in response to pT33.3, we performed a standard PCR assay in which we amplified from the leader to the first known spacer of *P. furiosus* CRISPR arrays and looked for longer bands corresponding to "expanded" arrays, ie arrays with newly integrated spacers (Figure 2.1A). A small shuttle vector, pYS3, was used as a control (as in previous studies [35, 40, 41]).

In the absence of priming, adaptation occurs at low frequencies in Pfu and normally requires several rounds of PCR to visualize a single integration event [35, 40]. Surprisingly, in the presence of pT33.3 we observed a striking increase in adaptation

events in wildtype Pfu strains, with expanded products easily visible after first-round PCR (Figure 2.1B, *=unexpanded array, +1=new spacer/expanded band). This "hyperadaptation" effect did not occur in the absence of pT33.3, nor with our pYS3 control plasmid. Moreover, this phenomenon occurred at 5 of the 7 CRISPR loci in *P. furiosus* (the CRISPR1 and CRISPR8 loci do not adapt as efficiently as the other 5 loci [35]) (Figure 2.1B).

New spacers are derived from pT33.3 and center around a potential protospacer

To identify the source of these new spacers, we did Illumina platform high-throughput sequencing (MiSeq) of the expanded array PCR amplicons, then extracted any new leader-adjacent spacer sequences from the reads and aligned them to a reference containing the pT33.3 and the *P. furiosus*. In our control adaptation assay where *P. furiosus* is challenged with the pYS3 plasmid, we observed that the majority of new spacers are acquired from the genome (Figure 2.2A) [35, 40]. In contrast, challenging *P. furiosus* with pT33.3 resulted in the vast majority (99% or more) of new spacers being acquired from the plasmid rather than the genome (Figure 2.2A).

We aligned these new spacers to the pT33.3 reference sequence to identify the locations and context sequences for the corresponding protospacers, ie the sources of the newly-acquired spacers. On pT33.3, protospacers were concentrated around a particular site in a pattern reminiscent of primed adaptation [40]. Specifically, the pattern was characteristic of Type I primed adaptation, wherein spacer acquisition occurs bidirectionally, emanating from the priming site in a predominately 3'-5' orientation Figure 2.2B). In Type I systems, Cas3 is the major nuclease involved in interference and

is required for primed adaptation in these systems [15, 19, 42]. With its 3' to 5' nuclease/helicase activity, Cas3 is able to degrade DNA of invading MGEs providing immunity, which in turn provides free DNA ends for subsequent primed spacer uptake. The directionality of the new spacers acquired from pT33.3 followed this pattern and alerted us to the possibility of priming against the plasmid.

We focused on the DNA at the center of the apparent priming region and compared it to all 200 spacers in *P. furiosus*'s seven CRISPR arrays. We identified one spacer (spacer 13 in the CRISPR1 array) with a near-perfect match against the plasmid (2 mismatches within the 37 bp putative protospacer) (Figure 2.2C). We thus hypothesized that hyperadaptation in the presence of pT33.3 was due to extensive priming against the plasmid instigated by this single CRISPR1 spacer.

Hyperadaptation requires the Type IB effector complex and a priming spacer

In previous studies, we determined that the type I-B effector complex is responsible for primed adaptation in *P. furiosus* [40]. To confirm that we are observing primed adaptation against pT33.3, we introduced the plasmid (by mating) into *P. furiosus* strains encoding only single effector complexes. We only observed hyperadaptation in strains encoding the type I-B CRISPR-Cas system (either wildtype or the type I-B-only strain) (Figure 2.3). This added further weight to our hypothesis that primed adaptation is the cause of pT33.3-induced hyperadaptation.

Next, we deleted the entire CRISPR1 array, which encodes the presumed priming spacer. In the absence of the CRISPR1 array, hyperadaptation was lost (Figure 2.4A), confirming that the CRISPR1-encoded spacer is required for this phenotype. We

sequenced the rare adaptation events captured in the absence of CRISPR1 and hyperadaptation, and found that the majority of new spacers were acquired from the *P. furiosus* genome rather than pT33.3, similar to that observed with the pYS3 plasmid (Figure 2.4B) [35, 40]. For the minority of spacers that did arise from pT33.3 (~5-7% of new spacer reads), the pattern across the plasmid was random, rather than centered around any specific site (Figure 2.4C, bottom panel). Thus, priming was eliminated in the absence of the CRISPR1 array (Figure 2.4). Together with the type I-B data, these results confirm that hyperadaptation against pT33.3 is likely due to priming by a naturally occurring spacer.

Evidence for processive acquisition of new spacers via primed adaptation

This new plasmid presented an opportunity to study patterns of spacer uptake during priming in a natural scenario. We noted that under priming conditions (I-B and CRISPR1 array present) a majority of new spacers (65.2%) were taken from the non-target strand, and the extent of uptake tapered off gradually moving out from the protospacer in a 3' to 5' manner (Figure 2.4C, top panel). This is consistent with priming patterns observed in other studies [15-17, 19, 43]. Spacers are also acquired from the target strand in a 3' to 5' directionality, though not as frequently (25.7%). In the absence of the CRISPR1 array (ΔCR1 array), and therefore no priming activity, the acquisition of spacers from both strands and directions is more evenly distributed (Figure 2.4C, bottom panel).

For all sequencing data described thus far, we focused on single, leader-adjacent, newly integrated spacers. However, many of the array amplicons we observed showed

evidence of more than one new leader-adjacent spacer; for example in Figure 1B and Figure 5A, faint bands of up to 400 bp are visible and these products would correspond to arrays with up to four newly integrated spacers (with each new spacer-repeat unit extending the amplicon by about 67bp). These products gave us the opportunity to examine arrays that had multiple contiguous new spacers. We isolated PCR products corresponding to arrays with three new leader-adjacent spacers, generated libraries, and sequenced on a MiSeq to generate single 300 bp reads. We then filtered for unique reads with exactly three new spacers which could be aligned (redundant reads, reads with low quality scores at the distal end, or shorter reads that did not span three new leader-adjacent spacers were excluded). The aligned spacers were analyzed for spatial relationships among the three positions: leader-adjacent (NS1), second (NS2), and distal (NS3) (Figure 5A). Since new integrations are thought to occur almost exclusively at the leader-adjacent position [15, 41, 44, 45], we assumed that NS3 spacers were integrated first in time, followed by NS2 and then NS1.

Our first observation was that the diversity of spacers was greatest at the leader-adjacent (NS1) position, followed by NS2 and then NS3. This was consistent with the model of sequential spacer addition described above: after a given integration event, if that cell proliferates, multiple progeny with that spacer could go on to new integrations. We did not find evidence of spatial correlations among the three contiguous spacers; for example, the three new spacers could be derived either from the same strand or different strands and the frequency of those two possibilities was roughly the same as what would arise from random events (Figure 5B-C). The three new spacers usually did not arise from contiguous plasmid DNA and protospacers could be found 10kb or more from one

another (Figure 5D, distances between protospacers). Interestingly, we saw a subtle trend suggesting that the oldest spacers (NS3) were usually closer to the priming target than subsequent spacers (NS2, NS1). This, together with the observations on positional differences in spacer diversity (Figure 5A) could support a model wherein priming takes place iteratively over multiple cellular divisions.

Discussion

In this study, we challenged *P. furiosus* with the conjugative plasmid, pT33.3, to study the response of the CRISPR-Cas systems to a naturally occurring MGE. Previous work revealed this plasmid can readily transfer to a variety of Thermococcales via a self-transmission mechanism resembling bacterial plasmid conjugation. In this study, we confirm that pT33.3 can transfer to, and replicate in, *P. furiosus*, following transfer from either *T. litoralis* or *T. kodakarensis* donors. It must be noted that the genetically tractable lab strain of Pfu used in this study is naturally competent for DNA uptake [46], thus is it impossible to determine the route of plasmid transfer occurring in these experiments.

There is conflicting data regarding spacer acquisition against conjugative plasmids in laboratory settings. In bacterial systems, MOB_F (Mobility) plasmids are more heavily targeted in the leading region (region adjacent to the oriT which enters the recipient cell first), while MOB_P plasmids are targeted in the lagging region (that enter the recipient cell last) [30]. In contrast, spacers acquired against plasmid pMGB1 in the archaeon *Saccharolobus solfataricus* P2 were evenly distributed across the plasmid [29]. There is very little data on how conjugation operates in archaea, but genetic homologies and similar experimental requirements (e.g. cell-to-cell contact) suggest it resembles

bacterial plasmid transfer, where DNA enters the cell in a single stranded state (except for a few enigmatic cases), through oriT-mediated transfer [47]. How spacers would be acquired from a plasmid such as this remains unknown.

We were surprised to observe a robust increase in the frequency of newly acquired spacers in *P. furiosus* strains in the presence of pT33.3 (Figure 2.1B). This hyperadaptation effect occurred in multiple CRISPR loci in transconjugant cells. All CRISPR loci in *P. furiosus* are active, although spacer uptake in the CRISPR1 and CRISPR8 loci is weaker than that of the other CRISPR loci [35]. We initially hypothesized that conjugation itself was the cause of this increase in adaptation. However, sequencing and analysis of the new spacers revealed a pattern reminiscent of primed adaptation (Figure 2.2B). We observed a clear preference for new spacers at a particular site of pT33.3, and adaptation radiating from this site in a 3'-5' direction. In Type I systems, Cas3 is the major nuclease involved in interference and functions via 3'-5' exonuclease activity. This activity generates substrates which are frequently used as substrates for further adaptation and is thus a major requirement of primed adaptation [36, 40, 48-54]. Due to the spacer acquisition pattern we observed, we predicted that there would be a sequence in the vicinity of this hotspot that would potentially match a spacer in one of the CRISPR arrays. Indeed, spacer 13 of CRISPR1 was a partial match to pT33.3, with only two nucleotides different.

P. furiosus has 7 CRISPR arrays encoding a total of 200 spacers. However, to date, the source of these spacers has remained mysterious. None of the 200 spacers match to any known MGE in sequence databases. It was thus novel to find a natural protospacer targeted by the *P. furiosus* CRISPR-Cas systems. The presence of this spacer suggests *P.*

furiosus may have encountered this, or a similar conjugative plasmid, in the wild and acquired a spacer that efficiently targeted and silenced this invader. This study is the first example of a natural invader for *P. furiosus*, with no known viruses or natural plasmids identified in *P. furiosus* isolates. So far, most sampling of Thermococcales has concentrated on isolating culturable cells, which almost by definition selects against superfluous MGEs, or those which confer a growth deficit such as viruses. Hence, while *P. furiosus* clearly has active CRISPR immunity and encounters numerous invaders, these are not reflected in culturing experiments.

From our previous studies, we know that the type I-B effector complex is the sole effector complex involved in primed adaptation in our system [40]. We confirm that pT33.3-induced hyperadaptation occurs only in wildtype and type I-B-only strains (Figure 2.3), and that this hyperadaptation requires the CRISPR1 array. We noted that the hyperadaptation was more evident in strains encoding only the type I-B effector complex. As hypothesized previously, this is likely due to dilution of the crRNAs among the type I-A and III-B effector complexes [40].

The elimination of priming allowed us to observe naïve adaptation against pT33.3. These results reveal no obvious regions of heightened adaptation. This contrasts to that observed with *E. coli* conjugative elements, mirroring more closely that observed with the Crenarchaeon *S. solfataricus*. We might expect spacer uptake to more heavily target regions adjacent to the oriT i.e. regions entering the recipient cell early, simply due to their prolonged exposure to adaptation machinery. However, in our system it is unclear if this conjugative plasmid is truly entering the cell via conjugation or transformation, or even what form the DNA is in when entering the cell. Bacterial conjugative plasmids

most often enter recipients through a Type IV secretion system, as linear single-stranded DNA; however, the entering plasmid does not provide a free DNA end for adaptation, being covalently bound to the nickase/relaxase protein [55-58]. It is possible that our assay is not ideal for detecting true hotspots resulting from a plasmid *P. furiosus* by conjugation. Future experiments to answer this question could include using the natural isolate of *P. furiosus*, DSM3638, which is not naturally competent for DNA uptake [31, 59]. Thus, any transfer of pT33.3 would require conjugation. However, the lack of competency makes this strain not amenable to genetic manipulation, and this strain also encodes the spacer responsible for pT33.3 priming in our lab *P. furiosus* strain. Thus, conjugation of pT33.3 to DSM3638 would likely induce hyperadaptation through priming, obscuring any data regarding conjugation-induced adaptation.

It is also interesting to consider that while pT33.3 is heavily targeted by the Type I-B effector complex under priming conditions, this plasmid persists in cultures (not shown). Even if the first mismatched spacer is ineffective at clearing pT33.3, all subsequent adaptation events (resulting from priming) should perfectly match pT33.3, inducing strong CRISPR targeting. Thus, it seems that pT33.3 may be able to render the CRISPR system ineffective at interference activity through an as yet unknown anti-CRISPR mechanism. To date, all known anti-CRISPRs are proteins that protect invading MGEs against the CRISPR-Cas defense system by abolishing CRISPR targeting [60, 61]. These proteins can act upon CRISPR effectors in various ways, including binding to complexes, preventing loading of crRNAs, inhibiting PAM recognition, etc. [60-72]. If such anti-CRISPRs are encoded by pT33.3, they could play a role in the persistence of the pT33.3 plasmid. It is clear that the *P. furiosus* Type I-B effector is able to target

pT33.3 (Figure 2.1B, Figure 2.3), at least initially, as priming requires an initial targeting even, and the downstream Cas3-mediated nuclease activity. Thus the relationship between pT33.3 and the CRISPR-Cas system(s) remains complex and unclear.

It would be interesting to further investigate how extensive priming against pT33.3 affects its survival in cultures. Further experiments could include performing a passaging assay to determine how long before the cells were cured of the plasmid in a variety of strains, including the wildtype and the priming null strain (deletion of the matching protospacer in the CRISPR1 array). We would hypothesize that priming could clear a plasmid relatively quickly in wildtype strains, especially in *P. furiosus* where the newly acquired spacers are used by all three effector complexes (IA, IB, and IIIB), and more slowly (if at all) in the priming null strain.

Overall, we have identified a natural invader of the model archaeon, *P. furiosus.* This will greatly enhance our ability to study CRISPR adaptation in archaea. Additionally, we reveal that the CRISPR system of *P. furiosus* has encountered either pT33.3, or a homologue thereof, in the past, encoding a partially matching CRISPR spacer in the CRISPR1 array. This spacer induces a priming response, with massive spacer uptake against pT33.3 in transconjugate cells. Identifying the natural source of a *P. furiosus* spacer starts a new era of CRISPR research with this species.

Material and methods

Strain growth conditions

For *P. furiosus* strains, we grew cultures with defined media (either uracil or tryptophan deficient for selection of plasmids with either the pyrF gene or the trpAB

gene) with cellobiose as the carbon source, or complex media with yeast extract and tryptone. We used the same complex media for growth with *Thermococcus kodakarensis* and *Thermococcus litoralis* strains. For liquid cultures, we grew overnight (~16 hours) at either 95°C (*P. furiosus*) or 85°C (*Thermococcus*). For plates, we either used defined media or complex media, and grew strains for 3 nights (~65 hours) at either 95°C (*P. furiosus*) or 85°C (*Thermococcus* strains).

Conjugation Assays

The *Thermococcus* donor strains were freshly grown in complex media at 85°C and the *Pyrococcus* recipient strains were grown in defined selective media at 95°C, all grown overnight for ~16 hours. Next 5 mL of each overnight culture was pipetted into a 15 mL conical tube. Cultures were centrifuged at ~3400 RPM for 20 minutes. Supernatant was discarded and all remaining liquid media was tapped out. The tip of the pipette tip was cut to prevent shearing of cells, and cells were resuspended in 50 μL of 1X Pfu Base Salts. Recipient cells were mixed with the appropriate donor in a 1.5 mL tube, for a total of 100 μL cell resuspension. The entire cell mixture was spotted in the center of a plate, which was then incubated at 95°C for 3 nights. After opening the chamber, and the spot was scraped into 5 mL of defined selective liquid media and grown overnight at 95°C.

Adaptation Assays

A population of cells were picked (~20 colonies), added to defined, selective, liquid media, and then allowed to grow at 95°C overnight (~16 hours). Following

overnight growth, 1 mL of the liquid culture was collected for genomic DNA extraction using the Quick-DNA miniprep kit (Zymo Research). Ten ng of genomic DNA was used in each PCR; adaptation primers spanned from within the leader to the first known spacer of the array. PCR products were then run on a 2.5% agarose gel (120 V for ~48 minutes) to separate unexpanded bands (no integration of new spacers) from expanded bands (successful integration of new spacer).

Illumina Sequencing

To prepare samples for Illumina sequencing, the same culturing and DNA isolation protocol was used as described above for adaptation assays. Following PCR, expanded bands were excised (or potential expanded PCR that was too faint to see on a gel in first round PCR) and DNA was recovered from the gel using the Gel Recovery Kit (Zymo Research) and eluted in 10 μL of nuclease free water. One μL from the elution was used in a second round of PCR, with primer containing adapters required for Illumina-platform sequencing. Expanded bands were sized selected on an agarose gel and recovered using the same kit, and eluted in 10 μL of nuclease free water. A 1:10 dilution of the elution was made, and 1 μL from that elution was added into a final PCR with primers containing sequencing barcodes. PCR products were run on the gel, expanded bands were excised, and gel recovered. Twenty μL of nuclease free water was used for the final elution.

Final gel-purified amplicon libraries were pooled according to band intensity, normalized according to concentration and number of samples represented in the pool, and sequenced on an Illumina MiSeq to yield 250 by 50 paired-end reads; the 250 base

read 1 sequences were used in this study. After sequencing, samples were de-multiplexed by index and the sequence corresponding to a new spacer was extracted from each read. Spacer sequences were then aligned and analyzed, as previously described [40, 41, 73], to characterize protospacer sources and patterns of protospacer distribution along the pT33.3 plasmid

To prepare the long, three-spacer array amplicons for sequencing, Illumina adapters were added to PCR products by a ligation method rather than the sequential PCRs described above. Briefly, expanded array products in the size range expected for 3 new spacers were size selected from an agarose gel and purified using the Gel Recovery Kit (Zymo Research). We then used the NEB Ultra II DNA library kit to ligate Illumina adapters/barcodes to the amplicons and enrich/purify the resulting libraries. Amplicon libraries were sequenced on an Illumina MiSeq instrument set to generate single 300bp reads.

Strain construction

To generate the strains with the CRISPR 1 array deleted, we used homologous recombination of PCR products, as described previously [31].

References

1. Barrangou, R., et al., *CRISPR provides acquired resistance against viruses in prokaryotes.* Science, 2007. **315**(5819): p. 1709-12.
2. Mojica, F.J., et al., *Intervening sequences of regularly spaced prokaryotic repeats derive from foreign genetic elements.* J Mol Evol, 2005. **60**(2): p. 174-82.
3. Brouns, S.J., et al., *Small CRISPR RNAs guide antiviral defense in prokaryotes.* Science, 2008. **321**(5891): p. 960-4.
4. Makarova, K.S., et al., *Evolutionary classification of CRISPR-Cas systems: a burst of class 2 and derived variants.* Nat Rev Microbiol, 2020. **18**(2): p. 67-83.
5. van der Oost, J., et al., *CRISPR-based adaptive and heritable immunity in prokaryotes.* Trends Biochem Sci, 2009. **34**(8): p. 401-7.
6. Makarova, K.S., et al., *Unification of Cas protein families and a simple scenario for the origin and evolution of CRISPR-Cas systems.* Biol Direct, 2011. **6**: p. 38.
7. Carte, J., et al., *Cas6 is an endoribonuclease that generates guide RNAs for invader defense in prokaryotes.* Genes Dev, 2008. **22**(24): p. 3489-96.
8. Deltcheva, E., et al., *CRISPR RNA maturation by trans-encoded small RNA and host factor RNase III.* Nature, 2011. **471**(7340): p. 602-7.
9. Sokolowski, R.D., S. Graham, and M.F. White, *Cas6 specificity and CRISPR RNA loading in a complex CRISPR-Cas system.* Nucleic Acids Res, 2014. **42**(10): p. 6532-41.

10. Wang, R., et al., *Interaction of the Cas6 riboendonuclease with CRISPR RNAs: recognition and cleavage.* Structure, 2011. **19**(2): p. 257-64.

11. Samai, P., et al., *Co-transcriptional DNA and RNA Cleavage during Type III CRISPR-Cas Immunity.* Cell, 2015. **161**(5): p. 1164-74.

12. Jinek, M., et al., *A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity.* Science, 2012. **337**(6096): p. 816-21.

13. Jore, M.M., et al., *Structural basis for CRISPR RNA-guided DNA recognition by Cascade.* Nat Struct Mol Biol, 2011. **18**(5): p. 529-36.

14. Gasiunas, G., et al., *Cas9-crRNA ribonucleoprotein complex mediates specific DNA cleavage for adaptive immunity in bacteria.* Proc Natl Acad Sci U S A, 2012. **109**(39): p. E2579-86.

15. Datsenko, K.A., et al., *Molecular memory of prior infections activates the CRISPR/Cas adaptive bacterial immunity system.* Nat Commun, 2012. **3**: p. 945.

16. Swarts, D.C., et al., *CRISPR interference directs strand specific spacer acquisition.* PLoS One, 2012. **7**(4): p. e35888.

17. Fineran, P.C., et al., *Degenerate target sites mediate rapid primed CRISPR adaptation.* Proc Natl Acad Sci U S A, 2014. **111**(16): p. E1629-38.

18. Li, M., et al., *Adaptation of the Haloarcula hispanica CRISPR-Cas system to a purified virus strictly requires a priming process.* Nucleic Acids Res, 2014. **42**(4): p. 2483-92.

19. Richter, C., et al., *Priming in the Type I-F CRISPR-Cas system triggers strand-independent spacer acquisition, bi-directionally from the primed protospacer.* Nucleic Acids Res, 2014.

20. Semenova, E., et al., *Highly efficient primed spacer acquisition from targets destroyed by the Escherichia coli type I-E CRISPR-Cas interfering complex.* Proc Natl Acad Sci U S A, 2016.

21. Xue, C., et al., *CRISPR interference and priming varies with individual spacer sequences.* Nucleic Acids Res, 2015. **43**(22): p. 10831-47.

22. Amitai, G. and R. Sorek, *CRISPR-Cas adaptation: insights into the mechanism of action.* Nat Rev Microbiol, 2016. **14**(2): p. 67-76.

23. Sternberg, S.H., et al., *Adaptation in CRISPR-Cas Systems.* Mol Cell, 2016. **61**(6): p. 797-808.

24. Shmakov, S.A., et al., *The CRISPR Spacer Space Is Dominated by Sequences from Species-Specific Mobilomes.* mBio, 2017. **8**(5).

25. Wang, H., et al., *Archaeal extrachromosomal genetic elements.* Microbiol Mol Biol Rev, 2015. **79**(1): p. 117-52.

26. Greve, B., et al., *Genomic comparison of archaeal conjugative plasmids from Sulfolobus.* Archaea, 2004. **1**(4): p. 231-9.

27. Gudbergsdottir, S., et al., *Dynamic properties of the Sulfolobus CRISPR/Cas and CRISPR/Cmr systems when challenged with vector-borne viral and plasmid genes and protospacers.* Mol Microbiol, 2011. **79**(1): p. 35-49.

28. Manica, A., et al., *In vivo activity of CRISPR-mediated virus defence in a hyperthermophilic archaeon.* Mol Microbiol, 2011. **80**(2): p. 481-91.

29. Erdmann, S., S.A. Shah, and R.A. Garrett, *SMV1 virus-induced CRISPR spacer acquisition from the conjugative plasmid pMGB1 in Sulfolobus solfataricus P2.* Biochem Soc Trans, 2013. **41**(6): p. 1449-58.

30. Westra, E.R., et al., *CRISPR-Cas systems preferentially target the leading regions of MOBF conjugative plasmids.* RNA Biol, 2013. **10**(5): p. 749-61.

31. Farkas, J., et al., *Recombinogenic properties of Pyrococcus furiosus strain COM1 enable rapid selection of targeted mutants.* Appl Environ Microbiol, 2012. **78**(13): p. 4669-76.

32. Bridger, S.L., et al., *Genome sequencing of a genetically tractable Pyrococcus furiosus strain reveals a highly dynamic genome.* J Bacteriol, 2012. **194**(15): p. 4097-106.

33. Majumdar, S., et al., *Three CRISPR-Cas immune effector complexes coexist in Pyrococcus furiosus.* RNA, 2015.

34. Terns, R.M. and M.P. Terns, *The RNA- and DNA-targeting CRISPR-Cas immune systems of Pyrococcus furiosus.* Biochem Soc Trans, 2013. **41**(6): p. 1416-21.

35. Shiimori, M., et al., *Role of free DNA ends and protospacer adjacent motifs for CRISPR DNA uptake in Pyrococcus furiosus.* Nucleic Acids Res, 2017.

36. Elmore, J., et al., *DNA targeting by the type I-G and type I-A CRISPR-Cas systems of Pyrococcus furiosus.* Nucleic Acids Res, 2015.

37. Elmore, J.R., et al., *Bipartite recognition of target RNAs activates DNA cleavage by the Type III-B CRISPR-Cas system.* Genes Dev, 2016. **30**(4): p. 447-59.

38. Hale, C.R., et al., *Target RNA capture and cleavage by the Cmr type III-B CRISPR-Cas effector complex.* Genes Dev, 2014. **28**(21): p. 2432-43.

39. Majumdar, S., et al., *Three CRISPR-Cas immune effector complexes coexist in Pyrococcus furiosus.* RNA, 2015. **21**(6): p. 1147-58.

40. Garrett, S., et al., *Primed CRISPR DNA uptake in Pyrococcus furiosus.* Nucleic Acids Res, 2020.

41. Shiimori, M., et al., *Cas4 Nucleases Define the PAM, Length, and Orientation of DNA Fragments Integrated at CRISPR Loci.* Mol Cell, 2018. **70**(5): p. 814-824 e6.

42. Vorontsova, D., et al., *Foreign DNA acquisition by the I-F CRISPR-Cas system requires all components of the interference machinery.* Nucleic Acids Res, 2015. **43**(22): p. 10848-60.

43. Rao, C., D. Chin, and A.W. Ensminger, *Priming in a permissive type I-C CRISPR-Cas system reveals distinct dynamics of spacer acquisition and loss.* RNA, 2017. **23**(10): p. 1525-1538.

44. Yosef, I., M.G. Goren, and U. Qimron, *Proteins and DNA elements essential for the CRISPR adaptation process in Escherichia coli.* Nucleic Acids Res, 2012. **40**(12): p. 5569-76.

45. Garrett, S.C., *Pruning and Tending Immune Memories: Spacer Dynamics in the CRISPR Array.* Front Microbiol, 2021. **12**: p. 664299.

46. Lipscomb, G.L., et al., *Natural competence in the hyperthermophilic archaeon Pyrococcus furiosus facilitates genetic manipulation: construction of markerless deletions of genes encoding the two cytoplasmic hydrogenases.* Appl Environ Microbiol, 2011. **77**(7): p. 2232-8.

47. Wagner, A., et al., *Mechanisms of gene flow in archaea.* Nat Rev Microbiol, 2017. **15**(8): p. 492-501.

48. Beloglazova, N., et al., *Structure and activity of the Cas3 HD nuclease MJ0384, an effector enzyme of the CRISPR interference.* EMBO J, 2011. **30**(22): p. 4616-27.

49. Cooper, L.A., A.M. Stringer, and J.T. Wade, *Determining the Specificity of Cascade Binding, Interference, and Primed Adaptation In Vivo in the Escherichia coli Type I-E CRISPR-Cas System.* MBio, 2018. **9**(2).

50. Fagerlund, R.D., et al., *Spacer capture and integration by a type I-F Cas1-Cas2-3 CRISPR adaptation complex.* Proc Natl Acad Sci U S A, 2017.

51. Huo, Y., et al., *Structures of CRISPR Cas3 offer mechanistic insights into Cascade-activated DNA unwinding and degradation.* Nat Struct Mol Biol, 2014. **21**(9): p. 771-7.

52. Kunne, T., et al., *Cas3-Derived Target DNA Degradation Fragments Fuel Primed CRISPR Adaptation.* Mol Cell, 2016. **63**(5): p. 852-64.

53. Loeff, L., S.J.J. Brouns, and C. Joo, *Repetitive DNA Reeling by the Cascade-Cas3 Complex in Nucleotide Unwinding Steps.* Mol Cell, 2018. **70**(3): p. 385-394 e3.

54. Musharova, O., et al., *Prespacers formed during primed adaptation associate with the Cas1-Cas2 adaptation complex and the Cas3 interference nuclease-helicase.* Proc Natl Acad Sci U S A, 2021. **118**(22).

55. Banuelos-Vazquez, L.A., G. Torres Tejerizo, and S. Brom, *Regulation of conjugative transfer of plasmids and integrative conjugative elements.* Plasmid, 2017. **91**: p. 82-89.

56. Christie, P.J., *Type IV secretion: intercellular transfer of macromolecules by systems ancestrally related to conjugation machines.* Mol Microbiol, 2001. **40**(2): p. 294-305.

57. Llosa, M., et al., *Bacterial conjugation: a two-step mechanism for DNA transport.* Mol Microbiol, 2002. **45**(1): p. 1-8.

58. Wallden, K., A. Rivera-Calzada, and G. Waksman, *Type IV secretion systems: versatility and diversity in function.* Cell Microbiol, 2010. **12**(9): p. 1203-12.

59. Farkas, J., et al., *Defining components of the chromosomal origin of replication of the hyperthermophilic archaeon Pyrococcus furiosus needed for construction of a stable replicating shuttle vector.* Appl Environ Microbiol, 2011. **77**(18): p. 6343-9.

60. Pawluk, A., A.R. Davidson, and K.L. Maxwell, *Anti-CRISPR: discovery, mechanism and function.* Nat Rev Microbiol, 2018. **16**(1): p. 12-17.

61. Wiedenheft, B., *In defense of phage: viral suppressors of CRISPR-mediated adaptive immunity in bacteria.* RNA Biol, 2013. **10**(5): p. 886-90.

62. Basgall, E.M., et al., *Gene drive inhibition by the anti-CRISPR proteins AcrIIA2 and AcrIIA4 in Saccharomyces cerevisiae.* Microbiology, 2018.

63. Bondy-Denomy, J., *Protein inhibitors of CRISPR-Cas9.* ACS Chem Biol, 2017.

64. Bondy-Denomy, J., et al., *Multiple mechanisms for CRISPR-Cas inhibition by anti-CRISPR proteins.* Nature, 2015. **526**(7571): p. 136-9.

65. Bondy-Denomy, J., et al., *Bacteriophage genes that inactivate the CRISPR/Cas bacterial immune system.* Nature, 2013. **493**(7432): p. 429-32.

66. Borges, A.L., et al., *Bacteriophage Cooperation Suppresses CRISPR-Cas3 and Cas9 Immunity.* Cell, 2018.

67. Hynes, A.P., et al., *An anti-CRISPR from a virulent streptococcal phage inhibits Streptococcus pyogenes Cas9.* Nat Microbiol, 2017. **2**(10): p. 1374-1380.

68. Ka, D., et al., *Crystal structure of an anti-CRISPR protein, AcrIIA1.* Nucleic Acids Res, 2017.

69. Koonin, E.V. and K.S. Makarova, *Anti-CRISPRs on the march.* Science, 2018. **362**(6411): p. 156-157.

70. Landsberger, M., et al., *Anti-CRISPR Phages Cooperate to Overcome CRISPR-Cas Immunity.* Cell, 2018.

71. Pawluk, A., et al., *Disabling a Type I-E CRISPR-Cas Nuclease with a Bacteriophage-Encoded Anti-CRISPR Protein.* MBio, 2017. **8**(6).

72. Touchon, M. and E.P. Rocha, *The small, slow and specialized CRISPR and anti-CRISPR of Escherichia and Salmonella.* PLoS One, 2010. **5**(6): p. e11126.

73. Shiimori, M., et al., *Role of free DNA ends and protospacer adjacent motifs for CRISPR DNA uptake in Pyrococcus furiosus.* Nucleic Acids Res, 2017. **45**(19): p. 11281-11294.

Figure 2.1. Hyperadaptation observed in the presence of pT33.3.

(A) Flowchart of spacer analysis after conjugation and adaptation assays. Both *Thermococcus kodakarensis* (Tko) and *Thermococcus litoralis* (T. lit) donors were used to conjugate wildtype *Pyrococcus furiosus* (Pfu) strains. Resulting transconjugate colonies were grown in selective media, and the isolated genomic DNA from these cells was used to amplify the CRISPR loci from the leader to the first known spacer on the array. Expanded bands were excised and subsequently sent for HTS. (**B)** Analysis of adaptation at multiple CRISPR loci. CRISPR loci were amplified from the CRISPR leader to the first known spacer on the array for all CRISPR loci using primers shown in **(A)** (***** = unexpanded array, +1 = expanded array).

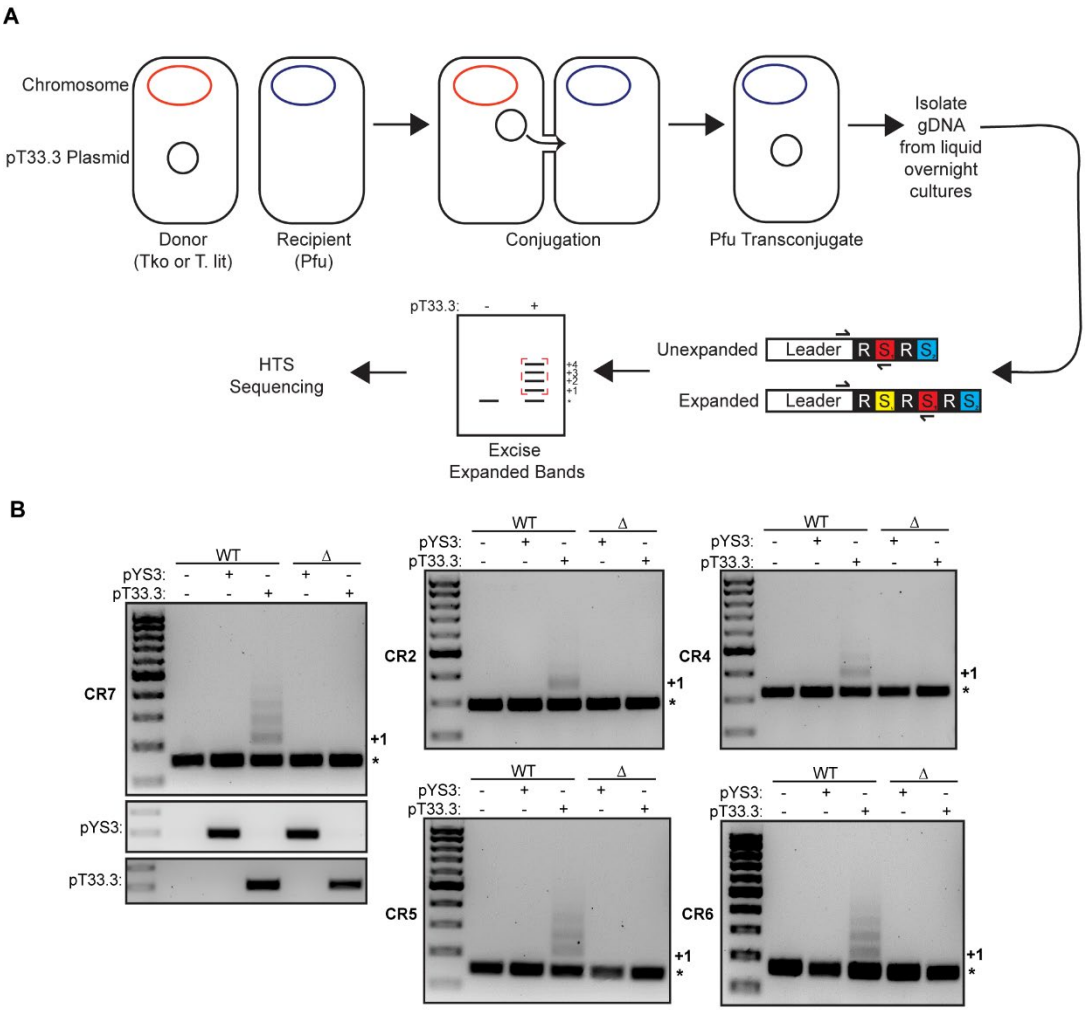
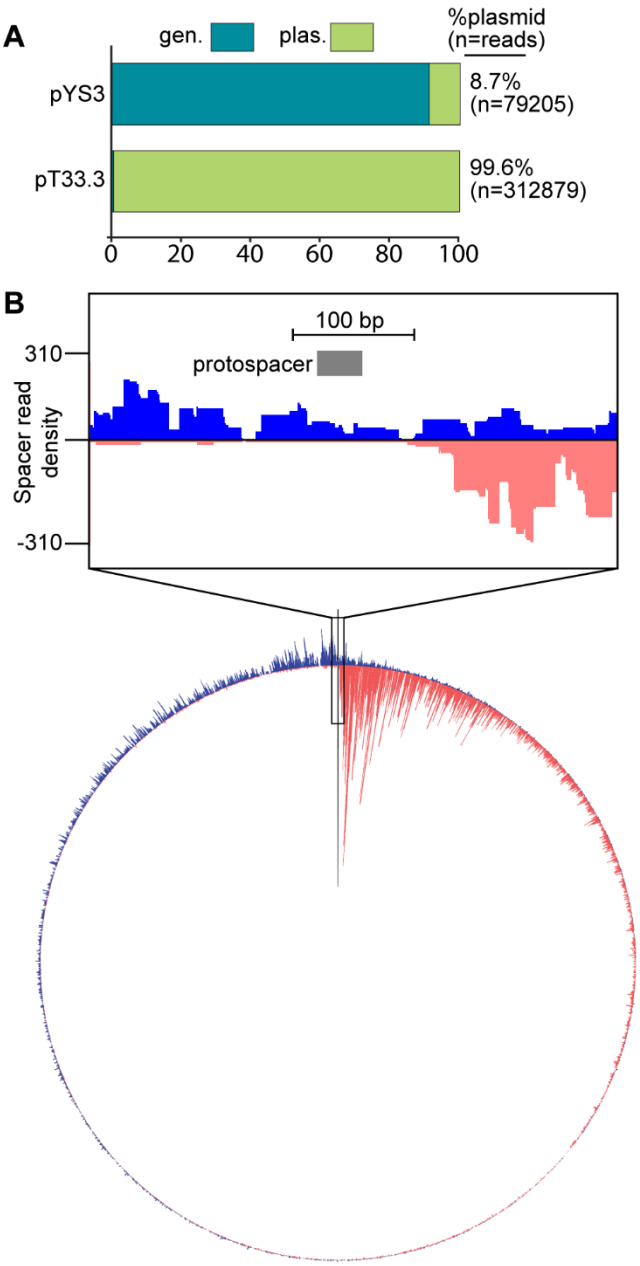

A
Chromosome
pT33.3 Plasmid
Donor
(Tko or T. lit)
Recipient
(Pfu)
Conjugation
Pfu Transconjugate
Isolate
gDNA
from liquid
overnight
cultures
Unexpanded
Leader
R S R S
Expanded
Leader
R S R S R S
pT33.3:
Excise
Expanded Bands
HTS
Sequencing
B
WT
Δ
pYS3:
pT33.3:
CR7
CR2
CR4
CR5
CR6
+1
*

Figure 2.2. The majority of new spacers are acquired from pT33.3.

(A) Bar graphs demonstrating the percentage of new spacers acquired from the pT33.3 plasmid compared to the well-studied pYS3-Pgdh plasmid. **(B**) Distribution of protospacers on the pT33.3 conjugative plasmid. Protospacers on the plus and minus strand are indicated in blue and pink respectively. Protospacers were deduced from aligning new spacers acquired into CRISPR loci of WT strain. Inset above plasmid shows the center of the hotspot, and the potential protospacer being targeted (gray). (**C)** Sequences in the hotspot region were aligned with all 200 spacers in Pfu, the partially matching protospacer was found in the CRISPR1 array with two mismatches (yellow highlight).



C

CR1: CTGGGCAAGTTCTGGCCTATACTGTCTCCTAATGTCT

pT33.3: CTGGGCGAGTTCGGGCCTATACTGTCTCCTAATGTCT

Figure 2.3. The hyperadaptation effect is isolated to the type I-B effector complex. Adaptation analysis of all CRISPR loci, amplifying from the leader to the first known spacer (as indicated in Figure 1A) for wildtype (WT), single effector complex strains (Type I-B only, Type I-A only, Type III-B only), and null strains (adaptation and interference null). (C is wildtype strain in the absence of pT33.3).

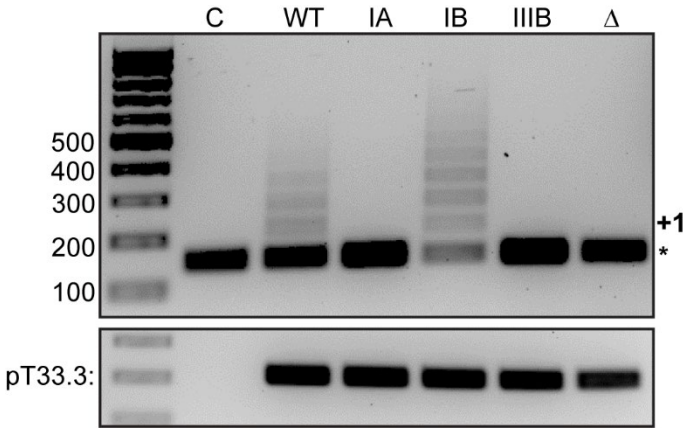

C
WT
IA
IB
IIIB
Δ
500
400
300
200
100
+1
*
pT33.3:

Figure 2.4. The hyperadaptation is caused by priming of the type I-B effector complex.

(A) Adaptation analysis of CRISPR loci between WT (wildtype) and IB (Type I-B only) strains, either in the presence or absence of the CRISPR1 array containing the partially matching protospacer. (**B)** Bar graph shows the percent of pT33.3-derived new spacers (versus genome-derived spacers) in WT and IB-only strains either with or without the CRISPR 1 array. **(C)** Genome browser tracks show the distribution of aligned new spacers across the pT33.3 plasmid in representative replicates for WT and WT-ΔCR1 strains. New spacers aligning to the top strand are shown in blue; bottom strand in pink. The x-axis corresponds to the linearized pT33.3 plasmid reference sequence; the gray bar indicates the position of the protospacer targeted by the CRISPR 1 spacer. The y-axis indicates cumulative read (spacer) depth. The linearized reference was divided into four quadrants and total, normalized spacer coverage was determined for each quadrant; percents shown are average values from eight replicates.

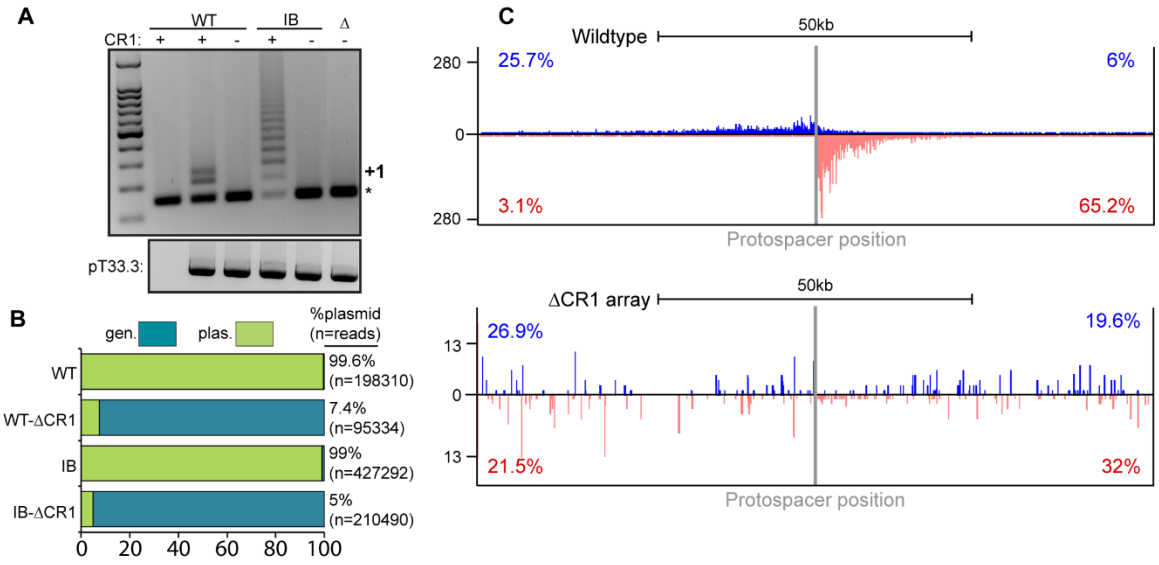
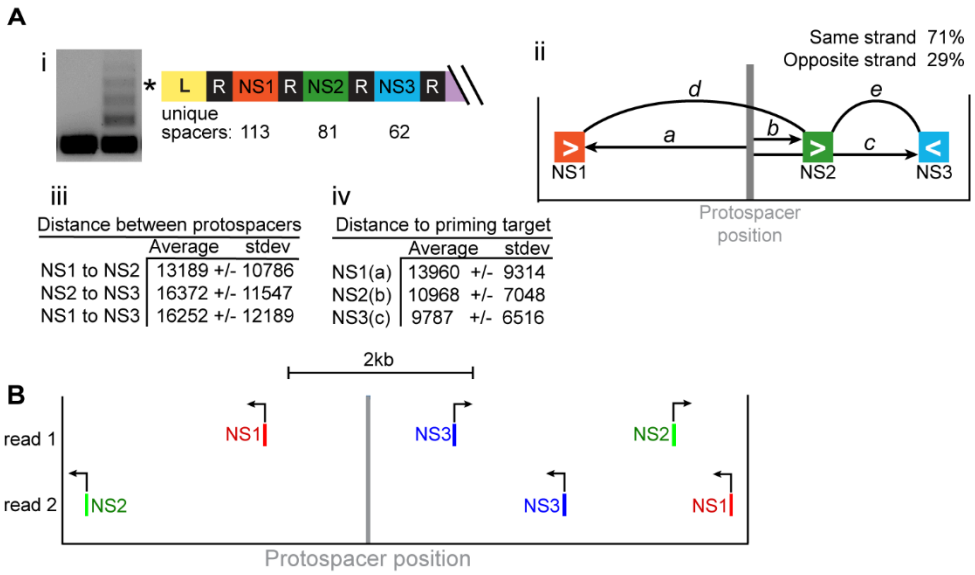

A
WT
IB
Δ
CR1:
+
+
-
+
-
-
+1
*
pT33.3:
B
gen.
plas.
%plasmid
(n=reads)
WT
99.6%
(n=198310)
WT-ΔCR1
7.4%
(n=95334)
IB
99%
(n=427292)
IB-ΔCR1
5%
(n=210490)
0
20
40
60
80
100
C
Wildtype
50kb
280
0
280
25.7%
6%
3.1%
65.2%
Protospacer position
ΔCR1 array
50kb
13
0
13
26.9%
19.6%
21.5%
32%
Protospacer position

Figure 2.5. Primed adaptation against pT33.3 plasmid.

(A) (i) PCR products corresponding to arrays with three contiguous new spacers were sequenced and analyzed. A total of 155 unique three-spacer reads were captured; among these reads, greater spacer diversity was observed at the leader-adjacent end as compared to the more distal positions. **(ii)** Cartoon schematic shows spacers from a hypothetical three-spacer read and the measurements made to summarize spacial relationships among them. **(iii)** The average distance between protospacers from a single read varied; spacial clustering was uncommon. **(iv)** In general, distal spacers (presumed to be captured earlier than spacers in the leader-adjacent position) were slightly closer to the protospacer. **(B)** Spacers from two representative three-spacer reads are shown as genome browser tracks to illustrate spacial relationships.


A
i
*
L
R
NS1
R
NS2
R
NS3
R
unique
spacers: 113 81 62
ii
Same strand 71%
Opposite strand 29%
d
e
a
b
c
NS1
NS2
NS3
Protospacer
position
iii
Distance between protospacers
Average stdev
NS1 to NS2 13189 +/- 10786
NS2 to NS3 16372 +/- 11547
NS1 to NS3 16252 +/- 12189
iv
Distance to priming target
Average stdev
NS1(a) 13960 +/- 9314
NS2(b) 10968 +/- 7048
NS3(c) 9787 +/- 6516
B
2kb
read 1
read 2
NS1
NS3
NS2
NS2
NS3
NS1
Protospacer position

CHAPTER 3

HISTONE DIRECTED INTEGRATION OF NEW SPACERS AT CRISPR ARRAYS IN *PYROCOCCUS FURIOSUS*[1]

[1]Elizabeth A. Watts, Sandra Garrett, Landon M. Clark, Brenton G. Graveley, Michael P. Terns. *To be submitted to Nucleic Acids Research.*

Abstract

The Clustered Regularly Interspaced Short Palindromic Repeat (CRISPR)-CRISPR associated (Cas) defense pathway is involved in adaptive immunity in prokaryotes. This pathway involves the acquisition of DNA from foreign invaders to integrate into the host genome in a region known as a CRISPR array in a polarized manner (adaptation), which is subsequently transcribed and processed to produce mature crRNA (CRISPR RNA) guides that associate with an effector complex to target and silence recurring invaders. This chapter seeks to explore the mechanisms behind the first step (adaptation) in which DNA is acquired from the invader and integrated into the CRISPR array at the leader adjacent repeat in our model organism, the hyperthermophilic archaeon *Pyrococcus furiosus*. Previous studies in our lab demonstrate that integration events were not polarized to the first repeat in the array adjacent to the leader *in vitro*, but rather integrated at every repeat, and repeat-like sequences. These results indicate that some factor is missing *in vitro* that is present *in vivo* to guide new integration events to the first repeat. The combination of high throughput sequencing of micrococcal nuclease digested DNA and adaptation studies allowed for us to identify archaeal histones as factors important in guiding integration events to the leader adjacent repeat in CRISPR arrays. These results reveal a new mechanism for the integration of new spacers into the CRISPR array in an archaeal CRISPR system.

Introduction

The CRISPR-Cas system is an adaptive immune system found in prokaryotes. This defense pathway is composed of three major steps. The first step is adaptation, in

which Cas proteins acquire DNA (spacer) from an invading mobile genetic element (MGE) and insert the spacer into a CRISPR array on the host genome at the first repeat adjacent to the CRISPR leader in a polarized manner [1]. The next step is crRNA biogenesis, in which the CRISPR array is transcribed via a promoter on the CRISPR leader and processed to produce mature crRNAs that then associate with an effector complex to carry out the final step, interference, to guide the complex and silence target invaders [2-7].

Adaptation has several steps that the mechanism behind which are still unclear. Integration of new spacers occurs in a polarized fashion at CRISPR arrays *in vivo* for most CRISPR-Cas systems [8-12]. However, for Type I systems, *in vitro* integration assays reveal integration occurring at all repeats [13, 14]. In previous studies in our lab, we demonstrated that integration events *in vitro* occurred at every repeat, and repeat-like sequences outside of the CRISPR array [14]. However, *in vivo*, integration occurs primarily at the leader adjacent repeat, with no ectopic events observed at downstream repeats. Therefore, there must be some factor that guides integration events to the first repeat *in vivo* that is not present in our *in vitro* reactions. In Type I-E and Type I-F systems, they discovered that integration host factor (IHF) was involved in the integration of new spacers at the first repeat *in vitro* and *in vivo* [15, 16]. It binds to a consensus sequence on the CRISPR leader, and bends the DNA 120°, and has been demonstrated to interact with the Cas1 integrase during integration of new spacers into the CRISPR array [15-18].

It is still unknown how polarized integration of new spacers occurs at archaeal CRISPR arrays *in vivo*. It has been demonstrated that in *Saccharolobus solfataricus*

(formerly *Sulfolobus solfataricus*), lysate added to *in vitro* integration assays allowed for specific integration at the first repeat, while integration occurred at various repeats in the absence of the lysate [19]. However, they were unable to identify which factor (or factors) were causing this specificity.

We sought out to seek the answer to this question in our model organism, *Pyrococcus furiosus* (*P. furiosus*). *P. furiosus* is an anaerobic hyperthermophile with three CRISPR-Cas systems (type I-A, type I-B, and type III-B), and seven CRISPR arrays on the genome [20-26]. We decided to analyze CRISPR leaders to find potential binding factors that may be involved in polarizing integration to the first repeat. Through a combination of micrococcal nuclease assays and sequencing of the DNA protected regions from these assays, we identified two major peaks on CRISPR leaders adjacent to the first repeat on the CRISPR array. The size and underlying sequence of these regions led us to identify a potential factor that may be binding to these leaders, archaeal histones.

Histones are found in most archaea and are highly conserved with eukaryotic histones [27-30]. While similar to eukaryotic histones in structure and amino acid sequence, they differ in that most lack the flexible N-terminal tails that protrude from eukaryotic histones which allow for post-translational modifications [27]. They also differ in that eukaryotic histones can only form heterodimers and primarily tightly wrap 147 bp of DNA, while archaeal histones can form homodimers and (due to the symmetry of the histone dimer structure it forms) continuously wrap DNA depending on the sequence it is binding to, wrapping 30 bp to >300 bp of DNA in what is referred to as a superhelical ramp [27, 29, 31, 32].

P. furiosus has two histones, archaeal histones A and B. In this study we perform a series of experiments to test our hypothesis that histones may be involved with polarizing integration events to the leader adjacent repeat. Here we study the influence of histones on integration events at CRISPR arrays both *in vivo* and *in vitro*. We find that the absence of histone A *in vivo* decreases the frequency of integration events at the first repeat. Additionally, the presence of these archaeal histones in *in vitro* integration assays polarize integration events to the leader adjacent repeat. Overall, we have found that histones appear play a substantial role in the polarized integration of new spacers at CRISPR arrays in *P. furiosus*.

Results

Potential conserved binding sites on CRISPR leaders

We sought to find any potential binding factors found on CRISPR leaders that could be involved with the integration of new spacers into the CRISPR array. To find these factors, we performed micrococcal nuclease assays on wildtype cell pellets from *Pyroccocus furiosus* (*P. furiosus*). We observed a laddering pattern similar to that observed in other archaea with histones (Figure 3.1A). This laddering pattern increased in 30 bp increments which indicates another histone dimer in complex with each other [27, 29, 33]. The smallest band observed was around 30 bp, which is the size of a single histone dimer binding to DNA. However, the 60 bp band was more intense, indicating that two histone dimers binding to DNA is more stable than one histone dimer, which was observed in other archaea as well [27, 33, 34]. The 60 bp and 90 bp bands were consistently the most prominent in multiple assays, likely indicating that either two or

three histone dimers are the most stable when binding to DNA in *P. furiosus* (Figure 3.1A). The largest micrococcal nuclease digested products observed were around 180 bp, but these were faint and indicated more rare events.

The 30, 60, 90, and 120 bp bands were excised and underwent high throughput sequencing (HTS) to observe what regions of DNA were protected in the micrococcal nuclease assay (Figure 3.1A). Due to our interest in potential binding factors on CRISPR leaders, this is where we focused our analysis. We identified two consistent binding regions on the majority of CRISPR leaders, specifically in the region spanning the leader-repeat region (L-R), and a region about 100 bp upstream of the first repeat (L) (Figure 3.1B, Figure 3.2, Figure 3.3). Each of these peaks were about 60 bp wide, which is reminiscent of two histone dimers binding to this region (Figure 3.3). Also, the underlying sequence of these peaks indicated that a histone may prefer to bind here, as demonstrated by the AT rich regions on the end of the peaks and GC rich content within. Due to the CRISPR leaders in *P. furiosus* being highly conserved between each other, it was predictable that the sequences of these binding regions would be similar amongst arrays, again indicating that it is potentially the sequence that is leading to binding/DNA protection in these regions (Figure 3.3) [26, 35, 36]. However, the CRISPR4 and CRISPR8 arrays vary the most from the other 5 arrays in the L region (Figure 3.3). There were also peaks of DNA protected reads in the CRISPR arrays, most of them 90 bp +, spanning spacers and repeats (Figure 3.1B, Figure 3.2). These variable spacers may contain sequences conducive to histone binding [31, 33, 37, 38].

These consistent peaks of DNA protection found on most CRISPR leaders led us to predict that histones may play a role in integration at the first repeat on *P. furiosus*

CRISPR arrays. Histones in archaea are closely conserved with eukaryotic histones, as observed with the H3-H4 dimer from eukaryotic histones and a homodimer of histone B from *Methanothermus fervidus* (Figure 3.4A). However, the symmetry of archaeal histones allows for them to stack as dimers, and form what has been referred to as a hypernucleosome or a superhelical ramp, explaining the laddering of DNA protected regions we observe from micrococcal nuclease assays (Figure 3.4B, Figure 3.1A) [27, 31]. Depending on the underlying sequence of the genome, they have been observed to bind anywhere from 30-300 bp. Archaeal histones have extremely conserved histone folds, and *P. furiosus*, *Thermococcus kodakarensis*, and *Methanothermus fervidus* individual histone monomers appear to be almost identical both in structure and amino acid sequence to each other (Figure 3.4A, Figure 3.4C, Figure 3.4D). They have been shown to form both homodimers and heterodimers, in contrast to eukaryotic histones that can only form heterodimers (Figure 3.4A) [27, 28, 31, 39].

Histones play a potential role in integration at the first repeat *in vivo*

To test if histones play a role in directing new integration events to the first repeat, we decided to make histone gene deletions in *P. furiosus*. We deleted the histone genes individually, however, we were unable to delete both histone A and B in the same strain. This suggested that histone-mediated events in *P. furiosus* were essential, and the deletion of both histones has proved to be an impossible task in *Thermococcus kodakarensis* as well (closely related to *P. furiosus*) [28]. Therefore, we tested strains with either histone A or histone B deleted. We made these deletions in wildtype strains of *P. furiosus* and a strain with only the type I-B effector complex (the other two effector

complexes deleted from this strain). We chose to include the type I-B only strain due to this effector complex's role in primed adaptation and we are utilizing the pT33.3 conjugative plasmid discussed in chapter 2. From our studies, the pT33.3 plasmid induces a robust primed adaptation response and causes hyperadaptation to occur at the first repeat *in vivo* (Chapter 2). This allows for us to observe any major changes in integration events at the first repeat *in vivo*. The type I-B only strain demonstrates an even larger effect of hyperadaptation, due to the crRNAs being diluted due to the other two effector complexes (type I-A and type III-B) in wildtype strains (Chapter 2) [21].

We performed a conjugation assay using the pT33.3 plasmid and mated donor cells with recipient wildtype, type I-B only strains, as well as with those strains with the individual histone genes deleted in each of those backgrounds (Chapter 2, Figure 2.1A). We then observed adaptation at the first repeat in multiple arrays by amplifying from the leader of the CRISPR array to the existing spacer 1 to observe any new repeat-spacer units added to the array. As expected, we saw hyperadaptation occurring in all strains with the pT33.3 conjugative plasmid present (Figure 3.5A). However, the deletion of histone A appears to significantly decrease integration events at the first repeat on multiple CRISPR arrays (Figure 3.5A). These deletions did not cause total loss of integration in any strain. Archaeal histones can also form homodimers in addition to heterodimers and therefore it makes sense that integration may not be lost if one histone is deleted [27, 40].

We then questioned if the deletion of these histones led to any ectopic integration events to occur at downstream repeats *in vivo*. We amplified from spacer 1 to spacer 3, and from spacer 3 to spacer 6 on the CRISPR7 array to observe any expanded bands to

indicate new integration events. We did not observe any new integration events, only unexpanded bands in all strains (Figure 3.5B). This indicates that integration is still occurring at the first repeat *in vivo* with these histone deletion strains. To detect any changes in DNA protected patterns on the CRISPR leaders due to these histone deletions (specifically at the L and L-R regions (Figure 3.3)), micrococcal nuclease assays were performed on wildtype strains with either histone A or B deleted. There were no significant changes in micrococcal nuclease digested DNA patterns when analyzing agarose gels or HTS data (Figure 3.5C, Figure 3.5D). However, we are potentially biasing results due to analysis of these protected 60, 90, and 120 bp reads, which are more than likely histone protected DNA regions (heterodimer or homodimer protected reads), and anything that binds will be shown due to the sensitivity of this assay, so it would be difficult to observe any minor changes at these leader regions. Our ability to assess if histones are involved in adaptation is hindered by the failure to delete both histones *in vivo*.

Histones guide new spacers to the first repeat *in vitro*

To circumvent the issue of the inability to delete both histones *in vivo*, we decided to try an *in vitro* approach. We have performed successful *in vitro* integration assays in previous studies [14], however now we are including histones in this assay. We are using a pCRISPR plasmid with a full CRISPR7 leader and a three repeat array [14] in either the absence or presence of archaeal histones (Figure 3.6A). The Cas1-Cas2 integrase complex (with a prespacer) will then be added to the reaction to facilitate integration events into the pCRISPR plasmid. In the absence of histones, we expect to observe no

bias and observe integration events at every repeat. However, if histones do play a role in directing integration events to the first repeat, this would create a bias for repeat 1 (Figure 3.6). To observe integration events on the plasmid, we are using a PCR based approach. In #1 reactions, amplification from the specific prespacer integrated into the array to a region beyond the CRISPR array will yield PCR products ranging from integration events at repeat 1 to be the largest size, to integration at repeat 3 to be the smallest size (Figure 3.6A, Figure 3.6B). The opposite reaction is performed for the #2 reactions, in which amplification occurs from the CRISPR leader to the specific prespacer which make repeat 1 integration events the smallest size PCR product and repeat 3 the largest (Figure 3.6A, Figure 3.6B).

Using this assay, we decided to perform our initial tests using purified *Thermococcus kodakarensis* histones A and B. These histones were readily available from our collaborators, and they are extremely conserved with *P. furiosus* histones (Figure 3.4C, Figure 3.4D, Figure 3.7). We tested combining *Thermococcus kodakarensis* histones A and B in one reaction (to form possible heterodimers) and A and B separately (to form possible homodimers) on a plasmid with the full CRISPR7 leader and an array with three repeats (Figure 3.8A). In the absence of histones (0 μM), we observed integration at all three repeats as we have seen in the past (Figure 3.8B). As the concentration of histones increased, there was a marked increase of the amount of integration occurring at the first repeat (blue box=R1), as opposed to the other repeats present in the array (Figure 3.8B). We observed this by amplifying from specific prespacer to beyond the CRISPR array (#1 reactions) and also from the leader to the specific spacer (#2 reactions), allowing us to observe these integration patterns from two

directions, with repeat 1 either being the largest or smallest PCR product observed, respectfully (Figure 3.8A, Figure 3.8B). We also quantified these data by measuring the intensity of each band (integration event) at each concentration of histones added using ImageJ software and calculating the percentage of each band when compared to the total intensity of all bands within a given lane. This aligned with our PCR/agarose gel results, as the concentration of histones increased, so did the preference for integration events at repeat 1 (Figure 3.9C).

Due to the success with using *Thermococcus kodakarensis* histones, we then proceeded to using purified *P. furiosus* histones A and B (Figure 3.7). The same *in vitro* integration assay was carried out as described previously, however we proceeded with the #1 reaction as it gave more clear results when compared to the #2 reactions (Figure 3.9A). In the absence of histones, we could observe integration events at all three repeats (Figure 3.9B). We then observed the same drive to the first repeat that we observed with the *Thermococcus kodakarensis* histones as we increased the concentration of *P. furiosus* histones (Figure 3.9B). This was also demonstrated with quantitative analysis (Figure 3.9C).

Preference for the first repeat is a histone specific effect

We then tested DNA binding protein TrmBL2 (Figure 3.7) to demonstrate that the guidance of integration events to repeat 1 is specific to histones. TrmBL2 has been found to compete with histones for binding DNA under certain conditions [37, 41] and can form stiff filamentous structures and prevent the packaging of DNA [41]. For these assays, we are using a pCRISPR plasmid with a full CRISPR7 leader and a longer CRISPR array

with 11 repeats and 10 spacers (Figure 3.10A). In the absence of histones, integration is observed at multiple repeats in #2 reactions (Figure 3.6B), with repeat 6 visualized as the largest product (Figure 3.10B). As demonstrated with Figure 3.10B, histones continue to influence integration events to occur at repeat 1 with this larger array, which is also demonstrated with quantitative analysis (Figure 3.10C). The addition of TrmBL2 to the *in vitro* integration reaction did not replicate this phenomenon we have been observing with the histones. Instead, TrmBL2 appears to inhibit integration at repeats with increasing concentration to fainter bands overall (Figure 3.10D). The quantitative analysis is consistent with PCR results, as preference for repeat 1 is not seen, and rather repeat preference is relatively the same with increasing concentrations of TrmBL2 (Figure 3.10E). This further enhances our hypothesis that histones are specifically involved with integration at the first repeat, and not just any DNA binding protein causes this effect.

Histone directed integration events at repeat 1 is observed in both supercoiled and linear substrates

We were curious as to why we were not observing integration at repeats past repeat 6 when amplifying from the leader to the specific spacer in #2 reactions (Figure 3.10B, right panel). It should be possible to observe integration at every repeat *in vitro*, but perhaps that particular reaction was not sensitive enough to capture those larger integration events. Therefore, we designed primers to amplify from spacer 10 on CRISPR array to the specific prespacer with the #1 reaction, to make integration events further downstream on the arrays the smallest PCR product rather than the larger one (Figure 3.11A). Another question we sought to answer was if the integration of new spacers

depended on the topology of the DNA being used as substrate for integration to occur. We performed *in vitro* integration assays on both supercoiled and linear DNA substrates, and also amplified from either spacer 5 or spacer 10 on the CRISPR array to the specific spacer on the pCRISPR plasmid. In this assay we decided to use 0 and 3.4 μM of histones because we have found from our previous assays that 3.4 μM was optimal for histone guided integration at the first repeat. As you can see in Figure 3.11B, both supercoiled and linear DNA substrates can be used to integrate new spacers into the array. We were also able to visualize integration events at repeats 1-10 in the absence of histones when amplifying from the specific prespacer to spacer 10 on the array (Figure 3.11B). The preference for repeat 1 in the presence of histones continues to be clear, demonstrated by both PCR products ran on agarose gels and quantitative analysis (Figure 3.11B (red asterisk), Figure 3.11C).

Preference for the first repeat in the presence of histones is not CRISPR7 locus specific

All experiments performed thus far have all been performed with the pCRISPR plasmid with the CRISPR7 leader and array. To make sure this was not a specific effect contained to the CRISPR7 plasmid, we performed this assay on multiple CRISPR array plasmids, all containing 11 repeats and 10 spacers in their arrays. As observed before, integration in the presence of higher concentrations of histones caused preference for repeat 1 to appear, demonstrated by both PCR products from #1 reactions and quantitative analysis of these gels (Figure 3.12A, Figure 3.12B, red asterisk). However, this is only one experiment and needs to be replicated to confirm results.

HTS data is consistent with previous observations of histone directed integration events

To validate this preference for the first repeat in the presence of histones that we have been observing *in vitro*, we went forward with HTS on these integrated plasmid samples to further confirm this phenomenon. Sequencing analysis indicated that integration was occurring at most repeats *in vitro* in the absence of histones (0 μM) (Figure 3.13A). Repeats 1, 2, 3, 4, and 11 appear to be integrated more frequently than others, with lower number of reads detected for the other repeats. As histone concentration increased, so did the preference for repeat 1 (Figure 3.13A-D). At 3.4 μM, integration events at repeats 2, 3, 4, and 11 have significantly decreased, while integration at repeat 1 remains high (Figure 3.13D). These data confirm that histones do guide integration events to the first repeat *in vitro*.

Discussion

In our attempts to understand how the integration of new spacers were polarized to the first repeat in archaea, we were led to analyze potential areas of DNA protection on the CRISPR leaders. We already know that IHF in bacterial systems binds to the CRISPR leader to bend the DNA at 120°, and also has interactions with the Cas1 integrase during this bending, allowing for the Cas1-Cas2 complex to successfully integrate new spacers into the first repeat on the CRISPR array [15, 16, 42, 43]. There is a consensus sequence (specific A-T rich sites) that was found on leaders in Type I-E and Type I-F systems [15,

16], among others as well. However, there were no known homologues of IHF in our archaeal system.

Through micrococcal nuclease assays, we found similar patterns of DNA protection that had been found in other archaeal organisms with histones (Figure 3.1A) [31, 33, 34, 40, 44, 45]. After sequencing of protected DNA reads on CRISPR leaders, we found the major peaks on the majority of CRISPR leaders in *P. furiosus*. One such peak spanned the region where the leader and the first repeat meet (L-R), and the other was about 100 bp upstream of the first repeat (L) (Figure 3.3). Both peaks were about 60 bp in length. We know from our work that 60 bp and 90 bp bands were dominant in our micrococcal nuclease assays (Figure 3.1A), and this is the predicted size of two or three histone dimers binding to the DNA [27, 28, 34, 45].

While it may not necessarily be a histone binding there, as micrococcal nuclease assays can reveal any regions of protection by DNA binding proteins, it's size and underlying sequence led us to believe that it was indeed a histone. Histones were widely believed to binding nonspecifically, but the literature states that they do have preferred sequences they like to bind to [30, 33, 46]. These include helical repeats of A/Ts and G/Cs dinucleotides, and AT-rich oligonucleotides appear to repel histone binding [46]. It is also found that the end sequences of histone protected regions are more AT rich, allowing for some prediction of whether there may be a histone binding site or not [33]. When analyzing the sequences of these two peaks, the underlying sequence appears to be favorable of histone binding, and the ends of these peaks are also relatively AT rich (Figure 3.3). However, it should be noted that there are a few stretches of A/T tracts within the L and L-R peaks we are observing on these leaders which are known to cause

rigidity of the DNA and potentially repel histones (Figure 3.3). However, these regions are G/C rich as well. Overall, these factors allow for us to predict that there are two histone dimers binding to the leader at these separate sites.

The CRISPR4 and CRISPR8 loci sequences were by far the least conserved when compared to the other loci at the L region (Figure 3.3). The binding peak at the CRISPR4 array at the L region is extremely low, but this is not reflected with the CRISPR8 array (Figure 3.3). This does not appear to affect the integration of spacers into the CRISPR4 array, however, we do know that the CRISPR8 array does not integrate new spacers at as high of a frequency when compared to the other arrays [22]. It is still unknown how important these sites are for integration to occur, but it is interesting to note. To understand the importance of these sites, future experiments could entail the mutation or deletion of these sites to observe if histones are repelled or unable to bind *in vivo*. Plasmids with these mutations could also be utilized in *in vitro* assays to observe any changes in integration patterns. These experiments could explain if the sequence itself is important for histone binding, or potentially if these sites cooperate with each other for successful integration at the first repeat.

Our *in vivo* approach was simply to delete the histones to observe any effects on integration at the CRISPR array. Unfortunately, we were unable to delete both histone A and B in the same strain of *P. furiosus*. This was the same case in *Thermococcus kodakarensis* [28]. They were, however, able to create single histone deletion strains, and we found that to be the case for our system as well. However, the presence of one histone gene still makes it difficult to see any effects on adaptation due to the ability of these histones to form homodimers. In *Thermococcus kodakarensis*, they found that when

single histone deletions were created there were differences in levels of transcription for certain genes, and the deletion of histone A no longer allowed for transformation of DNA into *Thermococcus kodakarensis* [28]. These differences allow for us to assume that there are indeed different roles for each histone within the cell, and that it may be the case for *P. furiosus* histones as well. The deletion of histone B did not cause any major effects of integration on the CRISPR arrays (perhaps slightly), however, the deletion of histone A did decrease the frequency of integration at the arrays significantly (Figure 3.5A). This provided evidence that perhaps there is a role for histones to be involved with integration at the first repeat *in vivo*, and we can further explore this *in vitro*.

Thermococcus kodakarensis histones A and B are extremely conserved with *P. furiosus* histones A and B (~90% identity) (Figure 3.4C, Figure 3.4D). Our collaborators work frequently with purified histones A and B from *Thermococcus kodakarensis* and were willing for us to use these proteins in our *in vitro* integration assays for our initial tests. Excitingly, we found that these histones (even with the combination of purified *P. furiosus* Cas1 and Cas2 proteins) caused a preference for the first repeat in the three repeat CRISPR array. It should also be noted that in the presence of histone A alone, that integration at the first repeat was stronger (more intense band) than with histone B alone (Figure 3.8B). This may be connected to the *in vivo* data, as the frequency of integration was significantly diminished in the absence of *P. furiosus* histone A (Figure 3.5A). Perhaps these data provide evidence that histone A plays a greater role in the integration of new spacers at the CRISPR arrays than histone B, although histone B still drives integration to the first repeat (Figure 3.5A, Figure 3.8B).

When using purified *P. furiosus* histones A and B, there was a striking difference in integration patterns in the absence vs the presence of histones, just like we had observed with histones A and B from *Thermococcus kodakarensis* (Figure 3.9B). The quantitation of bands with ImageJ software in these assays confirmed these results. Histones are known to wrap DNA to help compact the genome. In this case, perhaps it has a similar function to IHF, an architectural protein that bends the DNA at the leader adjacent to the first repeat to cause a signal for the Cas1-Cas2 integrase complex to recognize and integrate new spacers there [15, 16, 47].

TrmBL2 was introduced into the *in vitro* integration assays to provide evidence that this preference for the first repeat was a specific effect caused by the addition of histones, not just any DNA binding protein. As mentioned previously, TrmBL2 had been shown to compete with histone binding under certain conditions and forms stiff filamentous structures which antagonize the compaction of the genome [37, 41, 48]. During these *in vitro* integration assays it was found integration was not preferred at the first repeat in the presence of TrmBL2, in fact, integration overall was diminished at all repeats with increasing concentrations (Figure 3.10D). This decrease of integration events with increasing concentrations may be due to TrmBL2 coating the DNA all over the plasmid, including at the CRISPR array, and blocking integration from taking place. In *in vivo* studies, TrmBL2 is involved in gene repression, potentially blocking transcription factors or RNA polymerase by binding to promoters [41]. Also, as noted above, its stiff filamentous structures antagonize compaction of the genome, no longer allowing for DNA to bend, which may be deleterious for new integration events to occur.

We also observed that both supercoiled and linear DNA were able to be used as substrates for the Cas1-Cas2 integrase complex to integrate new spacers into the repeats. Both supercoiled and linear forms of DNA were found to be able to reproduce the same integration patterns in both the absence and presence of histones (Figure 3.11). In *E. coli*, DNA is maintained in a negatively supercoiled state while *P. furiosus* DNA is maintained in a positively supercoiled state [49]. The pCRISPR plasmids used in our *in vitro* integration assays are most likely in a negatively supercoiled state (as they were maintained in *E. coli*), but histones have been demonstrated to induce positive supercoiled substrates when bound [50]. Therefore, we had predicted that perhaps the histones were causing a more preferred substrate for the Cas1-Cas2 integrase complex to integrate into due to the typical positively supercoiled state of *P. furiosus* DNA and causing this first repeat preference. However, we observed the same first repeat preference with linear substrate, which cannot be made into a positively supercoiled state in the presence of histones. Therefore, we were able to confirm that this was not an artificial phenomenon due to the topology of the substrate, and the Cas1-Cas2 integrase complex appear to be influenced by histone specific effects.

These data provide evidence that histones do play a role in integrating new spacers at the first repeat, both *in vitro* and *in vivo*. However, there is still the question remaining, are these histones the specific DNA binding proteins creating those protected DNA peaks on the CRISPR leaders? Future studies could entail the addition of histones to these pCRISPR plasmids in our *in vitro* integration assays and performing micrococcal nuclease assays. By sequencing these protected reads, we could provide direct evidence that these peaks we observe on the CRISPR leaders are histones. Other experiments could

include adding histones to purified genomes and observing if we see the same DNA protection patterns and performing *in vitro* integration assays and subsequent PCR with all 7 CRISPR loci present on the genome.

While it is clear that histones do play a role in the integration of new spacers at the first repeat, it is not a perfect phenomenon. There are still integration events occurring at other CRISPR repeats in these *in vitro* integration assays. This suggests that there may be other proteins involved in this process that we are not aware of. In our *in vivo* experiments, integration has only ever been observed at the first repeat in *P. furiosus* (Figure 3.5A, Figure 3.5B, Chapter 2) [21-23]. It is possible that the cooperation of several proteins may be required to have sole integration at the first repeat *in vivo*, which we are unable to replicate *in vitro*.

There are other proteins in *P. furiosus* that could be involved in adaptation (other DNA binding proteins, proteins involved in DNA repair etc.) like there has been observed in other systems [15, 51, 52]. Now that we have utilized the novel pT33.3 conjugative plasmid, we can visualize adaptation events at the first repeat in first round PCR to the point where we can easily observe changes to the hyperadaptation pattern if we find another factor that could be involved in spacer acquisition or spacer integration into the CRISPR arrays (Chapter 2). For example, the NurA-HerA (Nuclease of Archaea, Helicase of Archaea) protein complex is involved in homologous recombination/DNA repair in archaea [53-57]. The nuclease and helicase of this complex targets double stranded DNA breaks, and resections the break back to a point where homologous recombination of the exposed single stranded DNA can undergo homologous recombination. This gene pair has also been shown to be in close association with several

CRISPR genes in the *Pyrobaculum* species [54]. This may indicate a function of the NurA-HerA complex in adaptation, perhaps serving a similar function that RecBCD has been shown in E. coli [51]. These genes are in *P. furiosus* and could be potential factors to test in further adaptation studies in combination with the pT33.3 conjugative plasmid. The results from this chapter broaden our understanding behind the mechanisms behind CRISPR adaptation in archaeal systems.

Material and Methods

Micrococcal nuclease assays for genomic DNA

Pyrococcus furiosus cultures were grown in 250 mL cultures in defined media to either exponential or stationary phase, pelleted, and flash frozen. These pellets were resuspended in MNase buffer (50 mM Tris-HCl pH 8.0, 100 mM NaCl and 1 mM CaCl2) then lysed mechanically by freeze/thawing with liquid nitrogen while grinding with a mortar and pestle. Lysate was then clarified by centrifuging down cell debris, and then adding RNaseA to a final concentration of 2 mg/500 uL and incubating at 37°C for one hour. 500 Units of MNase was then added to the lysate, and 100 uL fractions were taken out at various timepoints. MNase reactions were then stopped by the immediate addition of phenol/chloroform/isoamyl alcohol (25:24:1), shaken vigorously, and centrifuged at 4°C for 5 minutes at 15,000 rpm. The aqueous layer was then taken into a separate 1.5 mL tube with 200 μL of 1 M Tris-HCl pH 8.0. DNA was precipitated and recovered by ethanol washes. Resuspend DNA pellet in 10 mM Tris-HCl pH 8.0. DNA was then visualized on a 4% Nusieve gel.

Conjugation Assays

The Thermococcus donor strains were grown up in complex media at 85°C and the Pyrococcus recipient strains were grown in defined selective media at 95°C, all grown overnight for ~16 hours. We always use freshly grown strains for conjugation assays. 5 mL of each overnight culture was pipetted into a 15 mL conical tube. Cultures were centrifuged at ~3400 RPM for 20 minutes. Supernatant was discarded and all remaining liquid media was tapped out. The tip of the pipette tip was cut to prevent shearing of cells, and cells were resuspended in 50 μL of 1X ASW (if plating on complex media first) or 1X Pfu Base Salts (if going straight to selective media). Recipient cells were mixed with the appropriate donor in a 1.5 mL tube, for a total of 100 μL cell resuspension. The entire cell mixture was spotted in the center of the selective media plate. Plates were grown anaerobically for three nights at 95°C. The spot of growth was scraped into 5 mL of defined selective liquid media and grown overnight at 95°C.

Adaptation Assays

A population of cells are picked (~20 colonies) and added to defined, selective, liquid media and allowed to grow at 95°C overnight (~16 hours). 1 mL of the overnight culture is then used to extract genomic DNA from cells using the Quick-DNA miniprep kit (Zymo Research). 10 ng of genomic DNA is used in PCR reactions. To determine whether or not adaptation/integration of new spacers has taken place on the CRISPR arrays, primers that are located on the leader and first known spacer of the array are utilized. PCR products are then run on a 2.5% agarose gel (120 V for ~48 minutes) to

separate unexpanded bands (no integration of new spacers) from expanded bands (successful integration of new spacer).

Strain construction

To generate the strains with the histone deletions, we used homologous recombination of PCR products, as described previously [58].

***in vitro* integration assay**

Full CRISPR leaders and an array consisting of 10 full spacers and 11 repeats were added to Blunt II TOPO vectors via blunt end ligation. These plasmids (pCRISPR) were then used in *in vitro* integration assays. In a 500 μL tube, a final concentration of 1 μM Cas1, 1 μM Cas2, 1 mM DTT, 10 mM $MgCl_2$, and 50 nM of prespacer (5 nt overhangs) was incubated in reaction buffer (20 mM Tris, 100 mM KCl, 5% glycerol, pH 7.5) for 1 hour at 4°C. During this time, single reaction tubes for each titration of histones used in experiments were prepared. In 500 μL tubes, a final concentration of 50 nM pCRISPR and varying concentrations of histones or TrmBL2 were added (0-5.1 μM) were incubated at 70°C for 1 hour. After this incubation, the Cas1-Cas2 reaction mix was added to the pCRISPR tubes for a final volume of 20 μL. This combined reaction was then incubated for another hour at 70°C. The reactions were stopped, and proteins were degraded by addition of EDTA and proteinase K (Life Technologies). DNA was purified from reactions through use of the DNA Clean and Concentrator Kit (Zymo Research, Irvine, CA, USA), and eluted in a volume of 15 μL of TE Buffer, and 1 μL was used from this to add to a PCR reaction to detect where new spacers integrated on the CRISPR

arrays. Primers were made to amplify from the CRISPR leader to the prespacer, or from Spacer 5 or 10 on the CRISPR array to the prespacer. PCR products were then run on a 4% Nusieve agarose gel at 120 V for 45-50 minutes.

ImageJ Analysis

ImageJ software was downloaded via the NIH website (https://imagej.nih.gov/ij/download.html). The average intensity of each band (integration event at a repeat) in each lane was acquired. Depending on how bands were observed, the percentage of a given band was calculated by taking the band in question and dividing its average intensity by the total band intensities observed for that lane and then multiplied by 100 to get the percentage of the intensity of that specific band when compared to all other bands in that lane.

References

1. Barrangou, R., et al., *CRISPR provides acquired resistance against viruses in prokaryotes.* Science, 2007. **315**(5819): p. 1709-12.
2. Brouns, S.J., et al., *Small CRISPR RNAs guide antiviral defense in prokaryotes.* Science, 2008. **321**(5891): p. 960-4.
3. Deng, L., et al., *A novel interference mechanism by a type IIIB CRISPR-Cmr module in Sulfolobus.* Mol Microbiol, 2013. **87**(5): p. 1088-99.
4. Elmore, J.R., et al., *Bipartite recognition of target RNAs activates DNA cleavage by the Type III-B CRISPR-Cas system.* Genes Dev, 2016. **30**(4): p. 447-59.
5. Garneau, J.E., et al., *The CRISPR/Cas bacterial immune system cleaves bacteriophage and plasmid DNA.* Nature, 2010. **468**(7320): p. 67-71.
6. Hale, C.R., et al., *RNA-guided RNA cleavage by a CRISPR RNA-Cas protein complex.* Cell, 2009. **139**(5): p. 945-56.
7. Zetsche, B., et al., *Cpf1 Is a Single RNA-Guided Endonuclease of a Class 2 CRISPR-Cas System.* Cell, 2015. **163**(3): p. 759-71.
8. Datsenko, K.A., et al., *Molecular memory of prior infections activates the CRISPR/Cas adaptive bacterial immunity system.* Nat Commun, 2012. **3**: p. 945.
9. Diez-Villasenor, C., et al., *CRISPR-spacer integration reporter plasmids reveal distinct genuine acquisition specificities among CRISPR-Cas I-E variants of Escherichia coli.* RNA Biol, 2013. **10**(5): p. 792-802.

10. Levy, A., et al., *CRISPR adaptation biases explain preference for acquisition of foreign DNA.* Nature, 2015. **520**(7548): p. 505-10.

11. Swarts, D.C., et al., *CRISPR interference directs strand specific spacer acquisition.* PLoS One, 2012. **7**(4): p. e35888.

12. Yosef, I., M.G. Goren, and U. Qimron, *Proteins and DNA elements essential for the CRISPR adaptation process in Escherichia coli.* Nucleic Acids Res, 2012. **40**(12): p. 5569-76.

13. Nunez, J.K., et al., *Integrase-mediated spacer acquisition during CRISPR-Cas adaptive immunity.* Nature, 2015. **519**(7542): p. 193-8.

14. Grainy, J., et al., *CRISPR repeat sequences and relative spacing specify DNA integration by Pyrococcus furiosus Cas1 and Cas2.* Nucleic Acids Res, 2019.

15. Nunez, J.K., et al., *CRISPR Immunological Memory Requires a Host Factor for Specificity.* Mol Cell, 2016. **62**(6): p. 824-33.

16. Fagerlund, R.D., et al., *Spacer capture and integration by a type I-F Cas1-Cas2-3 CRISPR adaptation complex.* Proc Natl Acad Sci U S A, 2017.

17. Lee, H. and D.G. Sashital, *Creating memories: molecular mechanisms of CRISPR adaptation.* Trends Biochem Sci, 2022.

18. Yoganand, K.N., et al., *Asymmetric positioning of Cas1-2 complex and Integration Host Factor induced DNA bending guide the unidirectional homing of protospacer in CRISPR-Cas type I-E system.* Nucleic Acids Res, 2016.

19. Rollie, C., et al., *Prespacer processing and specific integration in a Type I-A CRISPR system.* Nucleic Acids Res, 2018. **46**(3): p. 1007-1020.

20. *<Shiimori-Cas4's function in Pyrococcus furiosus.pdf>.*

21. Garrett, S., et al., *Primed CRISPR DNA uptake in Pyrococcus furiosus.* Nucleic Acids Res, 2020.

22. Shiimori, M., et al., *Role of free DNA ends and protospacer adjacent motifs for CRISPR DNA uptake in Pyrococcus furiosus.* Nucleic Acids Res, 2017. **45**(19): p. 11281-11294.

23. Shiimori, M., et al., *Cas4 Nucleases Define the PAM, Length, and Orientation of DNA Fragments Integrated at CRISPR Loci.* Mol Cell, 2018. **70**(5): p. 814-824 e6.

24. Majumdar, S., et al., *Target DNA recognition and cleavage by a reconstituted Type I-G CRISPR-Cas immune effector complex.* Extremophiles, 2017. **21**(1): p. 95-107.

25. Majumdar, S., et al., *Three CRISPR-Cas immune effector complexes coexist in Pyrococcus furiosus.* RNA, 2015.

26. Terns, R.M. and M.P. Terns, *The RNA- and DNA-targeting CRISPR-Cas immune systems of Pyrococcus furiosus.* Biochem Soc Trans, 2013. **41**(6): p. 1416-21.

27. Henneman, B., et al., *Structure and function of archaeal histones.* PLoS Genet, 2018. **14**(9): p. e1007582.

28. Cubonovaa, L., et al., *An archaeal histone is required for transformation of Thermococcus kodakarensis.* J Bacteriol, 2012. **194**(24): p. 6864-74.

29. Laursen, S.P., S. Bowerman, and K. Luger, *Archaea: The Final Frontier of Chromatin.* J Mol Biol, 2021. **433**(6): p. 166791.

30. Struhl, K. and E. Segal, *Determinants of nucleosome positioning.* Nat Struct Mol Biol, 2013. **20**(3): p. 267-73.

31. *<2018-Science Structure of Histone-based-Luger.pdf>.*

32. Mattiroli, F., et al., *Structure of histone-based chromatin in Archaea.* Science, 2017. **357**(6351): p. 609-612.

33. Maruyama, H., et al., *An alternative beads-on-a-string chromatin architecture in Thermococcus kodakarensis.* EMBO Rep, 2013. **14**(8): p. 711-7.

34. *<Tko histone binding (Santangelo and Widom).pdf>.*

35. Norais, C., et al., *Diversity of CRISPR systems in the euryarchaeal Pyrococcales.* RNA Biol, 2013. **10**(5): p. 659-70.

36. Abreu, E., et al., *TIN2-tethered TPP1 recruits human telomerase to telomeres in vivo.* Mol Cell Biol, 2010. **30**(12): p. 2971-82.

37. Maruyama, H., et al., *Histone and TK0471/TrmBL2 form a novel heterogeneous genome architecture in the hyperthermophilic archaeon Thermococcus kodakarensis.* Mol Biol Cell, 2011. **22**(3): p. 386-98.

38. Morozov, A.V., *Sequence determinants of histone-DNA binding preferences: comment on "Cracking the chromatin code: precise rule of nucleosome positioning" by Edward N. Trifonov.* Phys Life Rev, 2011. **8**(1): p. 62-3; discussion 69-72.

39. Decanniere, K., et al., *Crystal structures of recombinant histones HMfA and HMfB from the hyperthermophilic archaeon Methanothermus fervidus.* J Mol Biol, 2000. **303**(1): p. 35-47.

40. *<1994-Sandman-Growth phase dependent synthesis of histones Mf.pdf>.*

41. Efremov, A.K., et al., *Transcriptional Repressor TrmBL2 from Thermococcus kodakarensis Forms Filamentous Nucleoprotein Structures and Competes with*

Histones for DNA Binding in a Salt- and DNA Supercoiling-dependent Manner. J Biol Chem, 2015. **290**(25): p. 15770-15784.

42. Wei, Y. and M.P. Terns, *CRISPR Outsourcing: Commissioning IHF for Site-Specific Integration of Foreign DNA at the CRISPR Array.* Mol Cell, 2016. **62**(6): p. 803-4.

43. Wright, A.V., et al., *Structures of the CRISPR genome integration complex.* Science, 2017. **357**(6356): p. 1113-1118.

44. Rojec, M., et al., *Chromatinization of Escherichia coli with archaeal histones.* Elife, 2019. **8**.

45. Sanders, T.J., et al., *Extended Archaeal Histone-Based Chromatin Structure Regulates Global Gene Expression in Thermococcus kodakarensis.* Front Microbiol, 2021. **12**: p. 681150.

46. *<13-Archaeal nucleosome positioning.pdf>*.

47. Wei, Y. and M.P. Terns, *CRISPR Outsourcing: Commissioning IHF for Site-Specific Integration of Foreign DNA at the CRISPR Array.* Mol Cell, 2016. **62**(6): p. 803-804.

48. Wierer, S., et al., *TrmBL2 from Pyrococcus furiosus Interacts Both with Double-Stranded and Single-Stranded DNA.* PLoS One, 2016. **11**(5): p. e0156098.

49. Peter, B.J., et al., *Genomic transcriptional response to loss of chromosomal supercoiling in Escherichia coli.* Genome Biol, 2004. **5**(11): p. R87.

50. Musgrave, D.R., K.M. Sandman, and J.N. Reeve, *DNA binding by the archaeal histone HMf results in positive supercoiling.* Proc Natl Acad Sci U S A, 1991. **88**(23): p. 10397-401.

51. Ivancic-Bace, I., et al., *Different genome stability proteins underpin primed and naive adaptation in E. coli CRISPR-Cas immunity.* Nucleic Acids Res, 2015. **43**(22): p. 10821-30.

52. Radovcic, M., et al., *CRISPR-Cas adaptation in Escherichia coli requires RecBCD helicase but not nuclease activity, is independent of homologous recombination, and is antagonized by 5′ ssDNA exonucleases.* Nucleic Acids Res, 2018. **46**(19): p. 10173-10183.

53. *NurA, a novel 5′-3′ nuclease gene linked to rad50 and mre11 homologs of thermophilic Archaea.*

54. Bernick, D.L., et al., *Comparative genomic and transcriptional analyses of CRISPR systems across the genus Pyrobaculum.* Front Microbiol, 2012. **3**: p. 251.

55. Blackwood, J.K., et al., *Structural and functional insights into DNA-end processing by the archaeal HerA helicase-NurA nuclease complex.* Nucleic Acids Res, 2012. **40**(7): p. 3183-96.

56. Byrne, R.T., et al., *Molecular architecture of the HerA-NurA DNA double-strand break resection complex.* FEBS Lett, 2014. **588**(24): p. 4637-44.

57. De Falco, M., et al., *NurA Is Endowed with Endo- and Exonuclease Activities that Are Modulated by HerA: New Insight into Their Role in DNA-End Processing.* PLoS One, 2015. **10**(11): p. e0142345.

58. Farkas, J., et al., *Recombinogenic properties of Pyrococcus furiosus strain COM1 enable rapid selection of targeted mutants.* Appl Environ Microbiol, 2012. **78**(13): p. 4669-76.

Figure 3.1. DNA protected regions on the CRISPR leaders.

(A) Separation of DNA fragments from micrococcal nuclease digestion of *P. furiosus* chromatin. Different colored boxes indicate different size distributions of potential histone dimer polymers. (**B)** HTS sequencing of DNA protected fragments on the CRISPR loci. The full CRISPR7 is shown in the top panel, and the insets shows the leaders (pale yellow), the first three repeats (black boxes), and 2 spacers (white space between black boxes) for the CRISPR7, CRISPR2, CRISPR4, and CRISPR6 loci.

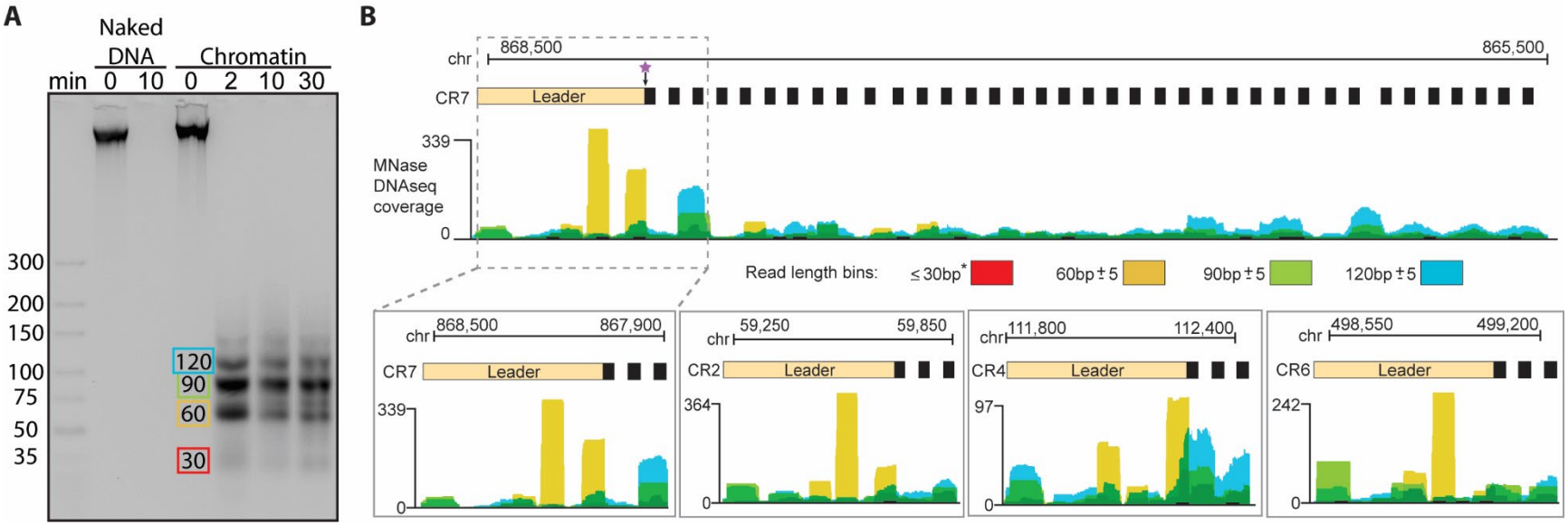

A
Naked DNA
Chromatin
min 0 10 0 2 10 30
300
200
150
100
75
50
35
120
90
60
30
B
chr
868,500
865,500
CR7
Leader
339
MNase DNAseq coverage
0
Read length bins:
≤30bp*
60bp±5
90bp±5
120bp±5
chr 868,500 867,900
CR7 Leader
339
0
chr 59,250 59,850
CR2 Leader
364
0
chr 111,800 112,400
CR4 Leader
97
0
chr 498,550 499,200
CR6 Leader
242
0

Figure 3.2. DNA protection on all CRISPR loci.

HTS of micrococcal nuclease digested DNA for the full CRISPR array for all seven CRISPR loci (leaders (pale yellow), the repeats (black boxes), and 2 spacers (white space between black boxes)).

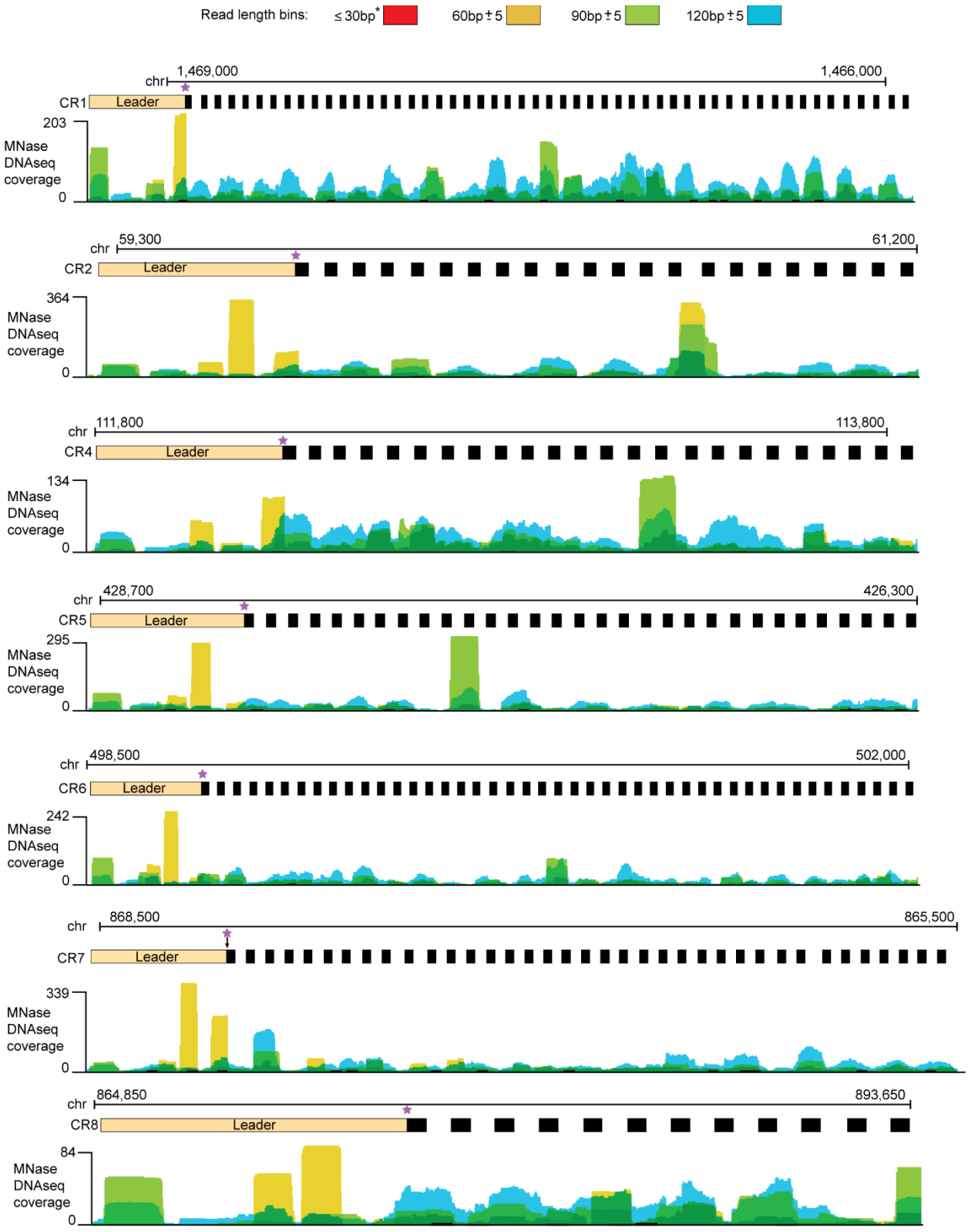

Read length bins:
≤ 30bp*
60bp ± 5
90bp ± 5
120bp ± 5
chr
1,469,000
1,466,000
CR1
Leader
203
MNase
DNAseq
coverage
0
chr
59,300
61,200
CR2
Leader
364
MNase
DNAseq
coverage
0
chr
111,800
113,800
CR4
Leader
134
MNase
DNAseq
coverage
0
chr
428,700
426,300
CR5
Leader
295
MNase
DNAseq
coverage
0
chr
498,500
502,000
CR6
Leader
242
MNase
DNAseq
coverage
0
chr
868,500
865,500
CR7
Leader
339
MNase
DNAseq
coverage
0
chr
864,850
893,650
CR8
Leader
84
MNase
DNAseq
coverage
0

Figure 3.3. Potential histone binding on CRISPR leaders.

Sequence alignment of the 60 bp protected regions on all seven CRISPR leaders. L=Leader conserved binding region, LR=Conserved binding region spanning the leader-repeat junction, +1=Transcription start site.

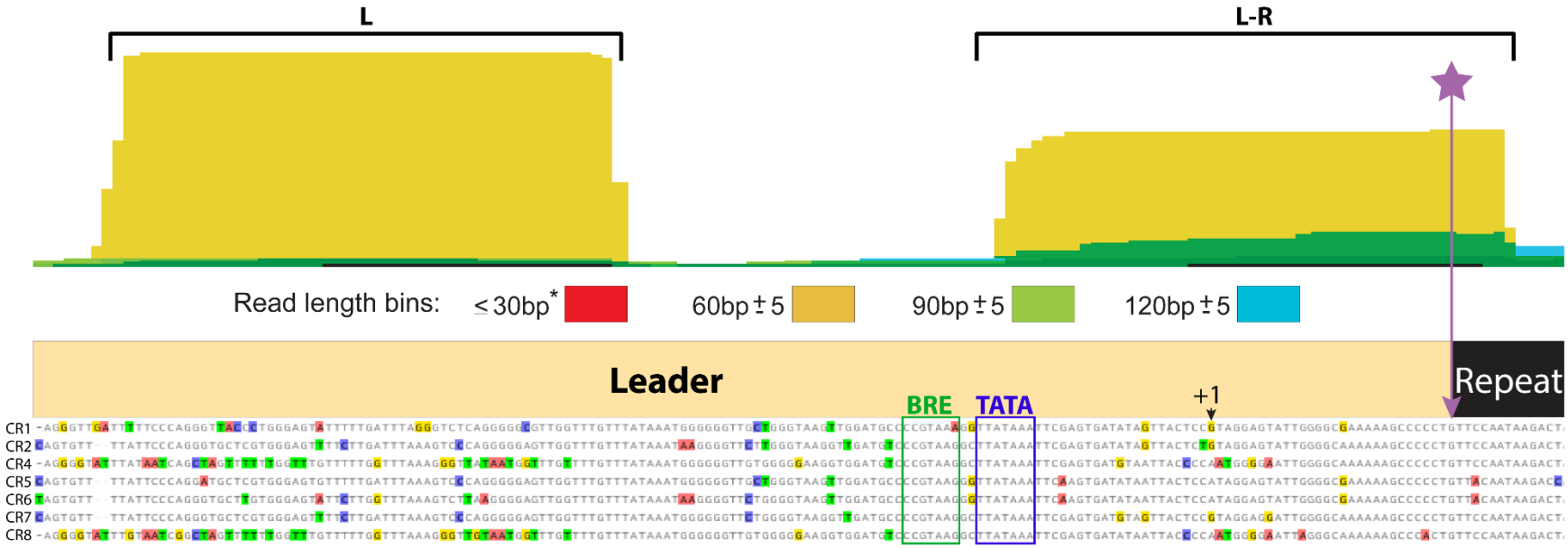
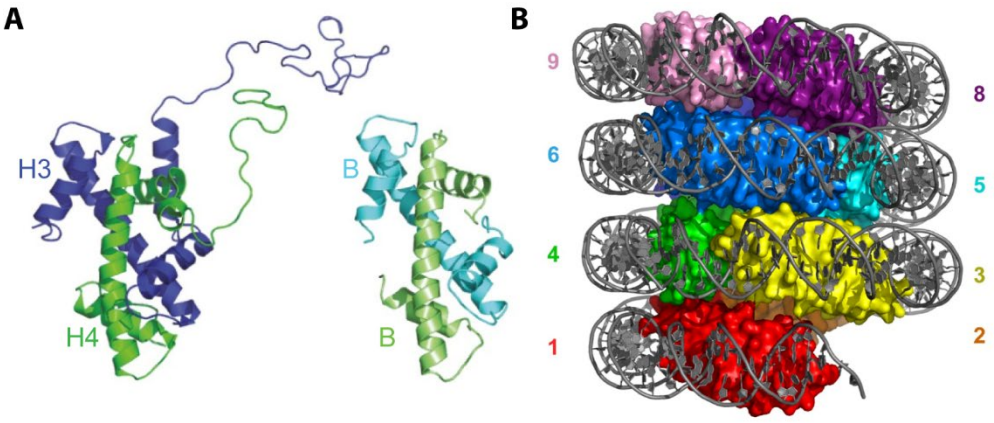
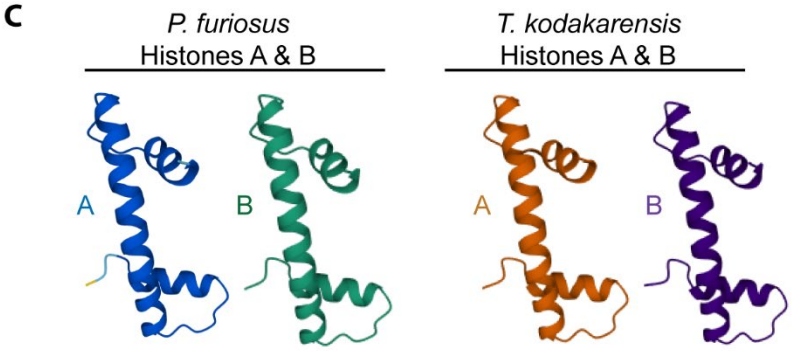

L
L-R
Read length bins:
≤ 30bp*
60bp ± 5
90bp ± 5
120bp ± 5
Leader
Repeat
BRE
TATA
+1
CR1
CR2
CR4
CR5
CR6
CR7
CR8

Figure 3.4. Conservation of histones among eukaryotic and archaeal organisms. **(A)** A eukaryotic (H3=dark blue, H4=dark green) dimer on the left, and an archaeal histone dimer on the right (homodimer of histone B from *Methanothermus fervidus*), **(B)** and a hypernucleosome structure formed from homodimers of histone B from *Methanothermus fervidus*. The numbering and the corresponding colors indicate the positions of each dimer (dimers 5 and 6 obscure view of dimer 7). DNA is shown as cartoon and gray in this image. Adapted from Henneman et al, 2018 [27]. **(C)** AlphaFold 3D structure attained from amino acid sequence for *P. furiosus* histone A (blue) and histone B (green), and *Thermococcus kodakarensis* histones A (orange) and B (purple). **(D)** Clustal Omega Multiple Sequence Alignment of histones from *P. furiosus* (Pfu A and Pfu B), *Thermococcus kodakarensis* (Tko A and Tko B), and *Methanothermus fervidus* (HMf A and HMf B). * = completely conserved residues, : = conservation between groups of strongly similar properties, . = conservation between groups of weakly similar properties.





D

```
Pfu A: MAELPIAPVDRLIRKAGAARVSEEAAKVLAEHLEEKALEIAKKAVALAQHAGRKTVKAEDIKLAIKS---    67
Pfu B: MAELPIAPVDRLIRKAGAERVSEDAAKVLAEYLEEYAIELSKKAVDFARHAGRKTVKAEDIKLAIKA---    67
Tko A: MGELPIAPVDRLIRKAGAERVSEQAAKVLAEYLEEYAIEVAKKAVEFARHAGRKTVKVEDIKLAIKS---    67
Tko B: MGELPIAPVDRLIRKAGAQRVSEQAAKVLAEHLEEKAIEIAKKAVDLAKHAGRKTVKVEDIKLAIKS---    67
HMf A: MGELPIAPIGRIIKNAGAERVSDDARIALAKVLEEMGEEIASEAVKLAKHAGRKTIKAEDIELARKMFK-    69
HMf B: -MELPIAPIGRIIKDAGAERVSDDARITLAKILEEMGRDIASEAIKLARHAGRKTIKAEDIELAVRRFKK    69
       ******:.*:*:.*** ***::*  .**: *** . :::.:*: :*:******:*.***:** :
```

Figure 3.5. Gene deletions demonstrate changes in integration at the first repeat in the absence of histone A, but no change in DNA protection.

(A) Adaptation assays (PCR from the CRISPR leader to the first repeat in the CRISPR array) on *in vivo* deletions of histones A or B in wildtype and type I-B only *P. furiosus* cells for several CRISPR loci. C = Absence of the pT33.3 conjugative plasmid, Δ = Deletion of the genes involved in adaptation (Cas1, Cas2, Cas4-1, Cas4-2), (* = unexpanded array, +1 = expanded array). **(B)** PCR products from amplifying downstream spacers from canonical spacer 1 and spacer 3, and spacer 3 to spacer 6 on the CRISPR7 array. **(C)** Micrococcal nuclease digestion on chromatin from wildtype and histone deletion strains. Numbers inside of colored boxes indicate different sizes of histone polymers observed on the agarose gel. **(D)** HTS sequencing of DNA fragments excised from **(C)** agarose gels (30, 60, 90, and 120 bp) of the CRISPR7 leader. Different colored peaks correspond to the size of coverage seen in **(C).**

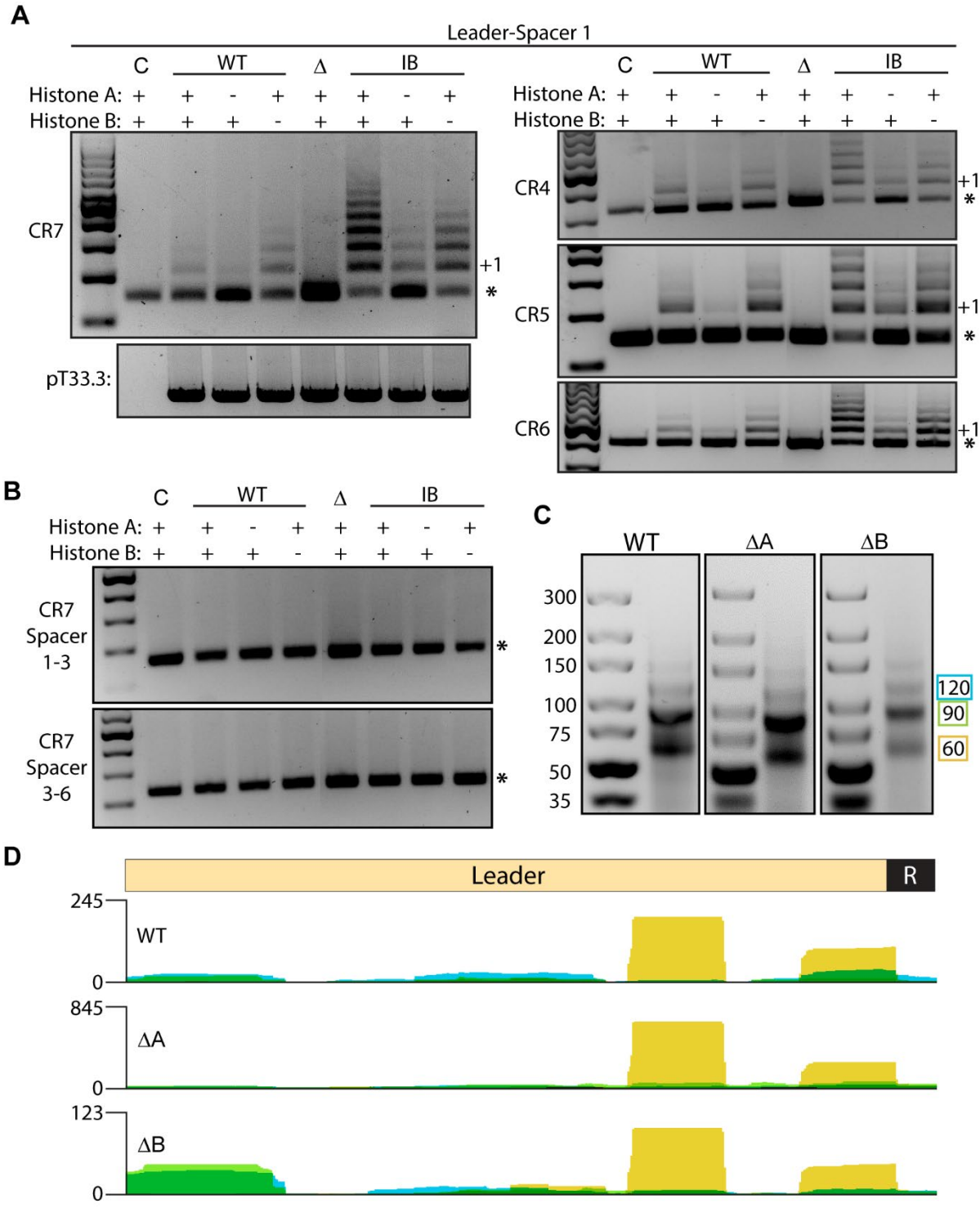

A
Leader-Spacer 1
C
WT
Δ
IB
Histone A: + + - + + + - +
Histone B: + + + - + + + -
CR7
+1
*
pT33.3:
CR4
CR5
CR6
B
CR7 Spacer 1-3
CR7 Spacer 3-6
C
WT
ΔA
ΔB
300
200
150
100
75
50
35
120
90
60
D
Leader
R
245
WT
0
845
ΔA
0
123
ΔB
0

Figure 3.6. Schematic of *in vitro* integration assay.

(A) In the absence of histones, a plasmid with a full CRISPR leader and 3 repeat array is added to a reaction with the Cas1-Cas2 integrase complex (orange and red respectively) with a prespacer (purple shades), integration events are demonstrated to occur at multiple repeats (No bias, #1 and #2). Same reaction in the presence of histones (blue and green dimers encircling the pCRISPR plasmid), it is predicted that if histones play a role in directing integration events at the first repeat, then there would be one main product at repeat 1 in both reactions (Bias, #1 and #2). **(B)** After the reaction, the plasmid undergoes a PCR, amplifying from the specific prespacer to beyond the CRISPR array (#1 reaction) or from the CRISPR leader to the specific prespacer (#2 reaction). Different sized expected PCR products are indicated below for integration at every repeat. Pale yellow = CRISPR leader, black (R#) = repeats, different colors = spacers. Primer locations are indicated with the half black arrows.

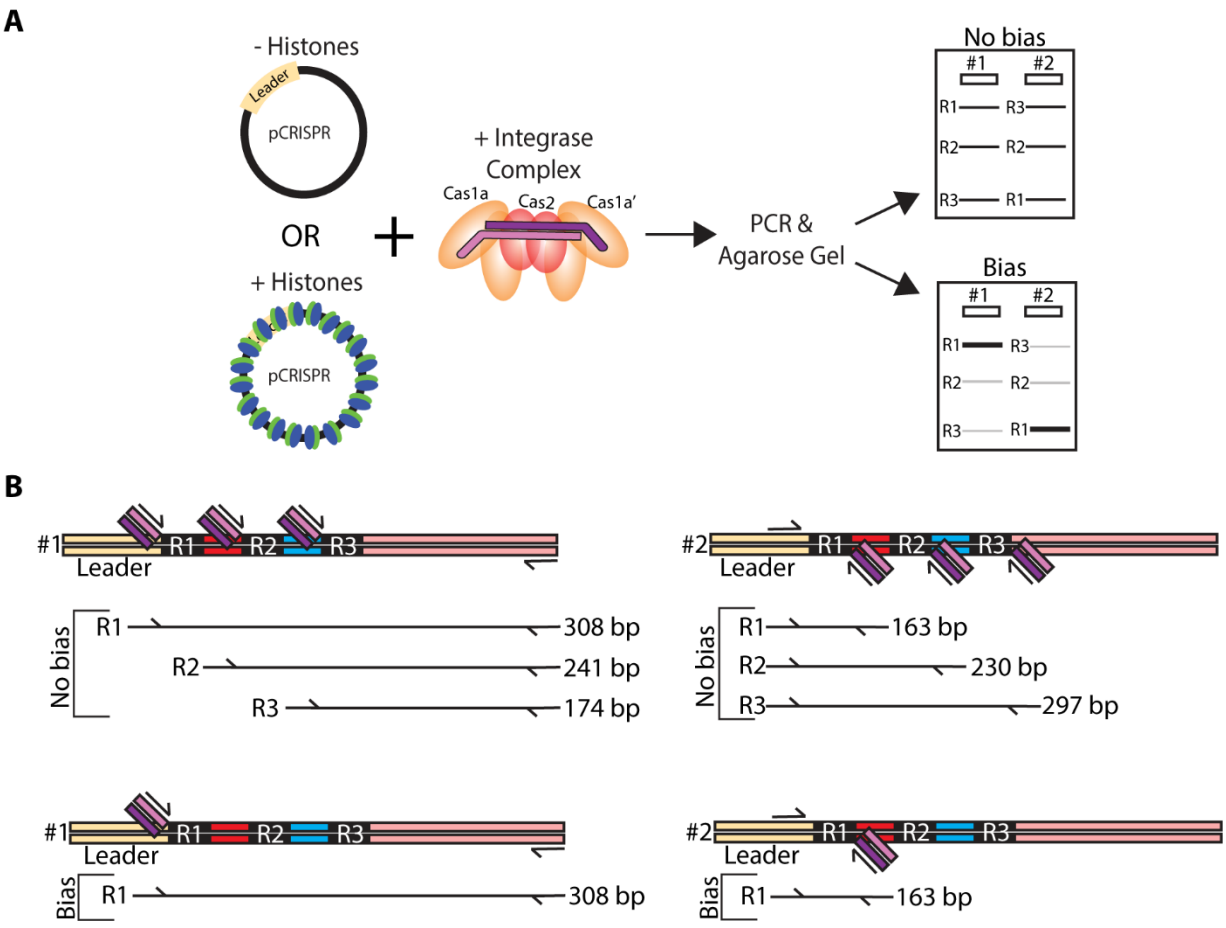

A
- Histones
Leader
pCRISPR
OR
+ Histones
pCRISPR
+ Integrase
Complex
Cas1a
Cas2
Cas1a'
PCR &
Agarose Gel
No bias
#1
#2
R1
R3
R2
R2
R3
R1
Bias
#1
#2
R1
R3
R2
R2
R3
R1
B
#1
R1
R2
R3
Leader
No bias
R1
308 bp
R2
241 bp
R3
174 bp
#2
R1
R2
R3
Leader
No bias
R1
163 bp
R2
230 bp
R3
297 bp
#1
R1
R2
R3
Leader
Bias
R1
308 bp
#2
R1
R2
R3
Leader
Bias
R1
163 bp

Figure 3.7. Purified recombinant proteins.

SDS PAGE gels with the purified proteins used in these assays. Cas1=37.5 kDa, Cas2=10 kDa, P. Furiosus histones A=7.4 kDa and B=7.3 kDa, T. kodakarensis histones A=7.3 kDa and B=7.1 kDa, TrmBL2=30.6 kDa.

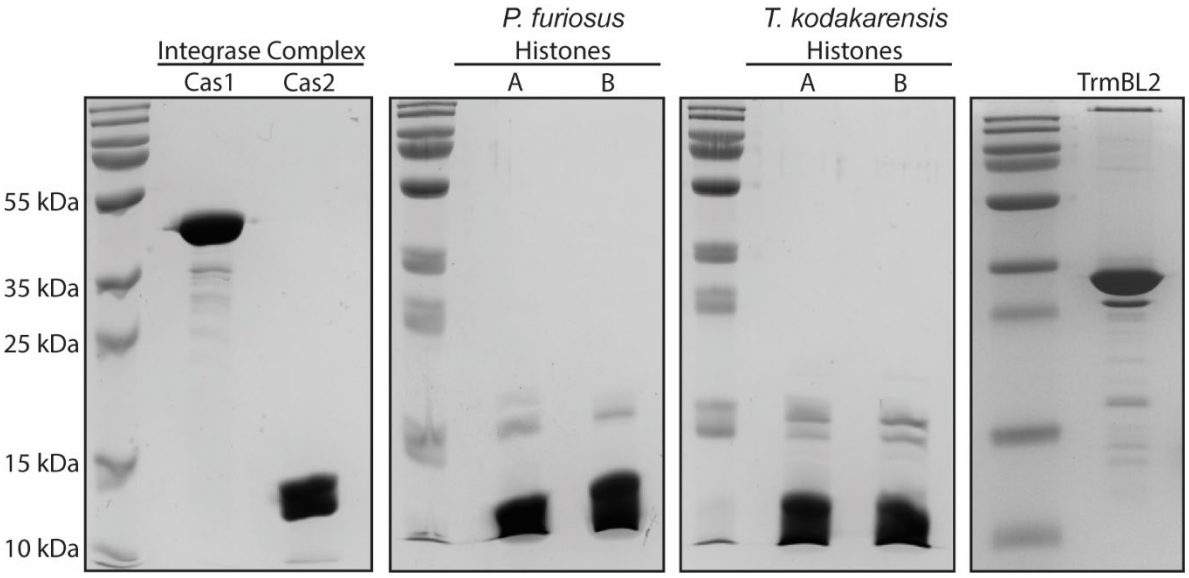

Integrase Complex
Cas1
Cas2
P. furiosus
Histones
A
B
T. kodakarensis
Histones
A
B
TrmBL2
55 kDa
35 kDa
25 kDa
15 kDa
10 kDa

Figure 3.8. *in vitro* integration assay reveals *T. kodakarensis* histone driven integration at the first repeat on CRISPR array.

(A) Schematic of expected products formed during in vitro integration assay using a plasmid with a full CRISPR7 leader and an array with 3 repeats amplifying from either the specific prespacer (purple shades) to beyond the CRISPR array (#1 reaction), or from the Leader (pale yellow) to the specific prespacer (#2 reaction). (**B)** PCR products from *in vitro* integration assays in the presence of *T. kodakarensis* histones run on ethidium bromide-stained agarose gels. R# = integration event at repeats, * = integration event between R3 and reverse primer on pCRISPR backbone in #1 reactions. (**C)** ImageJ analysis of each band intensity (integration event) as a percentage of all bands visualized for each concentration of histones added in **(B).**

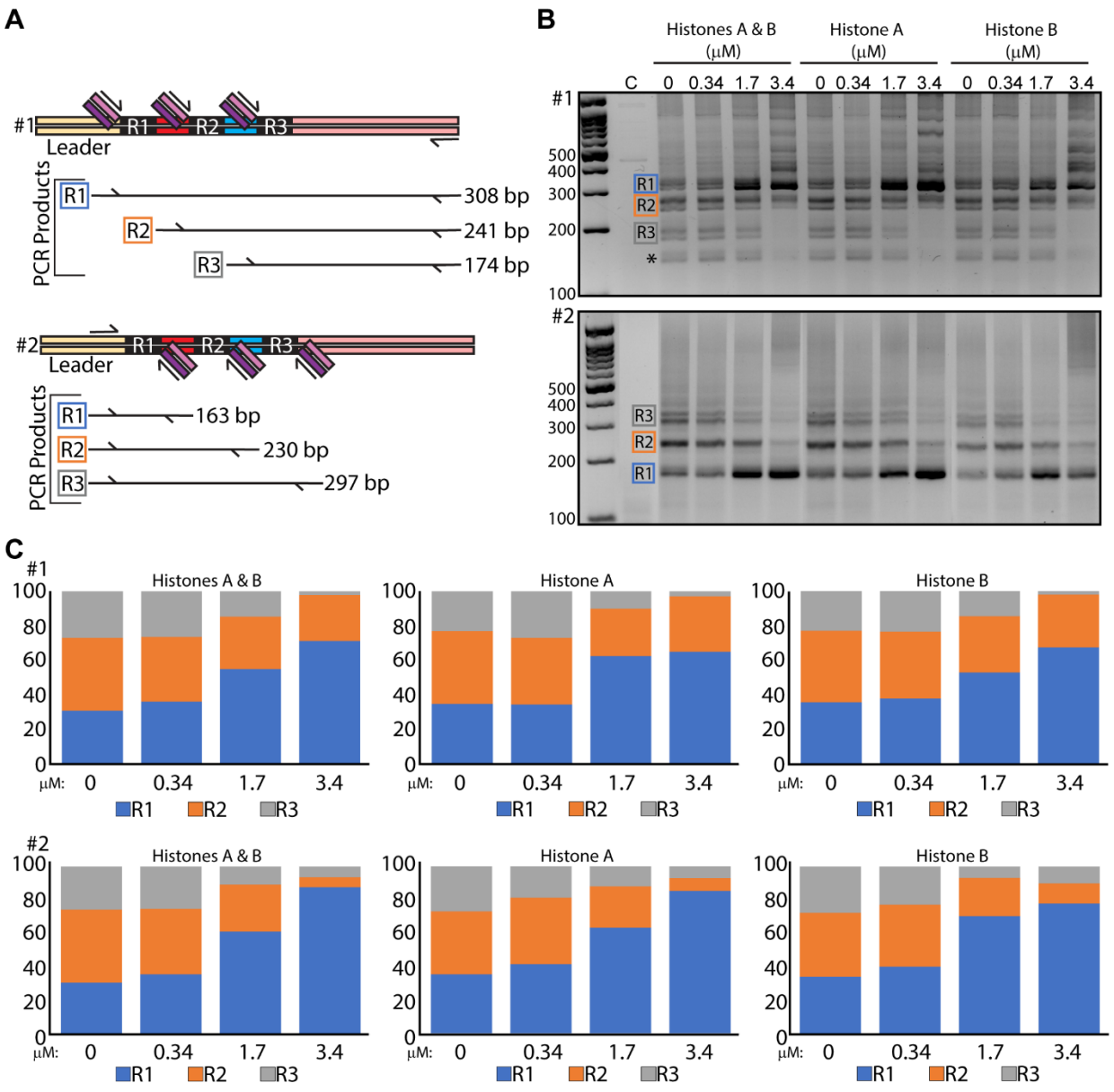

A
#1
Leader
R1
R2
R3
PCR Products
308 bp
241 bp
174 bp
#2
163 bp
230 bp
297 bp
B
Histones A & B (μM)
Histone A (μM)
Histone B (μM)
C 0 0.34 1.7 3.4 0 0.34 1.7 3.4 0 0.34 1.7 3.4
500
400
300
200
100
*
C
Histones A & B
Histone A
Histone B
μM: 0 0.34 1.7 3.4
R1 R2 R3

Figure 3.9. The presence of *P. furiosus* histones demonstrate same preference of integration at the first repeat.

(A) Schematic of expected products formed during in vitro integration assay using a plasmid with a full CRISPR7 leader and an array with 3 repeats amplifying from either the specific prespacer (purple shades) to beyond the CRISPR array (#1 reaction). (**B)** PCR products from *in vitro* integration assays in the presence of *P. furiosus* histones run on ethidium bromide-stained agarose gels. R# = integration event at repeats, * = integration event between R3 and reverse primer on pCRISPR backbone in #1 reactions. (**C)** ImageJ analysis of each band intensity (integration event) as a percentage of all bands visualized for each concentration of histones added in **(B).**

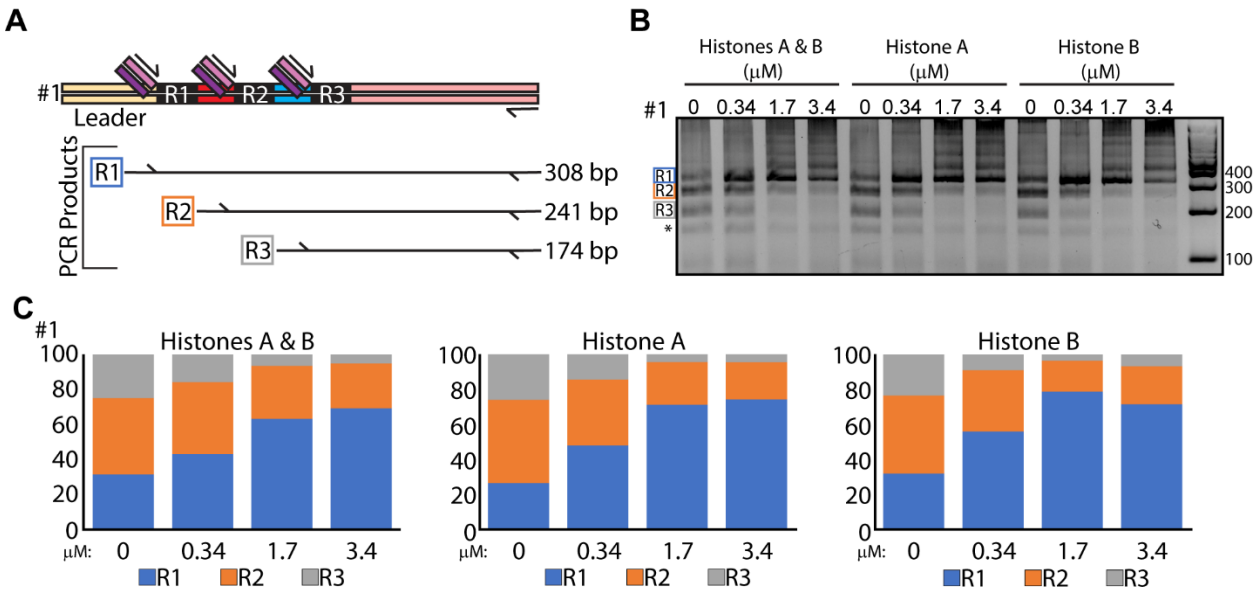

A
#1
R1
R2
R3
Leader
PCR Products
R1
R2
R3
308 bp
241 bp
174 bp
B
Histones A & B (μM)
Histone A (μM)
Histone B (μM)
#1
0 0.34 1.7 3.4 0 0.34 1.7 3.4 0 0.34 1.7 3.4
R1
R2
R3
*
400
300
200
100
C
#1
Histones A & B
Histone A
Histone B
100
80
60
40
20
0
μM: 0 0.34 1.7 3.4
R1 R2 R3

Figure 3.10. Preference for the first repeat is only observed in the presence of archaeal histones.

(A) Schematic of expected products formed during in vitro integration assay using a plasmid with a full CRISPR7 leader and an array with 11 repeats amplifying from the specific prespacer (purple shades) to spacer 5 the CRISPR array (#1 reaction). **(B)** PCR products from *in vitro* integration assays in the presence of *P. furiosus* histones run on ethidium bromide-stained agarose gels. R# = integration event at repeats. (**C)** ImageJ analysis of each band intensity (integration event) as a percentage of all bands visualized for each concentration of histones added from the agarose gel in **(B). (D)** PCR products from *in vitro* integration assays in the presence of TrmBL2 run on ethidium bromide-stained agarose gels. **(E)** ImageJ analysis of each band intensity (integration event) as a percentage of all bands visualized for each concentration of histones added from the agarose gel in **(D).**

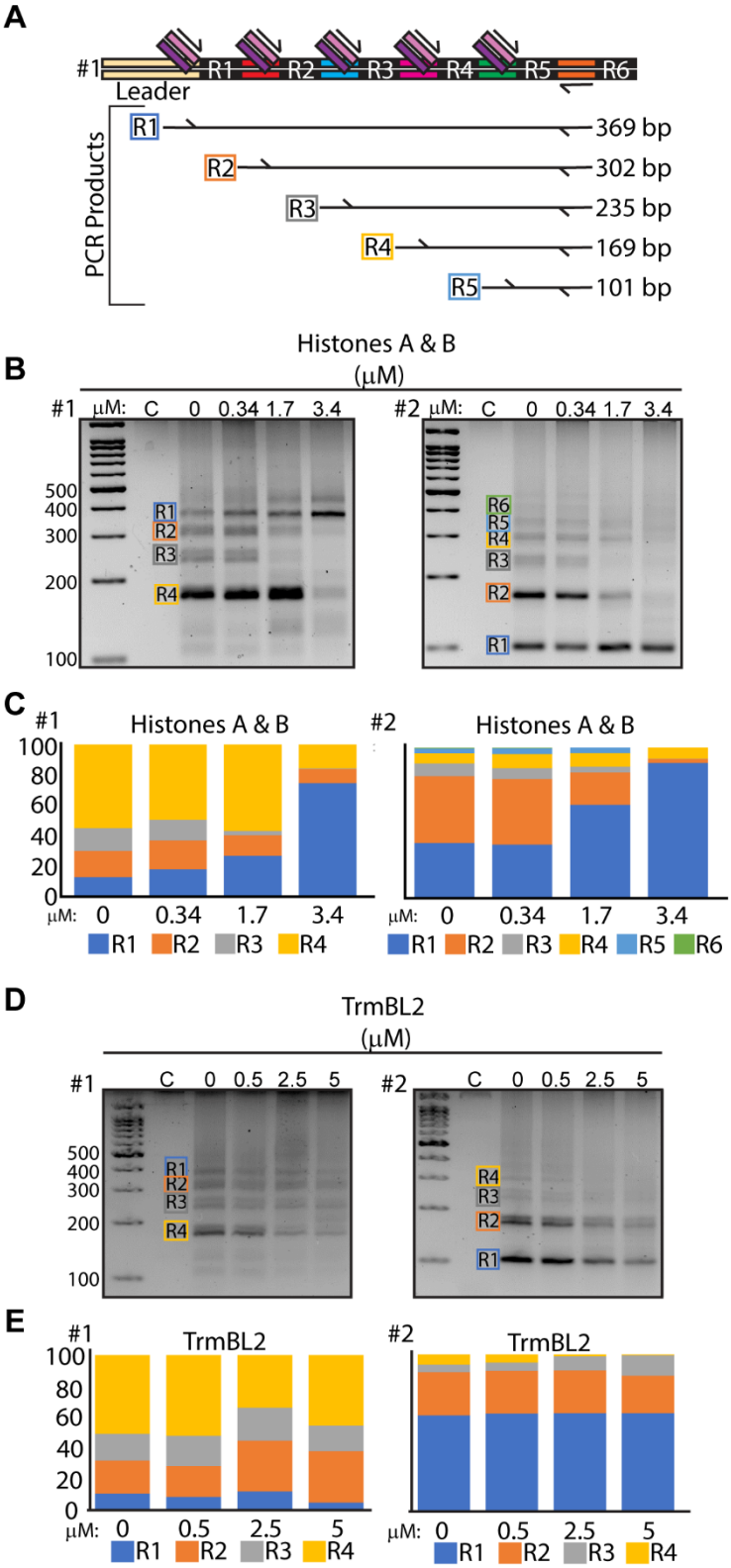

A
#1
R1
R2
R3
R4
R5
R6
Leader
PCR Products
R1
369 bp
R2
302 bp
R3
235 bp
R4
169 bp
R5
101 bp
B
Histones A & B
(μM)
#1
μM: C 0 0.34 1.7 3.4
#2
μM: C 0 0.34 1.7 3.4
500
400
300
200
100
R1
R2
R3
R4
R6
R5
R4
R3
R2
R1
C
#1
Histones A & B
#2
Histones A & B
100
80
60
40
20
0
μM: 0 0.34 1.7 3.4
R1 R2 R3 R4
μM: 0 0.34 1.7 3.4
R1 R2 R3 R4 R5 R6
D
TrmBL2
(μM)
#1
C 0 0.5 2.5 5
#2
C 0 0.5 2.5 5
500
400
300
200
100
R1
R2
R3
R4
R4
R3
R2
R1
E
#1
TrmBL2
#2
TrmBL2
100
80
60
40
20
0
μM: 0 0.5 2.5 5
R1 R2 R3 R4
μM: 0 0.5 2.5 5
R1 R2 R3 R4

Figure 3.11. Preference for the first repeat in the presence of histones is observed when amplifying from various spacers in both supercoiled and linear forms of the plasmid.

(A) Schematic of expected products formed during in vitro integration assay using a plasmid with a full CRISPR7 leader and an array with 11 repeats amplifying from the specific prespacer (purple shades) to spacer 10 (#1 reaction). **(B)** PCR products from *in vitro* integration assays amplifying from either specific prespacer to spacer 5 or specific prespacer to spacer 10 in the CRISPR arrays in the presence of *P. furiosus* histones run on ethidium bromide-stained agarose gels. Red asterisk = integration at repeat 1, R# = integration event at repeats, S_0=specific prespacer, S_5=spacer 5, S_{10}=spacer 10. **(C)** ImageJ analysis of each band intensity (integration event) as a percentage of all bands visualized for each concentration of histones added from the agarose gel in **(B).**

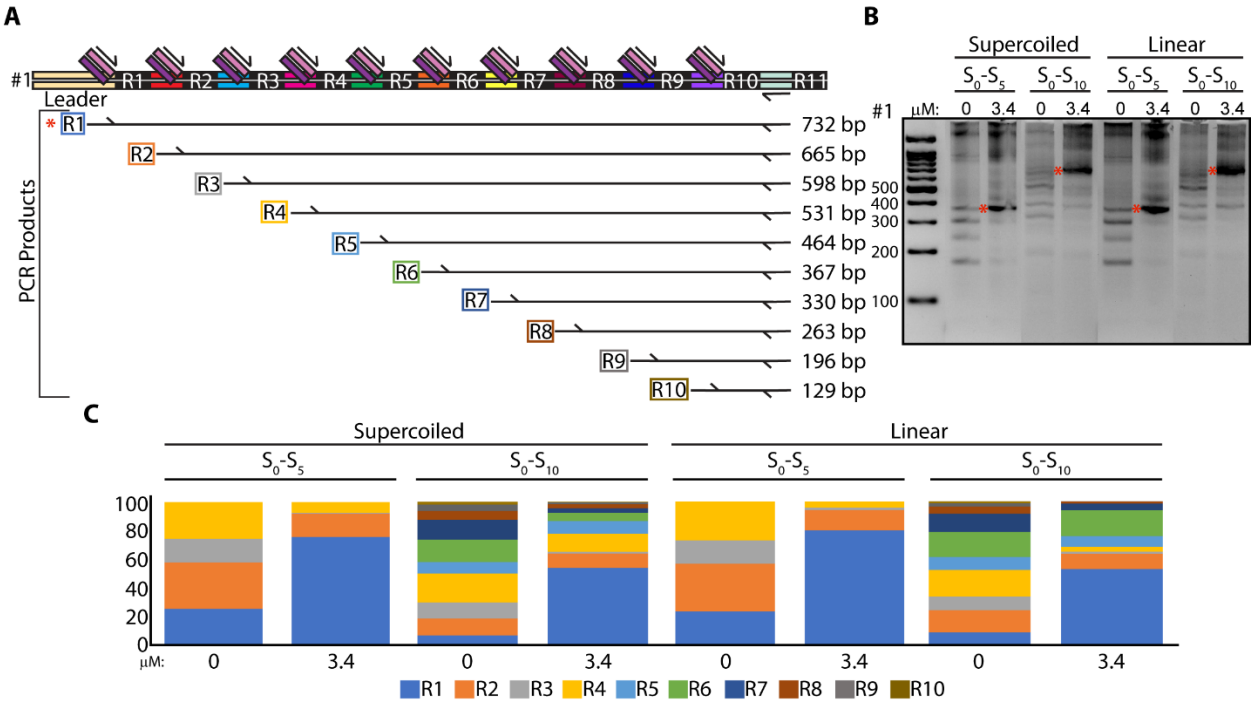

A
#1
R1
R2
R3
R4
R5
R6
R7
R8
R9
R10
R11
Leader
PCR Products
732 bp
665 bp
598 bp
531 bp
464 bp
367 bp
330 bp
263 bp
196 bp
129 bp
B
Supercoiled
Linear
S0-S5
S0-S10
μM:
0
3.4
500
400
300
200
100
C
100
80
60
40
20
0

Figure 3. 12. Multiple CRISPR loci demonstrate the same preference for the first repeat in the presence of histones.

(A) PCR products from *in vitro* integration assays performed as previously described, but utilizing the CRISPR2, CRISPR5, and CRISPR6 leaders and arrays. **(B)** ImageJ analysis of each band intensity (integration event) as a percentage of all bands visualized for each concentration of histones added from **(A).** Red asterisk = integration at repeat 1.

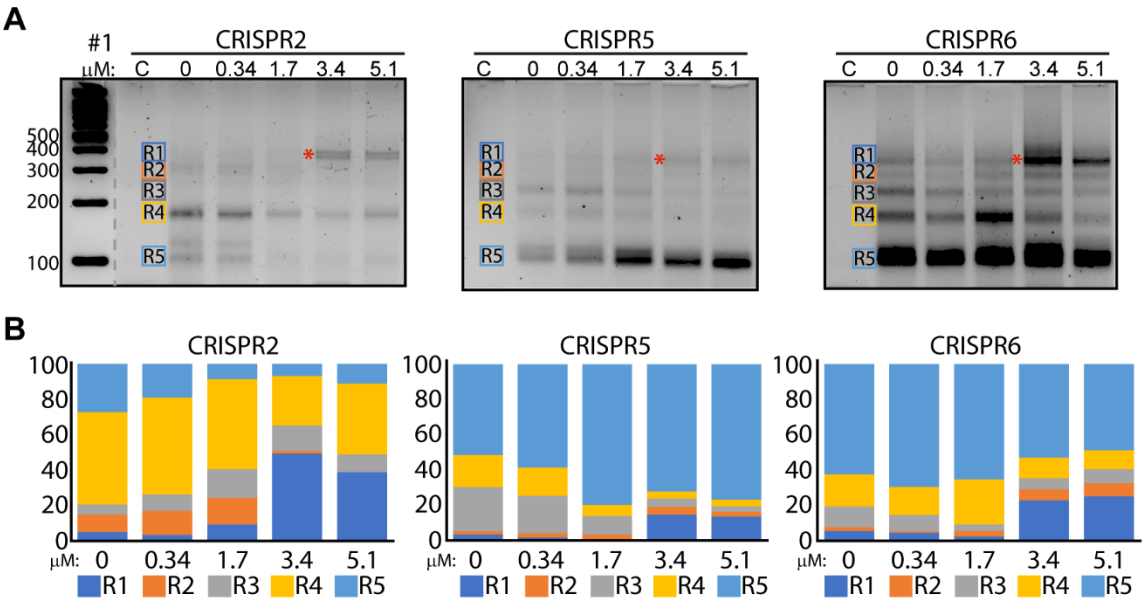
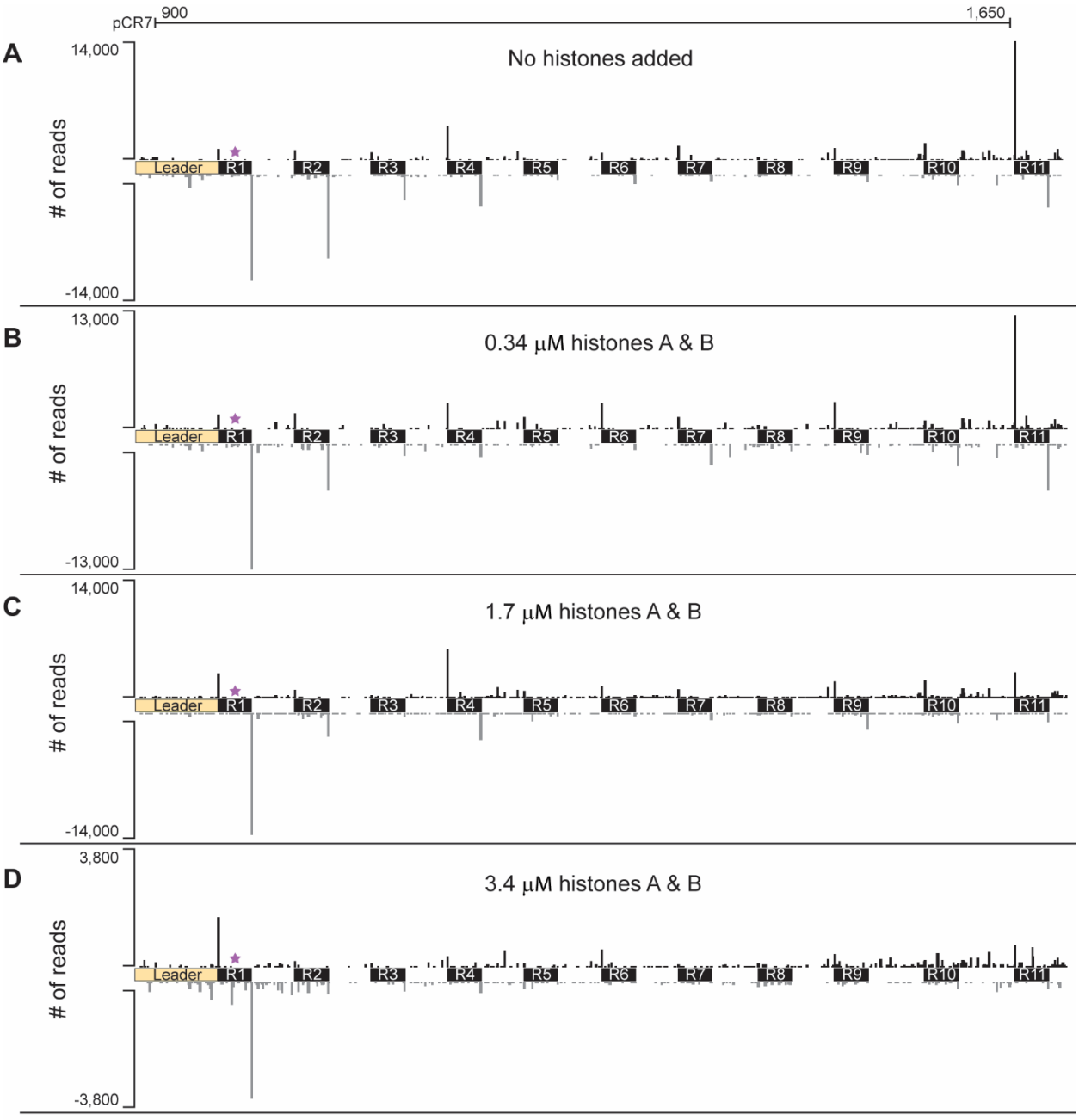

A
#1
CRISPR2
CRISPR5
CRISPR6
µM: C 0 0.34 1.7 3.4 5.1
500
400
300
200
100
R1
R2
R3
R4
R5
*
B
CRISPR2
CRISPR5
CRISPR6
100
80
60
40
20
0
µM: 0 0.34 1.7 3.4 5.1
R1 R2 R3 R4 R5

Figure 3.13. HTS validates histone driven integration events at the first repeat *in vitro*.

The CRISPR7 pCRISPR plasmid with 11 repeats underwent *in vitro* integration assays and was subsequently sequenced to observe how integration events were influenced in the presence of histones. **(A)** No addition of histones, **(B)** 0.34 μM of histones, **(C)** 1.7 μM of histones, and **(D)** 3.4 μM of histones were added to *in vitro* reactions. Leader = pale yellow, repeats = black, white spacing = spacers, purple star = repeat 1.


pCR7
900
1,650
A
No histones added
14,000
-14,000
# of reads
B
0.34 μM histones A & B
13,000
-13,000
C
1.7 μM histones A & B
14,000
-14,000
D
3.4 μM histones A & B
3,800
-3,800
Leader
R1
R2
R3
R4
R5
R6
R7
R8
R9
R10
R11

CHAPTER 4

DISCUSSION

Prokaryotic organisms, in an attempt to protect themselves against deleterious mobile genetic elements (MGEs), utilize both innate and adaptive immune systems. The work presented in my dissertation focuses on the only known immune system to produce adaptive, heritable immunity, the CRISPR-Cas system. My work specifically focuses on the first step in this system, adaptation, which gives rise to the heritable nature of CRISPR-Cas immunity by capturing DNA from invading MGEs. The CRISPR-Cas systems of our model organism, *Pyrococcus furiosus* (*P. furiosus*) have been extensively studied, yet without any isolated viruses or natural plasmids, we had previously been unable to study how these systems respond to a natural invader. In Chapter 2 of this dissertation, we describe the response of *P. furiosus* to a newly discovered conjugative plasmid. Much to our surprise, this natural plasmid causes a hyper-adaptation response at all CRISPR loci. We further investigate this response and identify the cause as priming-based adaptation. As with all studied CRISPR-Cas systems, the new spacers acquired against MGEs (including the priming-derived spacers targeting the conjugative plasmid) are integrated in the first repeat (relative to transcription initiation) of CRISPR loci. The mechanism behind this polarity of integration was previously unknown in Archaea. In Chapter 3 of this dissertation, I describe the first identified factor in an archaeal system that drives new spacer integration events to the first repeat on the CRISPR array. This factor is the histones responsible for compaction and structuring of the chromosome in

many archaeal species (and eukaryotes). This finding contributes an entirely new mechanism to the adaptation process.

Mobile Genetic Elements in Archaea

Viruses are the most abundant biological entities on Earth, present in all studied ecosystems, including the most extreme environments supporting only archaeal extremophiles. Over 60 viruses and plasmids have been isolated from archaea and have been found to have novel morphological and genomic qualities when compared to those of bacteria [1].

However, when it comes to conjugative plasmids, there is only one archaeal family found to host them, *Sulfolobacea* [1]. Not much is known about these archaeal conjugative plasmids, with only descriptive studies (no genetic studies) reported. Similar to the well-studied bacterial conjugative elements, these plasmids do require cell-to-cell contact for transfer and encode genes homologous to bacterial Type IV secretion systems [2-4].

In Chapter 2 of this dissertation, we discuss a novel conjugative plasmid found in the *Thermococcaceae* family, which includes *P. furiosus*. This natural conjugative plasmid, named pT33.3 after the host species in which it was identified, *Thermococcus sp. 33.3,* has been found to transfer between many Thermococcales species. We were excited to test this plasmid in *P. furiosus*, as this would be the first natural invader identified for our system. Upon transfer of pT33.3 to *P. furiosus,* we immediately observed a robust response against this foreign invader from the CRISPR-Cas system. Here, 5 of the 7 CRISPR loci integrated new spacers at a significantly higher frequency

than we typically observe against our lab-generated shuttle vectors (Figure 2.1). We initially hypothesized that the act of conjugation itself was leading to this extreme adaptation. Our classical adaptation assays use the natural competency of *P. furiosus* COM1 to introduce DNA, while pT33.3 is able to transform cells which are not naturally competent for DNA uptake. Thus, perhaps the plasmid-encoded entry mechanism could trigger adaptation more-so than the recipient-encoded transformation process. The vast majority of studied conjugative elements enter recipient cells in a linear, single stranded form, with the 5' end covalently linked to a transfer protein (and thus likely not a free-end susceptible to spacer uptake). It is difficult to imagine that Cas1/Cas2 would recognize this single-stranded DNA as a substrate for spacer acquisition. Moreover, host crRNPs would also be unable to target this incoming DNA in a single-stranded state. Thus, the timing of both new spacer acquisition and targeting by pre-existing spacers may depend upon (and indeed provide insight into) the timing of complementary strand synthesis of conjugative plasmids. It has been found in bacterial systems that spacer acquisition from conjugative plasmids was primarily preferred on either the leading region (first section entering the recipient cell) or the lagging region depending on the plasmid MOB classification [5], suggesting adaptation was responsive to conjugation mechanism. However, a study in *Saccharolobus solfataricus* P2 revealed that spacer acquisition appeared to be evenly spread throughout the plasmid [6].

The latter scenario described above in *S. solfataricus* reflects our findings in *P. furiosus* with pT33.3. The initial hyperadaptation events that we observed was due to a primed adaptation event against the pT33.3 (Figure 2.4), which we characterized thoroughly (Chapter 2). However, when we eliminated the priming protospacer, we found

that naïve spacer acquisition occurs evenly throughout the plasmid, no longer centered on any particular region (Figure 2.4). Additionally, the rate of spacer uptake from pT33.3 was no higher than that seen in previous transformation-based assays. This suggests that either adaptation in *P. furiosus* is not responsive to the conjugative mechanism of pT33.3; or that conjugation is not the main mechanism of pT33.3 entry into the hyper-competent strain *P. furiosus* COM1 used in these assays [7]. Indeed, we may expect active DNA uptake mechanisms such as natural competence to be at least partially protected from CRISPR targeting. Such uptake mechanisms likely evolved to benefit the cell by promoting homologous recombination and gene acquisition for rapid fitness gains, and thus immediate targeting of incoming DNAs would be antithetical. To finally determine whether conjugative transfer can trigger adaptation, we could transfer pT33.3 (or a derivative thereof without the priming spacer) to *P. furiosus* strain DSM3638, which is not naturally competent for DNA uptake. Any pT33.3 transconjugants of DSM3638 must result from plasmid self-transmission rather than recipient-encoded uptake, and thus could be assayed for elevated rates of adaptation.

The finding that there was a spacer which partially matches pT33.3, which causes primed adaptation, was exciting in itself (Figure 2.2). This is the first discovery of a source for one of the 200 spacers contained in the seven CRISPR arrays of *P. furiosus*. While a large number of Thermococcales strains have been isolated and grown in lab culture, very few harbor viruses or extra-chromosomal elements. Thus the mobilome of Thermococcales responsible for the many thousands of spacers encoded by these species remains mysterious. The presence of a spacer capable of targeting pT33.3 suggests that *P. furiosus* encountered this plasmid (or a similar conjugative plasmid) in the wild, prior to

isolation. The two mismatches between the spacer and pT33.3-encoded protospacer, along with the fact that *P. furiosus* and *T. sp. 33-3* were isolated ~7000 miles apart (Vulcano, Italy vs. East Pacific Rise, respectively), and from different environments (volcanic caldera sediments vs. tectonic plate boundary hydrothermal vent), likely indicates that the spacer was not acquired from pT33.3 itself, but rather a related element. The protospacer is found in a homologue of *traJ* – a gene involved in regulation of expression of transfer genes in bacterial conjugative plasmids. Thus the mismatched spacer in *P. furiosus* hints at the presence of other conjugative elements in Thermococcales which are yet to be identified. Combined with the 199 other unknown spacers, the CRISPR arrays of *P. furiosus* suggest a very active and diverse mobilome waiting to be discovered.

Adaptation in *P. furiosus*

Adaptation entails the acquisition of DNA from a foreign invader, which is then trimmed down to the appropriate size, and ultimately integrated into the CRISPR array in the correct orientation (with reference to the PAM (Protospacer Adjacent Motif) sequence) in a polarized manner at the first repeat adjacent to the CRISPR leader, in most cases.

It was found *in vitro* for our system using *P. furiosus* recombinant proteins and CRISPR loci, that integration of new spacers occurs at every repeat [8]. This was found to be the case in other type I systems as well, for example the type I-E and the type I-F system [9, 10]. However, in *P. furiosus*, not only was integration occurring at every repeat, it was occurring at repeat-like elements not on the CRISPR array [8]. It was found

that DNA sequences with those similar to the CRISPR repeat borders in *P. furiosus*, were found to have integration events with top and bottom strand events occurring about 30 bp apart (the size of a repeat in *P. furiosus*) [8]. A guanine was found to be a preference for sites of integration for the Cas1-Cas2 integrase complex in *P. furiosus*, as well as in type I-A, I-B, I-E, II-A, and V-C systems as well [9, 11-13]. There was also an adenine-rich sequence located in the central region of repeats and repeat-like sequences that is predicted to play a role in the orientation of the Cas1-Cas2 integration complex in the correct position to integrate new spacers, as mutations at this motif disrupted integration events at both borders of the repeat [8].

In contrast to the type I system's preference for repeat-like sequences for integration to occur, type II systems have been found to integrate specifically at the leader adjacent repeat. In type II studies, it has been found that the first 10 bp or so of the leader before the first repeat and 5 bp of the repeat are important for integration efficiency both *in vivo* and *in vitro* [13-15].

This lack of specificity for the first repeat in type I systems is not observed *in vivo*, with all integration events occurring at the first repeat adjacent to the leader, with no ectopic events observed (Figure 3.5) [16]. In chapter 3, I identify archaeal histones as factors that direct integration to the first repeat.

Histones in *P. furiosus*

Micrococcal nuclease assays are a common experiment performed to observe DNA protection patterns created by histones, or potentially some other DNA binding protein. We used this assay in our own work in chapter 3 to observe if any DNA

protection peaks were on the CRISPR leaders to potentially identify a factor binding to the leader and guiding integration at the first repeat. We observed a laddering of micrococcal nuclease digested DNA similar to that seen in other archaea, with 30 bp being the smallest band observed, and upwards of 150 bp+, bands going up in 30 bp increments, the same as an additional histone dimer bound to the DNA (Figure 3.1, Figure 3.4B). We then performed HTS on these DNA fragments and found that there were two 60 bp peaks observed on the CRISPR leader, one 100 bp upstream of the first repeat and the other spanning the leader-repeat junction (Figure 3.1, Figure 3.2, Figure 3.3). The size of these peaks, 60 bp, is the same size as two histone dimers binding to these areas. The underlying sequence also suggests that histones may bind there.

Another interesting note was that these 60 bp peaks were not found to be common throughout the genome of *P. furiosus*, only occurring every 1 kb or so. Most CRISPR leaders in *P. furiosus* were found to have these peaks (or some rather one or the other preferred) (Figure 3.3). It is interesting to think that the genome evolved to dictate that a histone dimer pair should bind to these CRISPR leaders and fill the role of a signal to guide integration events to the first repeat *in vivo*.

The deletion of these histones *in vivo* proved difficult, as was found in other organisms as well [17]. We were able to delete either histone A or histone B, but not both in the same strain. This made studying any potential role histones may play in integration *in vivo* difficult, as archaeal histones can form either homodimers or heterodimers. However, we know that although the amino acid sequences for both histone A and histone B are extremely conserved with each other (Figure 3.4C, Figure 3.4D), that can play different roles in the cell. For example, in *Thermococcus kodakarensis*, the deletion

of histone A prevented subsequent transformation of DNA into the cell, but the deletion of histone B did not [17]. With our *in vivo* studies, we did find that the deletion of histone A appeared to decrease the frequency of integration events at the first repeat, while the deletion of histone B did not have as significant an effect (Figure 3.5A). This could suggest that histone A plays a larger role in the integration of new spacers at the first repeat than histone B, although, it is difficult to say exactly what is happening there. We also know that deletion of one histone or the other in *Thermococcus kodakarensis* has different effects on gene expression in the cell [17, 18]. Perhaps the deletion of histone A in *P. furiosus* affects the expression levels of another protein that may cooperate with histones to guide integration to the first repeat *in vivo*.

To bypass the inability to delete both histones *in vivo*, we went *in vitro* to be able to control exactly what we are incorporating into the reaction. In our first studies, we utilized both *Thermococcus kodakarensis* and *P. furiosus* histones and performed assays in which we added both histones A and B, histone A alone, and histone B alone. Strikingly, all combinations appeared to drive integration events to the first repeat (Figure 3.8 and Figure 3.9). This collaborated with what we observed *in vivo*, as potential heterodimers and homodimers could guide integration events to the first repeat.

The purified recombinant histones in our *in vitro* studies have demonstrated that they direct integration at the first repeat in CRISPR array in a variety of experiments. However, we needed to confirm that this was a histone specific effect, not just any DNA binding protein. Therefore, we decided to experiment with another major DNA binding protein found in *P. furiosus*, TrmBL2. As described previously, TrmBL2 forms stiff filaments on DNA, and has been demonstrated to compete with histone preferred

sequences [19-21]. In our studies, we found that the addition of TrmBL2 did not have the same effect as histones in driving integration events to the first repeat. In fact, it appeared to inhibit integration at increasing concentrations (Figure 3.14). This further demonstrates that this histone directed effect we are observing is indeed a histone specific one. TrmBL2 antagonizes the compaction of DNA due to the stiff filaments it forms, while histones wrap and compact it. The wrapping of DNA by histones may be the reason why the Cas1-Cas2 integrase complex recognize the leader adjacent repeat, similar to the way IHF bends DNA and creates a signal while also bringing upstream sequence motifs in contact with Cas1 in type I-E and type I-F systems. IHF is also found to interact with Cas1, and we have also observed interactions of archaeal histones with Cas1 in *P. furiosus* with immunoprecipitation and mass spectrometry to identify potential factors associated with Cas1 *in vivo*, although we do not know if these are direct or indirect interactions (unpublished).

To further confirm that histones are indeed binding to the leader to create those two conserved DNA protection peaks, future experiments can involve the addition of histones to the pCRISPR plasmid and performing micrococcal nuclease assays on these substrates. In this controlled experiment, we can perform HTS on the micrococcal nuclease digested plasmid DNA and observe if those same peaks arise on the CRISPR leader in *in vitro* conditions with only histones present. In other studies, the addition of purified *Thermococcus kodakarensis* histones to purified genomic DNA created a similar binding profile as observed *in vivo* [22]. Another future experiment here involves the addition of histones to purified *P. furiosus* genomic DNA and performing the same assay as above to observe strictly histone binding at the CRISPR leaders. The profile of histone

binding to plasmid DNA may be different than that of genomic DNA, and the comparison between a 6 kb plasmid to a 1.9 Mb genome may have interesting differences in the way histones bind and wrap DNA given the size constraints of the plasmid.

During our studies in chapter 3, it was discovered that both supercoiled and linear substrates could be utilized for the Cas1-Cas2 integrase complex to integrate new spacers. This was surprising, due to type I systems in bacteria being unable to integrate new spacers into CRISPR arrays with linear substrate [11, 12]. Additionally, CRISPR-Cas targeting is dependent upon DNA supercoiling state [23], with at least one Acr inhibiting crRNP interference by relieving DNA torsion [24] . DNA topology is thus critically important for multiple steps of CRISPR-Cas immunity. In *E. coli*, DNA is maintained in a negatively supercoiled state while *P. furiosus* DNA is maintained in a positively supercoiled state by the action of reverse gyrase [25, 26]. This is thought to protect the genome from denaturation at the 100°C optimum growth temperature for *Pyrococcus* [26]. The pCRISPR plasmids used in our *in vitro* integration assays are most likely negatively supercoiled (as they were isolated from *E. coli*), and thus may be an unfamiliar substrate for *P. furiosus* Cas1-Cas2. Archaeal histones have been shown to induce positive supercoiling in plasmids by tightly wrapping DNA [27]. Hence, we wondered if rather than actively directing adaptation, histone binding was simply converting plasmids to a more natural substrate for the *P. furiosus* adaptation complex. However, both supercoiled and linear forms of DNA were found to be able to reproduce the same integration patterns (Figure 3.11), indicating that this adaptation polarizing effect of histones is not an indirect result of supercoiling, but rather a histone-specific effect on Cas1-Cas2.

Concluding remarks

The CRISPR-Cas adaptive immune system has been found to incorporate both its own Cas proteins and trans-acting factors in the host cell to accomplish successful defense against foreign invaders. In this dissertation, we describe the first, natural foreign invader discovered for *P. furiosus*. The hyperadaptation response to this element allowed for us to find the source (or related source) of one of the 200 spacers in the CRISPR arrays due to the robust primed adaptation event occurring in these cells. We then identify the factor that directs polarized integration of new spacers at the leader adjacent repeat in an archaeal CRISPR-Cas system, archaeal histones. This dissertation features a unique response against the first natural invader for *P. furiosus*, as well elucidating a new mechanism by which the polarized integration of new spacers at the CRISPR arrays can occur.

References

1. Wang, H., et al., *Archaeal extrachromosomal genetic elements.* Microbiol Mol Biol Rev, 2015. **79**(1): p. 117-52.
2. Wagner, A., et al., *Mechanisms of gene flow in archaea.* Nat Rev Microbiol, 2017. **15**(8): p. 492-501.
3. Guglielmini, J., et al., *Key components of the eight classes of type IV secretion systems involved in bacterial conjugation or protein secretion.* Nucleic Acids Res, 2014. **42**(9): p. 5715-27.
4. Alvarez-Martinez, C.E. and P.J. Christie, *Biological diversity of prokaryotic type IV secretion systems.* Microbiol Mol Biol Rev, 2009. **73**(4): p. 775-808.
5. Westra, E.R., et al., *CRISPR-Cas systems preferentially target the leading regions of MOBF conjugative plasmids.* RNA Biol, 2013. **10**(5): p. 749-61.
6. Erdmann, S., S.A. Shah, and R.A. Garrett, *SMV1 virus-induced CRISPR spacer acquisition from the conjugative plasmid pMGB1 in Sulfolobus solfataricus P2.* Biochem Soc Trans, 2013. **41**(6): p. 1449-58.
7. Farkas, J., et al., *Recombinogenic properties of Pyrococcus furiosus strain COM1 enable rapid selection of targeted mutants.* Appl Environ Microbiol, 2012. **78**(13): p. 4669-76.
8. Grainy, J., et al., *CRISPR repeat sequences and relative spacing specify DNA integration by Pyrococcus furiosus Cas1 and Cas2.* Nucleic Acids Res, 2019.

9. Nunez, J.K., et al., *Integrase-mediated spacer acquisition during CRISPR-Cas adaptive immunity.* Nature, 2015. **519**(7542): p. 193-8.

10. Fagerlund, R.D., et al., *Spacer capture and integration by a type I-F Cas1-Cas2-3 CRISPR adaptation complex.* Proc Natl Acad Sci U S A, 2017.

11. Rollie, C., et al., *Prespacer processing and specific integration in a Type I-A CRISPR system.* Nucleic Acids Res, 2018. **46**(3): p. 1007-1020.

12. Rollie, C., et al., *Intrinsic sequence specificity of the Cas1 integrase directs new spacer acquisition.* Elife, 2015. **4**.

13. Xiao, Y., et al., *How type II CRISPR-Cas establish immunity through Cas1-Cas2-mediated spacer integration.* Nature, 2017. **550**(7674): p. 137-141.

14. *<16-Wright-Protecting genome integrity.pdf>.*

15. McGinn, J. and L.A. Marraffini, *CRISPR-Cas Systems Optimize Their Immune Response by Specifying the Site of Spacer Integration.* Mol Cell, 2016.

16. Shiimori, M., et al., *Role of free DNA ends and protospacer adjacent motifs for CRISPR DNA uptake in Pyrococcus furiosus.* Nucleic Acids Res, 2017.

17. Cubonovaa, L., et al., *An archaeal histone is required for transformation of Thermococcus kodakarensis.* J Bacteriol, 2012. **194**(24): p. 6864-74.

18. Sanders, T.J., et al., *Extended Archaeal Histone-Based Chromatin Structure Regulates Global Gene Expression in Thermococcus kodakarensis.* Front Microbiol, 2021. **12**: p. 681150.

19. Efremov, A.K., et al., *Transcriptional Repressor TrmBL2 from Thermococcus kodakarensis Forms Filamentous Nucleoprotein Structures and Competes with*

Histones for DNA Binding in a Salt- and DNA Supercoiling-dependent Manner. J Biol Chem, 2015. **290**(25): p. 15770-15784.

20. Maruyama, H., et al., *Histone and TK0471/TrmBL2 form a novel heterogeneous genome architecture in the hyperthermophilic archaeon Thermococcus kodakarensis.* Mol Biol Cell, 2011. **22**(3): p. 386-98.
21. Wierer, S., et al., *TrmBL2 from Pyrococcus furiosus Interacts Both with Double-Stranded and Single-Stranded DNA.* PLoS One, 2016. **11**(5): p. e0156098.
22. Sanders, T.J., et al., *TFS and Spt4/5 accelerate transcription through archaeal histone-based chromatin.* Mol Microbiol, 2019. **111**(3): p. 784-797.
23. Szczelkun, M.D., et al., *Direct observation of R-loop formation by single RNA-guided Cas9 and Cascade effector complexes.* Proc Natl Acad Sci U S A, 2014.
24. Forsberg, K.J., et al., *The novel anti-CRISPR AcrIIA22 relieves DNA torsion in target plasmids and impairs SpyCas9 activity.* PLoS Biol, 2021. **19**(10): p. e3001428.
25. Peter, B.J., et al., *Genomic transcriptional response to loss of chromosomal supercoiling in Escherichia coli.* Genome Biol, 2004. **5**(11): p. R87.
26. Rodriguez, A.C. and D. Stock, *Crystal structure of reverse gyrase: insights into the positive supercoiling of DNA.* EMBO J, 2002. **21**(3): p. 418-26.
27. Musgrave, D.R., K.M. Sandman, and J.N. Reeve, *DNA binding by the archaeal histone HMf results in positive supercoiling.* Proc Natl Acad Sci U S A, 1991. **88**(23): p. 10397-401.