

TRANS-ACTING FACTORS OF CRISPR-CAS ADAPTATION IN THE
HYPERTHERMOPHILIC ARCHAEON *PYROCOCCUS FURIOSUS*

by

MICHAEL ANDREW ELLIS

(Under the Direction of MICHAEL P. TERNES)

ABSTRACT

Recognition and targeting of invaders by the prokaryotic CRISPR-Cas system depends on the acquisition of invader sequences within the host CRISPR locus. We are investigating the mechanisms of this unique biological process. The Cas1 and Cas2 proteins universal among the CRISPR-Cas systems – have been genetically linked to adaptation, as has the 5’-3’ exonuclease Cas4 that is associated with some systems. The *Pyrococcus furiosus* genome encodes three CRISPR-Cas systems (Cmr Type III-B, Csa Type I-A, and Cst Type I-G), two Cas4 proteins (Cas4-1 and Cas4-2), as well as Cas1 and Cas2 proteins. In this work, *in vivo* interactions among these adaptation proteins were analyzed. This study indicates the existence of a complex that includes the Cas1 and Cas4-1 proteins, and evidence of Cas1 and Cas2 association. Moreover, we demonstrate that Cas1 associates with a plethora of non-Cas proteins and nucleic acids, which may be relevant to CRISPR-Cas adaptation.

INDEX WORDS: CRISPR, Cas, Cas1, Cas2, Cas4, *Pyrococcus furiosus*

TRANS-ACTING FACTORS OF CRISPR-CAS ADAPTATION IN THE
HYPERTHERMOPHILIC ARCHAEON *PYROCOCCUS FURIOSUS*

by

MICHAEL ANDREW ELLIS

BA, Earlham College 2010

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

MASTER OF SCIENCE

ATHENS, GEORGIA

2017

© 2017

Michael Andrew Ellis

All Rights Reserved

THE TRANS-ACTING FACTORS OF CRISPR-CAS ADAPTATION IN THE
HYPERTHERMOPHILIC ARCHAEON *PYROCOCCUS FURIOSUS*

by

MICHAEL ANDREW ELLIS

Major Professor:	Michael P. Terns
Committee:	Rebecca M. Terns
	Sidney R. Kushner
	Claiborne V. C. Glover III

Electronic Version Approved:

Suzanne Barbour
Dean of the Graduate School
The University of Georgia
May 2017

DEDICATION

I would like to dedicate this work to my wife, Katie Putney. She is the person who has to put up with me each day while I deal with graduate school and all of the joys that come with it. Without her vital support, this would not have been possible. I would also dedicate this work to my parents who have always supported me no matter what the cost or situation.

ACKNOWLEDGEMENTS

A fundamental truth that I have learned during my ongoing development as a scientist is that no work is done in isolation. As it takes a village to raise a child, it also takes one to raise a scientist. I am no exception. In that spirit, I would like to take the time to thank and praise those who have significantly influenced my burgeoning career. First and foremost, I would like to thank Dr. Michael Terns and Dr. Rebecca Terns. Thank you for giving me an opportunity to work with so many intelligent and passionate people. Thank you for the opportunity to work in one of the most exciting fields of molecular biology: CRISPR-Cas. Dr. Rebecca Terns, I additionally thank you for your counsel and kind words of encouragement. You helped me find a path when I was unsure of myself, and I thank you for being a guide. I thank my additional committee members Dr. Sidney Kushner and Dr. Claiborne Glover for their support and helpful advice. As I previously stated, it takes a village to raise a scientist. As such, I would be remiss not to acknowledge the Terns lab members past and present, especially Joshua Elmore, Morgan Teachey, Julie Grainy, Nolan Sheppard, Walter Woodside, Kawanda Foster, Elizabeth Watts, Caryn Hale, Sonali Majumdar, and Jason Carte. I thank you for the times we learned together, supported one another, and worked together. It was my privilege to work with and train my undergraduate researchers Justin Dumrongkulraksa and Chip Chambers. I hope you learned from me as I learned from you, and I hope our work provided a gateway into a larger world of learning for each of you. Lastly, I would thank our collaborators Dr. Lance Wells, Dr. Linda Zhao, Dr. Chris Stephens, and Dr. Mark Compton. Thank you for your hard and invaluable efforts!

TABLE OF CONTENTS

	Page
Acknowledgements	#v
 CHAPTER	
1) Introduction & literature review	#1
References	#18
Figures and Tables	#26
 2) The Trans-acting Factors of CRISPR-Cas Adaptation in the Hyperthermophilic Archaeon <i>Pyrococcus furiosus</i>	
.....	#34
Introduction	#34
Results	#38
Discussion	#47
Materials and methods	#56
References	#62
Figures and Tables	#66
 3) Appendix	
Strains and Plasmids	#147
Abbreviations	#147

CHAPTER 1

INTRODUCTION AND LITERATURE REVIEW

Introduction and Relevance of CRISPR-Cas Systems

Prokaryotic organisms are an abundant and diverse source of life on this planet. These microorganisms are the ancestors of eukaryotic life, and akin to their predecessors, are subject to invasion from viruses and other mobile genetic elements. To persist and survive, prokaryotes and phages compete in a constant struggle of infection and defense. This competition creates an evolutionary arms race between weapons of attack and defense among these organisms. To this end, prokaryotes, akin to many eukaryotic organisms, have evolved both innate and adaptive immune systems to fight infection (Chopin et al., 2005; Cook et al., 2013; Labrie et al., 2010; Stern and Sorek, 2011).

In prokaryotes, innate immunity consists of restriction-modification systems, abortive infection, prevention of phage adsorption, and toxin/antitoxin systems (Chopin et al., 2005; Cook et al., 2013; Labrie et al., 2010; Stern and Sorek, 2011). CRISPR (Clustered Regularly Interspaced Short Palindromic Repeats)-Cas (CRISPR-Associated) systems comprise a newly discovered, adaptive mode of prokaryotic immunity (Bolotin et al., 2005; Godde and Bickerton, 2006; Grissa et al., 2007a). Present in 40% of all sequenced bacterial genomes and ~90% of archaeal genomes, these systems rely on a sequence-based RNA-guided targeting of a recognized invader (Grissa et al., 2007b). The immunity is specific to nucleic acids with the specified sequence and is heritable by subsequent generations.

In addition to the immediate relevance of CRISPR-Cas to the ability of prokaryotes to survive infection, the scientific study of CRISPR-Cas has had far-reaching technological impacts. CRISPR's programmable immunity has benefitted many industries via the protection of valuable strains from phage infection (Barrangou and Horvath, 2012; Terns and Terns, 2014). Sequence-guided nuclease components of CRISPR-Cas, most notably the Cas9 protein, have gained attention and renown as genome-editing tools (Cong et al., 2013; DiCarlo et al., 2013; Hsu et al., 2014; Jinek et al., 2012; Jinek et al., 2013; Mali et al., 2013; Yang et al., 2014). Cas9, with the aid of crRNA (CRISPR RNA) targeting, can cheaply and efficiently make highly specific alterations to genomes. Cas9 has therefore contributed to new methods in gene knockout and mutation across many different model organisms, speeding the study of countless biological questions. The continued study of the CRISPR-Cas systems and their components promises to yield additional understanding about the interplay of prokaryotes and their viral invaders as well as additional biological tools of untold use.

CRISPR Locus Structure

The hallmark of CRISPR-Cas immune systems is a unique genetic structure known as the CRISPR locus (Figure 1.1). The locus has a specialized structure consisting of repeated DNA elements alternated with invader-derived DNA sequences called spacers. *In silico* analyses of CRISPR loci resulted in the mapping of these spacer sequences to phage genomes, thus provoking the hypothesis that CRISPR loci may in fact function as immune systems (Bolotin et al., 2005; Mojica et al., 2005; Pourcel et al., 2005). Spacers are separated by repeating sequence elements of approximately twenty to forty base pairs called repeats (Jansen et al., 2002). An AT-rich leader region is located upstream of the spacer and repeat sequences. The leader contains the locus promoter sequence, as well as additional sequences critical for the acquisition of new

invader sequences (Jansen et al., 2002; Pougach et al., 2010; Yosef et al., 2012). The integration of new spacers is highly directional, occurring in a leader-proximal fashion (Yosef et al., 2012). Modules of *cas* genes that encode Cas proteins flank the CRISPR locus (Jansen et al., 2002). These proteins make up the crucial factors that drive CRISPR immunity. Transcription of the CRISPR locus leads to the production of CRISPR RNAs (crRNAs), which guide sequence-specific recognition of invaders with the aid of effector Cas proteins. Cas proteins act as essential machinery for CRISPR-Cas processes, providing for the capture of foreign DNA sequences or protospacers, the biogenesis and processing of crRNAs, and the execution of invader silencing.

CRISPR-Cas Mechanisms

Research into CRISPR-Cas systems has provided a model that immunity occurs in three stages: adaptation, crRNA biogenesis, and invader silencing or defense (Figure 1.2). The first stage, known as adaptation, can be thought of as the acquisition of immunity. During this stage, foreign DNA sequences derived from the invading phage are captured by Cas proteins and integrated into the CRISPR locus (Bolotin et al., 2005; Mojica et al., 2005). Phage DNA sequences targeted for integration at this step are referred to as protospacers. Protospacers are preferentially selected from sequences containing a short sequence motif of several nucleotides called a PAM or protospacer adjacent motif (Datsenko et al., 2012; Deveau et al., 2008; Mojica et al., 2009; Swarts et al., 2012). The protospacer adjacent motif is a 2-7 nucleotide element that serves as a recognition site for the processing and acquisition of nearby protospacers as well as in defense. Integration of a new spacer occurs in a directional manner, occurring at the leader proximal end of the CRISPR array (Barrangou, 2013; Barrangou et al., 2007). This step is accompanied by a duplication of the first repeat sequence ensuring that each new spacer is flanked by a repeat and maintaining the structure of the locus (Barrangou et al., 2007).

The second phase, called crRNA biogenesis, refers to the transcription and processing of the crRNA. As the CRISPR locus is transcribed, the nascent transcript is full-length and contains numerous different spacer sequences. To ensure precise targeting to a single invader, the long immature crRNA transcript is processed into small single mature crRNAs, each containing a single spacer sequence. Cas6 is responsible for the primary processing in this step, cleaving within repeat sequences of the nascent crRNA and yielding smaller crRNAs now specific to a single invader (Carte et al., 2010; Carte et al., 2008; Gesner et al., 2011; Haurwitz et al., 2010; Haurwitz et al., 2012; Nam et al., 2012; Niewoehner et al., 2014; Sternberg et al., 2012; Wiedenheft et al., 2011).

In the final step, crRNAs are sorted into effector complexes of Cas proteins, which then carry out the targeted destruction of the invader. The crRNAs are coordinated by interacting Cas protein complexes during this step and provide the sequence-specific targeting of the invader by base pairing (Brouns et al., 2008; Gasiunas et al., 2012; Jinek et al., 2012; Jore et al., 2011; Wiedenheft et al., 2011). PAM sequences serve as recognition motifs for distinguishing self from non-self during this step and are essential for the cleavage of target DNA (Mojica et al., 2009). Interference complexes can take two distinct forms. Multisubunit complexes such as Cascade, Csm, and Cmr are hallmarks of Type I and Type III CRISPR-Cas systems (Makarova et al., 2011b; Makarova et al., 2015), whereas Type II CRISPR systems carry out crRNA coordination and immunity via the Cas9 protein (Gasiunas et al., 2012; Jinek et al., 2012). Interference or cleavage of the target can occur by different mechanisms. For some CRISPR systems, nuclease activity may come from the effector complex itself as in Cas9 systems (Gasiunas et al., 2012; Jinek et al., 2012). However, in others, an outside protein is recruited to perform destruction of

the targeted phage DNA, as in Cas3/Type I systems (Beloglazova et al., 2011; Brouns et al., 2008; Sinkunas et al., 2011).

CRISPR-Cas Classification and Biodiversity

CRISPR-Cas systems exhibit remarkable diversity in both mechanism and components. The continual discovery of new systems has repeatedly prompted the expansion of the existing classification system regarding CRISPR-Cas. Currently, there are two classes and six overall types (Classes 1 and 2, Types I-VI), and these may or may not contain further subtypes (Makarova et al., 2015; Mohanraju et al., 2016). CRISPR systems containing a multiprotein effector complex as in Type I and Type III systems are noted as class 1 (Types I, III, and IV) and are observed in both bacteria and archaea. Conversely, systems containing a single “all in one” effector nuclease such as Cas9 from the type II-A and II-B systems are referred to as class 2 and are mostly restricted to bacteria.

In total, there are now 19 CRISPR-Cas system subtypes, each classified under a primary type system with the different subtypes distinguished by letters (Type I A-G, Type II A-C, Type III A-D, Type IV, Type V, and Type VI). Each type is distinguished by a signature Cas protein, Cas3 for Type I, Cas9 for Type II, Cas10 for Type III, Cpf1 for Type V, and C2c2 for Type VI. Type IV systems contain some elements from the Cascade effector complex such as the Cas5 and Cas7 proteins but lacks a CRISPR array and the Cas1/Cas2 proteins. Because of the absence of crRNA-based defense in the Type IV system, it has been proposed that this system may function more akin to an innate immune system recognizing preset sequences (Rath et al., 2015). Overall, the Cas proteins exhibit remarkable diversity and are spread over 45 Cas gene families, although a single organism will only possess a given number or subset of these proteins (Haft et al., 2005).

CRISPR-Cas in *Pyrococcus furiosus*

CRISPR loci and Cas proteins

Pyrococcus furiosus (Pfu) is a hyperthermophilic archaeon. This microbe was originally isolated from a marine thermal sediment and has an optimal growth temperature of 100°C (Fiala, 1986). The organism is notable for the isolation and use of its thermostable DNA polymerase for carrying out PCR (polymerase chain reaction). The COM1 strain of *Pyrococcus furiosus* (used as a base strain in this study) contains seven CRISPR loci notated as CRISPRs 1, 2, 4, 5, 6, 7, and 8, skipping CRISPR locus three because of a previous mis-annotation (Figure 1.3) (Terns and Terns, 2013). Interestingly, a single, 30 base pair repeat sequence is shared amongst all of the CRISPR loci despite the presence of multiple CRISPR immune systems. Spacer content varies between the seven loci with CRISPR 1 having the most at 50 spacers and CRISPR 8 possessing only 11 spacers. In total, 200 spacers are encoded by the CRISPR loci of this organism. The Cas protein machinery for these CRISPR loci is encoded in two separate loci bordering the CRISPR 5 and CRISPR 7 loci. Together, these two gene clusters encode 27 Cas proteins employed across the across three immune systems.

*crRNA Biogenesis & Processing in *Pyrococcus furiosus**

Studies from our laboratory have revealed much about crRNA biogenesis and processing in *P. furiosus*. Each one of the seven CRISPR loci is actively transcribed and produces crRNAs (Hale et al., 2008; Hale et al., 2009; Terns and Terns, 2013). Cas6 is the primary enzyme responsible for the necessary processing and maturation of crRNAs from each of the CRISPR loci, functioning as a metal-independent endoribonuclease (Carte et al., 2010; Carte et al., 2008; Terns and Terns, 2013; Wang et al., 2011). Cas6 acts by cleaving within the multiple repeat sequences present within long immature crRNAs. Prior to this action by Cas6, nascent crRNA

transcripts contain multiple spacer units that correspond to various invaders. Cleavage by Cas6 yields individual crRNAs that are now capable of targeting a single specific invader, and can be sorted into distinct effector complexes. Initial cleavage by Cas6 yields what is referred to as the 1X intermediate crRNA, which contains an eight nucleotide segment of the repeat on the 5' end of the crRNA (5'-AUUGAAAG-3'), the spacer-related/guide sequence of the crRNA, and 22 nucleotides of unprocessed flanking repeat sequence on the 3' end. As mentioned previously, the repeat sequences within the CRISPR loci are 30 base pairs and are thus responsible for this eight and 22 nucleotide split on the 5' and 3' ends of the 1X intermediate, respectively. Further trimming of the 1X intermediate occurs at the 3' end, eliminating the 22 nucleotides of repeat (Carte et al., 2010; Carte et al., 2008). The mechanism for this additional processing, however, is still unknown. The final processing product is a 45 nucleotide crRNA with a 37 nucleotide guide sequence and eight nucleotide 5' tag. The mature crRNAs are then incorporated into the various effector nuclease complexes, which can then carry out invader silencing (Elmore et al., 2015; Hale et al., 2012; Majumdar et al., 2015).

Pyrococcus furiosus CRISPR Immune systems

Together, the Cas genes and proteins of *Pyrococcus furiosus* comprise three separate and independent CRISPR immune systems that utilize the 7 CRISPR loci (Elmore et al., 2015; Hale et al., 2012; Majumdar et al., 2015). Two systems, Type I-A, and Type I-G, fall under the type one classification hallmarked by a Cas3 signature protein. The third system is a Type III-B immune system with a Cas10 signature protein. The two Type I systems have been characterized as DNA targeting systems, while the Type III-B system can be best described as a transcription-dependent DNA targeting immune system (Elmore et al., 2015; Elmore et al., 2016; Majumdar et al., 2015). Thus, DNA cleavage in the Type III-B system is dependent upon the

transcription of a foreign or target RNA. The crRNA binding to the target RNA then serves to activate the effector cleavage of invader DNA (Elmore et al., 2016). Each of these independent CRISPR immune systems can associate with crRNAs produced from any of the seven CRISPR loci (Terns and Terns 2013).

CRISPR-Cas Adaptation

Adaptation in CRISPR-Cas systems serves as a necessary first step in developing immunity to a particular invader and is the focus of this study. Adaptation crucially generates a molecular memory of past infections. It is in this process that a piece of invader or viral DNA termed a “protospacer” is identified, processed, and integrated into the CRISPR locus at the leader/repeat junction. The integration of a new viral sequence later allows for the generation of a crRNA targeting this same sequence, conferring immunity. The model systems of *E. coli* (I-E) together with *S. thermophilus* (II-A) and *S. pyogenes* (type II-A) have served as benchmarks for investigation of this topic in the CRISPR field (Barrangou et al., 2007; Heler et al., 2015; Nunez et al., 2014; Wei et al., 2015a; Wei et al., 2015b; Yosef et al., 2012). It was in the *Streptococcus thermophilus* II-A system where spacer acquisition was equated with a gain of immunity for the first time (Barrangou et al., 2007). The I-E *E. coli* system then provided crucial early examples of the *in vivo* adaptation requirements for both cis-acting sequences and trans-acting factors (Datsenko et al., 2012; Yosef et al., 2012). However, adaptation studies in Type-III systems and many others remain underrepresented (Amitai and Sorek, 2016; Sternberg et al., 2016). The involvement of the universal Cas proteins Cas1 and Cas2, however, suggests a possible conserved mechanism across different CRISPR systems at the integration step (Makarova et al., 2011b; Makarova et al., 2015).

Protospacer Selection

The first step in the capture of any new protospacer is the initial selection of foreign DNA to undergo integration into the CRISPR locus. Various studies in CRISPR adaptation indicate that protospacer selection is a targeted process. The protospacer adjacent motif or PAM is a 2-7 nucleotide element that serves as a recognition site for the processing and acquisition of nearby protospacers from the invading genome or plasmid (Deveau et al., 2008; Mojica et al., 2009). PAMs have been demonstrated to be crucial in the selection of new protospacers. Several studies mapping protospacers against plasmid and viral genomes have highlighted a distinct correlation of neighboring PAMs flanking protospacers (Datsenko et al., 2012; Deveau et al., 2008; Erdmann and Garrett, 2012; Swarts et al., 2012; Yosef et al., 2012). Beyond their roles in adaptation, PAMs are also vital in distinguishing self from non-self in Type I and Type II CRISPR-Cas immune systems.

In addition to PAM biases, a recent *E. coli* genome-wide study of protospacer acquisition found that dsDNA breaks could be a major source of protospacers (Levy et al., 2015). Typical sites of stalled DNA replication such as termination sites can be areas especially prone to dsDNA breaks. Deep sequencing analyses of millions of protospacers illustrated protospacer “hotspots” correlating to these stalled fork regions reaching to the nearest *Chi* site. As the RecBCD complex is responsible for trimming dsDNA breaks to *Chi* sites in *E. coli*, it was theorized that intermediates of RecB cleavage make up a potential pool of protospacers. The lack of *Chi* sites on foreign DNA resulting in increased RecB cleavages is further hypothesized to contribute to self-nonsel discrimination. Importantly, this study linked the processes of DNA repair and recombination to adaptation, as these two pathways influence both protospacer generation and p additional factors necessary for the adaptation process of trans-acting factors and in the

theoretical pool of protospacers (Levy et al., 2015). Another recent study from the same system, however, suggests that the potential spacer pool results directly from cleavage products arising from the Type I Cas3 effector nuclease during primed adaptation (Musharova et al., 2017).

Naïve and Primed Adaptation

Two general modes of adaptation have been described in Type I CRISPR-Cas systems. Naïve adaptation can be described as a form of immune system-independent acquisition. In the Type I-E system, naïve adaptation only requires the Cas1 and Cas2 proteins putatively thought to compose the integrase responsible for protospacer integration into the CRISPR locus (Yosef et al., 2012). The acquisition of new spacers under naïve adaptation is biased towards extrachromosomal DNA, and is inefficient when compared to primed adaptation. (Diez-Villasenor et al., 2013; Levy et al., 2015). In contrast, primed adaptation is a more efficient, immune system-dependent form of spacer acquisition. Priming is reliant upon a preexisting spacer as well as all the components of the CRISPR immune system, such as crRNAs, and the effector complexes/nucleases. The preexisting spacer, often with mismatches in PAM or targeted phage sequence allows for an enhanced acquisition of new spacers from a previously encountered invader. Additionally, spacer acquisition is enhanced *in cis* in regards to the crRNA/immune system target sequence. This creates strand biases of spacer acquisition, where under primed conditions, the non-target strand is preferred (Datsenko et al., 2012; Savitskaya et al., 2013; Swarts et al., 2012). Under naïve adaptation, no such biases exist (Yosef et al., 2012).

Cas Proteins in Adaptation

Cas1, Cas2, and Spacer Integration

United across nearly all CRISPR-Cas systems is the ubiquitous presence of the Cas1 and Cas2 proteins (Makarova et al., 2011b; Makarova et al., 2015). These proteins have been shown

through various studies not to be associated with or required for the crRNA biogenesis or invader silencing steps (Heler et al., 2015; Makarova et al., 2011b; Makarova et al., 2015). However, their degree of conservation has suggested a common model of adaptation across the many separate CRISPR-Cas systems (Makarova et al., 2011b; Makarova et al., 2015).

Initial studies of these proteins *in vitro* suggested that these proteins were nucleases but whose activities were not clearly related to the adaptation process. Specifically, studies demonstrated Cas1 as a metal-dependent ssDNA and dsDNA nuclease, and Cas2 as an endoribonuclease (Babu et al., 2011; Beloglazova et al., 2008; Wiedenheft et al., 2009). These cleavage activities lacked sequence specificity. However, Cas1 in *E. coli* did appear to demonstrate a preference for branched DNA substrates (Babu et al., 2011). Despite these studies, the true activity of these proteins *in vivo* remained initially undetermined.

As time progressed, *in vivo* evidence began to build linking Cas1 and Cas2 inexorably to spacer acquisition. Overexpression of the Cas1 and Cas2 proteins was shown to provide increased rates of spacer acquisition *in vivo* in the *E. coli* Type I-E system (Yosef et al., 2012). Moreover, in multiple studies, the presence of both Cas1 and Cas2 was essential to the acquisition of new spacers *in vivo* (Datsenko et al., 2012; Heler et al., 2015; Li et al., 2014; Nunez et al., 2014; Yosef et al., 2012). The nuclease domain of Cas1 was shown to be essential for spacer acquisition, but the catalytic activity of Cas2 appeared to be nonessential (Nunez et al., 2014; Yosef et al., 2012). Thus, Cas1 was implicated as the catalytic component in spacer integration with Cas2 playing a non-enzymatic role in the Type I-E CRISPR system.

An additional significant gain of knowledge came from the realization that Cas1 and Cas2 were later shown to form a complex that had with integrase activity. Together, Cas1 and Cas2 form an asymmetrical heterohexameric complex comprised of two Cas1 dimers bridged by

a Cas2 dimer (Moch et al., 2016; Nunez et al., 2015a; Nunez et al., 2014; Wang et al., 2015) (Figure 1.4A). *In vitro*, these two proteins make up the minimal machinery responsible for the final step of adaptation, the insertion of protospacers into the leader proximal end of the CRISPR locus in a two-step transesterification reaction (Nunez et al., 2015b; Rollie et al., 2015). This reaction occurs via two nucleophilic attacks of a 3'OH at the leader/repeat junction of the CRISPR locus. In the first step, a covalent intermediate is formed when the protospacer's 3'OH carries out a nucleophilic attack on the first strand of the repeat. The second step is performed by a second attack on the opposite DNA strand. The product of the full reaction is an expanded CRISPR array with single-stranded gaps on either side of the newly integrated spacer (Figure 1.4B) (Arslan et al., 2014). These sequence gaps are filled in by unknown DNA repair machinery. DNA polymerase I has been implicated as an essential factor for adaptation in the Type I-E system. Therefore, it is a possible candidate for the necessary polymerase in locus repair (Ivancic-Bace et al., 2015). The cis-acting sequences of the leader and repeat are both critical in the integration process. Mutations in the first 60 base pairs of the leader abrogate adaptation in *E. coli* (Yosef et al., 2012). A minimum of a single repeat is also required for integration at the CRISPR locus (Wei et al., 2015a) (Yosef et al., 2012). Mutations in the Type II-A system (Wei et al., 2015a). In Type I-E systems, Cas1 and Cas2 are responsible for PAM recognition, but this is not invariant in other systems (Wang et al., 2015).

Type II-A System Cas9 & Csn2

Apart from the Cas1 and Cas2 proteins, there are additional proteins implicated in the adaptation process. In the Type II-A system, the Csn2 protein has been linked genetically to adaptation, but its exact role is still unknown (Barrangou et al., 2007; Heler et al., 2015). Csn2 is a DNA binding protein that forms a cylindrical tetramer that lacks apparent nuclease activity in

studies performed thus far. *In vitro* studies have indicated that Csn2 binds DNA through the center of this cylindrical structure, with the ability to bind and slide upon free DNA ends (Arslan et al., 2013; Nam et al., 2011). This behavior bears similarities to conserved Ku protein, which facilitates non-homologous end joining (NHEJ), when binding to free DNA ends (Arslan et al., 2013). Csn2 was recently observed to form a large adaptation complex with the Cas1, Cas2, and Cas9 proteins in *S.pyogenes* (Heler et al., 2015). The entirety of this complex is required for adaptation *in vivo*. Cas9 serves a dual role being integral in both defense as an effector nuclease and as a part of the adaptation complex. In adaptation, Cas9 is able to provide a mechanism of PAM recognition. This result was confirmed via the study of *S.pyogenes* Cas9, where mutation of the PAM recognition domains in Cas9 for the interference pathway resulted in a consequent loss of PAM motif selection for incoming spacers (Heler et al., 2015).

Cas4

Cas4 is another likely Cas trans-acting factor linked to adaptation but with a currently unknown role. Cas4 is associated with multiple CRISPR-Cas subtypes (I-A, I-B, I-C, I-U, I-D, II-B). Additionally, Cas4 is often found genetically associated in operons with the Cas1 and Cas2 proteins (Makarova et al., 2011b; Makarova et al., 2015). In some cases, Cas1 and Cas4 genes have been found fused together (as in *Thermoproteus tenax* and *Myxococcus xanthus*), suggesting that they may function together (Plagens et al., 2012; Viswanathan et al., 2007).

In vitro studies have attributed a number of different nuclease activities to Cas4 proteins. While there is a dearth of information overall, *in vitro* research performed on Cas4 in the hyperthermophiles *Sulfolobus solfataricus* (SSO0001 & SSO1391) and *Pyrobaculum calidifontis* (Pcal 0546) have attributed a variety of nuclease activities to this protein (Lemak et al., 2013; Lemak et al., 2014; Zhang et al., 2012). SSO0001, a Cas4 from *S. solfataricus*, demonstrates

metal-dependent endonuclease and 5'-3' exonucleolytic activities (Zhang et al., 2012). It also shows a low level of ATP-independent unwinding activity. This protein further exhibits a decameric toroid quaternary structure (Lemak et al., 2013). Pcal 0546 (*P. calidifontis* Cas4), like SSO0001, also demonstrates unwinding and 5'-3' exonuclease activity (Lemak et al., 2014). Unlike the others, SSO1391 exhibits a bidirectional exonuclease activity (Lemak et al., 2014). These various activities have been shown to be generated by a RecB-like nuclease domain, common among the Cas4 proteins, in each case. RecB is a part of the RecBCD complex acting to perform homologous recombination in gram-negative bacteria (Dillingham and Kowalczykowski, 2008). RecB will trim broken DNA ends using its exonuclease activity to a Chi site, creating DNA overhangs primed for recombination.

Another structural aspect shared among Cas4 proteins is the presence of Fe-S clusters similar to AddB nucleases which act in homologous recombination in gram-positive bacteria (Lemak et al., 2013; Lemak et al., 2014). Fe-S cluster proteins commonly function in the regulation of transcription, DNA repair, electron transport, and iron storage (Wu and Brosh, 2012; Zhang, 2014). *P. furiosus* possesses two separate Cas4 proteins noted as Cas4-1 and Cas4-2 (Figure 1.3C) (Terns and Terns, 2013). One is genetically associated with the CRISPR locus (notated in this study as Cas4-1 PF1119), while the other is not associated with a CRISPR locus (Cas4-2 PF1793). The function and role of both proteins in adaptation are a primary focus of this study.

Non-Cas Proteins in Adaptation

Recent studies in the *E. coli* Type I-E system have resulted in new insights regarding critical non-Cas factors in adaptation. This information can be further used to understand possible factors needed in other model systems. One particular study addresses how integration

specificity to the leader proximal end of the CRISPR locus is achieved. The factor IHF (integration host factor) induces sequence-specific DNA-bending directly upstream of the first repeat within the leader sequence. This bending allows greater access to the scissile phosphate at the first repeat conveying increased specificity (Nunez et al., 2016). This work indicates that it is possible that DNA topology and the ability of Cas1 to access the DNA could impact adaptation. IHF is limited to gram-negative bacteria (Dillon and Dorman, 2010). Thus, it is possible that other factors induce conformational changes in DNA at the CRISPR locus in other systems. These changes in topology may play a role in conveying leader/repeat specificity and localizing integration events.

Additionally, the repair of the CRISPR locus after integration is still an on-going topic of research. DNA repair of the duplicated but still single-stranded repeat after integration is likely to be carried out via a non-Cas mechanism. In *E. coli*, DNA polymerase I has been implicated as essential for both primed and naïve adaptation (Ivancic-Bace et al., 2015). It is possible that this protein is the fill-in polymerase in this organism. Unfortunately, the ligases and polymerases needed to repair the relevant nicked DNA and the CRISPR repeat in other systems have not yet been studied.

Non-Cas trans-acting factors may also play roles in protospacer selection and processing. Specifically, proteins implicated in DNA replication and repair appear to be linked to adaptation. As noted previously, RecB is one such candidate. In *E. coli*, sites of replication fork stalling and double-stranded DNA breaks are hotspots of protospacer acquisition under naïve conditions (Levy et al., 2015). These sites extend to the borders of *Chi* sites, implying that the intermediates of RecB cleavage are possible adaptation substrates. Interestingly, deletions of each of the RecBCD proteins all result in the decrease of protospacer uptake and loss of these hotspots

(Levy et al., 2015). Further work in *E. coli* illustrates that RecB has a large effect on naïve, but not primed adaptation, likely implicating a difference in substrate generation between the two proteins (Ivancic-Bace et al., 2015). Cas1 also has been shown to interact with RecB in *E. coli* (Babu et al., 2011). Taken together, it is clear that the RecBCD enzyme may process and create substrates for integration. The similarities between Cas4 and RecB in their nuclease domain and exonuclease activity could indicate a similar role in adaptation across different CRISPR types and subtypes. Under primed conditions, additional proteins such as the RecG helicase and PriA have been shown to be essential for adaptation and may be candidates for further study (Ivancic-Bace et al., 2015).

Context and Significance of this Study

While the role and activity of Cas1 and Cas2 during integration has more recently come to light in the *E. coli* Type I-E system, there are many aspects of adaptation that remain to be fully understood. These areas include the mechanistic steps of how protospacers are selected and processed before integration by Cas1 and Cas2. The additional non-Cas factors influencing the adaptation process with regard to both protospacer selection and integration are also not well identified. This is particularly the case outside of the *E. coli* model system where most of the recent knowledge gains have been focused. In addition, the role of Cas4 in adaptation has not been addressed.

In this work, we sought to understand adaptation in the archaeal model system *Pyrococcus furiosus* by studying the *in vivo* interactions of adaptation Cas proteins Cas1, Cas2, and Cas4-1. We also sought to generate non-Cas protein candidates for further study that may be influential in the adaptation process. Based on available data from the Type I-E system (Nunez et al., 2015a; Nunez et al., 2014; Nunez et al., 2015b), and given the overall conservation of Cas1

and Cas2 across most CRISPR systems, we hypothesized that Cas1 and Cas2 likely form a complex and acted as an integrase in *P. furiosus*. The genetic association between Cas1 and Cas4 as well as the existence of fusion proteins may also indicate that Cas1 and Cas4 may function together (Jansen et al., 2002; Makarova et al., 2011a; Makarova et al., 2015). We also predicted that proteins implicated in DNA repair, topology, and recombination may be found associated with the adaptation proteins to generate and process potential substrates for Cas1 and Cas2.

References

- Amitai, G., and Sorek, R. (2016). CRISPR-Cas adaptation: insights into the mechanism of action. *Nat Rev Microbiol* 14, 67-76.
- Arslan, Z., Hermanns, V., Wurm, R., Wagner, R., and Pul, U. (2014). Detection and characterization of spacer integration intermediates in type I-E CRISPR-Cas system. *Nucleic Acids Res* 42, 7884-7893.
- Arslan, Z., Wurm, R., Brener, O., Ellinger, P., Nagel-Steger, L., Oesterhelt, F., Schmitt, L., Willbold, D., Wagner, R., Gohlke, H., *et al.* (2013). Double-strand DNA end-binding and sliding of the toroidal CRISPR-associated protein Csn2. *Nucleic Acids Res* 41, 6347-6359.
- Babu, M., Beloglazova, N., Flick, R., Graham, C., Skarina, T., Nocek, B., Gagarinova, A., Pogoutse, O., Brown, G., Binkowski, A., *et al.* (2011). A dual function of the CRISPR-Cas system in bacterial antiviral immunity and DNA repair. *Mol Microbiol* 79, 484-502.
- Barrangou, R. (2013). CRISPR-Cas systems and RNA-guided interference. *Wiley Interdiscip Rev RNA* 4, 267-278.
- Barrangou, R., Fremaux, C., Deveau, H., Richards, M., Boyaval, P., Moineau, S., Romero, D.A., and Horvath, P. (2007). CRISPR provides acquired resistance against viruses in prokaryotes. *Science* 315, 1709-1712.
- Barrangou, R., and Horvath, P. (2012). CRISPR: new horizons in phage resistance and strain identification. *Annu Rev Food Sci Technol* 3, 143-162.
- Beloglazova, N., Brown, G., Zimmerman, M.D., Proudfoot, M., Makarova, K.S., Kudritska, M., Kochinyan, S., Wang, S., Chruszcz, M., Minor, W., *et al.* (2008). A novel family of sequence-specific endoribonucleases associated with the clustered regularly interspaced short palindromic repeats. *J Biol Chem* 283, 20361-20371.
- Beloglazova, N., Petit, P., Flick, R., Brown, G., Savchenko, A., and Yakunin, A.F. (2011). Structure and activity of the Cas3 HD nuclease MJ0384, an effector enzyme of the CRISPR interference. *EMBO J* 30, 4616-4627.
- Bolotin, A., Quinkis, B., Sorokin, A., and Ehrlich, S.D. (2005). Clustered regularly interspaced short palindrome repeats (CRISPRs) have spacers of extrachromosomal origin. *Microbiology* 151, 2551-2561.
- Brouns, S.J., Jore, M.M., Lundgren, M., Westra, E.R., Slijkhuis, R.J., Snijders, A.P., Dickman, M.J., Makarova, K.S., Koonin, E.V., and van der Oost, J. (2008). Small CRISPR RNAs guide antiviral defense in prokaryotes. *Science* 321, 960-964.

- Carte, J., Pfister, N.T., Compton, M.M., Terns, R.M., and Terns, M.P. (2010). Binding and cleavage of CRISPR RNA by Cas6. *RNA* 16, 2181-2188.
- Carte, J., Wang, R., Li, H., Terns, R.M., and Terns, M.P. (2008). Cas6 is an endoribonuclease that generates guide RNAs for invader defense in prokaryotes. *Genes Dev* 22, 3489-3496.
- Chopin, M.C., Chopin, A., and Bidnenko, E. (2005). Phage abortive infection in lactococci: variations on a theme. *Curr Opin Microbiol* 8, 473-479.
- Cong, L., Ran, F.A., Cox, D., Lin, S., Barretto, R., Habib, N., Hsu, P.D., Wu, X., Jiang, W., Marraffini, L.A., *et al.* (2013). Multiplex genome engineering using CRISPR/Cas systems. *Science* 339, 819-823.
- Cook, G.M., Robson, J.R., Frampton, R.A., McKenzie, J., Przybilski, R., Fineran, P.C., and Arcus, V.L. (2013). Ribonucleases in bacterial toxin-antitoxin systems. *Biochim Biophys Acta* 1829, 523-531.
- Datsenko, K.A., Pougach, K., Tikhonov, A., Wanner, B.L., Severinov, K., and Semenova, E. (2012). Molecular memory of prior infections activates the CRISPR/Cas adaptive bacterial immunity system. *Nat Commun* 3, 945.
- Deveau, H., Barrangou, R., Garneau, J.E., Labonte, J., Fremaux, C., Boyaval, P., Romero, D.A., Horvath, P., and Moineau, S. (2008). Phage response to CRISPR-encoded resistance in *Streptococcus thermophilus*. *J Bacteriol* 190, 1390-1400.
- DiCarlo, J.E., Norville, J.E., Mali, P., Rios, X., Aach, J., and Church, G.M. (2013). Genome engineering in *Saccharomyces cerevisiae* using CRISPR-Cas systems. *Nucleic Acids Res* 41, 4336-4343.
- Diez-Villasenor, C., Guzman, N.M., Almendros, C., Garcia-Martinez, J., and Mojica, F.J. (2013). CRISPR-spacer integration reporter plasmids reveal distinct genuine acquisition specificities among CRISPR-Cas I-E variants of *Escherichia coli*. *RNA Biol* 10, 792-802.
- Dillingham, M.S., and Kowalczykowski, S.C. (2008). RecBCD enzyme and the repair of double-stranded DNA breaks. *Microbiol Mol Biol Rev* 72, 642-671,
- Dillon, S.C., and Dorman, C.J. (2010). Bacterial nucleoid-associated proteins, nucleoid structure and gene expression. *Nat Rev Microbiol* 8, 185-195.
- Elmore, J., Deighan, T., Westpheling, J., Terns, R.M., and Terns, M.P. (2015). DNA targeting by the type I-G and type I-A CRISPR-Cas systems of *Pyrococcus furiosus*. *Nucleic Acids Res* 43, 10353-10363.
- Elmore, J.R., Sheppard, N.F., Ramia, N., Deighan, T., Li, H., Terns, R.M., and Terns, M.P. (2016). Bipartite recognition of target RNAs activates DNA cleavage by the Type III-B CRISPR-Cas system. *Genes Dev*.

- Erdmann, S., and Garrett, R.A. (2012). Selective and hyperactive uptake of foreign DNA by adaptive immune systems of an archaeon via two distinct mechanisms. *Mol Microbiol* 85, 1044-1056.
- Fiala, G., Stetter, K.O. (1986). *Pyrococcus furiosus* sp. nov. represents a novel genus of marine heterotrophic Archaeobacteria growing optimally at 100. *Archives of Microbiology* 145, 56-61.
- Gasiunas, G., Barrangou, R., Horvath, P., and Siksnys, V. (2012). Cas9-crRNA ribonucleoprotein complex mediates specific DNA cleavage for adaptive immunity in bacteria. *Proc Natl Acad Sci U S A* 109, E2579-2586.
- Gesner, E.M., Schellenberg, M.J., Garside, E.L., George, M.M., and Macmillan, A.M. (2011). Recognition and maturation of effector RNAs in a CRISPR interference pathway. *Nat Struct Mol Biol* 18, 688-692.
- Godde, J.S., and Bickerton, A. (2006). The repetitive DNA elements called CRISPRs and their associated genes: evidence of horizontal transfer among prokaryotes. *J Mol Evol* 62, 718-729.
- Grissa, I., Vergnaud, G., and Pourcel, C. (2007a). The CRISPRdb database and tools to display CRISPRs and to generate dictionaries of spacers and repeats. *BMC Bioinformatics* 8, 172.
- Grissa, I., Vergnaud, G., and Pourcel, C. (2007b). CRISPRFinder: a web tool to identify clustered regularly interspaced short palindromic repeats. *Nucleic Acids Res* 35, W52-57.
- Haft, D.H., Selengut, J., Mongodin, E.F., and Nelson, K.E. (2005). A guild of 45 CRISPR-associated (Cas) protein families and multiple CRISPR/Cas subtypes exist in prokaryotic genomes. *PLoS Comput Biol* 1, e60.
- Hale, C., Kleppe, K., Terns, R.M., and Terns, M.P. (2008). Prokaryotic silencing (psi)RNAs in *Pyrococcus furiosus*. *RNA* 14, 2572-2579.
- Hale, C.R., Majumdar, S., Elmore, J., Pfister, N., Compton, M., Olson, S., Resch, A.M., Glover, C.V., 3rd, Graveley, B.R., Terns, R.M., *et al.* (2012). Essential features and rational design of CRISPR RNAs that function with the Cas RAMP module complex to cleave RNAs. *Mol Cell* 45, 292-302.
- Hale, C.R., Zhao, P., Olson, S., Duff, M.O., Graveley, B.R., Wells, L., Terns, R.M., and Terns, M.P. (2009). RNA-guided RNA cleavage by a CRISPR RNA-Cas protein complex. *Cell* 139, 945-956.
- Haurwitz, R.E., Jinek, M., Wiedenheft, B., Zhou, K., and Doudna, J.A. (2010). Sequence- and structure-specific RNA processing by a CRISPR endonuclease. *Science* 329, 1355-1358.
- Haurwitz, R.E., Sternberg, S.H., and Doudna, J.A. (2012). Csy4 relies on an unusual catalytic dyad to position and cleave CRISPR RNA. *EMBO J* 31, 2824-2832.

- Heler, R., Samai, P., Modell, J.W., Weiner, C., Goldberg, G.W., Bikard, D., and Marraffini, L.A. (2015). Cas9 specifies functional viral targets during CRISPR-Cas adaptation. *Nature* *519*, 199-202.
- Hsu, P.D., Lander, E.S., and Zhang, F. (2014). Development and Applications of CRISPR-Cas9 for Genome Engineering. *Cell* *157*, 1262-1278.
- Ivancic-Bace, I., Cass, S.D., Wearne, S.J., and Bolt, E.L. (2015). Different genome stability proteins underpin primed and naive adaptation in *E. coli* CRISPR-Cas immunity. *Nucleic Acids Res* *43*, 10821-10830.
- Jansen, R., Embden, J.D., Gastra, W., and Schouls, L.M. (2002). Identification of genes that are associated with DNA repeats in prokaryotes. *Mol Microbiol* *43*, 1565-1575.
- Jinek, M., Chylinski, K., Fonfara, I., Hauer, M., Doudna, J.A., and Charpentier, E. (2012). A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science* *337*, 816-821.
- Jinek, M., East, A., Cheng, A., Lin, S., Ma, E., and Doudna, J. (2013). RNA-programmed genome editing in human cells. *Elife* *2*, e00471.
- Jore, M.M., Lundgren, M., van Duijn, E., Bultema, J.B., Westra, E.R., Waghmare, S.P., Wiedenheft, B., Pul, U., Wurm, R., Wagner, R., *et al.* (2011). Structural basis for CRISPR RNA-guided DNA recognition by Cascade. *Nat Struct Mol Biol* *18*, 529-536.
- Labrie, S.J., Samson, J.E., and Moineau, S. (2010). Bacteriophage resistance mechanisms. *Nat Rev Microbiol* *8*, 317-327.
- Lemak, S., Beloglazova, N., Nocek, B., Skarina, T., Flick, R., Brown, G., Popovic, A., Joachimiak, A., Savchenko, A., and Yakunin, A.F. (2013). Toroidal structure and DNA cleavage by the CRISPR-associated [4Fe-4S] cluster containing Cas4 nuclease SSO0001 from *Sulfolobus solfataricus*. *J Am Chem Soc* *135*, 17476-17487.
- Lemak, S., Nocek, B., Beloglazova, N., Skarina, T., Flick, R., Brown, G., Joachimiak, A., Savchenko, A., and Yakunin, A.F. (2014). The CRISPR-associated Cas4 protein Pcal_0546 from *Pyrobaculum calidifontis* contains a [2Fe-2S] cluster: crystal structure and nuclease activity. *Nucleic Acids Res.*
- Levy, A., Goren, M.G., Yosef, I., Auster, O., Manor, M., Amitai, G., Edgar, R., Qimron, U., and Sorek, R. (2015). CRISPR adaptation biases explain preference for acquisition of foreign DNA. *Nature* *520*, 505-510.
- Li, M., Wang, R., Zhao, D., and Xiang, H. (2014). Adaptation of the *Haloarcula hispanica* CRISPR-Cas system to a purified virus strictly requires a priming process. *Nucleic Acids Res* *42*, 2483-2492.

- Majumdar, S., Zhao, P., Pfister, N.T., Compton, M., Olson, S., Glover, C.V., 3rd, Wells, L., Graveley, B.R., Terns, R.M., and Terns, M.P. (2015). Three CRISPR-Cas immune effector complexes coexist in *Pyrococcus furiosus*. *RNA* 21, 1147-1158.
- Makarova, K.S., Aravind, L., Wolf, Y.I., and Koonin, E.V. (2011a). Unification of Cas protein families and a simple scenario for the origin and evolution of CRISPR-Cas systems. *Biol Direct* 6, 38.
- Makarova, K.S., Haft, D.H., Barrangou, R., Brouns, S.J., Charpentier, E., Horvath, P., Moineau, S., Mojica, F.J., Wolf, Y.I., Yakunin, A.F., *et al.* (2011b). Evolution and classification of the CRISPR-Cas systems. *Nat Rev Microbiol* 9, 467-477.
- Makarova, K.S., Wolf, Y.I., Alkhnbashi, O.S., Costa, F., Shah, S.A., Saunders, S.J., Barrangou, R., Brouns, S.J., Charpentier, E., Haft, D.H., *et al.* (2015). An updated evolutionary classification of CRISPR-Cas systems. *Nat Rev Microbiol* 13, 722-736.
- Mali, P., Yang, L., Esvelt, K.M., Aach, J., Guell, M., DiCarlo, J.E., Norville, J.E., and Church, G.M. (2013). RNA-guided human genome engineering via Cas9. *Science* 339, 823-826.
- Moch, C., Fromant, M., Blanquet, S., and Plateau, P. (2016). DNA binding specificities of *Escherichia coli* Cas1-Cas2 integrase drive its recruitment at the CRISPR locus. *Nucleic Acids Res.*
- Mohanraju, P., Makarova, K.S., Zetsche, B., Zhang, F., Koonin, E.V., and van der Oost, J. (2016). Diverse evolutionary roots and mechanistic variations of the CRISPR-Cas systems. *Science* 353, aad5147.
- Mojica, F.J., Diez-Villasenor, C., Garcia-Martinez, J., and Almendros, C. (2009). Short motif sequences determine the targets of the prokaryotic CRISPR defence system. *Microbiology* 155, 733-740.
- Mojica, F.J., Diez-Villasenor, C., Garcia-Martinez, J., and Soria, E. (2005). Intervening sequences of regularly spaced prokaryotic repeats derive from foreign genetic elements. *J Mol Evol* 60, 174-182.
- Musharova, O., Klimuk, E., Datsenko, K.A., Metlitskaya, A., Logacheva, M., Semenova, E., Severinov, K., and Savitskaya, E. (2017). Spacer-length DNA intermediates are associated with Cas1 in cells undergoing primed CRISPR adaptation. *Nucleic Acids Res.*
- Nam, K.H., Haitjema, C., Liu, X., Ding, F., Wang, H., DeLisa, M.P., and Ke, A. (2012). Cas5d protein processes pre-crRNA and assembles into a cascade-like interference complex in subtype I-C/Dvulg CRISPR-Cas system. *Structure* 20, 1574-1584.
- Nam, K.H., Kurinov, I., and Ke, A. (2011). Crystal structure of clustered regularly interspaced short palindromic repeats (CRISPR)-associated Csn2 protein revealed Ca²⁺-dependent double-stranded DNA binding activity. *J Biol Chem* 286, 30759-30768.

- Niewoehner, O., Jinek, M., and Doudna, J.A. (2014). Evolution of CRISPR RNA recognition and processing by Cas6 endonucleases. *Nucleic Acids Res* 42, 1341-1353.
- Nunez, J.K., Bai, L., Harrington, L.B., Hinder, T.L., and Doudna, J.A. (2016). CRISPR Immunological Memory Requires a Host Factor for Specificity. *Mol Cell* 62, 824-833.
- Nunez, J.K., Harrington, L.B., Kranzusch, P.J., Engelman, A.N., and Doudna, J.A. (2015a). Foreign DNA capture during CRISPR-Cas adaptive immunity. *Nature* 527, 535-538.
- Nunez, J.K., Kranzusch, P.J., Noeske, J., Wright, A.V., Davies, C.W., and Doudna, J.A. (2014). Cas1-Cas2 complex formation mediates spacer acquisition during CRISPR-Cas adaptive immunity. *Nat Struct Mol Biol* 21, 528-534.
- Nunez, J.K., Lee, A.S., Engelman, A., and Doudna, J.A. (2015b). Integrase-mediated spacer acquisition during CRISPR-Cas adaptive immunity. *Nature* 519, 193-198.
- Plagens, A., Tjaden, B., Hagemann, A., Randau, L., and Hensel, R. (2012). Characterization of the CRISPR/Cas subtype I-A system of the hyperthermophilic crenarchaeon *Thermoproteus tenax*. *J Bacteriol* 194, 2491-2500.
- Pougach, K., Semenova, E., Bogdanova, E., Datsenko, K.A., Djordjevic, M., Wanner, B.L., and Severinov, K. (2010). Transcription, processing and function of CRISPR cassettes in *Escherichia coli*. *Mol Microbiol* 77, 1367-1379.
- Pourcel, C., Salvignol, G., and Vergnaud, G. (2005). CRISPR elements in *Yersinia pestis* acquire new repeats by preferential uptake of bacteriophage DNA, and provide additional tools for evolutionary studies. *Microbiology* 151, 653-663.
- Rath, D., Amlinger, L., Rath, A., and Lundgren, M. (2015). The CRISPR-Cas immune system: Biology, mechanisms and applications. *Biochimie*.
- Rollie, C., Schneider, S., Brinkmann, A.S., Bolt, E.L., and White, M.F. (2015). Intrinsic sequence specificity of the Cas1 integrase directs new spacer acquisition. *Elife* 4.
- Savitskaya, E., Semenova, E., Dedkov, V., Metlitskaya, A., and Severinov, K. (2013). High-throughput analysis of type I-E CRISPR/Cas spacer acquisition in *E. coli*. *RNA Biol* 10, 716-725.
- Sinkunas, T., Gasiunas, G., Fremaux, C., Barrangou, R., Horvath, P., and Siksnys, V. (2011). Cas3 is a single-stranded DNA nuclease and ATP-dependent helicase in the CRISPR/Cas immune system. *EMBO J* 30, 1335-1342.
- Stern, A., and Sorek, R. (2011). The phage-host arms race: shaping the evolution of microbes. *Bioessays* 33, 43-51.
- Sternberg, S.H., Haurwitz, R.E., and Doudna, J.A. (2012). Mechanism of substrate selection by a highly specific CRISPR endoribonuclease. *RNA* 18, 661-672.

- Sternberg, S.H., Richter, H., Charpentier, E., and Qimron, U. (2016). Adaptation in CRISPR-Cas Systems. *Mol Cell* 61, 797-808.
- Swarts, D.C., Mosterd, C., van Passel, M.W., and Brouns, S.J. (2012). CRISPR interference directs strand specific spacer acquisition. *PLoS One* 7, e35888.
- Terns, R.M., and Terns, M.P. (2013). The RNA- and DNA-targeting CRISPR-Cas immune systems of *Pyrococcus furiosus*. *Biochem Soc Trans* 41, 1416-1421.
- Terns, R.M., and Terns, M.P. (2014). CRISPR-based technologies: prokaryotic defense weapons repurposed. *Trends Genet* 30, 111-118.
- Viswanathan, P., Murphy, K., Julien, B., Garza, A.G., and Kroos, L. (2007). Regulation of dev, an operon that includes genes essential for *Myxococcus xanthus* development and CRISPR-associated genes and repeats. *J Bacteriol* 189, 3738-3750.
- Wang, J., Li, J., Zhao, H., Sheng, G., Wang, M., Yin, M., and Wang, Y. (2015). Structural and Mechanistic Basis of PAM-Dependent Spacer Acquisition in CRISPR-Cas Systems. *Cell* 163, 840-853.
- Wang, R., Preamplume, G., Terns, M.P., Terns, R.M., and Li, H. (2011). Interaction of the Cas6 ribonuclease with CRISPR RNAs: recognition and cleavage. *Structure* 19, 257-264.
- Wei, Y., Chesne, M.T., Terns, R.M., and Terns, M.P. (2015a). Sequences spanning the leader-repeat junction mediate CRISPR adaptation to phage in *Streptococcus thermophilus*. *Nucleic Acids Res* 43, 1749-1758.
- Wei, Y., Terns, R.M., and Terns, M.P. (2015b). Cas9 function and host genome sampling in Type II-A CRISPR-Cas adaptation. *Genes Dev* 29, 356-361.
- Wiedenheft, B., Lander, G.C., Zhou, K., Jore, M.M., Brouns, S.J., van der Oost, J., Doudna, J.A., and Nogales, E. (2011). Structures of the RNA-guided surveillance complex from a bacterial immune system. *Nature* 477, 486-489.
- Wiedenheft, B., Zhou, K., Jinek, M., Coyle, S.M., Ma, W., and Doudna, J.A. (2009). Structural basis for DNase activity of a conserved protein implicated in CRISPR-mediated genome defense. *Structure* 17, 904-912.
- Wu, Y., and Brosh, R.M., Jr. (2012). DNA helicase and helicase-nuclease enzymes with a conserved iron-sulfur cluster. *Nucleic Acids Res* 40, 4247-4260.
- Yang, L., Mali, P., Kim-Kiselak, C., and Church, G. (2014). CRISPR-Cas-mediated targeted genome editing in human cells. *Methods Mol Biol* 1114, 245-267.
- Yosef, I., Goren, M.G., and Qimron, U. (2012). Proteins and DNA elements essential for the CRISPR adaptation process in *Escherichia coli*. *Nucleic Acids Res* 40, 5569-5576.

Zhang, C. (2014). Essential functions of iron-requiring proteins in DNA replication, repair and cell cycle control. *Protein Cell* 5, 750-760.

Zhang, J., Kasciukovic, T., and White, M.F. (2012). The CRISPR associated protein Cas4 Is a 5' to 3' DNA exonuclease with an iron-sulfur cluster. *PLoS One* 7, e47232.

Figure 1.1 The Structure of the CRISPR Locus

CRISPR-Cas immune systems are characterized by a unique genetic structure called the CRISPR locus. It is comprised of a repeated constant sequence element of 20-40 base pairs (black) interspaced by variable invader-derived sequences (colored). Upstream of the CRISPR locus is an AT-rich leader sequence responsible for harboring the promoter machinery of the locus. The leader sequence is also important for adaptation, as new spacers are integrated in a leader proximal manner. Modules of *Cas* genes encoding Cas proteins also are associated with CRISPR loci (Gray boxes). The expressed Cas proteins are responsible for carrying out CRISPR processes. (Adapted from Terns & Terns, 2013.)

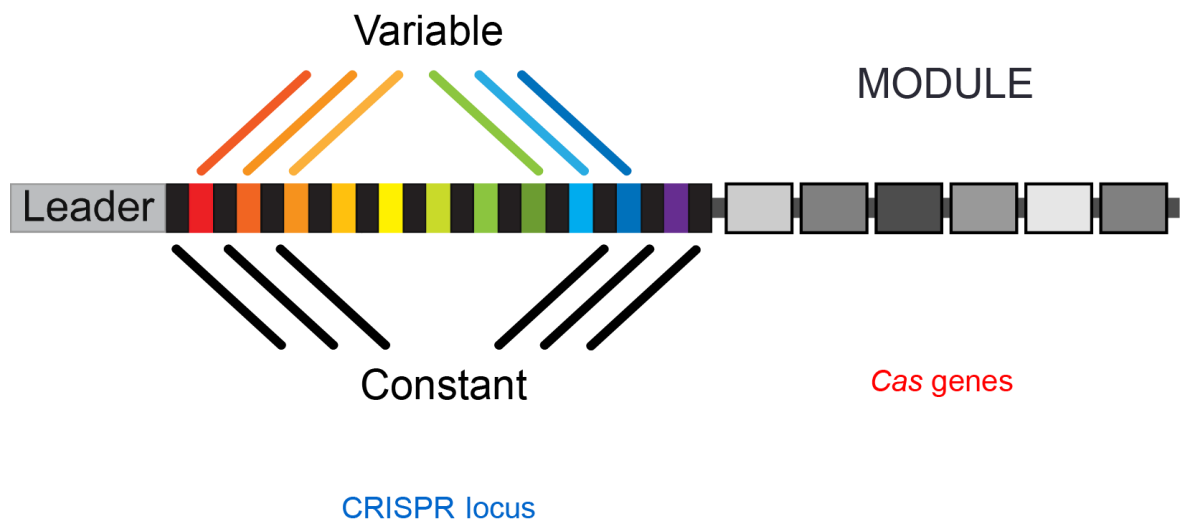


Figure 1.2 The Three Stages of CRISPR-Cas Immunity

CRISPR-Cas immunity occurs in three distinct stages. In the first stage, DNA originating from a viral invader, referred to as a protospacer, is integrated in a leader-proximal manner into the CRISPR locus. Protospacers are preferentially selected from sequences containing a short sequence motif of several nucleotides called a PAM (protospacer adjacent motif). During the second stage of CRISPR-mediated defense, crRNAs are expressed from the CRISPR locus and processed into mature CRISPR RNAs. A mature crRNA contains a single invader-matching sequence. The crRNA is left with a short 5' repeat sequence tag after processing. In the interference stage, mature crRNAs are loaded into effector complexes composed of Cas proteins that carry out crRNA--mediated destruction of the corresponding invader. (Adapted from Terns & Terns 2013.)

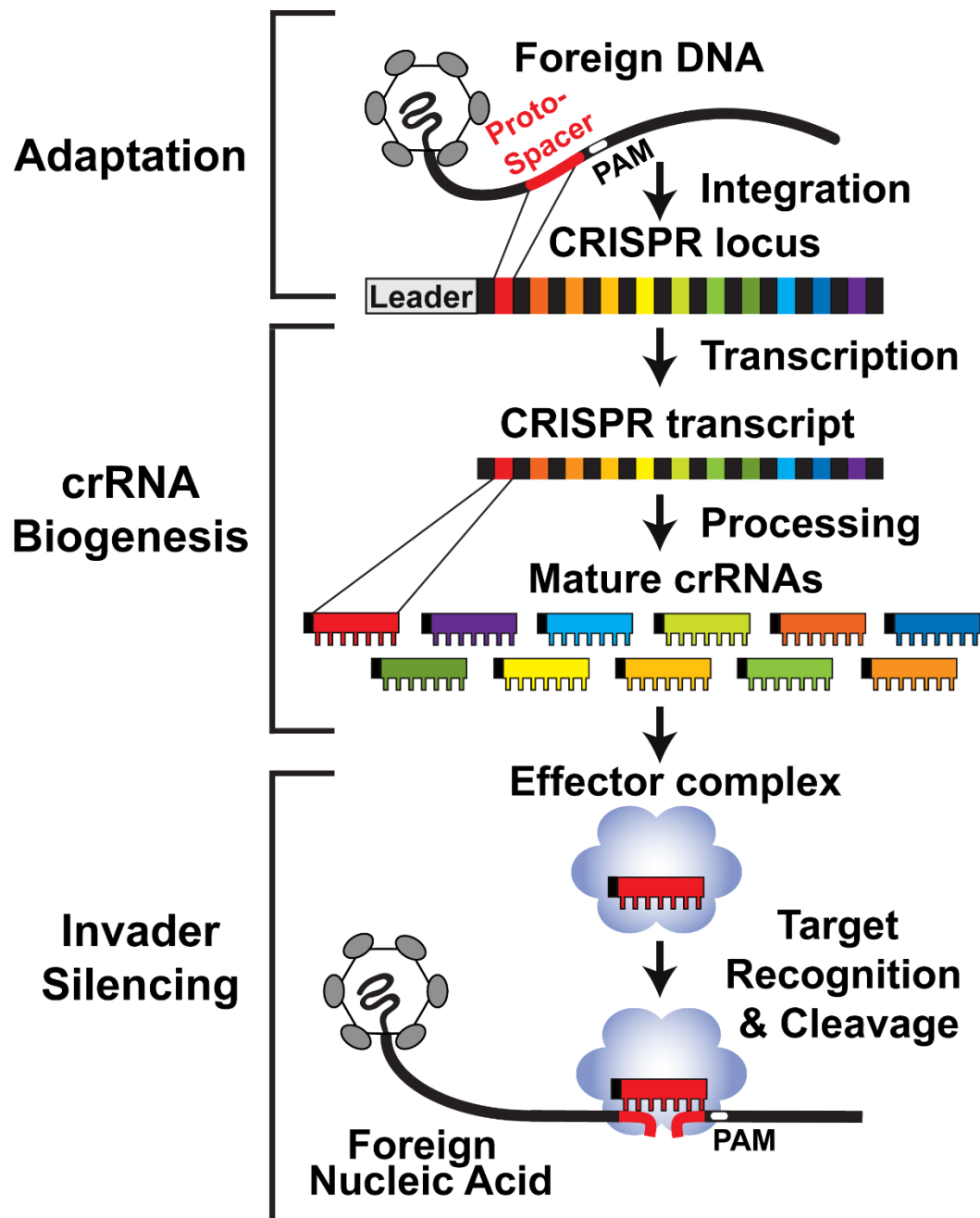


Figure 1.3 CRISPR-Cas systems in the Hyperthermophilic Archaeon *Pyrococcus furiosus*

A) *Pyrococcus furiosus* possesses a 1.9 Mb genome containing seven distinct CRISPR loci and two Cas protein gene clusters. B) The seven CRISPR loci and the number of spacers associated with each locus. C) The genomic organization of Cas protein-encoding genes in *Pyrococcus furiosus*. Biogenesis of crRNAs is controlled by Cas6 (gray). The proteins associated with each of the three immune systems are indicated as follows: Cmr (light blue Type III-B), Csa (green Type I-A), Cst (yellow Type I-G). The adaptation-associated genes are shown in red. Cas1, Cas2, and Cas4-1 are expressed in a single gene cassette. However, Cas4-2 is independent of the other CRISPR genes. (Adapted from Terns & Terns 2013.)

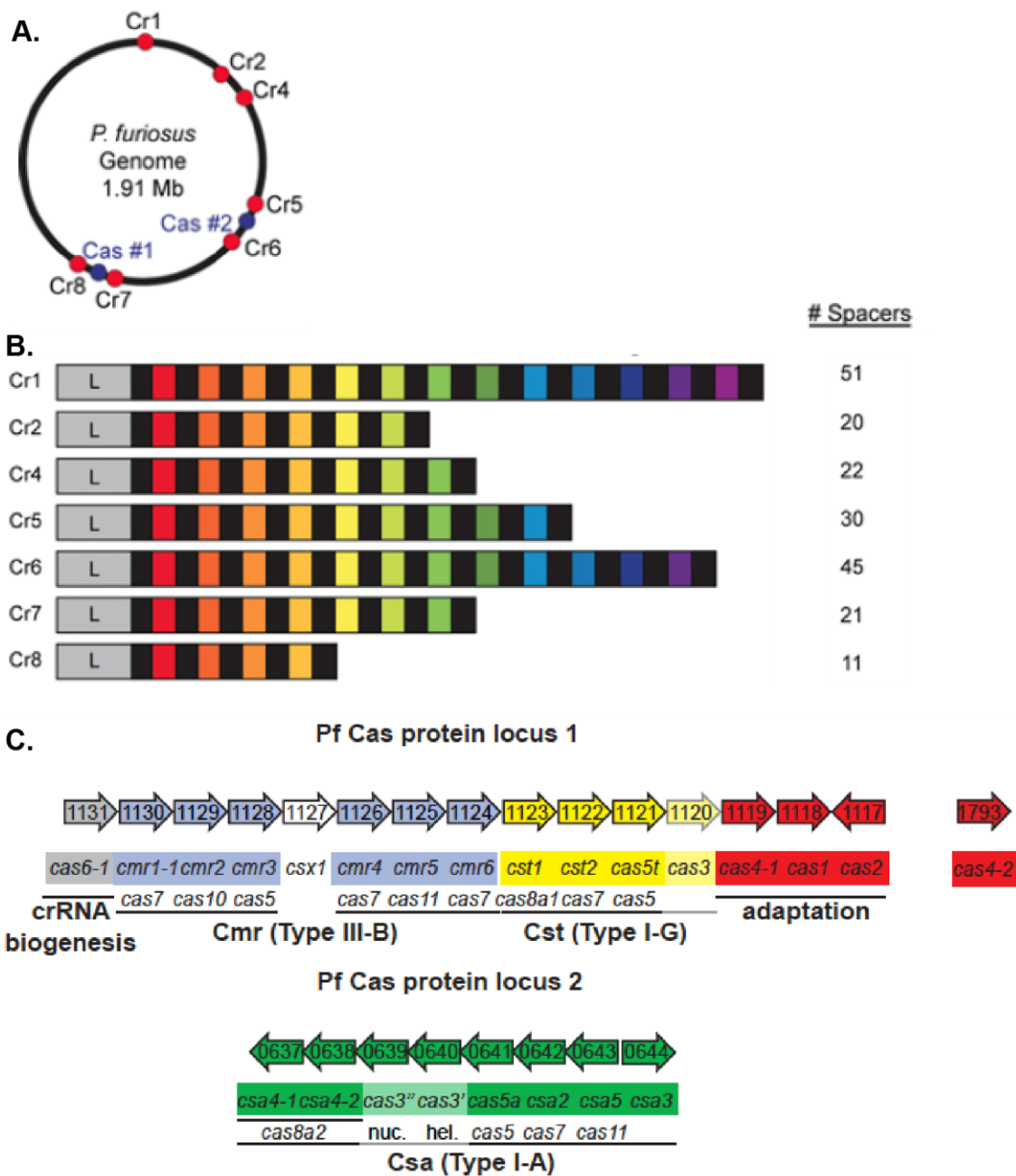
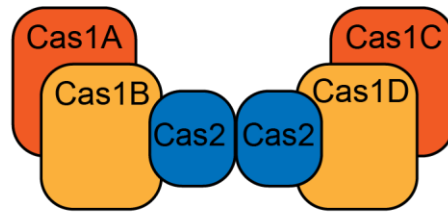


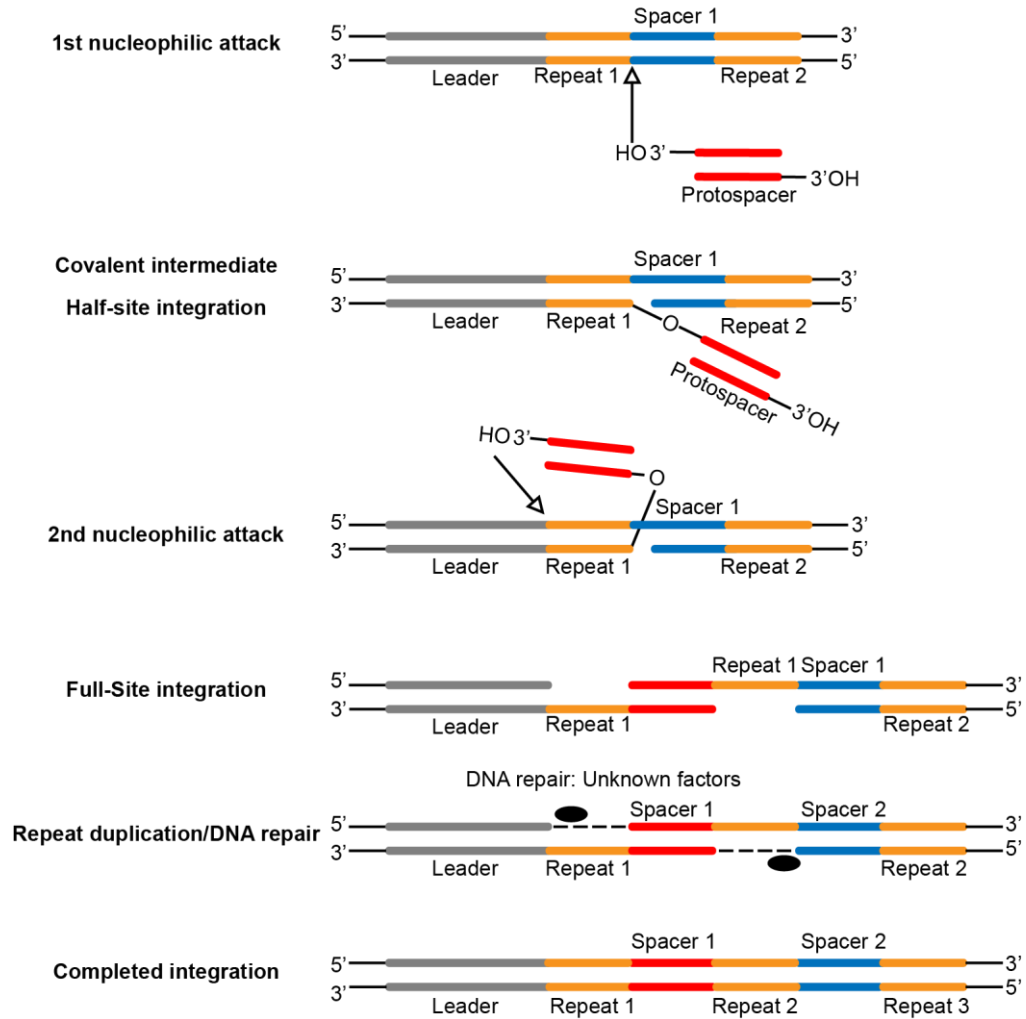
Figure 1.4 Mechanistic Model of Cas1/Cas2-Mediated Spacer Integration

A) Cartoon diagram of Cas1/Cas2 integrase complex as modeled from the *E. coli* Type I-E system. Cas1 and Cas2 form a heterohexameric structure composed of two identical Cas1 (A/B) dimers. The Cas1 dimers are bridged by a single Cas2 dimer, which plays a crucial structural role in the complex. B) A mechanistic model of Cas1/Cas2 spacer integration. Cas1 and Cas2 position an incoming protospacer for integration at the leader proximal repeat of the CRISPR locus. The reaction consists of two separate transesterification reactions performed by the nucleophilic attack of the 3'OH present on the incoming protospacer. The two attacks occur on opposite strands of the first repeat. The first attack, occurring on the minus strand in this case, forms a covalent intermediate in the half-site reaction. The second 3'OH then carries out the next nucleophilic attack on the plus strand. This results in the full-site integration of the new protospacer. Currently unknown enzymes then repair the CRISPR locus by filling in the repeats flanking the newly integrated spacer. Model based upon (Arslan et al., 2014; Nunez et al., 2015a; Nunez et al., 2014).

A.



B.



CHAPTER 2

Trans-acting Factors of CRISPR-Cas Adaptation in the Hyperthermophilic Archaeon *Pyrococcus furiosus*

Introduction

CRISPR-Cas systems comprise a heritable adaptive immune system in prokaryotes and provide defense against phages and other invading mobile genetic elements. A CRISPR locus acts as a sequence-specific memory bank that can be used to target virtually any invader nucleic acid sequence. This crucial sequence information is contained in specialized genetic structures called CRISPR loci. CRISPR loci contain arrays of identical repetitive sequence elements of 20 to 40 base pairs called repeats, which alternate with invader-derived sequences referred to as spacers (Jansen et al., 2002). Additionally, an AT-rich leader sequence containing cis-acting sequences relevant for transcription and the acquisition of new spacers is found upstream of the repeat/spacer array (Jansen et al., 2002; Yosef et al., 2012). The invader-derived information contained within a CRISPR locus is paramount in the adaptative response. A given CRISPR locus will, upon transcription, produce an arsenal of CRISPR RNAs able to guide and target Cas proteins in the destruction of foreign nucleic acids.

CRISPR-Cas systems exhibit remarkable ubiquity and diversity. CRISPRs have been found in ~40% of sequenced bacterial genomes and ~90% of archaeal genomes (Grissa et al., 2007). Throughout the diverse prokaryotes that utilize CRISPR systems, Cas proteins serve as the engines that drive the CRISPR immune response. There are multiple CRISPR-Cas systems,

each of which contains a different array of Cas proteins to mediate the molecular mechanisms of adaptation, crRNA biogenesis, and invader silencing. In total, there are six broad types of CRISPR-Cas systems (Type I-VI), and these are further divided into multiple subtypes. Each system contains a signature Cas protein (Cas3 Type I, Cas9 Type II, and Cas10 Type III, among others) distinguishing the individual system from its peers (Makarova et al., 2015).

The Cas proteins responsible for carrying out immune functions do so via three equally important processes. Adaptation involves the capture and integration of foreign DNA sequences (protospacers) into the CRISPR locus and acts as the first step in developing CRISPR immunity. It is during the adaptation process that immunity against a particular invader is acquired. Adaptation is followed by crRNA biogenesis, involving the transcription and processing of crRNAs. Mature crRNAs are then utilized by Cas effector complexes to carry out sequence-guided destruction in the invader silencing step.

The first step in CRISPR-Cas immunity is the capture and integration of foreign protospacers into the host CRISPR locus in adaptation. Despite the crucial importance of the adaptation step, it remains poorly understood in comparison to other aspects of CRISPR immunity (Amitai and Sorek, 2016). Its mechanism must involve several equally critical steps to allow for the capture of foreign DNA. To begin with, foreign DNA must be recognized and selected for integration. Mapping studies of acquired spacers have revealed the existence of short 3-7 nucleotide motifs known as PAMs (protospacer adjacent motif) flanking protospacers. PAM motifs vary among organisms and are important in both adaptation and immunity. Protospacers flanked by PAMs are highly preferred for integration. Moreover, PAMs are a crucial element in self-non-self recognition during immunity (Datsenko et al., 2012; Deveau et al., 2008; Mojica et al., 2009; Swarts et al., 2012). After selection, a protospacer must also be

further processed to a size usually between 30-40 base pairs prior to integration (Barrangou, 2013). Finally, the protospacer must be integrated into the CRISPR locus. The integration step occurs in a highly directional manner, occurring proximal to the leader sequence end. The final integration step is also accompanied by the duplication of the repeat sequence leading to the expansion of CRISPR array by one spacer/repeat unit (Barrangou et al., 2007; Yosef et al., 2012).

The study of the trans-acting factors involved in adaptation is still on-going. However, recent insights from the *E. coli* type I-E system have substantially increased the field's overall understanding of adaptation components. The most important of these insights is the implication of Cas1 and Cas2 as the lynchpin integrase in adaptation (Nunez et al., 2015a). The two proteins form a heterohexameric complex of two Cas1 dimers bridged by a Cas2 dimer (Nunez et al., 2014). The Cas1/Cas2 complex is responsible for the integration of new spacers via a two-step transesterification reaction akin to DNA transposition and retroviral integration. Integration is carried out by nucleophilic attacks of the 3'-OH groups of the protospacer on the ends of the first repeat (Nunez et al., 2015a). The dual reaction results in a newly integrated spacer flanked by single-stranded regions derived from the former first repeat, which must be filled in by unknown DNA repair mechanisms (Arslan et al., 2014).

Recent studies have also linked several critically important non-Cas proteins to adaptation. IHF (host integration factor) has been implicated as a specificity factor further targeting integration to the leader/repeat junction of the CRISPR locus (Nunez et al., 2016). In addition, the RecBCD complex, responsible for template-driven dsDNA repair and homologous recombination, has been implicated in naïve adaptation. The complex has been found to be responsible for “hotspot” sources of protospacer where DNA is preferentially captured for

adaptation (Levy et al., 2015). The primosomal protein PriA and the RecG helicase responsible for the resolution of stalled replication forks have also been implicated in primed adaptation in the Type I-E CRISPR system (Ivancic-Bace et al., 2015).

The universal nature of the Cas1 and Cas2 proteins found across all CRISPR systems may suggest a common mechanism for the integration step of adaptation (Makarova et al., 2011; Makarova and Koonin, 2015). However, for historical reasons, research in both spacer integration and adaptation has been mostly limited to the *E. coli* type I-E CRISPR system (Amitai and Sorek, 2016). The only essential Cas proteins implicated in adaptation in this system are the Cas1 and Cas2 proteins. However, other CRISPR systems have multiple Cas proteins involved in adaptation such as Cas4, Csn2, and Cas9. Moreover, non-Cas proteins likely to play a role in influencing protospacer selection, processing, and locus repair may differ in other systems. Therefore, it is of vital importance to understand both Cas and non-Cas trans-acting factors in additional model organisms.

The hyperthermophilic archaeon *Pyrococcus furiosus* has three CRISPR-Cas immune systems (Cmr Type III-B, Csa Type I-A, and Cst Type I-G). In addition, its putative adaptation machinery consists of two Cas4 proteins (Cas4-1 and Cas4-2), as well as the universal Cas1 and Cas2 proteins (Terns and Terns, 2013). Therefore, *P. furiosus* is an attractive model system to address adaptation in archaeal systems containing Cas4 proteins. In this study, we examined *P. furiosus* adaptation by investigating the adaptation Cas proteins *in vivo*. We addressed interactions among these Cas proteins by immunoprecipitation and additional techniques to identify *in vivo* protein partners. We increased the expression level of specific Cas proteins to stimulate the adaptation process with the goal of isolating additional protein complexes relevant

to adaptation. Using tandem mass spectrometry, we further attempted to screen for previously unidentified non-Cas proteins potentially implicated in *P. furiosus* adaptation.

Results

Production of Antibodies Against Cas1, Cas2, and Cas4-1

To experimentally address the native Cas proteins involved in *Pyrococcus furiosus* (*Pfu*) adaptation *in vivo*, we took steps to make and test antibodies as tools for protein isolation and detection. Thus, multiple antibodies were raised against the Cas proteins Cas1, Cas2, and Cas4-1. The majority of the antibodies were polyclonal (pAb) IgY antibodies raised and purified from Leghorn hens (*Gallus gallus*) (Figure 2.1). However, for Cas1, both polyclonal IgY antibodies and monoclonal (mAb) IgG antibodies produced from mouse hybridoma cell lines were constructed.

To produce IgY antibodies, recombinant polyhistidine-tagged proteins against each antigen were expressed and purified from *E. coli*. The purified proteins were injected as adjuvants into hens or used in the creation of mouse hybridomas, respectively. Antibodies were later harvested and purified from the egg yolk of these hens both before (pre-immune) and after (immune) antigen injection. Purification of IgY antibodies was carried out in a two-step polyethylene glycol (PEG) precipitation (Figure 2.1B) as described (Polson et al., 1980a; Polson et al., 1980b). Figure 2.1C demonstrates the results of a typical PEG antibody purification. Prominent bands were present at 68 kDa and 22 kDa, matching the molecular weights of IgY heavy chain and light chains, respectively. Albumins comprised the primary source of contaminant proteins in IgY purifications and were present at ~75 and 45 kDa (Figure 2.1C).

Once antibody purification was complete, antibodies were tested to verify their ability to recognize native and recombinant antigens as a measure of specificity and efficacy. A simple

western blot screen where recombinant polyhistidine-tagged protein purified from *E.coli* was tested alongside *Pyrococcus furiosus* whole cell lysates overexpressing Cas1, Cas2, and Cas4-1 (Figure 2.2). Antibodies made against each protein were able to visualize both recombinant and native antigens in immunoblotting. The overexpression of each protein was required for detection because of low native expression levels of the adaptation gene cassette. For example, Cas4-1 was not detected in lysate without overexpression (Figure 2.2E). The lack of detection of other proteins is indicative of specificity and no cross-reactivity with other proteins from *P. furiosus*. Therefore, we determined that this series of antibodies were suitable for further use in co-immunoprecipitation (Co-IP) studies to understand components vital to CRISPR-Cas adaptation.

Experimental Setup Adaptation Cas Co-IPs

To gain an understanding of *in vivo* adaptation and the associated proteins involved in the acquisition of new spacers, we used a co-immunoprecipitation (Co-IP) approach targeting Cas1, Cas2, and Cas4-1. Based on the knowledge from the previous testing that each antibody recognizes its associated antigen with specificity (Figure 2.2), we performed multiple Co-IP experiments targeting each Cas protein. A technical overview detailing the Co-IP experiments is shown in Figure 2.3. In order to stimulate the adaptation process and increase protein yield, Cas1, Cas2, and Cas4-1 were overexpressed in *P. furiosus* as an operon under a single promoter (TPF 41, most experiments) or under three separate promoters (TPF 68, Table 2.6 only). Whole cell lysates produced from *P. furiosus* overexpressing the adaptation Cas proteins were subjected to Co-IPs targeting each protein (Strains TPF-41 or TPF-68). Non-immune or pre-immune antibodies were used as control antibodies that do not recognize other Cas or *Pyrococcus furiosus* proteins, determining background proteins in the assay. Sample material produced via

Co-IP was either separated by 15% silver nitrate-stained sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE) or visualized by western blotting (Figure 2.3). Ultimately, tandem mass spectrometry analyses (MS) were performed using IP material obtained with each antibody to confirm protein identity in a given sample (Figure 2.3).

As a guide to the Cas interaction results across multiple experiments, Table 2.1 summarizes the adaptation Cas proteins that were identified in each immunoprecipitation by western blotting or mass spectrometry. The presence or visualization of protein in at least a single experimental replicate by western blot (WB) or mass spectrometry (MS) was the criteria for a positive identification. However, proteins may have been identified in multiple experiments, as is indicated in further experiments. In mass spectrometry, the protein in question must have met the requirement of having a two-fold enrichment of spectra in the immune (anti-Cas) compared to the IP control antibody. Because of cross-reactivity between the IP and western blot secondary antibody at the molecular weight of Cas4, Cas4 western blotting was indeterminate.

Cas1 Interacts with Cas4-1 and Cas2

Three antibodies targeting Cas1 were used in immunoprecipitation studies: two mouse monoclonal antibodies (A5D9 and A6.C11 hybridoma cell lines) and one polyclonal IgY. Figure 2.4 details both the Co-IP SDS-PAGE and western blotting results available for each Cas1 antibody. Strikingly, a similar three-band profile emerged in Co-IP SDS-PAGE studies from each of the three antibodies. These protein bands corresponded to the overall molecular weights of Cas1 (37 kDa), Cas4-1 (20 kDa), and Cas2 (10 kDa) were present and unique to anti-Cas1 immune antibodies (Figure 2.4 colored asterisks). It was also clear that other unknown bands that did not correspond to an expected Cas protein molecular weight were also present, perhaps

indicating the possibility of non-Cas proteins associated with Cas1 (Figure 2.4 black asterisks). When A5D9 Co-IP material was subjected to increasing concentrations of NaCl (Figure 2.4A), it was apparent that the potential complex between Cas1 and Cas4-1 was stable up to 500 mM NaCl. The possible interaction observed between Cas1 and Cas2 was less stable, only apparent up to 300 mM NaCl.

Immunoblotting of samples from the each of the two Cas1 mAb Co-IPs (Figure 2.4E & F) verified the presence of both Cas1 and Cas4-1. However, Cas2 was only visualized in the A5D9 mAb analysis (Figure 2.4F). These data, coupled with the loss of the 10 kDa band at 500 mM NaCl (Figure 2.4A), suggested that the interaction between Cas1 and Cas2 was less stable. A similar complex was observed at lower NaCl conditions (100 and 300 mM) in the reciprocal Cas2 Co-IP/western blot analyses (Figure 2.9B) lending further evidence of this complex.

Mass Spectrometry Confirmed Cas Interactions and Identified Non-Cas Factors

To verify Cas protein identity in Cas1 immunoprecipitations and identify non-Cas factors, mass spectrometry was performed on immunoprecipitated material obtained with each of the Cas1 antibodies (Tables 2.2-2.6). In general, two types of mass spectrometry experiments were performed: “shotgun” and band excision (Figure 2.5). “Shotgun” analyses were intended to identify all proteins in a given sample, while band excision was used to identify specific proteins. Protein identifications were based on the total amount of associated spectra and peptides in a given experiment between the anti-Cas1 IP (immune) antibody and a control antibody (pre-immune or non-immune) to account for background proteins. The fold change value in spectra in the immune compared to control antibody indicates the specificity of a given protein identification. A two-fold minimum immune enrichment of spectra was the criteria for verification of an immune-specific protein confirmation.

As expected, Cas1 was identified with greater than two-fold immune spectral enrichment in all Cas1 co-immunoprecipitations (Tables 2.2-2.6). The lowest level of immune enrichment when compared to the control antibody was 3.3-fold (Cas1 pAb IP Table 2.5). However, in most cases, Cas1 was a unique peptide, and was not detectable in non-immune controls in this series of experiments (Tables 2.2, 2.3, 2.4 and 2.6). Cumulatively, Cas1 was nearly 90-fold enriched regarding spectra, and 17-fold enriched regarding total peptides across all Cas1 mAb immunoprecipitations in the immune condition, indicating extreme specificity in anti-Cas1 immunoprecipitations (Table 2.11).

Cas4-1 was also identified across in Cas1 Co-IP mass spectrometry analyses (Tables 2.2-2.6). Cas4-1 was 540-fold enriched regarding spectra across all Cas1 mAb immunoprecipitations when analyzed cumulatively (Table 2.11). Cas2 was not identified nearly as often despite its presence being indicated in western blotting (Figure 2.4F). However, the protein was positively identified in a Co-IP/MS experiment performed under anaerobic conditions (Table 2.6). As Cas2 was not found in most Co-IP MS experiments, including the one shown in Figure 2.4A, another 10 kDa protein, Alba, was found to have the most spectra in the band excision analysis at 10 kDa. Cas1 and Cas4-1 were top identifications in their respective bands as shown (Table 2.4).

In addition to the positive identifications made against Cas proteins in this data set, many non-Cas factors were present across the Cas1 IP analyses. Most numerous were identifications of ribosomal proteins and other components of the translation machinery (Table 2.11). This trend was particularly prevalent in Cas1 mAb Co-IP analyses with a total of 84 different ribosomal and translation proteins identified across all studies. In addition, a set of identifications in the form of DNA-binding proteins was also present. Several were found in each Cas1 Co-IP, and in total 19 distinct DNA-binding proteins were observed. These included many proteins impacting DNA

topology such as Alba, Histone A1, and A2, and Topoisomerases I and IV. Several more identified proteins were implicated in DNA replication, repair, and recombination such as Hef nuclease, radA, DNA helicases, and replication factor (Forterre et al., 1996; Fujikane et al., 2010; Goyal et al., 2016; Kayoko Komori, 2000; Nishino et al., 2005a; Reeve et al., 2004; Tse-Dinh, 1998; Wardleworth et al., 2002). Other identifications included proteins without a known function or annotation (36 proteins) and proteins of other functions such as metabolism and cell division (42 proteins) (Table 2.11)

Cas2 and Cas4-1 Co-IPs isolate antigens, but Non-Specifically Precipitate Cas1

To continue Co-IP analyses of the adaptation proteins, further studies using both Cas2 and Cas4-1 IgY polyclonal antibodies were performed. As with Cas1, immunoprecipitations were performed with overexpressed whole cell lysates produced from TPF-41 (overexpressing Cas1, Cas2, Cas4-1) under different NaCl wash conditions. Eluted Co-IP material was then subjected to SDS-PAGE, western blotting, and mass spectrometry.

In Cas4-1 Co-IP/SDS-PAGE analyses (Figure 2.7A), an immune-specific enrichment of a band corresponding to the molecular weight of Cas4-1 (20 kDa) was observed. This band was present at 100, 300, and 500 mM NaCl conditions (Figure 2.7A). However, unlike Cas1 immunoprecipitations, no obvious enrichment of other Cas or non-Cas proteins was immediately apparent. As a follow-up, western blotting was performed to further elucidate Cas protein content in Cas4-1 immunoprecipitations (Figure 2.7B). Unfortunately, blotting results were uninterpretable for Cas4-1. As mentioned previously, this was caused by unwanted elution of IP antibody. Cross-reactivity between blotting secondary antibody and the light chain of IP antibody produced a strong signal at the molecular weight of Cas4-1 (20 kDa). Thus Cas4-1 was masked from detection. This was illustrated by Figure 2.7B where the strong band present at 20

kDa in IP/western blotting persisted regardless of whether or not *P. furiosus* lysate was applied to the IP (Figure 2.7B). Thus, the signal was produced independent of the presence of Cas4-1. Thus, confirmation of antigen pulldown was instead obtained via mass spectrometry (Table 2.7). While other antibody-correlated bands were present in blotting, these did not mask the presence of Cas1 or Cas2. Cas1 was observed at a nearly equal signal intensity in both pre-immune and immune conditions illustrating a lack of specificity not observed previously in Cas1 IPs. Cas2 was not observed in Cas4-1 Co-IP western blotting (Figure 2.7B).

Cas1 was identified in both pre-immune and immune conditions in Cas4-1 Co-IP mass spectrometry. In most replicates (Table 2.7 and 2.8), Cas1 was below the 2-fold minimum immune spectra criterion required for an immune-specific identification. Cas1 was often below a fold change value of 1 pointing to a higher level of Cas1 spectra in the pre-immune compared to the immune. While there was a single condition (Cas4-1 100 mM NaCl pAb IP in solution, Table 2.6) where Cas1 exhibited immune specificity, in the aggregate Cas1 was a non-specific protein. Thus, Cas1 co-precipitation could not be attributed to the presence of Cas4-1. Mass spectrometry performed on the 20 kDa SDS-PAGE band potentially representing Cas4-1 (Figure 2.8, Table 2.8) confirmed the presence of Cas4-1 in the excised area. Cas4-1 was present at an expected molecular weight. The presence of Cas4 was also confirmed in all other experimental replicates (Table 2.7 and 2.8).

Cas2 Co-IP analyses isolated Cas2 as expected. In Co-IP SDS-PAGE, a notable immune-specific enrichment of a 10 kDa protein was observed in all NaCl wash conditions (Figure 2.9A). Similar to Cas4-1 immunoprecipitations, no distinct bands corresponding to Cas1 or Cas4-1 were present. Follow-up studies using Co-IP western blotting were performed as before to determine Cas protein content (Figure 2.9B). In this case, Cas4-1 immunoblots were

bypassed due to their inconclusive nature as explained previously. The presence of Cas2, as well as Cas1, was confirmed in these Co-IP/western experiments. While Cas1 was found in pre-immunes to a lesser extent, clear enrichment in immune IPs was demonstrated at 100 and 300 mM NaCl (Figure 2.9B). This result supported the previous interaction observed in Cas1 Co-IPs, as well as the instability of the complex above 300 mM NaCl.

In Cas2 Co-IP mass spectrometry studies, Cas2 was found across all analyses as expected (Tables 2.9 and 2.10). Cas2 was also present at the expected 10 kDa molecular weight as verified by band excision mass spectrometry (Figure 2.10, Table 2.10). Cas1 was present in shotgun MS results at 100, 300, and 500 mM NaCl. However, in all cases, Cas1 was again below the 2-fold threshold of immune spectra enrichment compared to the control Co-IP antibody consistent with other IgY immunoprecipitations. Cas4-1 was also present but non-specific in 100 and 300 mM NaCl conditions, a likely cause being its association with Cas1, which was shown to be non-specifically associating with pre-immune samples.

Parallel Analyses of Adaptation Cas Protein Interactions

Complementary analyses to understand any physical associate between Cas1, Cas2, and Cas4-1 additional studies were conducted. To this end, both size exclusion chromatography (Figure 2.11) and Cas1 affinity purification via nickel chromatography (Figures 2.12, 2.13, 2.14) were completed as additional tests. To parallel Co-IP analyses, size exclusion chromatography was performed on *Pyrococcus furiosus* lysates overexpressing the Cas1, Cas2, and Cas4-1 proteins (TPF41). Prior to running Cas overexpressed lysates, a solution of five standard proteins (thyroglobulin 670 kDa, γ -globulin 158 kDa, ovalbumin 44 kDa, myoglobin 17 kDa, and vitamin B12 1.35 kDa) was run to provide molecular weight standards for protein elution (Figure 2.11 black arrows). Fractions were then assayed for the presence of Cas1, Cas2, and Cas4-1 via

western blotting. Interestingly, the peak fractions of Cas1 and Cas4-1 overlapped between the molecular weights of the 44 and 158 kDa standards. Such a result could be indicative of a Cas1 and Cas4-1 interaction, as both proteins eluted together at molecular weights larger than their expected monomer size (37 and 20 kDa). Cas1's elution profile also exhibited a larger series of elutions at molecular weights above 670 kDa, indicative of an unknown larger complex. Cas2 however, did not co-elute as expected with Cas1 and Cas4-1. Instead, the fractions where Cas2 was most present ranged from 17 kDa and below, likely indicative of a dimer or monomer fractionation profile given its molecular weight (10 kDa).

In a separate experiment, a C-terminal polyhistidine-tagged version of Cas1 expressed in *Pyrococcus furiosus* was subjected to batch nickel affinity chromatography (Figure 2.12). As in previous Co-IP experiments, Cas2 and Cas4-1 were also co-overexpressed in the experiment but left untagged. *Pfu* lysates were incubated with nickel resin and washed in increasing concentrations of imidazole to purify and elute Cas1 (Figures 2.13 and 2.14). Notably, SDS-PAGE of the eluted fractions at 500 mM imidazole displayed a similar band profile to that observed in Cas1 Co-IP experiments (Figure 2.13). The most prominent bands were once again featured at 37, 20, and 10 kDa, possibly representing Cas1, Cas4-1, and Cas2. These bands were enriched when compared to an untagged Cas1 control experiment run in parallel. To further understand protein identity, the same fractions were subject to western blotting (Figure 2.14). In both elution fractions, Cas1 (Figure 2.14A) and Cas4-1 (Figure 2.14B) were heavily enriched compared to an untagged control purification and the wash fractions. Cas2 was also slightly co-eluted with Cas1 and Cas4-1 compared to the untagged control.

Cas1 Physically Associates with Nucleic Acids

To understand the potential interaction of Cas1 and adaptation complexes with nucleic acid substrates *in vivo*, we performed chromatin immunoprecipitations against Cas1 using monoclonal antibodies (Figure 2.15). *P. furiosus* cell cultures were crosslinked via the addition of formaldehyde and then subjected to lysis and immunoprecipitation as before. Following nucleic acid precipitation and isolation after the IP, isolated nucleic acids were run on a 15% denaturing PAGE gel and visualized with SYBR. Nucleic acids were either run as originally eluted from the IP or subjected to RNase A digestion to identify the nucleic acid as either DNA or RNA. Interestingly, both Cas1 ChIP experiments demonstrated the isolation of large DNAs barely able to resolve in the gel. Smaller RNAs were also present and identified by their lack of resistance to the RNase A digest. No stainable quantity of nucleic acids was visualized in IPs lacking Cas1. Additional unshown DNase I digests were performed on these samples confirmed the larger bands were in fact DNA. Thus, Cas1 appears physically associated with nucleic acids *in vivo*. As ChIP experiments lacking Cas1 demonstrated no SYBR detectable yield of nucleic acids, nucleic acid isolation is specific to the presence of Cas1 (Figure 2.15).

Discussion

Cas1 Interacts with Cas2 and Cas4-1 *in vivo*.

In this work, data was obtained demonstrating the existence of protein complexes among Cas1 and Cas4-1 and Cas1 and Cas2. Cas1 demonstrated the co-isolation/co-elution of Cas4-1 in several different experiments including many independent Cas1 Co-IPs (Figure 2.4), Cas1 nickel chromatography (Figures 2.12-2.14), and size exclusion chromatography (Figure 2.11). Cas2 was present and identified in Cas1 Co-IPs by western blotting and mass spectrometry (Figure 2.4, Table 2.6). Additionally, Cas1 was found in reciprocal Cas2 Co-IPs supporting the

interaction between the two proteins (Figures 2.4 and 2.9). Mass spectrometry protein identifications corroborated the pulldown of Cas1, Cas2, and Cas4-1 in Co-IP studies against Cas1 (Tables 2.2-2.6). While Cas2 was only identified in a single Cas1 Co-IP/MS replicate (Table 2.6), it was in this replicate under anaerobic conditions, where the cellular environment of an anaerobic hyperthermophile was most strictly maintained. Moreover, due to its small size and lack of tryptic peptides, Cas2 was an innately weak candidate for mass spectrometry analyses and exhibited poor sequence coverage. However, the association between the Cas1 and Cas2 proteins did appear to be more transient and less consistent when compared to the interaction between the Cas1 and Cas4-1 proteins.

Nickel chromatography and size exclusion chromatography experiments generally supported the overall data trends observed by Co-IP. Nickel chromatography experiments demonstrated the co-isolation of Cas1 and Cas4-1 specifically enriched in the 500 mM imidazole elution, consistent with Cas1 Co-IP experiments (Figure 2.13 and 2.14). Cas2 was also enriched in the same elution fractions as Cas1, but to a much less extent, illustrating a less stable interaction.

In size exclusion chromatography, Cas1 and Cas4 co-eluted at molecular weights between 44 and 158 kDa with the peak roughly halfway between the two standards (Figure 2.11). The overlap of the two proteins supported the possibility of interaction. Cas2, however, did not co-elute with either protein, behaving roughly as a dimer (~20 kDa). The lack of Cas2 co-elution with Cas1 and Cas4-1 is a possible by-product of a less stable interaction not able to withstand gel filtration. This result was consistent with Co-IP results where Cas2 was repeatedly but sporadically observed in anti-Cas1 experiments (Figure 2.4). The Cas1/Cas2 interaction was also lost at lower ionic strengths (300 mM) than the Cas1/Cas4-1 interaction. It is also possible

that the Cas1 and Cas4-1 proteins simply have oligomeric states, and that their co-fractionation is not due to physical interaction. For example, perhaps Cas1 elutes as a dimer of ~ 75 kDa. A repeat of this experiment with lysates lacking Cas1 or Cas4-1 could serve as a control to test this possibility: if the two protein's physically associate, they should elute differently in the absence of one another. This could serve to illustrate the link between the two proteins more definitively.

While this work does not address the stoichiometry of the possible Cas1 and Cas2 complex, complex formation between Cas1 and Cas2 was consistent with current models implicating Cas1 and Cas2 as the central integrase for spacer integration (Nunez et al., 2014; Nunez et al., 2015b; Wang et al., 2015). In the type I-E CRISPR system, Cas1 and Cas2 support the fundamental two-step transesterification reaction of protospacer integration at the leader/repeat junction(Nunez et al., 2015b). In so doing, they form a stable asymmetrical heterohexameric complex composed to two flanking Cas1 homodimers bridged by a Cas2 homodimer. Each Cas1 homodimer comprises a single catalytic subunit, with both homodimers being necessary for a full integration utilizing both required free 3'-OH groups on either side of the incoming protospacer. While Cas2 does not contribute to the activity of the complex, it is essential for creating an area for binding the incoming protospacer and complex structure (Nunez et al., 2015a; Nunez et al., 2014; Wang et al., 2015). Additionally, the two proteins also form a larger complex in type II-A CRISPR-Cas systems with Cas1, Cas2, Csn2, and Cas9 (Heler et al., 2015). Taken together, it is logical that a similar complex and activity may be present in *Pyrococcus furiosus*.

While a plethora of evidence exists to support the existence of a Cas1 and Cas2 complex, the knowledge regarding Cas4 proteins is much more limited, as their overall role in the CRISPR mechanism has not been established. Cas4 proteins are often in the same gene cluster as Cas1

and Cas2; in addition, Cas4 and Cas1 gene fusions also have been previously reported, as in *Thermoproteus tenax* (Makarova et al., 2011; Makarova et al., 2015; Plagens et al., 2012; Viswanathan et al., 2007). However, evidence for the two proteins existing as a complex has not yet been demonstrated *in vivo*. The evidence presented here illustrates such a possible interaction.

The reported activity of Cas4 proteins could play roles in several different processes of adaptation, especially in protospacer substrate generation and processing. Cas4 has been reported to have a 5'-3' exonuclease activity *in vitro*, with a limited ATP-independent unwinding activity. Moreover, the active site of Cas4 alongside the presence of Fe-S clusters are reminiscent of the *E. coli* RecB nuclease and Gram Positive AddAB nucleases implicated in template-driven DNA repair (Lemak et al., 2013; Lemak et al., 2014; Zhang et al., 2012). It has been previously shown using the *E. coli* type I-E Cas1/Cas2 integrase that the preferred protospacer integration substrate is one with 3' overhangs, which further exposes the protospacer's 3'OH for nucleophilic attack on the leader/repeat junction (Nunez et al., 2015a; Rollie et al., 2015; Wang et al., 2015). Cas4 proteins might be responsible for creating such overhangs via exonuclease trimming thereby making a more preferred integration substrate.

In addition, it is unknown how larger pieces of DNA targeted by the adaptation machinery are processed into mature protospacers prior to integration. Previous findings indicate that substrates likely selected for integration arise from cleavage products generated from Cas3 CRISPR defense and RecBCD-based DNA repair and DNA replication intermediates (Levy et al., 2015; Musharova et al., 2017). Cas4 proteins could be responsible for trimming the single-stranded overhangs in these substrates and processing protospacers down to integration sizes. RecB was found to be essential for naïve adaptation in *E. coli* (Ivancic-Bace et al., 2015; Levy et

al., 2015). The Cas4 RecB-like nuclease domain and activity may also suggest that Cas4 replaces RecB in adaptation in organisms lacking RecB though the exact role RecB plays in adaptation is not fully understood (Lemak et al., 2013; Lemak et al., 2014; Levy et al., 2015; Zhang et al., 2012). Therefore, a Cas1 and Cas4 complex for selecting and processing protospacers could then logically function together. Unfortunately, because of the lack of an antibody, Cas4-2 could not be studied directly via Co-IP and was not identified in any Cas1 Co-IP/MS studies. In most cases, however, it was not co-overexpressed with the rest of the adaptation Cas genes (exception in cases where TPF68 was used). A priority for future experiments is to ascertain how this protein may interact and function with the other adaptation Cas proteins in *Pyrococcus furiosus*.

Cas1 Associations with Non-Cas Proteins

The primary goal of the Co-IP mass spectrometry analyses performed in this work was to act as a screening study to understand and identify unknown accessory proteins impacting the adaptation process. To analyze results in total, we pooled the data produced from all replicates of Cas1 monoclonal antibody Co-IP/MS analyses (Table 2.11). Across all experiments, one hundred eighty four proteins were determined to be specific to the anti-Cas1 antibodies as determined by a minimum two-fold spectral count enrichment throughout all Cas1 monoclonal IPs (Table 2.11). To further classify these proteins, three were Cas proteins, (Cas1, Cas2, and Cas4-1), 19 were DNA-binding proteins, 84 were ribosomal/translation associated, 36 were uncharacterized (no annotation), and 42 were linked to various other cellular processes.

This proteomic screen was intended to include Co-IP/MS analyses of Cas2 and Cas4-1 in addition to Cas1. However, because Cas1 was isolated consistently in pre-immune controls in polyclonal IgY IPs, those analyses were not included. The presence of Cas1 in pre-immune controls could have had an adverse impact on the specificity of other proteins associated with

Cas1 proteins, causing them to be unfairly classified as non-specific in addition to Cas1 itself. Possible examples of this were observed was the non-specific precipitation of Cas4-1 in a few experiments (Table 2.7 and Table 2.9) Therefore, we limited our cumulative analysis to Cas1 mAb Co-IP/MS experiments which illustrated significant specificity for Cas1.

While additional experiments are required to fully link candidate proteins to the adaptation process with more certainty, it is notable that many DNA-associating proteins were identified in this analysis (Table 2.11). As Cas1 was shown to associate with DNA not only in the various functional studies of other CRISPR-Cas systems by virtue of its its protospacer integrase and nuclease activities but also in the ChIP analysis presented here, (Figure 2.15) two models exist for the isolation of these DNA-binding proteins.

One possibility is that Cas1 may have indirectly interacted with these proteins through an associated piece of DNA. Under this model, Cas1 and associated DNA-binding proteins such as Hef nuclease or RadA would be co-isolated by binding to the same piece of DNA. This would possibly indicate that Cas1 may have a preference for DNA substrates influenced by such proteins. These DNA substrates may serve as “hotspots” for protospacer acquisition. In the case of Hef nuclease, a replication fork resolvase protein responsible for the clearing of stalled replication forks, Cas1 would be associated with forked DNA substrates (Fujikane et al., 2010; Lestini et al., 2015; Nishino et al., 2005b). Studies of non-Cas factors in the Type I-E system integral to adaptation support this as a possible model. The RecG and PriA proteins acting in a similar pathway to Hef in the restart of stalled replication were recently shown to be essential for primed adaptation (Ivancic-Bace et al., 2015). Moreover, Cas1 itself has been shown to interact with branched DNA substrates as intermediates in the integration process (Babu et al., 2011; Rollie et al., 2015).

Another possible model is that Cas1 interacts directly with these proteins during the adaptation process. While it is unlikely that direct physical interaction is plausible for all 19 DNA binding proteins, in the Type I-E *E. coli* system Cas1 was reported to interact with the RecB protein of the RecBCD complex (Babu et al., 2011). Thus, it is plausible that Cas1 may directly interact with some proteins implicated in DNA repair, replication, or recombination. As previously mentioned, the nucleic acid substrates of DNA breakage, replication, and repair processes have been demonstrated as sources of potential protospacers (Levy et al., 2015). The Hef nuclease (holliday junction resolvase/forked DNA resolvase), RadA (strand exchange protein), replication factor A (single strand binding protein), DNA polymerase II, and DNA helicases are all proteins implicated in such processes (Fujikane et al., 2010; Kayoko Komori, 2000; Lestini et al., 2015; Nishino et al., 2005a; Nishino et al., 2005b). Cas1 could interact with these proteins and utilize a physical association to gain access to protospacer substrates. Whether the nature of the interaction between these proteins and Cas1 is direct or indirect, it is plausible that Cas1 may have a preference for DNA substrates influenced by DNA repair and recombination processes, or that it gains access to them via a potential interaction with one of the proteins in question. To better understand the role these candidates play in adaptation, additional genetic studies of proteins through knockouts is recommended.

The other notable DNA-binding proteins that were identified in Cas1 Co-IP/MS were proteins implicated in DNA topology such as histones and Alba responsible for the bending and compaction of DNA. Alba is a highly expressed, archaeal, histone-like protein known for bridging DNA duplexes as well as binding cooperatively along DNA to form filament-like structures (Goyal et al., 2016; Wardleworth et al., 2002). Other than the Cas1 and Cas4-1 proteins, Alba was the most numerous protein in terms of spectra and was identified in numerous

experiments (Table 2.11). The fact that Alba proteins are highly expressed and have a wide-spread binding across the genomes of archaea may simply support a non-specific and indirect model of interaction (Laurens et al., 2012). However, the DNA topology at the CRISPR locus may be vital to spacer integration for both access and specificity. It was recently shown that the DNA-bending protein IHF in *E. coli* was important in conveying leader/repeat junction integration specificity (Nunez et al., 2016). IHF introduces sequence-specific bending within the leader region allowing for Cas1/Cas2 recognition of this specific integration site (Nunez et al., 2016). While IHF is only present in gram-negative bacteria, other organisms may have unique candidate factors responsible for this process. Thus, DNA topology and unknown factors may be vital in conveying site-specific integration *in vivo*.

While DNA-binding proteins may have a more immediate functional significance in adaptation, many other proteins were highlighted by the Co-IP MS screen when analyzed cumulatively (Table 2.11). The most numerous of these were the 84 identifications of ribosomal proteins. A possible explanation for this is isolation of Cas1 during translation on the ribosome. This is also a possible alternative explanation for the Cas interactions repeatedly observed in Co-IP studies as Cas proteins were overexpressed as operons in most IP experiments (TPF 41). Therefore, Cas1, Cas4-1 and Cas2 could simply be associated by being translated together rather than interacting physically in the cell.

We contend that this is not likely to be the case however, as the Cas proteins were observed to interact in separate scenarios where they were not overexpressed as operons. A primary example of this is in the nickel chromatography experiment, where Cas1 was expressed separately on a plasmid, whereas Cas4-1 and Cas2 were expressed on the genome (Figure 2.13

and 2.14). Additionally, anaerobic Cas1 A5D9 co-immunoprecipitations utilized strain TPF 68, which expressed Cas1, Cas2, Cas4-1, and Cas4-2 under separate promoters (Table 2.6).

The overexpression of the target Cas proteins in this study might have further contributed to the phenomenon of frequent ribosomal identifications, as immunoprecipitations performed under wildtype expression did not isolate ribosomes (data not shown). For undetermined reasons, this isolation of ribosomal proteins was specific to mouse IgG monoclonal Cas1 immunoprecipitations, and was not observed in other IP data sets. Figure 2.15 illustrates that Cas1 also co-isolates RNAs in addition to DNA. Ribosomal co-precipitation of rRNA and tRNA could be a plausible source of these RNAs. It is difficult to rationalize a functional linkage for other proteins identified in this study from the “unannotated” and “other” categories. While valuable insights may be gleaned from the further study of these candidates, the significance of DNA replication and repair to protospacer integration makes the study of those candidate proteins more attractive.

Materials and Methods

Preparation of polyclonal antibodies against Cas proteins

Recombinant 6X his-tagged protein antigens Cas1, Cas2, and Cas4-1 were expressed in *E.coli* via pET24D or pET200D vectors in BL21. Cultures were grown to mid-log phase (OD 0.6) and induced via the addition of 1mM IPTG for 16 hours overnight. Cultures were pelleted via centrifugation and lysed by sonication. The Cas proteins were then purified via nickel affinity chromatography. 200 ug of purified protein was then injected into the breast of Leghorn hens as an adjuvant. Antibodies were then raised and purified as described in Carte et al., 2010 and Polson et al., 1980.

***P. furiosus* strain construction**

P. furiosus strains used in this study are listed in Appendix Table A.2. Strains were constructed using a marker pop-in/pop-out technique for gene replacement or gene deletion. Details are previously described in (Lipscomb et al. 2011; Farkas et al. 2012).

***P. furiosus* strains and growth conditions**

The strains and plasmids used in this study are listed in Appendix Table A.2 and A.3 respectively. *P. furiosus* strains were grown under strictly anaerobic conditions at 90°C in complex medium unless otherwise noted. If selection for the maintenance of a plasmid was needed, as was the case in for *P. furiosus* nickel chromatography experiments, cultures were grown in selective defined medium (- uracil). Cultures and media were prepared as described previously (Lipscomb et al. 2011), with medium pH adjusted to ~6.5. Cultures were inoculated with 1%–2% inoculum or a single colony and grown anaerobically. Medium was supplemented with 20 µM uracil as needed for *pyrF* selections.

Preparation of *P. furiosus* S20 cell lysates

Pfu cells were grown to an O.D 600 between 0.3 and 0.4, pelleted, and then suspended in a 1:4 w/v ratio with lysis buffer (50 mM Tris pH 8.0 with one tablet per 10 mL EDTA-free Complete Mini Protease Inhibitor (Roche)). Cells were lysed via sonication on ice for two minutes total pulse time (10-second pulse, 20-second rest). Cell lysate were then centrifuged at 20,000 x g for 20 minutes. The soluble supernatants were then separated from the insoluble pellet and quantified in triplicate via the Qubit protein assay (Life Technologies). Soluble lysates were then used for Co-IP and chromatography studies.

Western blotting

Typically either 50-100 ng purified protein, 50 ug of *P. furiosus* S20 cell lysates, or ¼ of a total immunoprecipitation, were loaded on SDS-PAGE for western blotting. The amount of antibody used in probing blots was empirically determined and varied per antibody and preparation. Typical dilutions ranged from 1:12,500-25,000 for chicken pAb IgYs and 1:25,000-50,000 for mouse mAb IgG (Cas1 mAb only) in 5% milk/TBS solution. Blotting was performed via standard procedures. For chicken IgY blotting, 1:25,000 diluted Gallus Immunotech 1mg/ml goat anti-IgY HRP secondary antibody was used. For mouse IgG, 1:3,000 diluted Bio-Rad goat anti-mouse HRP conjugate was used. Blots were visualized by ECL Prime western blot chemiluminescent detection reagents (GE Life Sciences).

Cas1 mAb IgG Co-IPs

Ten ug of mouse A5D9 cell line or A6.C11 cell line Cas1 IgGs were coupled to 60 ul protein G sepharose slurry (Thermo Fisher) or 50 ul dynabeads protein G (Thermo Fisher) according to manufacturer's standard procedures. After coupling, the resin was incubated with 2 mg *Pfu* S20 lysates from the appropriate strain (TPF41 or TPF68). The reaction was then brought to 1 mL

with IPP buffer (10 mM Tris pH 8.0, 0.1% Igepal, 100 mM-1 M NaCl, 50 U SUPERase-IN) and incubated for two hours with end-over-end mixing at room temperature. If Co-IPs were performed under anaerobic conditions, then this procedure took place entirely in an anaerobic chamber except for the cell lysis step. Following incubation, beads were pelleted by centrifugation for two minutes at 3,000 x g or using Dynamag 6 for Dynabead IP. Reactions were washed 4X by pelleting beads, removing the supernatant, and bringing reactions back to 1 mL with IPP buffer. Elutions were carried out with 40 ul of non-reducing SDS loading buffer (-BME), and heating at 65°C for 5 minutes. ¼ of a Co-IP was then subject to 15% SDS-PAGE, or western blotting, while the remaining 50% was analyzed via mass spectrometry.

Cas1, Cas2, Cas4-1, IgY Co-IPs

One hundred twenty five ug chicken IgYs for the appropriate antigen were covalently coupled to 60 ul Ultralink resin (Pierce) according to manufacturer's standard protocols. After coupling, beads were incubated with 2 mg Pfu S20 lysates from the appropriate strain. The reaction was then brought to 1mL with IPP buffer (10 mM Tris pH 8.0, 0.1% Igepal, 100 mM-1 M NaCl, 50 U SUPERase-IN) and incubated for 2 hours with end-over-end mixing at room temperature. Following incubation, beads were pelleted by centrifugation for 2 minutes at 3000 x g. Reactions were washed 4X by pelleting the beads, removing the supernatant, and bringing the reactions back to 1 mL volume with IPP buffer. Elutions were carried out via 40 ul of non-reducing SDS loading buffer (-BME), and heating at 65°C for 5 minutes. ¼ of a Co-IP was then subject to 15% SDS-PAGE, or western blotting, while the remaining 50% was analyzed via mass spectrometry.

Size Exclusion Chromatography

P. furiosus TPF41 cell lysates were prepared as described previously and then separated via gel filtration on a Sephacryl, 16/60 S200 HR column. The column was equilibrated with 2 column volumes of IPP buffer without Igepal detergent (10 mM Tris pH 8.0, 300 mM NaCl) and then run in the same buffer to mimic Co-IP conditions. A gel filtration standard (Bio-Rad) containing a mix of 5 proteins (thyroglobulin 670 kDa, γ -globulin 158 kDa, ovalbumin 44 kDa, myoglobin 17 kDa, and vitamin B12 1.35 kDa) was run as a parallel column to TPF41 lysates to determine molecular weight. Fractions of 3 mL were collected, and 10 μ L of each fraction was subjected to western blotting to determine fraction protein content.

Nickel Affinity Chromatography

Pyrococcus furiosus strain TPF68 was transformed with pEAW002 expressing a polyhistidine-tagged Cas1. Lysates from this strain were then subjected to batch affinity purification using a Hispur Ni-NTA resin (Thermo Scientific). Untransformed TPF68 lysates were used as a tagless control. Lysates and resin were incubated together for 2 hours with rotation in an anaerobic chamber in the absence of oxygen. The nickel resin was washed with five 1 mL volume (10 column volume) washes with increasing amounts of imidazole to remove contaminant proteins. Three washes were repeated with 20 mM imidazole, followed by two washes at 50 and 100 mM imidazole. Nickel affinity purified proteins are then eluted twice with high imidazole concentration at 500 mM (200 μ L/ 2 column volumes). Next, purified proteins were subject to both SDS-PAGE and western blotting analysis to examine protein content.

Cas1 mAb ChIP

100 mL Pfu cells grown to an OD 600 of 0.3 as previously described were cooled to room temperature and then fixed for 15 minutes with mixing via the introduction of 1% ChIP grade formaldehyde (Pierce, 16%, methanol free). Crosslinking was quenched via the introduction of glycine to a concentration of 125mM incubated with mixing for an additional 5 minutes. Cells were then pelleted via centrifugation, and excess media was removed. Cell pellets were weighed and resuspended 1:4 in lysis buffer (50 mM Tris pH 8.0 with one tablet per 10 mL EDTA-free Complete Mini protease inhibitor (Roche) and sonicated for 2 minutes pulse time (10 second pulse, 20-second rest). Cell lysates were clarified via centrifugation at 20,000 x g for 20 minutes. The soluble portion of the lysate was then quantified in triplicate via Qubit protein assay. 2 mg total protein was loaded per IP, and IP was performed with A5D9 and A6C11 Cas1 mAbs as described above. IP material was eluted in 50 mM Tris, 10mM EDTA 1% SDS buffer, and the crosslinking was reversed overnight by heating at 65C. After reversal of crosslinks, ChIP material was treated with proteinase K (Thermo Scientific) and subject to phenol-chloroform extraction. The material was further analyzed via denaturing PAGE and staining with SYBR gold.

Sample Preparation for Tandem Mass Spectrometry

Samples were processed both by in-gel and in-solution methods. In each method, ½ of a Co-IP reaction was prepared for analysis. For in-gel methods, samples were either processed by a shotgun method and by band cutting. For shotgun methodology, Co-IP samples eluted in SDS loading buffer were run into a gel to remove SDS detergent. Entire lanes were then excised from the SDS-PAGE gel for a given sample. For band cutting analyses (Cas1 mAb IP 200 mM NaCl) the gel was silver stained, and then the sample was excised from the relevant area. Pre-immune,

non-immune, and samples from deletion strains were processed as specificity controls alongside immunes. Samples processed in solution were eluted via 8 M urea and then analyzed.

Protein Identification by Tandem Mass Spectrometry

Co-IP samples were prepared by either in-gel or in-solution methods as previously detailed. In-gel and in-solution tryptic digests were performed as described previously (Hale et al., 2009; Lim et al., 2008; Wells et al., 2002). Tryptic peptides were analyzed via nano ESI-LC-MS/MS on a hybrid ion trap mass spectrometer (LTQ-XL-Orbitrap Thermo Scientific). Peptide data were searched against a *Pyrococcus furiosus*-specific database (Pyrococcus furiosus COM1 UniProt) using SEQUEST (Proteome Discoverer 1.4, Thermo Fisher Scientific). For control Co-IP samples (non-immune or pre-immune) search outputs were filtered to achieve a 5% protein false discovery rate at 99% confidence using the ProValT algorithm as deployed in PROTEOIQ (www.nusep.com). Conversely, for immune samples, search output was filtered to a 1% false discovery rate at 99% confidence. Thus, the resulting peptide identifications were additionally scrutinized for immune specificity.

References

- Amitai, G., and Sorek, R. (2016). CRISPR-Cas adaptation: insights into the mechanism of action. *Nat Rev Microbiol* 14, 67-76.
- Arslan, Z., Hermanns, V., Wurm, R., Wagner, R., and Pul, U. (2014). Detection and characterization of spacer integration intermediates in type I-E CRISPR-Cas system. *Nucleic Acids Res* 42, 7884-7893.
- Babu, M., Beloglazova, N., Flick, R., Graham, C., Skarina, T., Nocek, B., Gagarinova, A., Pogoutse, O., Brown, G., Binkowski, A., *et al.* (2011). A dual function of the CRISPR-Cas system in bacterial antiviral immunity and DNA repair. *Mol Microbiol* 79, 484-502.
- Barrangou, R. (2013). CRISPR-Cas systems and RNA-guided interference. *Wiley Interdiscip Rev RNA* 4, 267-278.
- Barrangou, R., Fremaux, C., Deveau, H., Richards, M., Boyaval, P., Moineau, S., Romero, D.A., and Horvath, P. (2007). CRISPR provides acquired resistance against viruses in prokaryotes. *Science* 315, 1709-1712.
- Datsenko, K.A., Pougach, K., Tikhonov, A., Wanner, B.L., Severinov, K., and Semenova, E. (2012). Molecular memory of prior infections activates the CRISPR/Cas adaptive bacterial immunity system. *Nat Commun* 3, 945.
- Deveau, H., Barrangou, R., Garneau, J.E., Labonte, J., Fremaux, C., Boyaval, P., Romero, D.A., Horvath, P., and Moineau, S. (2008). Phage response to CRISPR-encoded resistance in *Streptococcus thermophilus*. *J Bacteriol* 190, 1390-1400.
- Forterre, P., Bergerat, A., and Lopez-Garcia, P. (1996). The unique DNA topology and DNA topoisomerases of hyperthermophilic archaea. *FEMS Microbiol Rev* 18, 237-248.
- Fujikane, R., Ishino, S., Ishino, Y., and Forterre, P. (2010). Genetic analysis of DNA repair in the hyperthermophilic archaeon, *Thermococcus kodakaraensis*. *Genes Genet Syst* 85, 243-257.
- Goyal, M., Banerjee, C., Nag, S., and Bandyopadhyay, U. (2016). The Alba protein family: Structure and function. *Biochim Biophys Acta* 1864, 570-583.
- Grissa, I., Vergnaud, G., and Pourcel, C. (2007). CRISPRFinder: a web tool to identify clustered regularly interspaced short palindromic repeats. *Nucleic Acids Res* 35, W52-57.
- Hale, C.R., Zhao, P., Olson, S., Duff, M.O., Graveley, B.R., Wells, L., Terns, R.M., and Terns, M.P. (2009). RNA-guided RNA cleavage by a CRISPR RNA-Cas protein complex. *Cell* 139, 945-956.

- Heler, R., Samai, P., Modell, J.W., Weiner, C., Goldberg, G.W., Bikard, D., and Marraffini, L.A. (2015). Cas9 specifies functional viral targets during CRISPR-Cas adaptation. *Nature* 519, 199-202.
- Ivancic-Bace, I., Cass, S.D., Wearne, S.J., and Bolt, E.L. (2015). Different genome stability proteins underpin primed and naive adaptation in *E. coli* CRISPR-Cas immunity. *Nucleic Acids Res* 43, 10821-10830.
- Jansen, R., Embden, J.D., Gaastra, W., and Schouls, L.M. (2002). Identification of genes that are associated with DNA repeats in prokaryotes. *Mol Microbiol* 43, 1565-1575.
- Kayoko Komori, T.M., Jocelyne DiRuggiero, Rhonda Holley-Shanks, Ikuko Hayashi, Isaac K. O. Cann, Kota Mayanagi, Hideo Shinagawa and Yoshizumi Ishino (2000). Both RadA and RadB are involved in homologous recombination in *Pyrococcus furiosus*. *The Journal of Biological Chemistry* 275.
- Laurens, N., Driessen, R.P.C., Heller, I., Vorselen, D., Noom, M.C., Hol, F.J.H., White, M.F., Dame, R.T., and Wuite, G.J.L. (2012). Alba shapes the archaeal genome using a delicate balance of bridging and stiffening the DNA. *Nature Communications* 3, 1328.
- Lemak, S., Beloglazova, N., Nocek, B., Skarina, T., Flick, R., Brown, G., Popovic, A., Joachimiak, A., Savchenko, A., and Yakunin, A.F. (2013). Toroidal structure and DNA cleavage by the CRISPR-associated [4Fe-4S] cluster containing Cas4 nuclease SSO0001 from *Sulfolobus solfataricus*. *J Am Chem Soc* 135, 17476-17487.
- Lemak, S., Nocek, B., Beloglazova, N., Skarina, T., Flick, R., Brown, G., Joachimiak, A., Savchenko, A., and Yakunin, A.F. (2014). The CRISPR-associated Cas4 protein Pcal_0546 from *Pyrobaculum calidifontis* contains a [2Fe-2S] cluster: crystal structure and nuclease activity. *Nucleic Acids Res*.
- Lestini, R., Delpech, F., and Myllykallio, H. (2015). DNA replication restart and cellular dynamics of Hef helicase/nuclease protein in *Haloferax volcanii*. *Biochimie* 118, 254-263.
- Levy, A., Goren, M.G., Yosef, I., Auster, O., Manor, M., Amitai, G., Edgar, R., Qimron, U., and Sorek, R. (2015). CRISPR adaptation biases explain preference for acquisition of foreign DNA. *Nature* 520, 505-510.
- Lim, J.M., Sherling, D., Teo, C.F., Hausman, D.B., Lin, D., and Wells, L. (2008). Defining the regulated secreted proteome of rodent adipocytes upon the induction of insulin resistance. *J Proteome Res* 7, 1251-1263.
- Makarova, K.S., Haft, D.H., Barrangou, R., Brouns, S.J., Charpentier, E., Horvath, P., Moineau, S., Mojica, F.J., Wolf, Y.I., Yakunin, A.F., *et al.* (2011). Evolution and classification of the CRISPR-Cas systems. *Nat Rev Microbiol* 9, 467-477.
- Makarova, K.S., and Koonin, E.V. (2015). Annotation and Classification of CRISPR-Cas Systems. *Methods Mol Biol* 1311, 47-75.

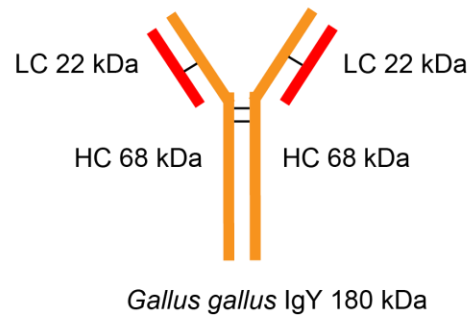
- Makarova, K.S., Wolf, Y.I., Alkhnbashi, O.S., Costa, F., Shah, S.A., Saunders, S.J., Barrangou, R., Brouns, S.J., Charpentier, E., Haft, D.H., *et al.* (2015). An updated evolutionary classification of CRISPR-Cas systems. *Nat Rev Microbiol* 13, 722-736.
- Malina, A., Mills, J.R., Cencic, R., Yan, Y., Fraser, J., Schippers, L.M., Paquet, M., Dostie, J., and Pelletier, J. (2013). Repurposing CRISPR/Cas9 for in situ functional assays. *Genes Dev* 27, 2602-2614.
- Mojica, F.J., Diez-Villasenor, C., Garcia-Martinez, J., and Almendros, C. (2009). Short motif sequences determine the targets of the prokaryotic CRISPR defence system. *Microbiology* 155, 733-740.
- Musharova, O., Klimuk, E., Datsenko, K.A., Metlitskaya, A., Logacheva, M., Semenova, E., Severinov, K., and Savitskaya, E. (2017). Spacer-length DNA intermediates are associated with Cas1 in cells undergoing primed CRISPR adaptation. *Nucleic Acids Res* 45, 3297-3307.
- Nishino, T., Komori, K., Ishino, Y., and Morikawa, K. (2005a). Structural and functional analyses of an archaeal XPF/Rad1/Mus81 nuclease: asymmetric DNA binding and cleavage mechanisms. *Structure* 13, 1183-1192.
- Nishino, T., Komori, K., Tsuchiya, D., Ishino, Y., and Morikawa, K. (2005b). Crystal structure and functional implications of *Pyrococcus furiosus* hef helicase domain involved in branched DNA processing. *Structure* 13, 143-153.
- Nunez, J.K., Bai, L., Harrington, L.B., Hinder, T.L., and Doudna, J.A. (2016). CRISPR Immunological Memory Requires a Host Factor for Specificity. *Mol Cell* 62, 824-833.
- Nunez, J.K., Harrington, L.B., Kranzusch, P.J., Engelman, A.N., and Doudna, J.A. (2015a). Foreign DNA capture during CRISPR-Cas adaptive immunity. *Nature* 527, 535-538.
- Nunez, J.K., Kranzusch, P.J., Noeske, J., Wright, A.V., Davies, C.W., and Doudna, J.A. (2014). Cas1-Cas2 complex formation mediates spacer acquisition during CRISPR-Cas adaptive immunity. *Nat Struct Mol Biol* 21, 528-534.
- Nunez, J.K., Lee, A.S., Engelman, A., and Doudna, J.A. (2015b). Integrase-mediated spacer acquisition during CRISPR-Cas adaptive immunity. *Nature* 519, 193-198.
- Plagens, A., Tjaden, B., Hagemann, A., Randau, L., and Hensel, R. (2012). Characterization of the CRISPR/Cas subtype I-A system of the hyperthermophilic crenarchaeon *Thermoproteus tenax*. *J Bacteriol* 194, 2491-2500.
- Polson, A., von Wechmar, M.B., and Fazakerley, G. (1980a). Antibodies to proteins from yolk of immunized hens. *Immunol Commun* 9, 495-514.
- Polson, A., von Wechmar, M.B., and van Regenmortel, M.H. (1980b). Isolation of viral IgY antibodies from yolks of immunized hens. *Immunol Commun* 9, 475-493.

- Reeve, J.N., Bailey, K.A., Li, W.T., Marc, F., Sandman, K., and Soares, D.J. (2004). Archaeal histones: structures, stability and DNA binding. *Biochem Soc Trans* 32, 227-230.
- Rollie, C., Schneider, S., Brinkmann, A.S., Bolt, E.L., and White, M.F. (2015). Intrinsic sequence specificity of the Cas1 integrase directs new spacer acquisition. *Elife* 4.
- Salzberg, S.L., Sommer, D.D., Schatz, M.C., Phillippy, A.M., Rabinowicz, P.D., Tsuge, S., Furutani, A., Ochiai, H., Delcher, A.L., Kelley, D., *et al.* (2008). Genome sequence and rapid evolution of the rice pathogen *Xanthomonas oryzae* pv. *oryzae* PXO99A. *BMC Genomics* 9, 204.
- Swarts, D.C., Mosterd, C., van Passel, M.W., and Brouns, S.J. (2012). CRISPR interference directs strand specific spacer acquisition. *PLoS One* 7, e35888.
- Terns, R.M., and Terns, M.P. (2013). The RNA- and DNA-targeting CRISPR-Cas immune systems of *Pyrococcus furiosus*. *Biochem Soc Trans* 41, 1416-1421.
- Tse-Dinh, Y.C. (1998). Bacterial and archeal type I topoisomerases. *Biochim Biophys Acta* 1400, 19-27.
- Viswanathan, P., Murphy, K., Julien, B., Garza, A.G., and Kroos, L. (2007). Regulation of dev, an operon that includes genes essential for *Myxococcus xanthus* development and CRISPR-associated genes and repeats. *J Bacteriol* 189, 3738-3750.
- Wang, J., Li, J., Zhao, H., Sheng, G., Wang, M., Yin, M., and Wang, Y. (2015). Structural and Mechanistic Basis of PAM-Dependent Spacer Acquisition in CRISPR-Cas Systems. *Cell* 163, 840-853.
- Wardleworth, B.N., Russell, R.J., Bell, S.D., Taylor, G.L., and White, M.F. (2002). Structure of Alba: an archaeal chromatin protein modulated by acetylation. *EMBO J* 21, 4654-4662.
- Wells, L., Vosseller, K., Cole, R.N., Cronshaw, J.M., Matunis, M.J., and Hart, G.W. (2002). Mapping sites of O-GlcNAc modification using affinity tags for serine and threonine post-translational modifications. *Mol Cell Proteomics* 1, 791-804.
- Yosef, I., Goren, M.G., and Qimron, U. (2012). Proteins and DNA elements essential for the CRISPR adaptation process in *Escherichia coli*. *Nucleic Acids Res* 40, 5569-5576.
- Zhang, J., Kasciukovic, T., and White, M.F. (2012). The CRISPR associated protein Cas4 Is a 5' to 3' DNA exonuclease with an iron-sulfur cluster. *PLoS One* 7, e47232.

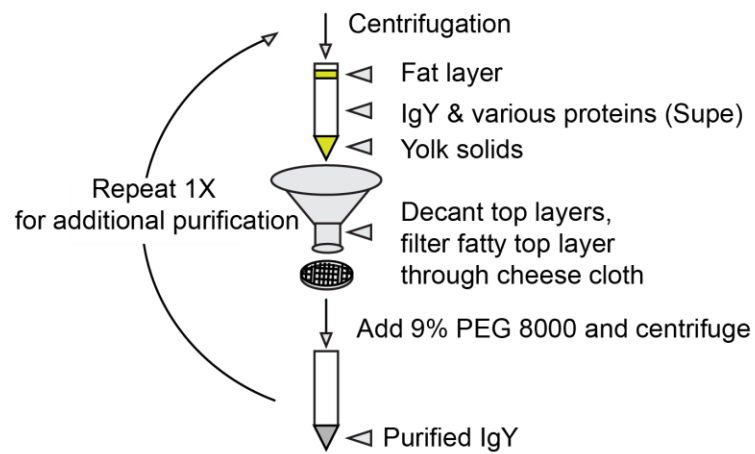
Figure 2.1 Structure and Purification of *Gallus gallus* IgY Antibodies

A) The basic structure of a *Gallus gallus* IgY with two heavy chains (HC, orange) of 68 kDa and two light chains (LC/red) of 22 kDa. The antibody has a total molecular weight of 180 kDa. B) IgY purification by precipitation using polyethylene glycol (PEG). Yolk material containing IgY polyclonal antibodies was drained from Leghorn hen eggs, mixed with a lysis buffer with 3.5% PEG, and subjected to centrifugation. Centrifugation yielded a phase separation between yolk solids, lipids, and IgY/proteins. The aqueous layer (supernatant) was filtered through a funnel and absorbent cotton/cheesecloth to remove the upper phase composed of fats. The filtrate was subjected to a second 9% PEG precipitation via centrifugation. The entire process was repeated an additional time for further purification (see Materials and Methods for further details). C) SDS-PAGE of purified IgYs from various hens. The final purification contains antibody heavy chains and light chains decoupled from one another under reducing conditions in SDS-PAGE (red and orange arrows, respectively). Albumins(black arrows) make up the major contaminants present in a purification. .

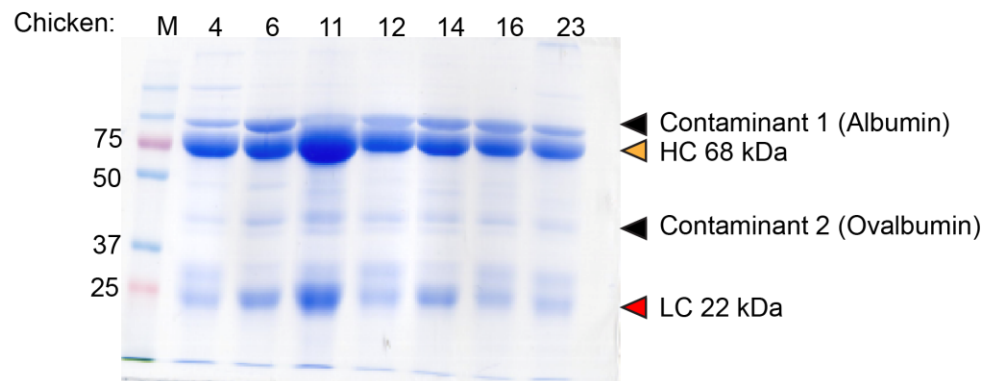
A.



B. Input: 1part egg yolk + 4parts Lysis buffer+ 3.5% PEG 8000



C.



**Figure 2.2 Cas1, Cas2, & Cas4-1 IgY Antibodies Recognize Native and Recombinant
Antigens**

Polyclonal IgY antibodies (pAbs) Figure 2.1 raised against Cas1, Cas2, and Cas4-1, and Cas1 monoclonal IgG (mAbs) antibodies raised against Cas1 were subjected to western blotting to test reactivity and specificity. A) Cas1 mAb IgG A6.C11 was used to probe immunoblots loaded with 100 ng purified recombinant protein (R) and two lanes loaded with 50 ug total protein of *Pyrococcus furiosus* cell lysates made from strains TPF40 and TPF41 overexpressing native Cas1, Cas2, and Cas4-1 proteins (L+). B) Cas1 mAb IgG A5D9-probed western blots set up as in panel A. C-E) Cas1, Cas2, and Cas4-1 pAb IgYs were tested in western blots set up as in the previous panels. Cas4-1 blot contains a single lane of WT lysate from strain JFW02 without overexpression of Cas1, Cas2, and Cas4-1 (L-).

Figure 2.3 Experimental Design of Adaptation Cas Protein Co-IPs

Two milligrams of *Pyrococcus furiosus* (*Pfu*) lysates were subjected to co-immunoprecipitation (Co-IP) assays to assess Cas and non-Cas associated factors using antibodies described in Figure 2.2. Lysates produced from strains overexpressing the adaptation Cas proteins Cas1, Cas2, and Cas4-1 (TPF41) or Cas1, Cas2, Cas4-1, and Cas4-2 (TPF68) were used in most cases to stimulate the adaptation process and increase assay yields due to low native expression levels of target Cas proteins. IgG- or IgY-coupled beads were incubated with lysates for two hours and then subjected to four 1 mL volume washes with IPP buffer before being eluted. Eluted IP material was then subjected to silver stain SDS-PAGE, immunoblotting, and tandem mass spectrometry (MS) analyses to determine protein content and identity.

Input: 2mg Pfu lysates, adaptation Cas overexpressed

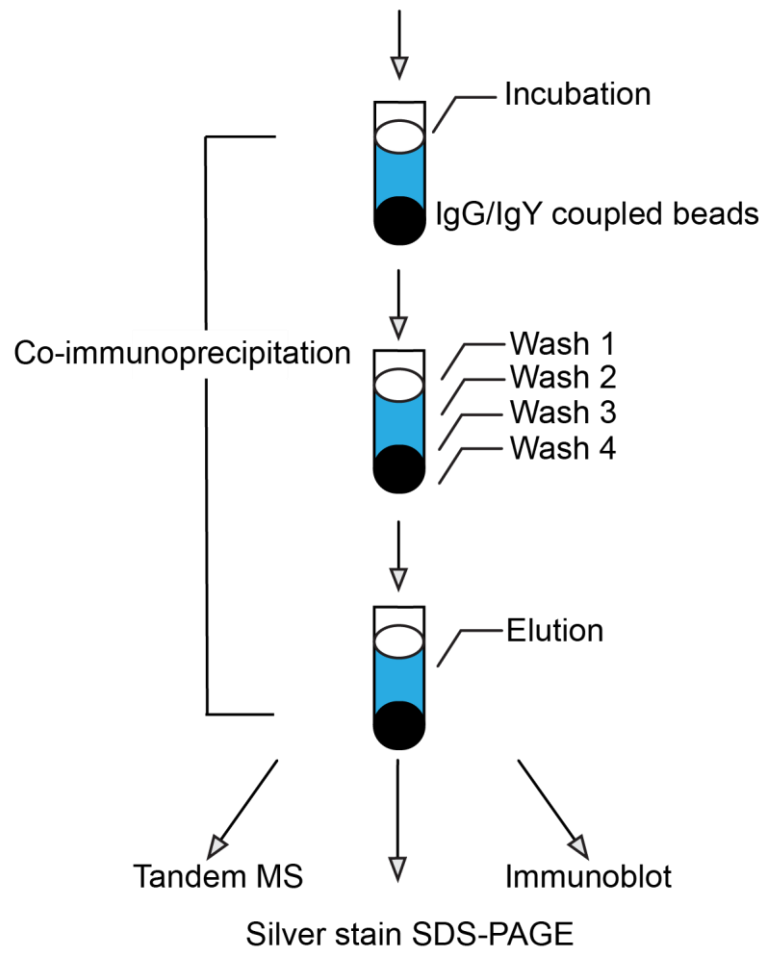


Table 2.1 Summary Co-IP Results by Antibody & Cas protein

Immunoprecipitation results carried out against lysates overexpressing (TPF41 or TPF68) adaptation Cas proteins are summarized in the table below. Anti-Cas protein IP antibodies (top bars) against Cas1, Cas2, and Cas4-1 were used to test for adaptation Cas protein complexes. The ability of an IP to precipitate another Cas protein (left heading) is indicated in the corresponding box. Check marks indicate the protein was identified in at least a single experiment using a particular technique while X denotes a lack of identification from any experiment. While a single experimental identification is the criterion used in this summary, a protein may have been identified in more than one experiment as shown in later figures. In particular, Cas2 was identified in a single Cas1 IP, whereas Cas1 and Cas4-1 were identified in multiple experiments. Methods of protein identification used in these studies consist of both western blotting (WB) and mass spectrometry (MS). In protein identification using MS, the protein in question must have had at least a two-fold spectral count enrichment when compared to a control antibody in the experiment where it was identified. If no or indeterminate data exists for a particular protein or condition, it is notated as “no data” (ND). Gray boxes indicate expected isolation of an antigen such as Cas1 being identified in a Cas1 immunoprecipitation.

		IP Antibody									
		Cas1 mAb A5D9		Cas1 mAb A6.C11		Cas1 pAb		Cas2 pAb		Cas4-1 pAb	
Technique:		MS	WB	MS	WB	MS	WB	MS	WB	MS	WB
	Cas1:	✓	✓	✓	✓	✓	ND	X	✓	X	X
Cas Proteins:	Cas2:	✓	✓	X	X	ND	ND	✓	✓	X	X
	Cas4-1:	✓	✓	✓	✓	ND	ND	X	ND	✓	ND

Legend:

WB: Identified in Co-IP western blotting in at least a single IP experiment

MS: Identified in Co-IP mass spec with 2-fold minimum spectral count enrichment vs a control antibody in at least a single experiment

ND: No data (experiment not performed or indeterminate data)

X: Protein not identified in above techniques

Gray: Antigen identification

Figure 2.4 Cas1 Co-IP Immunoprecipitates Cas4-1 and Cas2/Alba Complexes Along with Unknown Factors

A) Cas1 IgG monoclonal antibody A5D9 Co-IP salt curve. Anti-Cas1 Co-IP performed with A5D9 antibody against lysates with overexpressed Cas1, Cas2, and Cas4-1 (TPF41) was subject to different salt wash conditions from 100 mM to 1 M NaCl. In each salt condition, a band indicative of the IP target/antigen Cas1 was present and specific at 37 kDa (green asterisks) with the anti-Cas1 immune antibody (I) compared to a non-immune control (N) on a 15% SDS-PAGE. Also present were prominent bands at 20 and 10 kDa (blue and red asterisks), possibly indicative of Cas4-1 and Alba, respectively. Additional immune-specific bands that highlight the co-precipitation of other factors are indicated by black asterisks. A similar profile of protein bands was observed when different Cas1 monoclonal and polyclonal antibodies were used in similar experiments (B, C, and D). In polyclonal Co-IPs, (C & D) preimmune antibodies made before hen immunization (P) were used as controls. In this figure, both experiments in A and C were subject to tandem mass spectrometry (MS) analysis to identify protein content. Cas1, Cas4-1, and Alba were isolated and identified by MS in experiment A, while Cas1 and Cas4-1 were identified in experiment C (see MS tables). In separate experiments under the same conditions, Cas1 A6.C11 (E) and A5D9 mAb Co-IPs (F) were subjected to western blotting, confirming the immune-specific presence of Cas4-1 and Cas2 (A5D9 Cas1 IP only).

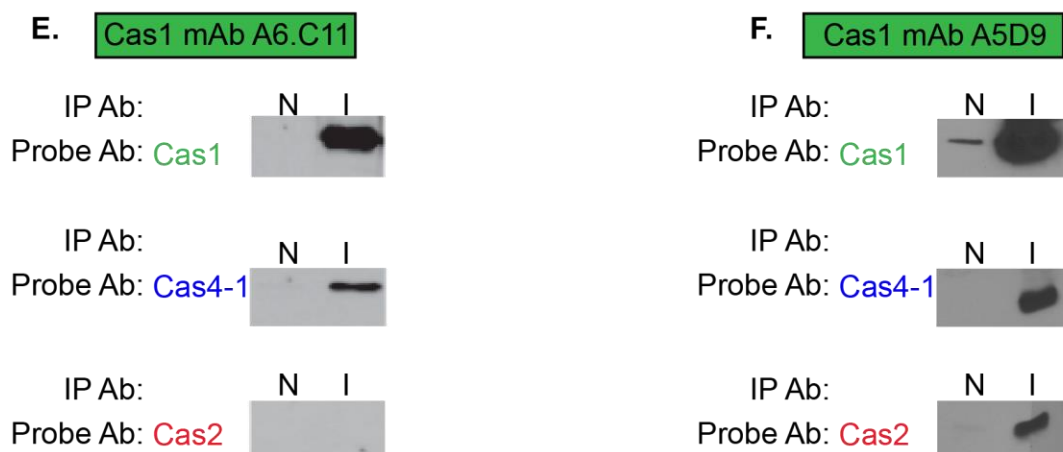
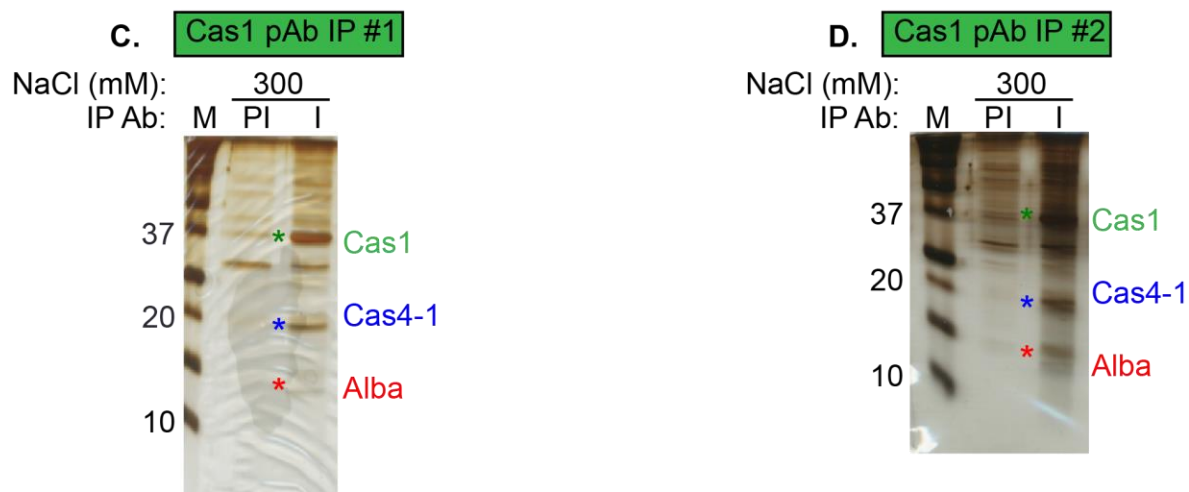
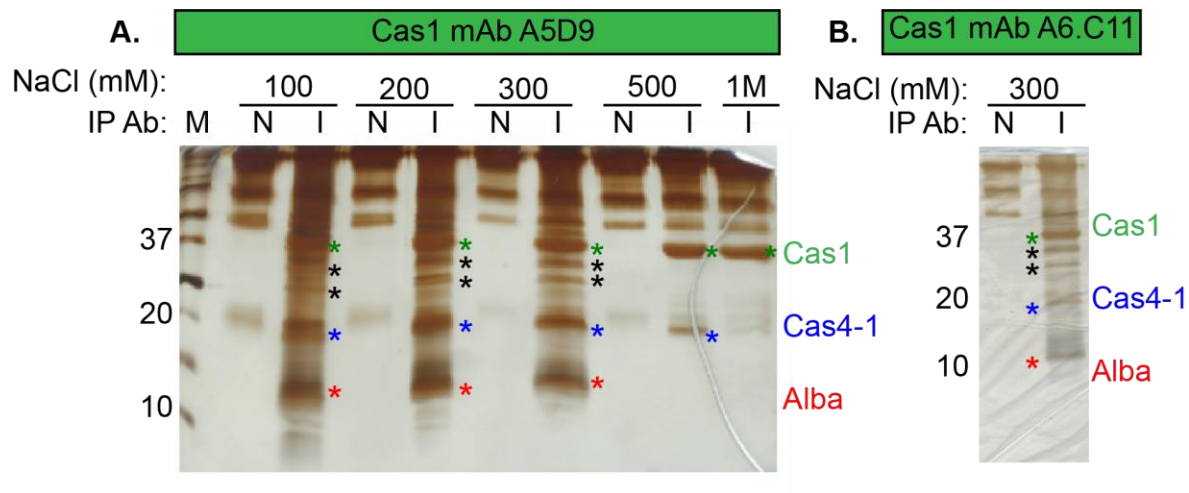


Figure 2.5 Sample Preparation and Analysis of Immunoprecipitated Material by Mass Spectrometry

For each co-immunoprecipitation assay, half of the eluted material from an individual immunoprecipitation was analyzed via tandem mass spectrometry. Samples were analyzed in two distinct fashions, shotgun (right panel) and band/gel cutting (left panel). In shotgun analyses, the entirety of the eluted material was analyzed together simultaneously. Shotgun-style analysis was achieved by two separate techniques of sample preparation: in-solution, and in-gel (bottom right panel). For in-solution analysis, immunoprecipitated material eluted in 8M urea was immediately subjected to proteolytic digestion and analysis. Samples subjected to in-gel analysis, however, were eluted in SDS Laemmli buffer. While this was a strong condition for the elution of immunoprecipitated material, Laemmli buffer was incompatible with mass spectrometry analysis. Therefore, samples were run on an SDS-PAGE gel to remove Laemmli sample buffer, and an entire gel lane was excised and subject to in-gel digest. Conversely, in band cutting analyses (bottom left panel), specific bands or areas of gel (black boxes) were excised and analyzed separately for the purpose of better understanding band identity and protein molecular weight. One IP sample may yield one or many distinct bands to cut for analysis as shown.

Immunoprecipitated Proteins: Cas1, Cas2, Cas4-1 & Others

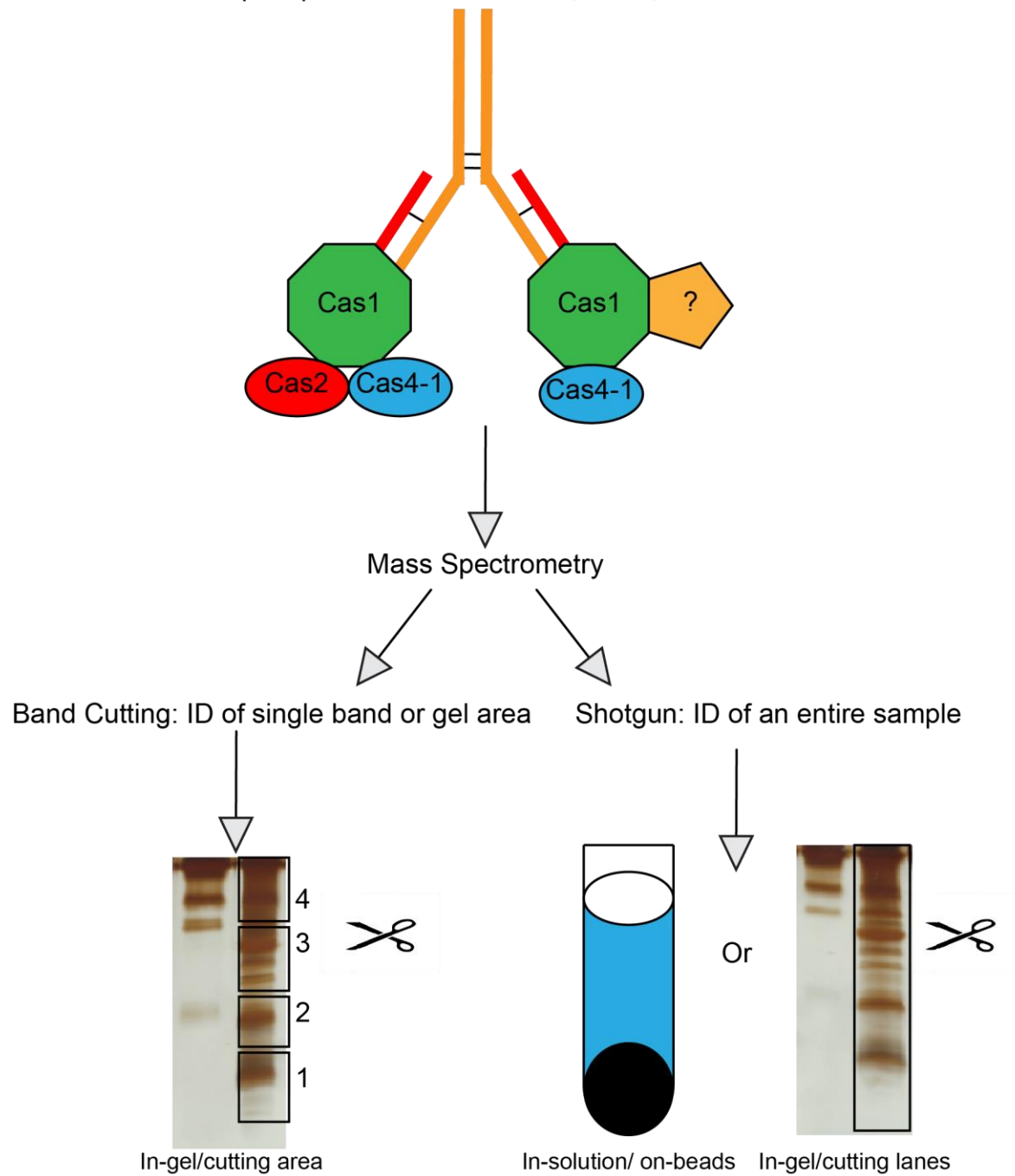


Table 2.2 Shotgun Mass Spectrometry Confirms the Presence of Cas1, Cas4-1, and Alba in Cas1 A6.C11 mAb Co-IP

In the table below, shotgun mass spectrometry results from Cas1 mAb A6.C11 IP experiments are detailed. Immunoprecipitations were performed using lysates overexpressing Cas1, Cas2, and Cas4-1 (TPF41). These experiments were separate from those shown in Figure 2.4. The analysis was conducted via the shotgun in-gel method as detailed in Figure 2.5. Three separate immunoprecipitations and their respective mass spectrometry runs are presented. For each protein identified, the annotation, gene, and locus tabs (left three headings) detail information regarding the proteins identity and function as detailed from the *Pyrococcus furiosus* COM1 strain FASTA database. The molecular weight (Salzberg et al.) indicates the size of the protein. The four right tabs show both the total peptides and spectra associated with the non-immune (NI) control antibody and the immune (IM) anti-Cas1 antibody. The fold change spectra tab is shown as a measure of specificity between the non-immune and immune immunoprecipitations and is calculated simply as immune spectra divided by the non-immune spectra. If a protein has no associated spectra in the non-immune control, it is designated as “unique.” All unique proteins are colored as deep green and are completely immune-specific for a given experiment. Lesser degrees of specificity (unique to two-fold) are colored as light green. Peptides present at two-fold or less are colored as yellow, and one-fold or less are colored as red. Proteins that fall into either of those two categories are considered non-specific and are attributed to the background of the assay. The table is sorted according to the fold change spectra value.

Cas1 A6.C11 mAb Co-IP Shotgun MS Analysis Run#1 (in-gel) 9/18/13								
Annotation	Gene	Locus	MW (kDa)	Fold Change Spectra	Total Peptides NI	Total Peptides IM	Total Spectra NI	Total Spectra IM
Translation initiation factor 2 subunit gamma	eif2g	PF1717	44.99	UNIQUE	0	4	0	11
Proteasome-activating nucleotidase	pan	PF0115	44.76	UNIQUE	0	3	0	6
Cas1	cas1	PF1118	37.44	6.88	8	16	52	358
Cas4-1	PF1119	PF1119	20.14	2.60	2	2	5	13
Uncharacterized protein	PF0124	PF0124	39.38	0.00	2	0	2	0
Cas1 A6.C11 mAb Co-IP Shotgun MS Analysis Run#2 (in-gel) 9/25/13								
Annotation	Gene	Locus	MW (kDa)	Fold Change Spectra	Total Peptides NI	Total Peptides IM	Total Spectra NI	Total Spectra IM
Cas4-1	PF1119	PF1119	20.14	UNIQUE	0	2	0	15
Cas1	cas1	PF1118	37.44	35.77	3	15	13	465
Cas1 A6.C11 mAb Co-IP Shotgun MS Analysis Run#3 (in-gel) 9/25/13								
Annotation	Gene	Locus	MW (kDa)	Fold Change Spectra	Total Peptides NI	Total Peptides IM	Total Spectra NI	Total Spectra IM
Cas1	cas1	PF1118	37.44	UNIQUE	0	9	0	50
Cas4-1	PF1119	PF1119	20.14	UNIQUE	0	2	0	5
DNA/RNA-binding protein Alba	alba	PF1881	10.36	UNIQUE	0	2	0	9

**Table 2.3 Shotgun Mass Spectrometry Confirms Presence of Cas1, Cas4-1, and Alba in
Cas1 A5.D9 mAb Co-IPs**

This Table illustrates results from shotgun A5D9 Cas1 mAb immunoprecipitations using lysates overexpressing Cas1, Cas2, and Cas4-1 (TPF41). Both in-gel and in-solution style shotgun analyses are detailed (see Figure 2.5). In-gel analyses directly correlate with the Figure 2.4 panel A NaCl titration. However, different salts were also analyzed by the in-solution method. Four separate immunoprecipitations (2 in-solution and two in-gel) for both 100 mM and 300 mM NaCl wash conditions are shown. The table is setup as is detailed in Table 2.2.

Cas1 A5D9 mAb Co-IP 100 mM NaCl Shotgun MS Analysis (in-solution)									
Annotation	Gene	Locus/ ORF	MW (kDa)	Total % Seq Coverage	Fold Change Spectra	Total Peptides NI	Total Peptides IM	Total Spectra NI	Total Spectra IM
*Cas4-1	PF1119	PF1119	20.14	86.98	UNIQUE	0	25	0	452
Thermosome, single subunit	PF1974	PF1974	59.93	7.47	UNIQUE	0	5	0	147
ATPase	PFC_04 440	PFC_0 4440	68.14	8.97	UNIQUE	0	5	0	81
50S ribosomal protein L32e	rpl32e	PF1807	15.51	20.00	UNIQUE	0	3	0	81
50S ribosomal protein L2	rpl2	PF1822	26.00	15.48	UNIQUE	0	3	0	73
30S ribosomal protein S10	rps10p	PF1376	11.72	41.18	UNIQUE	0	5	0	66
30S ribosomal protein S12	rps12P	PF1559	16.41	45.58	UNIQUE	0	8	0	60
30S ribosomal protein S8e	rps8e	PF1069	14.24	40.16	UNIQUE	0	5	0	58
50S ribosomal protein L4	rpl4lp	PF1824	28.67	14.90	UNIQUE	0	5	0	53
50S ribosomal protein L10e	rpl10e	PF1279	20.90	9.94	UNIQUE	0	3	0	33
Uncharacterized protein	PF0797	PF0797	40.81	5.76	UNIQUE	0	4	0	29
30S ribosomal protein S3	rps3p	PF1819	23.40	30.95	UNIQUE	0	6	0	26
50S ribosomal protein L6	rpl6	PF1808	20.83	14.67	UNIQUE	0	3	0	26
50S ribosomal protein L15P	rpl15p	PF1802	16.37	16.33	UNIQUE	0	3	0	21
Uncharacterized protein	PF1625	PF1625	14.85	9.09	UNIQUE	0	2	0	20
50S ribosomal protein L39e	rpl39e	PF0379	6.28	17.65	UNIQUE	0	2	0	20
50S ribosomal protein L3	rpl3	PF1825	41.35	5.75	UNIQUE	0	3	0	17
30S ribosomal protein S27e	rps27e	PF0218	6.85	63.49	UNIQUE	0	3	0	16
50S ribosomal protein L29P	rpl29p	PF1818	8.51	22.22	UNIQUE	0	1	0	13
Inosine-5'- monophosphate dehydrogenase	PF1794	PF1794	21.32	14.97	UNIQUE	0	2	0	11
Uncharacterized protein	PF0537	PF0537	90.32	7.40	UNIQUE	0	3	0	10
Uncharacterized protein	PFC_05 155	PFC_0 5155	38.96	6.20	UNIQUE	0	1	0	6
DNA/RNA-binding protein Alba	albA	PF1881	10.36	22.58	UNIQUE	0	2	0	5
*Cas1	cas1	PF1118	37.44	5.59	UNIQUE	0	2	0	5
Cas1 A5D9 mAb Co-IP 100 mM NaCl Shotgun MS Analysis (in-gel)									
Annotation	Gene	Locus/ ORF	MW (kDa)	Total % Seq Coverage	Fold Change Spectra	Total Peptides NI	Total Peptides IM	Total Spectra NI	Total Spectra IM
*Cas1	cas1	PF1118	37.44	68.32	UNIQUE	0	27	0	784
*Cas4-1	PF1119	PF1119	20.14	74.56	UNIQUE	0	15	0	353
C/D box methylation guide ribonucleoprotein complex aNOP56 subunit	PF0060	PF0060	46.74	46.52	UNIQUE	0	22	0	229

DNA/RNA-binding protein Alba	albA	PF1881	10.36	72.04	32.5	2	9	6	195
Fibrillarin-like rRNA/tRNA 2'-O-methyltransferase	flpA	PF0059	25.73	75.77	UNIQUE	0	20	0	144
Uncharacterized protein	PF0490	PF0490	27.82	48.15	UNIQUE	0	11	0	109
tRNA-splicing ligase RtcB	PFC_10530	PFC_10530	108.72	22.04	UNIQUE	0	20	0	106
Uncharacterized protein	PF0496	PF0496	30.57	22.35	UNIQUE	0	8	0	97
50S ribosomal protein L4	rpl4lp	PF1824	28.67	50.20	UNIQUE	0	13	0	95
Putative tRNA(Met) cytidine acetyltransferase	PF0504	PF0504	94.37	26.23	UNIQUE	0	18	0	89
Translation initiation factor 2 subunit gamma	eif2g	PF1717	44.99	45.50	UNIQUE	0	17	0	87
30S ribosomal protein S7	rps7	PF1558	24.62	53.02	UNIQUE	0	13	0	82
Probable tRNA pseudouridine synthase B	truB	PF1785	38.52	25.88	UNIQUE	0	10	0	75
50S ribosomal protein L3	rpl3	PF1825	41.35	31.78	UNIQUE	0	14	0	71
Nucleoside-triphosphatase PFC_01625	PF0501	PF0501	20.83	52.72	UNIQUE	0	11	0	66
Uncharacterized protein	PF0797	PF0797	40.81	31.70	UNIQUE	0	11	0	65
30S ribosomal protein S3	rps3p	PF1819	23.40	52.86	UNIQUE	0	12	0	58
30S ribosomal protein S5	rps5p	PF1804	26.52	44.92	UNIQUE	0	12	0	56
30S ribosomal protein S2	rps2P	PF1640	22.97	50.99	UNIQUE	0	10	0	56
30S ribosomal protein S4	rps4p	PF1649	21.31	32.78	UNIQUE	0	8	0	53
50S ribosomal protein L22P	rpl22p	PF1820	17.61	21.29	UNIQUE	0	4	0	50
50S ribosomal protein L12	rpl12p	PF1994	11.22	71.03	UNIQUE	0	5	0	48
AsnC family transcriptional regulator	PF2053	PF2053	17.19	17.57	UNIQUE	0	3	0	46
50S ribosomal protein L31e	rpl31e	PF0378	11.03	40.00	UNIQUE	0	4	0	44
50S ribosomal protein L15e	rpl15e	PF0876	22.56	15.46	UNIQUE	0	5	0	39
50S ribosomal protein L10	rpl10	PF1993	37.09	16.52	UNIQUE	0	6	0	37
30S ribosomal protein S6e	rps6e	PF0488	13.95	61.60	UNIQUE	0	7	0	36
50S ribosomal protein L30P	rpl30p	PF1803	17.68	43.23	UNIQUE	0	6	0	35
30S ribosomal protein S11	rps11p	PF1648	14.70	25.55	UNIQUE	0	3	0	35
Histone a1	PF1831	PF1831	7.38	32.84	UNIQUE	0	2	0	34

30S ribosomal protein S15	rps15p	PF2056	18.57	41.77	UNIQUE	0	6	0	31
50S ribosomal protein L18P	rpl18p	PF1805	23.15	32.51	UNIQUE	0	5	0	31
30S ribosomal protein S17P	rps17p	PF1815	13.11	37.50	UNIQUE	0	4	0	29
50S ribosomal protein L13P	rpl13p	PF1645	16.28	28.17	UNIQUE	0	5	0	28
Histone a2	PF1722	PF1722	7.26	23.88	UNIQUE	0	2	0	27
50S ribosomal protein L10e	rpl10e	PF1279	20.90	19.34	UNIQUE	0	4	0	25
Ribosomal protein s6 modification protein	PF1682	PF1682	30.82	19.41	UNIQUE	0	4	0	24
Uncharacterized protein	PF1927	PF1927	23.42	19.90	UNIQUE	0	5	0	23
30S ribosomal protein S19e	rps19e	PF1499	17.33	18.67	UNIQUE	0	3	0	23
30S ribosomal protein S17e	rps17E	PF1491	7.90	53.73	UNIQUE	0	5	0	19
ATPase	PFC_04440	PFC_04440	68.14	9.14	UNIQUE	0	7	0	19
Probable Brix domain-containing ribosomal biogenesis protein	PF2010	PF2010	26.04	17.04	UNIQUE	0	3	0	19
30S ribosomal protein S3Ae	rps3ae	PF2054	22.91	22.73	UNIQUE	0	6	0	17
DNA-directed RNA polymerase	PFC_07045	PFC_07045	103.38	6.04	UNIQUE	0	5	0	17
50S ribosomal protein L2	rpl2	PF1822	26.00	24.69	UNIQUE	0	7	0	16
50S ribosomal protein L1	rpl1P	PF1992	23.82	25.46	UNIQUE	0	4	0	16
Ribosomal RNA small subunit methyltransferase Nep1	nep1	PF1262	25.92	17.94	UNIQUE	0	3	0	16
Acetyl-lysine deacetylase	lysK	PF1686	35.89	26.07	UNIQUE	0	7	0	15
Uncharacterized protein	PF0124	PF0124	39.38	19.35	UNIQUE	0	7	0	15
30S ribosomal protein S19P	rps19p	PF1821	15.30	14.39	UNIQUE	0	2	0	15
tRNA/rRNA methyltransferase	PF0251	PF0251	44.99	11.56	UNIQUE	0	5	0	14
50S ribosomal protein L35Ae	rpl35ae	PF1872	9.71	39.08	UNIQUE	0	2	0	14
Tryptophan synthase beta chain	trpB	PF1706	42.48	10.05	UNIQUE	0	5	0	13
30S ribosomal protein S13	rps13p	PF1650	16.89	15.54	UNIQUE	0	3	0	13
Uncharacterized protein	PF0213	PF0213	30.26	20.74	UNIQUE	0	5	0	12
Uncharacterized protein	PF0070	PF0070	36.82	8.67	UNIQUE	0	4	0	12
30S ribosomal protein S12	rps12P	PF1559	16.41	23.13	UNIQUE	0	3	0	12
50S ribosomal protein L18e	rpl18e	PF1646	13.71	21.67	UNIQUE	0	3	0	12

Exosome complex component Rrp41	rrp41	PF1568	27.81	12.00	UNIQUE	0	3	0	12
50S ribosomal protein L14P	rpl14p	PF1814	15.21	10.64	UNIQUE	0	2	0	12
Translation initiation factor 2 subunit alpha	eif2a	PF1140	31.88	17.45	UNIQUE	0	6	0	11
50S ribosomal protein L6	rpl6	PF1808	20.83	14.13	UNIQUE	0	2	0	11
DEXX-box atpase	PF0634	PF0634	54.56	8.73	UNIQUE	0	2	0	10
Uncharacterized protein	PF1641	PF1641	37.68	9.44	UNIQUE	0	3	0	9
Phosphoenolpyruvate synthase	PF0043	PF0043	90.41	6.49	UNIQUE	0	3	0	9
30S ribosomal protein S8e	rps8e	PF1069	14.24	22.05	UNIQUE	0	2	0	9
Uncharacterized protein	PFC_04875	PFC_04875	30.65	6.44	UNIQUE	0	2	0	9
DNA repair and recombination protein RadA	radA	PF1926	38.36	3.72	UNIQUE	0	2	0	9
Type 2 DNA topoisomerase 6 subunit A	top6A	PF1578	44.03	7.85	UNIQUE	0	3	0	8
30S ribosomal protein S4e	rps4e	PF1812	28.11	9.47	UNIQUE	0	2	0	7
Elongation factor 1-alpha	tuf	PF1375	47.59	6.31	UNIQUE	0	3	0	7
Uncharacterized protein	PF0829	PF0829	7.55	26.56	UNIQUE	0	2	0	7
Uncharacterized protein	PF2018	PF2018	31.12	13.01	UNIQUE	0	3	0	6
Beta-glucosidase	PF0442	PF0442	49.71	4.99	UNIQUE	0	2	0	6
50S ribosomal protein L18Ae	rplX	PF0376	9.30	29.87	UNIQUE	0	2	0	3
DNA-directed RNA polymerase	PF1564	PF1564	126.91	1.70	UNIQUE	0	2	0	3
Uncharacterized protein	PF0206	PF0206	30.52	6.55	0	1	0.00	24	0.00
Chromosome partition protein Smc	smc	PF1843	134.91	3.82	0	4	0.00	16	0

Cas1 A5D9 mAb Co-IP 300 mM NaCl Shotgun MS Analysis (in-solution)

Protein Name	Gene	Locus/ORF	MW (kDa)	Total % Seq Coverage	Fold Change Spectra	Total Peptides NI	Total Peptides IM	Total Spectra NI	Total Spectra IM
Cas4-1	PF1119	PF1119	20.14	91.12	UNIQUE	0.00	24	0	435
Thermosome, single subunit	PF1974	PF1974	59.93	7.47	UNIQUE	0.00	5	0	152
Cell division protein CDC48	PF0963	PF0963	94.04	4.30	UNIQUE	0.00	4	0	143
50S ribosomal protein L32e	rpl32e	PF1807	15.51	29.23	UNIQUE	0.00	4	0	95
50S ribosomal protein L15P	rpl15p	PF1802	16.37	27.89	UNIQUE	0.00	6	0	64
50S ribosomal protein L2	rpl2	PF1822	26.00	15.48	UNIQUE	0.00	4	0	58

50S ribosomal protein L4	rpl4p	PF1824	28.67	14.12	UNIQUE	0.00	5	0	56
30S ribosomal protein S12	rps12P	PF1559	16.41	49.66	UNIQUE	0.00	8	0	50
30S ribosomal protein S10	rps10p	PF1376	11.72	41.18	UNIQUE	0.00	5	0	50
50S ribosomal protein L6	rpl6	PF1808	20.83	18.48	UNIQUE	0.00	4	0	48
30S ribosomal protein S8e	rps8e	PF1069	14.24	51.18	UNIQUE	0.00	6	0	36
50S ribosomal protein L3	rpl3	PF1825	41.35	5.75	UNIQUE	0.00	4	0	36
50S ribosomal protein L29P	rpl29p	PF1818	8.51	22.22	UNIQUE	0.00	1	0	24
ATPase	PFC_04440	PFC_04440	68.14	7.31	UNIQUE	0.00	4	0	16
50S ribosomal protein L39e	rpl39e	PF0379	6.28	17.65	UNIQUE	0.00	2	0	16
Uncharacterized protein	PF1625	PF1625	14.85	9.09	UNIQUE	0.00	2	0	14
30S ribosomal protein S19P	rps19p	PF1821	15.30	6.06	UNIQUE	0.00	1	0	12
30S ribosomal protein S14 type Z	rps14P	PF1810	6.59	30.36	UNIQUE	0.00	2	0	6
Cas1	cas1	PF1118	37.44	6.52	UNIQUE	0.00	2	0	5
Uncharacterized protein	PF0206	PF0206	30.52	13.82	0	2	0	17	0
Leucine--tRNA ligase	leuS	PF0890	113.63	3.21	0	3	0	651	0
ATP-dependent helicase	PF1051	PF1051	100.16	5.66	0	3	0	36	0
NADH oxidase	PF2006	PF2006	45.45	3.39	0	1	0	7	0
Cas1 A5D9 mAb Co-IP 300 mM NaCl Shotgun MS Analysis (in-gel)									
Protein Name	Gene	Locus/ORF	MW (kDa)	Total % Seq Coverage	Fold Change Spectra	Total Peptides NI	Total Peptides IM	Total Spectra NI	Total Spectra IM
Cas1	cas1	PF1118	37.44	70.81	UNIQUE	0	29	0	967
Cas4-1	PF1119	PF1119	20.14	74.56	UNIQUE	0	14	0	335
C/D box methylation guide ribonucleoprotein complex aNOP56 subunit	PF0060	PF0060	46.74	46.52	UNIQUE	0	23	0	220
Fibrillarin-like rRNA/tRNA 2'-O-methyltransferase	flpA	PF0059	25.73	74.45	UNIQUE	0	19	0	167
Uncharacterized protein	PF0490	PF0490	27.82	48.15	UNIQUE	0	12	0	157
DNA/RNA-binding protein Alba	alba	PF1881	10.36	69.89	UNIQUE	0	7	0	108
Putative tRNA(Met) cytidine acetyltransferase	PF0504	PF0504	94.37	25.12	UNIQUE	0	17	0	99
50S ribosomal protein L3	rpl3	PF1825	41.35	43.56	UNIQUE	0	18	0	94

50S ribosomal protein L4	rpl4lp	PF1824	28.67	56.86	UNIQUE	0	14	0	92
30S ribosomal protein S7	rps7	PF1558	24.62	40.47	UNIQUE	0	10	0	67
50S ribosomal protein L15e	rpl15e	PF0876	22.56	16.49	UNIQUE	0	5	0	63
Probable tRNA pseudouridine synthase B	truB	PF1785	38.52	17.65	UNIQUE	0	7	0	58
50S ribosomal protein L22P	rpl22p	PF1820	17.61	27.74	UNIQUE	0	6	0	57
50S ribosomal protein L10	rpl10	PF1993	37.09	21.83	UNIQUE	0	8	0	56
30S ribosomal protein S4	rps4p	PF1649	21.31	32.78	UNIQUE	0	8	0	55
30S ribosomal protein S5	rps5p	PF1804	26.52	43.64	UNIQUE	0	11	0	46
50S ribosomal protein L18P	rpl18p	PF1805	23.15	25.62	UNIQUE	0	5	0	43
50S ribosomal protein L30P	rpl30p	PF1803	17.68	41.94	UNIQUE	0	6	0	42
50S ribosomal protein L12	rpl12p	PF1994	11.22	71.03	UNIQUE	0	5	0	37
30S ribosomal protein S6e	rps6e	PF0488	13.95	52.80	UNIQUE	0	6	0	33
50S ribosomal protein L2	rpl2	PF1822	26.00	38.08	UNIQUE	0	9	0	31
50S ribosomal protein L10e	rpl10e	PF1279	20.90	19.34	UNIQUE	0	4	0	31
50S ribosomal protein L1	rpl1P	PF1992	23.82	29.17	UNIQUE	0	5	0	26
50S ribosomal protein L13P	rpl13p	PF1645	16.28	28.87	UNIQUE	0	5	0	25
Nucleoside-triphosphatase PFC_01625	PF0501	PF0501	20.83	27.17	UNIQUE	0	6	0	24
30S ribosomal protein S15	rps15p	PF2056	18.57	33.54	UNIQUE	0	5	0	23
50S ribosomal protein L31e	rpl31e	PF0378	11.03	33.68	UNIQUE	0	4	0	22
50S ribosomal protein L15P	rpl15p	PF1802	16.37	25.17	UNIQUE	0	5	0	19
30S ribosomal protein S19P	rps19p	PF1821	15.30	9.85	UNIQUE	0	2	0	18
50S ribosomal protein L5	rpl5	PF1811	21.20	12.37	UNIQUE	0	3	0	18
50S ribosomal protein L32e	rpl32e	PF1807	15.51	33.08	UNIQUE	0	4	0	17
Uncharacterized protein	PF1927	PF1927	23.42	15.05	UNIQUE	0	3	0	17
30S ribosomal protein S17e	rps17E	PF1491	7.90	44.78	UNIQUE	0	4	0	16
50S ribosomal protein L23P	rpl23p	PF1823	9.85	45.35	UNIQUE	0	4	0	16
30S ribosomal protein S2	rps2P	PF1640	22.97	17.82	UNIQUE	0	4	0	15
30S ribosomal protein S3Ae	rps3ae	PF2054	22.91	19.70	UNIQUE	0	5	0	13

Probable Brix domain-containing ribosomal biogenesis protein	PF2010	PF2010	26.04	8.52	UNIQUE	0	2	0	13
50S ribosomal protein L35Ae	rpl35ae	PF1872	9.71	47.13	UNIQUE	0	3	0	11
50S ribosomal protein L6	rpl6	PF1808	20.83	14.13	UNIQUE	0	2	0	11
30S ribosomal protein S13	rps13p	PF1650	16.89	15.54	UNIQUE	0	3	0	11
Uncharacterized protein	PF0829	PF0829	7.55	26.56	UNIQUE	0	2	0	11
30S ribosomal protein S11	rps11p	PF1648	14.70	27.01	UNIQUE	0	4	0	9
ATPase	PFC_04440	PFC_04440	68.14	6.64	UNIQUE	0	4	0	7
50S ribosomal protein L18e	rpl18e	PF1646	13.71	14.17	UNIQUE	0	2	0	7
30S ribosomal protein S17P	rps17p	PF1815	13.11	20.54	UNIQUE	0	2	0	6
Uncharacterized protein	PFC_04875	PFC_04875	30.65	6.82	UNIQUE	0	2	0	4
Uncharacterized protein	PFC_04935	PFC_04935	29.95	4.12	UNIQUE	0	1	0	2
Type II secretion system protein	PFC_04200	PFC_04200	129.90	2.72	0	1	0	6	0

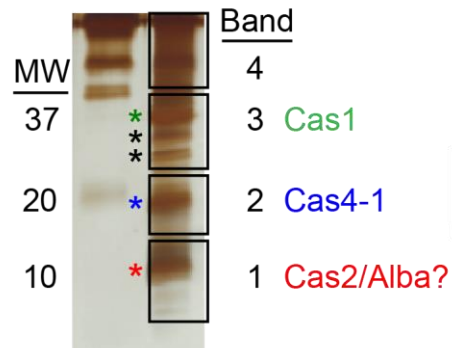
Figure 2.6 Band Excision Mass Spectrometry of Cas1 A5D9 mAb Immunoprecipitations

Cas1 A5D9 200 mM NaCl immunoprecipitated material from Figure 2.4 panel A was subject to mass spectrometry analyses. Rather than a shotgun approach, band excision was used to identify proteins in particular regions of the gel. Non-immune control antibody lanes (N) and anti-Cas1 antibody lanes (I) are indicated. The bands/regions cut are indicated by black boxes. Cuts are numbered 1-4 from smallest to largest molecular weight. Bands likely correlating to immune-specific Cas proteins are indicated by colored asterisks. Bands associated with unknown proteins are indicated by black asterisks. The mass spectrometry results from each region cut are indicated in the next table (Table 2.4).

Cas1 mAb A5D9

NaCl (mM): 200

IP Ab: N I



**Table 2.4 Cas1, Cas4-1, and Alba are Identified at the Expected Molecular Weights in Cas1
mAb Co-IPs**

The table below illustrates protein identifications from band excision mass spectrometry of A5D9 Cas1 mAb immunoprecipitation material. The immunoprecipitation was performed at 200 mM NaCl wash conditions with TPF41 lysates, and is the exact material from the Figure 2.4 SDS-PAGE (Figure 2.4, panel A, 200 mM lane). The cutting of each band/gel area is as indicated in Figure 2.6. The table headers indicate which band each analysis correlates to (bands1-4 Figure 2.6). Other information and results present in the table are set up as in Table 2.2.

Cas1 A5D9 mAb Co-IP 200 mM NaCl SDS-PAGE Band Analysis (band #1)									
Annotation	Gene	Locus/ ORF	MW (kDa)	Total % Seq Coverage	Fold Change Spectra	Total Peptides NI	Total Peptides IM	Total Spectra NI	Total Spectra IM
DNA/RNA-binding protein Alba	albA	PF1881	10.36	78.49	UNIQUE	0	11	0	855
30S ribosomal protein S19e	rps19e	PF1499	17.33	60.00	UNIQUE	0	12	0	336
50S ribosomal protein L18e	rpl18e	PF1646	13.71	70.83	UNIQUE	0	12	0	316
50S ribosomal protein L13P	rpl13p	PF1645	16.28	54.93	UNIQUE	0	11	0	311
30S ribosomal protein S17P	rps17p	PF1815	13.11	68.75	UNIQUE	0	10	0	264
30S ribosomal protein S11	rps11p	PF1648	14.70	50.36	UNIQUE	0	8	0	186
50S ribosomal protein L7Ae	rpl7ae	PF1367	13.37	60.98	UNIQUE	0	6	0	176
50S ribosomal protein L31e	rpl31e	PF0378	11.03	47.37	UNIQUE	0	7	0	167
30S ribosomal protein S13	rps13p	PF1650	16.89	60.81	UNIQUE	0	12	0	154
50S ribosomal protein L12	rpl12p	PF1994	11.22	71.03	UNIQUE	0	5	0	152
50S ribosomal protein L30P	rpl30p	PF1803	17.68	65.16	UNIQUE	0	10	0	149
30S ribosomal protein S6e	rps6e	PF0488	13.95	62.40	UNIQUE	0	8	0	136
30S ribosomal protein S9	rps9p	PF1644	15.26	41.48	UNIQUE	0	5	0	110
30S ribosomal protein S8	rps8p	PF1809	14.63	44.62	UNIQUE	0	7	0	104
50S ribosomal protein L24P	rpl24p	PF1813	14.30	23.14	UNIQUE	0	5	0	98
Uncharacterized protein	PF1625	PF1625	14.85	63.64	UNIQUE	0	9	0	95
30S ribosomal protein S10	rps10p	PF1376	11.72	35.29	UNIQUE	0	4	0	94
30S ribosomal protein S17e	rps17E	PF1491	7.90	55.22	UNIQUE	0	7	0	90
50S ribosomal protein L18Ae	rplX	PF0376	9.30	49.35	UNIQUE	0	4	0	82
Uncharacterized protein	PF0206	PF0206	30.52	10.91	UNIQUE	0	2	0	77
50S ribosomal protein L21e	rpl21e	PF1035	11.32	45.36	UNIQUE	0	4	0	72
H/ACA RNA-protein complex component Gar1	PFC_09635	PFC_09635	11.45	41.24	UNIQUE	0	5	0	65
30S ribosomal protein S8e	rps8e	PF1069	14.24	50.39	UNIQUE	0	6	0	59
30S ribosomal protein S24e	rps24e	PF0253	11.65	44.44	UNIQUE	0	5	0	58
AsnC family transcriptional regulator	PF2053	PF2053	17.19	51.35	UNIQUE	0	10	0	57
50S ribosomal protein L23P	rpl23p	PF1823	9.85	56.98	UNIQUE	0	7	0	57

Translation initiation factor 2 subunit beta	eif2b	PF0481	16.21	63.57	UNIQUE	0	9	0	56
50S ribosomal protein L14P	rpl14p	PF1814	15.21	58.87	UNIQUE	0	7	0	55
50S ribosomal protein L14e	rpl14e	PF0819	8.94	54.22	UNIQUE	0	5	0	55
30S ribosomal protein S19P	rps19p	PF1821	15.30	34.85	UNIQUE	0	6	0	52
Ornithine carbamoyltransferase	PF0594	PF0594	35.14	11.43	UNIQUE	0	2	0	51
Histone a2	PF1722	PF1722	7.26	23.88	UNIQUE	0	3	0	48
Cas4-1	PF1119	PF1119	20.14	26.63	UNIQUE	0	7	0	45
50S ribosomal protein L37e	rpl37e	PF1541	7.20	40.32	UNIQUE	0	3	0	45
50S ribosomal protein L35Ae	rpl35ae	PF1872	9.71	47.13	UNIQUE	0	4	0	41
50S ribosomal protein L39e	rpl39e	PF0379	6.28	29.41	UNIQUE	0	3	0	41
Uncharacterized protein	PFC_04545	PFC_04545	9.03	37.50	UNIQUE	0	5	0	39
Histone a1	PF1831	PF1831	7.38	32.84	UNIQUE	0	2	0	37
50S ribosomal protein L29P	rpl29p	PF1818	8.51	63.89	UNIQUE	0	4	0	28
Type II secretion system protein	PFC_04200	PFC_04200	129.90	4.38	UNIQUE	0	3	0	27
Probable transcription termination protein NusA	nusA	PF1560	16.43	33.79	UNIQUE	0	5	0	24
Uncharacterized protein	PF0829	PF0829	7.55	26.56	UNIQUE	0	2	0	24
Uncharacterized protein	PFC_09020	PFC_09020	18.20	30.97	UNIQUE	0	5	0	23
50S ribosomal protein L30e	rpl30e	PF1561	10.69	29.29	UNIQUE	0	3	0	23
30S ribosomal protein S14 type Z	rps14P	PF1810	6.59	30.36	UNIQUE	0	2	0	22
30S ribosomal protein S27e	rps27e	PF0218	6.85	50.79	UNIQUE	0	2	0	19
Putative snRNP Sm-like protein	PF1542	PF1542	8.42	39.47	UNIQUE	0	3	0	19
50S ribosomal protein L32e	rpl32e	PF1807	15.51	26.15	UNIQUE	0	3	0	18
Cas1	cas1	PF1118	37.44	16.46	UNIQUE	0	5	0	17
30S ribosomal protein S28e	rps28e	PF1368	8.08	26.76	UNIQUE	0	2	0	16
Periplasmic sugar binding protein	PF0119	PF0119	61.21	5.73	UNIQUE	0	2	0	15
Translation initiation factor 1A	eif1a	PFC_07140	13.05	36.28	UNIQUE	0	3	0	14
UPF0201 protein PFC_05700	PFC_05700	PFC_05700	15.71	22.22	UNIQUE	0	4	0	13
Ribonuclease P protein component 2	rnp2	PF1378	13.81	20.00	UNIQUE	0	2	0	13

Uncharacterized protein	PF1498	PF1498	16.49	10.79	UNIQUE	0	2	0	12
50S ribosomal protein L22P	rpl22p	PF1820	17.61	10.32	UNIQUE	0	2	0	9
Protein translation factor SUI1 homolog	PF1817	PF1817	11.39	17.35	UNIQUE	0	2	0	8
Cas1 A5D9 mAb Co-IP 200 mM NaCl SDS-PAGE Band Analysis									
Annotation	Gene	Locus/ORF	MW (kDa)	Total % Seq Coverage	Fold Change Spectra	Total Peptides NI	Total Peptides IM	Total Spectra NI	Total Spectra IM
Cas4-1	PF1119	PF1119	20.14	79.88	UNIQUE	0	18	0	824
Fibrillarin-like rRNA/tRNA 2'-O-methyltransferase	flpA	PF0059	25.73	75.77	UNIQUE	0	20	0	365
30S ribosomal protein S7	rps7	PF1558	24.62	65.12	UNIQUE	0	16	0	222
30S ribosomal protein S5	rps5p	PF1804	26.52	68.22	UNIQUE	0	18	0	209
30S ribosomal protein S15	rps15p	PF2056	18.57	61.39	UNIQUE	0	10	0	204
50S ribosomal protein L18P	rpl18p	PF1805	23.15	67.98	UNIQUE	0	13	0	202
50S ribosomal protein L1	rpl1P	PF1992	23.82	57.87	UNIQUE	0	10	0	173
50S ribosomal protein L4	rpl4lp	PF1824	28.67	57.25	UNIQUE	0	14	0	167
Uncharacterized protein	PF0490	PF0490	27.82	60.49	UNIQUE	0	15	0	145
Cell division protein CDC48	PF0963	PF0963	94.04	3.35	UNIQUE	0	3	0	135
Nucleoside-triphosphatase PFC_01625	PF0501	PF0501	20.83	72.28	UNIQUE	0	15	0	122
50S ribosomal protein L12	rpl12p	PF1994	11.22	71.96	UNIQUE	0	6	0	120
Probable Brix domain-containing ribosomal biogenesis protein	PF2010	PF2010	26.04	54.71	UNIQUE	0	11	0	117
50S ribosomal protein L6	rpl6	PF1808	20.83	56.52	UNIQUE	0	11	0	116
30S ribosomal protein S4	rps4p	PF1649	21.31	36.11	UNIQUE	0	11	0	115
30S ribosomal protein S2	rps2P	PF1640	22.97	42.57	UNIQUE	0	11	0	112
30S ribosomal protein S3	rps3p	PF1819	23.40	60.00	UNIQUE	0	13	0	112
50S ribosomal protein L15e	rpl15e	PF0876	22.56	35.05	UNIQUE	0	11	0	101
50S ribosomal protein L11	rpl11p	PF1991	17.61	58.54	UNIQUE	0	7	0	101
50S ribosomal protein L2	rpl2	PF1822	26.00	47.28	UNIQUE	0	12	0	99
Uncharacterized protein	PF1927	PF1927	23.42	48.54	UNIQUE	0	9	0	87
30S ribosomal protein S3Ae	rps3ae	PF2054	22.91	55.05	UNIQUE	0	14	0	85

50S ribosomal protein L5	rpl5	PF1811	21.20	39.78	UNIQUE	0	8	0	82
50S ribosomal protein L22P	rpl22p	PF1820	17.61	50.97	UNIQUE	0	8	0	80
Uncharacterized protein	PF0206	PF0206	30.52	6.55	UNIQUE	0	1	0	77
50S ribosomal protein L10e	rpl10e	PF1279	20.90	38.67	UNIQUE	0	8	0	75
50S ribosomal protein L15P	rpl15p	PF1802	16.37	57.14	UNIQUE	0	9	0	71
30S ribosomal protein S4e	rps4e	PF1812	28.11	46.50	UNIQUE	0	9	0	58
Ribosomal RNA small subunit methyltransferase Nep1	nep1	PF1262	25.92	39.91	UNIQUE	0	7	0	56
Cas1	cas1	PF1118	37.44	37.27	UNIQUE	0	11	0	46
Translation initiation factor 6	eif6	PF0377	24.56	40.09	UNIQUE	0	7	0	41
Uncharacterized protein	PF1201	PF1201	19.34	54.27	UNIQUE	0	7	0	38
30S ribosomal protein S12	rps12P	PF1559	16.41	34.01	UNIQUE	0	5	0	37
50S ribosomal protein L19e	rpl19e	PF1806	17.84	25.33	UNIQUE	0	4	0	36
AsnC family transcriptional regulator	PF2053	PF2053	17.19	34.46	UNIQUE	0	6	0	34
50S ribosomal protein L32e	rpl32e	PF1807	15.51	26.15	UNIQUE	0	3	0	30
Uncharacterized protein	PF1827	PF1827	31.77	14.34	UNIQUE	0	4	0	29
Inosine-5'-monophosphate dehydrogenase	PF1794	PF1794	21.32	28.34	UNIQUE	0	5	0	14
Exosome complex RNA-binding protein Csl4	PF0052	PF0052	21.94	17.26	UNIQUE	0	4	0	12
tRNA-splicing endonuclease	endA	PF0266	19.89	10.59	UNIQUE	0	2	0	12
30S ribosomal protein S19e	rps19e	PF1499	17.33	16.00	UNIQUE	0	2	0	11
30S ribosomal protein S19P	rps19p	PF1821	15.30	9.85	UNIQUE	0	2	0	11
30S ribosomal protein S8e	rps8e	PF1069	14.24	23.62	UNIQUE	0	3	0	9
Uncharacterized protein	PF1038	PF1038	23.55	15.12	UNIQUE	0	3	0	9
BtpA family protein	PF0860	PF0860	28.56	7.63	UNIQUE	0	2	0	9
Uncharacterized protein	PFC_09020	PFC_09020	18.20	18.71	UNIQUE	0	3	0	8
Uncharacterized protein	PF1498	PF1498	16.49	5.04	UNIQUE	0	1	0	7
Phosphorylase	PF1535	PF1535	97.63	4.41	UNIQUE	0	2	0	5
Uncharacterized protein	PF0051	PF0051	25.68	9.59	UNIQUE	0	2	0	5
Uncharacterized protein	PF1184	PF1184	28.40	10.83	UNIQUE	0	2	0	3

Uncharacterized protein	PFC_04 395	PFC_0 4395	45.23	6.98	UNIQUE	0	1	0	3
Arginase	PF0499	PF0499	27.26	3.80	UNIQUE	0	1	0	2
Uncharacterized protein	PFC_05 845	PFC_0 5845	18.54	17.79	0	1	0	17	0
Uncharacterized protein	PFC_04 845	PFC_0 4845	19.16	15.15	0	1	0	2	0
Cas1 A5D9 mAb Co-IP 200 mM NaCl SDS-PAGE Band Analysis (band #3)									
Annotation	Gene	Locus/ ORF	MW (kDa)	Total % Seq Coverage	Fold Change Spectra	Total Peptides NI	Total Peptides IM	Total Spectra NI	Total Spectra IM
Cas1	cas1	PF1118	37.44	91.93	UNIQUE	0	41	0	2385
C/D box methylation guide ribonucleoprotein complex aNOP56 subunit	PF0060	PF0060	46.74	73.88	UNIQUE	0	36	0	391
50S ribosomal protein L3	rpl3	PF1825	41.35	54.25	UNIQUE	0	22	0	295
Probable tRNA pseudouridine synthase B	truB	PF1785	38.52	73.24	UNIQUE	0	23	0	246
Uncharacterized protein	PF0797	PF0797	40.81	58.21	UNIQUE	0	20	0	196
Uncharacterized protein	PF0070	PF0070	36.82	46.33	UNIQUE	0	15	0	161
50S ribosomal protein L4	rpl4lp	PF1824	28.67	58.82	UNIQUE	0	15	0	147
50S ribosomal protein L10	rpl10	PF1993	37.09	40.41	UNIQUE	0	13	0	129
50S ribosomal protein L12	rpl12p	PF1994	11.22	71.03	UNIQUE	0	5	0	96
Translation initiation factor 2 subunit gamma	eif2g	PF1717	44.99	49.39	UNIQUE	0	17	0	94
Uncharacterized protein	PF0206	PF0206	30.52	6.55	UNIQUE	0	1	0	85
Uncharacterized protein	PF0496	PF0496	30.57	24.62	UNIQUE	0	10	0	83
Exosome complex component Rrp42	rrp42	PFC_0 7065	30.47	34.66	UNIQUE	0	9	0	80
Ornithine carbamoyltransferase	PF0594	PF0594	35.14	20.00	UNIQUE	0	3	0	69
Uncharacterized protein	PF1101	PF1101	49.03	33.03	UNIQUE	0	14	0	59
30S ribosomal protein S4e	rps4e	PF1812	28.11	41.98	UNIQUE	0	9	0	54
Exosome complex component Rrp41	rrp41	PF1568	27.81	38.40	UNIQUE	0	10	0	50
Replication factor A	PF2020	PF2020	41.00	29.33	UNIQUE	0	10	0	46
Exosome complex component Rrp4	rrp4	PFC_0 7075	29.50	40.53	UNIQUE	0	9	0	41
Uncharacterized protein	PF0213	PF0213	30.26	27.04	UNIQUE	0	6	0	35
Type 2 DNA topoisomerase 6 subunit A	top6A	PF1578	44.03	21.20	UNIQUE	0	9	0	33

Translation initiation factor 2 subunit alpha	eif2a	PF1140	31.88	22.91	UNIQUE	0	7	0	32
Beta-glucosidase	PF0442	PF0442	49.71	21.14	UNIQUE	0	9	0	30
Uncharacterized protein	PF2018	PF2018	31.12	23.79	UNIQUE	0	6	0	30
DNA polymerase II large subunit	polC	PF0019	143.06	3.56	UNIQUE	0	4	0	30
Uncharacterized protein	PFC_04875	PFC_04875	30.65	25.00	UNIQUE	0	7	0	29
RNA 3'-terminal phosphate cyclase	rtcA	PF1549	36.76	24.27	UNIQUE	0	6	0	27
Cas4-1	PF1119	PF1119	20.14	25.44	UNIQUE	0	5	0	27
Tryptophan synthase beta chain	trpB	PF1706	42.48	22.42	UNIQUE	0	9	0	25
RNA-processing protein	PF1580	PF1580	25.46	17.65	UNIQUE	0	4	0	22
Uncharacterized protein	PF1826	PF1826	29.23	22.31	UNIQUE	0	5	0	15
Acetyl-lysine deacetylase	lysK	PF1686	35.89	22.39	UNIQUE	0	6	0	14
Fibrillar-like rRNA/tRNA 2'-O-methyltransferase	flpA	PF0059	25.73	23.79	UNIQUE	0	5	0	12
30S ribosomal protein S3	rps3p	PF1819	23.40	29.05	UNIQUE	0	5	0	12
Uncharacterized protein	PF0099	PF0099	36.66	12.66	UNIQUE	0	5	0	12
tRNA/tRNA methyltransferase	PF0251	PF0251	44.99	11.56	UNIQUE	0	5	0	11
DNA repair and recombination protein RadA	radA	PF1926	38.36	9.74	UNIQUE	0	4	0	11
Uncharacterized protein	PF1027	PF1027	43.48	11.26	UNIQUE	0	4	0	9
DNA-directed RNA polymerase subunit A"	rpoA2	PF1562	44.36	7.56	UNIQUE	0	4	0	9
Proteasome-activating nucleotidase	pan	PF0115	44.76	4.55	UNIQUE	0	2	0	6
50S ribosomal protein L2	rpl2	PF1822	26.00	9.62	UNIQUE	0	2	0	5
L-isoaspartyl protein carboxyl methyltransferase	PF1896	PF1896	28.46	3.56	UNIQUE	0	1	0	4
Uncharacterized protein	PF1641	PF1641	37.68	3.83	UNIQUE	0	1	0	2
Probable cobyrinic acid synthase	cobQ	PF0301	54.49	2.07	0	1	0	22	0
Uncharacterized protein	PF1967	PF1967	47.43	7.26	0	1	0	2	0
Cas1 A5D9 mAb Co-IP 200 mM NaCl SDS-PAGE Band Analysis (band #4)									
Annotation	Gene	Locus/ ORF	MW (kDa)	Total % Seq Coverage	Fold Change Spectra	Total Peptides NI	Total Peptides IM	Total Spectra NI	Total Spectra IM
Putative tRNA(Met) cytidine acetyltransferase	PF0504	PF0504	94.37	15.32	UNIQUE	0	19	0	262

Cas1	cas1	PF1118	37.44	13.35	UNIQUE	0	5	0	136
Dna2-nam7 helicase family protein	PF0572	PF0572	74.52	5.50	UNIQUE	0	5	0	115
Phosphoenolpyruvate synthase	PF0043	PF0043	90.41	8.08	0.635714	10	7	140	89
S1p family ribosomal protein	PF0399	PF0399	83.14	10.95	UNIQUE	0	12	0	82
ATPase	PFC_0440	PFC_0440	68.14	9.80	UNIQUE	0	7	0	81
Hef nuclease	PFC_08485	PFC_08485	86.74	8.78	UNIQUE	0	9	0	63
DNA-directed RNA polymerase	PF1564	PF1564	126.91	7.79	UNIQUE	0	11	0	56
DNA-directed RNA polymerase	PFC_07045	PFC_07045	103.38	4.84	UNIQUE	0	7	0	48
DNA helicase	PFC_07780	PFC_07780	156.90	3.92	UNIQUE	0	7	0	46
Proteasome subunit alpha	psmA	PF1571	28.97	6.15	UNIQUE	0	2	0	35
Uncharacterized protein	PF1393	PF1393	74.43	7.91	UNIQUE	0	7	0	22
Uncharacterized protein	PF1119	PF1119	20.14	3.55	UNIQUE	0	1	0	21
DNA topoisomerase I	topA	PF0494	121.09	4.91	UNIQUE	0	5	0	9
DNA primase DnaG	dnaG	PF1725	51.02	1.77	UNIQUE	0	1	0	1

Table 2.5 Cas1 IgY Co-IP/shotgun MS Confirms the Immune Enrichment of Cas1

The table below illustrates results from shotgun MS performed on Cas1 IgY pAb IPs via in-gel sample preparation (see Figure 2.5). The in-gel analysis correlates with the experiment shown in Figure 2.4 panel C performed with TPF41 lysates. Only a single sample was tested for each IP condition in this case (pre-immune control (PI) and anti-Cas1 immune (I)). Table information is setup as indicated in Table 2.2 with the difference of a pre-immune control (PI) compared to a non-immune control antibody.

Cas1 pAb Co-IP 300 mM NaCl Shotgun MS Analysis (in-gel)								
Annotation	Locus/ORF	MW (kDa)	Total % Seq Coverage	Fold Change Spectra	Total Peptides PI	Total Peptides IM	Total Spectra PI	Total Spectra IM
chromosome segregation protein	PF1167	103.76	7.26	UNIQUE	0	5	0	98
isoleucyl-tRNA ligase	PF1096	125.66	6.29	UNIQUE	0	5	0	44
reverse gyrase	PF0495	139.88	3.38	UNIQUE	0	4	0	55
Cas1	PF1118	37.44	35.40	3.30	12	14	164	542
Cas4-1	PF1119	20.14	26.04	1.53	6	5	36	55
tryptophan synthase subunit alpha	PF1705	27.46	10.89	1.08	3	3	12	13
cell division protein CDC48	PF0963	94.04	17.08	1.07	13	7	58	62
glutamine synthetase	PF0450	50.13	19.13	0.85	9	8	53	45
aminohydrolase	PF0818	36.99	13.73	0.77	3	3	13	10
tryptophan synthase subunit beta	PF1706	42.48	43.56	0.50	17	8	121	61
elongation factor 1-alpha	PF1375	47.59	21.03	0.45	9	4	47	21
hypothetical protein	PF0797	40.81	8.93	0.33	4	4	15	5
3-methyl-2-oxobutanoate hydroxymethyltransferase	PF1143	31.85	36.04	0.29	11	7	241	71
phosphoenolpyruvate synthase	PF0043	90.41	9.42	0.00	8	0	56	0
glutamate synthase subunit alpha	PF0205	54.40	15.34	0.00	5	0	28	0
hypothetical protein	PF0204	42.63	13.28	0.00	6	0	42	0
methionine aminopeptidase	PF0541	32.80	15.93	0.00	5	0	34	0
hypothetical protein	PF1142	29.95	13.48	0.00	4	0	28	0
acetyl-lysine deacetylase	PF1686	35.89	4.29	0.00	1	0	1	0
type II secretion system protein	PF0994	129.90	2.72	0.00	1	0	13	0
hypothetical protein	PF0224	23.08	8.82	0.00	1	0	52	0

**Table 2.6 Shotgun Mass Spectrometry Confirms Presence of Cas1, Cas4-1, Alba, and Cas2
in Anaerobic Co-IPs**

The table below illustrates results from shotgun mass spectrometry analysis of A5D9 Cas1 mAb immunoprecipitations using lysates overexpressing Cas1, Cas2, Cas4-1, and Cas4-2 (TPF68) at 300 mM NaCl (gel not shown). To better recreate cellular conditions and avoid oxygen-based reduction of Cas4 iron-sulfur clusters, the experiment was performed in an anaerobic chamber to limit oxygen exposure. This experiment also had the benefit of a better representation of natural CRISPR adaptation conditions via the expression of Cas4-2. As in past analyses, spectra associated with the non-immune control antibody are compared to the immune as a measure of specificity (see legend of Table 2.2 for details).

Cas1 A5D9 mAb Anaerobic Co-IP 300 mM NaCl Shotgun MS Analysis (in-gel)								
Sequence Name	Locus/ORF	MW (kDa)	% Seq Coverage	Fold Change Spectra	NI Peptides	IM Peptides	Spectral Count NI	Spectral Count IM
Cas1	PF1118	37.44	54.97	UNIQUE	0	26	0	604
DNA helicase	PF0504	94.37	43.26	UNIQUE	0	35	0	303
proteasome-activating nucleotidase	PF0115	44.76	55.05	UNIQUE	0	22	0	111
50S ribosomal protein L3P	PF1825	41.35	42.47	UNIQUE	0	18	0	103
H/ACA RNA-protein complex component Cbf5p	PF1785	38.52	42.06	UNIQUE	0	18	0	129
fibrillarin	PF0059	25.73	67.40	UNIQUE	0	16	0	137
translation initiation factor IF-2	PF1717	44.99	30.90	UNIQUE	0	10	0	71
hypothetical protein	PF0490	27.82	47.74	UNIQUE	0	11	0	58
hypothetical protein	PF1867	33.11	35.99	UNIQUE	0	10	0	49
30S ribosomal protein S5	PF1804	26.52	52.12	UNIQUE	0	12	0	43
Cas4-1	PF1119	20.14	37.87	UNIQUE	0	9	0	175
acidic ribosomal protein P0	PF1993	37.09	28.32	UNIQUE	0	9	0	46
30S ribosomal protein S13	PF1650	16.89	38.51	UNIQUE	0	9	0	75
50S ribosomal protein L30	PF1803	17.68	65.16	UNIQUE	0	10	0	67
50S ribosomal protein L4P	PF1824	28.67	43.14	UNIQUE	0	10	0	48
hypothetical protein	PF0341	38.04	29.50	UNIQUE	0	9	0	46
30S ribosomal protein S19e	PF1499	17.33	36.67	UNIQUE	0	8	0	51
ATPase	PF1041	68.14	13.12	UNIQUE	0	8	0	46
50S ribosomal protein L1P	PF1992	23.82	32.87	UNIQUE	0	7	0	53
30S ribosomal protein S7	PF1558	24.62	37.67	UNIQUE	0	8	0	59
DNA repair and recombination protein RadA	PF1926	38.36	20.63	UNIQUE	0	6	0	26
50S ribosomal protein L13	PF1645	16.28	36.62	UNIQUE	0	6	0	42
hypothetical protein	PF1927	23.42	33.98	UNIQUE	0	6	0	30
enolase	PF0215	46.79	17.44	UNIQUE	0	6	0	30
50S ribosomal protein L5	PF1811	21.20	26.88	UNIQUE	0	5	0	39
nol1-nop2-sun family nucleolar protein IV	PF0666	52.16	11.06	UNIQUE	0	5	0	18
50S ribosomal protein L12	PF1994	11.22	57.01	UNIQUE	0	4	0	28
exosome complex exonuclease Rrp41	PF1568	27.81	25.20	UNIQUE	0	6	0	21
30S ribosomal protein S4	PF1649	21.31	26.11	UNIQUE	0	5	0	28
50S ribosomal protein L11	PF1991	17.61	43.90	UNIQUE	0	6	0	23
50S ribosomal protein L18	PF1805	23.15	35.47	UNIQUE	0	6	0	56
hypothetical protein	PF2018	31.12	20.82	UNIQUE	0	5	0	21
hypothetical protein	PF1615	108.72	5.72	UNIQUE	0	5	0	22
50S ribosomal protein L2	PF1822	26.00	23.43	UNIQUE	0	5	0	31
50S ribosomal protein L7Ae	PF1367	13.37	23.58	UNIQUE	0	5	0	34
ribosomal biogenesis protein	PF2010	26.04	28.25	UNIQUE	0	6	0	27
hypothetical protein	PF1625	14.85	45.45	UNIQUE	0	5	0	29

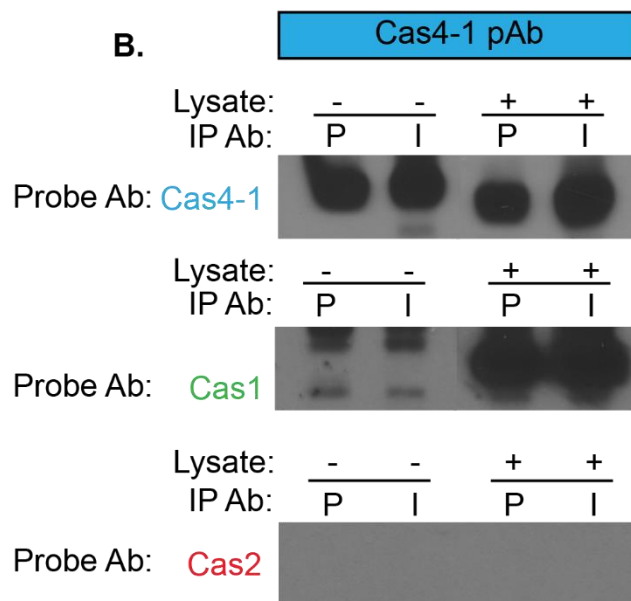
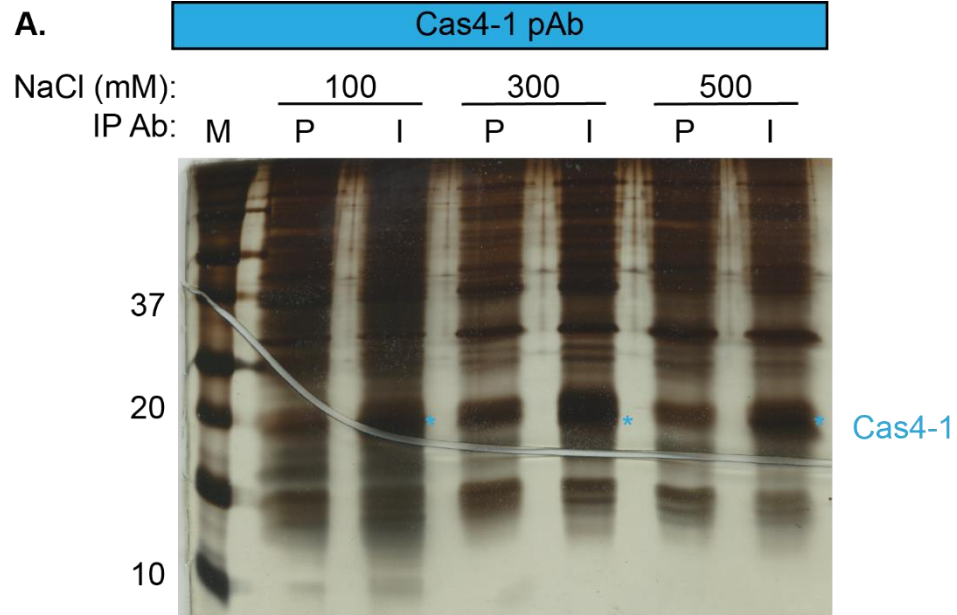
RNA-binding protein FAU-1	PF0022	53.62	13.86	UNIQUE	0	6	0	21
replication factor A	PF2020	41.00	12.29	UNIQUE	0	4	0	22
pyruvate formate-lyase activating enzyme-like protein	PF1397	40.36	18.68	UNIQUE	0	5	0	16
50S ribosomal protein L14e	PF0819	8.94	42.17	UNIQUE	0	4	0	32
hypothetical protein	PF0797	40.81	14.12	UNIQUE	0	5	0	19
translation initiation factor IF-6	PF0377	24.56	29.07	UNIQUE	0	5	0	19
50S ribosomal protein L24	PF1813	14.30	29.75	UNIQUE	0	5	0	24
50S ribosomal protein L6	PF1808	20.83	25.00	UNIQUE	0	4	0	19
DNA topoisomerase I	PF0494	121.09	4.43	UNIQUE	0	5	0	13
50S ribosomal protein L22	PF1820	17.61	26.45	UNIQUE	0	5	0	34
30S ribosomal protein S4e	PF1812	28.11	20.58	UNIQUE	0	4	0	13
hypothetical protein	PF1059	45.21	14.58	UNIQUE	0	5	0	14
30S ribosomal protein S15	PF2056	18.57	25.32	UNIQUE	0	4	0	40
50S ribosomal protein L15e	PF0876	22.56	15.98	UNIQUE	0	4	0	30
30S ribosomal protein S8e	PF1069	14.24	23.62	UNIQUE	0	3	0	18
proteasome subunit alpha	PF1571	28.97	15.77	UNIQUE	0	4	0	27
DNA-directed RNA polymerase subunit B	PF1564	126.91	3.85	UNIQUE	0	4	0	15
DNA topoisomerase VI subunit B	PF1579	64.37	8.63	UNIQUE	0	4	0	15
exosome complex RNA-binding protein Rrp4	PF1569	29.50	18.18	UNIQUE	0	4	0	11
hypothetical protein	PF1826	29.23	18.46	UNIQUE	0	4	0	20
H/ACA RNA-protein complex component Gar1	PF1791	11.45	35.05	UNIQUE	0	4	0	24
30S ribosomal protein S3Ae	PF2054	22.91	16.67	UNIQUE	0	4	0	16
aspartate carbamoyltransferase catalytic subunit	PF0599	34.68	12.01	UNIQUE	0	4	0	11
50S ribosomal protein L14	PF1814	15.21	21.99	UNIQUE	0	3	0	12
hypothetical protein	PF0560	33.88	11.33	UNIQUE	0	3	0	12
30S ribosomal protein S12	PF1559	16.41	23.81	UNIQUE	0	3	0	13
glutamine synthetase	PF0450	50.13	7.97	UNIQUE	0	3	0	10
30S ribosomal protein S9	PF1644	15.26	25.19	UNIQUE	0	3	0	9
30S ribosomal protein S2	PF1640	22.97	14.85	UNIQUE	0	3	0	13
30S ribosomal protein S11	PF1648	14.70	20.44	UNIQUE	0	3	0	18
50S ribosomal protein L19e	PF1806	17.84	16.00	UNIQUE	0	3	0	23
nol1-nop2-sun family nucleolar protein II	PF1265	50.39	7.85	UNIQUE	0	3	0	9
hypothetical protein	PF1027	43.48	8.85	UNIQUE	0	3	0	8
30S ribosomal protein S17e	PF1491	7.90	35.82	UNIQUE	0	3	0	19
30S ribosomal protein S10P	PF1376	11.72	28.43	UNIQUE	0	3	0	11
RNA-processing protein	PF1580	25.46	13.57	UNIQUE	0	3	0	25
50S ribosomal protein L15	PF1802	16.37	19.73	UNIQUE	0	3	0	20
50S ribosomal protein L10e	PF1279	20.90	13.81	UNIQUE	0	3	0	11
Cas2	PF1117	9.95	58.80	UNIQUE	0	2	0	18
translation initiation factor IF-2	PF0481	16.21	17.86	UNIQUE	0	2		9

ADP-dependent glucokinase	PF0312	51.22	7.03	UNIQUE	0	3	0	15
NDP-sugar synthase	PF0868	47.17	8.96	UNIQUE	0	3	0	10
30S ribosomal protein S6e	PF0488	13.95	20.00	UNIQUE	0	2	0	12
ornithine carbamoyltransferase	PF0594	35.14	8.25	UNIQUE	0	2	0	5
seryl-tRNA ligase	PF1204	53.15	6.37	UNIQUE	0	2	0	6
30S ribosomal protein S24e	PF0253	11.65	37.37	UNIQUE	0	3	0	9
hydrogenase maturation protein HypF	PF0559	87.34	4.14	UNIQUE	0	3	0	9
translation initiation factor IF-2	PF1140	31.88	8.73	UNIQUE	0	2	0	5
30S ribosomal protein S19	PF1821	15.30	9.85	UNIQUE	0	2	0	13
acetyl-CoA acetyltransferase	PF0973	40.94	7.73	UNIQUE	0	3	0	11
DNA topoisomerase VI subunit A	PF1578	44.03	6.28	UNIQUE	0	2	0	7
30S ribosomal protein S3	PF1819	23.40	10.95	UNIQUE	0	2	0	4
50S ribosomal protein L18e	PF1646	13.71	23.33	UNIQUE	0	2	0	9
2-ketoisovalerate ferredoxin oxidoreductase subunit beta	PF0968	34.73	8.04	UNIQUE	0	2	0	4
histone a1	PF1831	7.38	37.31	UNIQUE	0	2	0	7
H/ACA RNA-protein complex component Nop10p	PF1141	7.20	30.00	UNIQUE	0	2	0	15
RNA 3'-terminal-phosphate cyclase	PF1549	36.76	5.85	UNIQUE	0	2	0	9
GTP-binding protein	PF0998	40.64	6.48	UNIQUE	0	2	0	7
acetyl-lysine deacetylase	PF1686	35.89	7.36	UNIQUE	0	2	0	4
histone a2	PF1722	7.26	28.36	UNIQUE	0	2	0	11
hypothetical protein	PF1101	49.03	4.56	UNIQUE	0	2	0	3
aldehyde ferredoxin oxidoreductase	PF0464	73.88	3.83	UNIQUE	0	2	0	5
50S ribosomal protein LX	PF0376	9.30	29.87	UNIQUE	0	2	0	7
hypothetical protein	PF1038	23.55	9.76	UNIQUE	0	2	0	7
ribulose biphosphate carboxylase	PF1156	47.30	7.86	UNIQUE	0	2	0	5
AsnC family transcriptional regulator	PF0739	24.11	9.22	UNIQUE	0	2	0	11
bifunctional phosphomannomutase/phosphoglucosyltransferase	PF0588	49.58	5.71	UNIQUE	0	2	0	8
50S ribosomal protein L31e	PF0378	11.03	23.16	UNIQUE	0	2	0	7
50S ribosomal protein L23	PF1823	9.85	24.42	UNIQUE	0	2	0	9
aldehyde ferredoxin oxidoreductase	PF1203	68.71	3.07	UNIQUE	0	2	0	5
exosome complex RNA-binding protein Rrp42	PF1567	30.47	6.14	UNIQUE	0	2	0	10
C/D box methylation guide ribonucleoprotein complex aNOP56 subunit	PF0060	46.74	58.96	46.83	2	29	6	281
DNA/RNA-binding protein alba	PF1881	10.36	77.42	19.83	2	8	6	119
DNA primase	PF1725	51.02	26.33	16.67	2	11	3	50
hypothetical protein	PF1142	29.95	37.45	8.50	2	9	4	34
hypothetical protein	PF1191	38.76	6.09	7.00	2	2	2	14
bacteriochlorophyll synthase, 43 kDa subunit	PF0454	40.76	13.78	6.50	3	5	4	26

AsnC family transcriptional regulator	PF2053	17.19	31.08	5.33	3	5	3	16
elongation factor 1-alpha	PF1375	47.59	17.29	4.67	3	7	9	42
phosphoenolpyruvate synthase	PF0043	90.41	31.82	4.10	15	24	41	168
cell division protein CDC48	PF0963	94.04	9.92	3.92	7	7	13	51
thermosome, single subunit	PF1974	59.93	50.09	3.92	19	31	51	200
NTPase	PF0501	20.83	16.85	3.50	2	3	2	7
thiamine biosynthesis/tRNA modification protein ThiI	PF1288	40.69	6.34	2.67	2	2	3	8
pyruvate ferredoxin oxidoreductase subunit beta	PF0965	36.22	23.26	1.87	5	6	15	28
oxidoreductase	PF0276	53.71	12.22	1.86	3	5	7	13
replication factor C small subunit	PF0093	97.71	2.35	1.75	2	2	4	7
trehalose/maltose binding protein	PF1739	50.34	7.33	0.73	3	3	11	8
thymidine phosphorylase	PF1607	54.29	6.36	0.00	2	0	5	0
hypothetical protein	PF0807	64.19	3.02	0.00	2	0	8	0
flavoprotein	PF0751	47.35	4.11	0.00	2	0	4	0
DNA polymerase sliding clamp	PF0983	27.97	8.03	0.00	2	0	6	0
2-amino-3-ketobutyrate coenzyme A ligase	PF0265	43.77	8.86	0.00	3	0	3	0

Figure 2.7 Cas4-1 Co-IPs Isolate Cas4-1 but Non-Specifically Precipitate Cas1

A) Anti-Cas4-1 IgY co-immunoprecipitation (Co-IP) 15% SDS-PAGE. Co-immunoprecipitations were performed against Cas4-1 using the polyclonal IgY antibody from Figure 2.2. Each Co-IP was subject to three different salt wash conditions at 100, 300, and 500 mM salt using lysates from TPF41. The background of the assay is controlled for via pre-immune antibodies (P) purified from hens before immunization compared to the anti-Cas4-1 immune antibody (I) made after hen immunization. While the background is high, an immune-specific enrichment around 20 kDa molecular weight of Cas4-1 (blue asterisks) is present. This experiment was later subjected to tandem MS where the protein's presence was confirmed in each wash condition. B) Western blotting performed on anti-Cas4-1 immunoprecipitations. In a separate experiment, Cas4-1 immunoprecipitations performed in a similar fashion to panel A were subject to western blotting. IPs were performed at 300 mM NaCl. Due to anticipated background reactivity between co-eluted IP antibody and western secondary antibody, immunoprecipitations were performed + and – lysates to ascertain the background created by the antibody alone in western blotting. While there is little to no antibody-associated background from the IP detected in the blot at the molecular weights of Cas2 and Cas1 (middle and bottom), there is a strong cross-reactive band corresponding to the Co-IP antibody light chain at the molecular weight of Cas4-1 (top).



**Table 2.7 Cas4-1 Co-IP/MS Confirms Isolation of Cas4-1 but the Non-Specific
Identification of Cas1**

The table below represents results from shotgun Cas4-1 pAb immunoprecipitations performed with both in-gel and in-solution techniques (see Figure 2.5). In-gel analysis correlates with the salt analysis experiment shown in Figure 2.7 panel A using lysates overexpressing Cas1, Cas2, and Cas4-1 (TPF41) (100 and 300 mM NaCl). Table information is setup as indicated in Table 2.2 with the difference of a pre-immune control (PI) compared to a non-immune control antibody.

Cas4-1 pAb Co-IP 100 mM NaCl Shotgun MS Analysis (in-gel)								
Annotation	Locus/ORF	MW (kDa)	Total % Seq Coverage	Fold Change Spectra	Total Peptides PI	Total Peptides IM	Total Spectra NI	Total Spectra IM
hypothetical protein	PF0206	30.52	6.55	UNIQUE	0	1	0	57
hypothetical protein	PF1615	108.69	6.44	1.71	6	3	21	36
Cas4-1	PF1119	20.14	29.59	1.07	6	8	72	77
hypothetical protein	PF1142	29.95	10.11	1.00	4	2	25	25
Cas1	PF1118	37.44	49.38	0.42	15	12	153	65
3-methyl-2-oxobutanoate hydroxymethyltransferase	PF1143	31.85	36.75	0.40	13	9	179	72
glutamate synthase subunit alpha	PF0205	54.40	41.04	0.36	18	9	134	48
ribosomal protein s6 modification protein	PF1682	30.82	42.86	0.31	9	3	74	23
acetyl-lysine deacetylase	PF1686	35.89	27.91	0.22	8	3	36	8
glutamine synthetase	PF0450	50.13	41.23	0.18	15	5	80	14
tryptophan synthase subunit beta	PF1706	42.48	49.74	0.13	16	8	180	24
aspartate carbamoyltransferase catalytic subunit	PF0599	34.68	55.52	0.13	18	6	119	15
cell division protein CDC48	PF0963	94.04	51.14	0.00	43	0	310	0
acetylornithine/acetyl-lysine aminotransferase	PF1685	41.06	47.97	0.00	15	0	63	0
hypothetical protein	PF0204	42.63	30.89	0.00	13	0	85	0
phospho-sugar mutase	PF1729	26.46	45.73	0.00	11	0	51	0
translation initiation factor IF-2	PF1717	44.99	35.04	0.00	12	0	30	0
elongation factor 1-alpha	PF1375	47.59	22.20	0.00	8	0	35	0
cytochrome-c3 hydrogenase subunit gamma	PF0892	33.01	34.25	0.00	8	0	51	0
phosphoenolpyruvate synthase	PF0043	90.41	9.67	0.00	8	0	39	0
30S ribosomal protein S5	PF1804	26.52	31.36	0.00	7	0	13	0
tryptophan synthase subunit alpha	PF1705	27.46	25.40	0.00	7	0	41	0
methionine aminopeptidase	PF0541	32.80	23.39	0.00	7	0	27	0
hypothetical protein	PF1393	74.43	7.44	0.00	4	0	14	0
aminohydrolase	PF0818	36.99	20.60	0.00	4	0	21	0
enolase	PF0215	46.79	19.07	0.00	5	0	14	0
NDP-sugar synthase	PF0868	47.17	19.61	0.00	6	0	16	0
sulfhydrogenase subunit alpha	PF0894	48.27	13.79	0.00	6	0	12	0
bifunctional phosphoribosyl-AMP cyclohydrolase/phosphoribosyl-ATP pyrophosphatase	PF1664	24.20	24.40	0.00	5	0	15	0
30S ribosomal protein S3	PF1819	23.40	21.43	0.00	5	0	16	0
fibrillarlin	PF0059	25.73	19.38	0.00	4	0	13	0
isopentenyl pyrophosphate isomerase	PF0856	41.18	11.50	0.00	4	0	12	0
alpha-amylase	PF0272	76.21	8.78	0.00	5	0	12	0
proteasome-activating nucleotidase	PF0115	44.76	7.58	0.00	3	0	21	0
pyrolysine	PF0287	154.36	2.65	0.00	3	0	8	0
acetylornithine deacetylase	PF1185	39.76	10.00	0.00	3	0	12	0

hypothetical protein	PF0902	38.64	9.06	0.00	2	0	6	0
N-ethylammelline chlorohydrolase	PF1538	46.63	9.31	0.00	3	0	5	0
hypothetical protein	PF0381	33.80	10.78	0.00	3	0	5	0
hypothetical protein	PF1326	18.53	28.22	0.00	4	0	13	0
DNA primase	PF1725	51.02	7.96	0.00	3	0	15	0
thermosome, single subunit	PF1974	59.93	5.10	0.00	3	0	8	0
3-isopropylmalate dehydratase large subunit	PF1679	41.13	7.11	0.00	2	0	4	0
N-acetyl-gamma-glutamyl-phosphate reductase	PF1683	37.31	6.06	0.00	2	0	4	0
ATPase	PF1041	68.14	4.15	0.00	2	0	5	0
hypothetical protein	PF1438	69.86	3.79	0.00	2	0	7	0
aldehyde ferredoxin oxidoreductase	PF0464	73.88	3.37	0.00	2	0	5	0
30S ribosomal protein S2	PF1640	22.97	12.87	0.00	2	0	6	0
proteasome subunit alpha	PF1571	28.97	8.08	0.00	2	0	7	0
3-dehydroquinate synthase	PF1691	37.34	5.67	0.00	2	0	4	0
ATP dependent DNA ligase	PF0353	43.67	6.07	0.00	2	0	2	0
tRNA/rRNA methyltransferase	PF0251	44.99	4.77	0.00	2	0	40	0
50S ribosomal protein L1P	PF1992	23.82	6.02	0.00	1	0	2	0
50S ribosomal protein L12	PF1994	11.22	13.08	0.00	1	0	9	0
cleavage and polyadenylation specificity factor protein	PF1405	73.47	1.70	0.00	1	0	3	0
C/D box methylation guide ribonucleoprotein complex aNOP56 subunit	PF0060	46.74	2.99	0.00	1	0	5	0
ribosomal protein L15e	PF0876	22.56	6.70	0.00	2	0	4	0
hypothetical protein	PF0883	28.18	11.55	0.00	1	0	3	0

Cas4-1 pAb Co-IP 100 mM NaCl Shotgun MS Analysis (in-solution)

Annotation	Locus/ORF	MW (kDa)	Total % Seq Coverage	Fold Change Spectra	Total Peptides PI	Total Peptides IM	Total Spectra NI	Total Spectra IM
phosphoenolpyruvate synthase	PF0043	90.41	25.09	UNIQUE	0	17.0	0	78
argininosuccinate synthase	PF0207	45.89	30.73	UNIQUE	0	9.0	0	73
elongation factor 1-alpha	PF1375	47.59	20.33	UNIQUE	0	6.0	0	35
Cas4-1	PF1119	20.14	37.87	UNIQUE	0	7.0	0	56
ferredoxin	PF1909	7.28	88.06	UNIQUE	0	3.0	0	47
translation initiation factor IF-2	PF1717	44.99	17.03	UNIQUE	0	6.0	0	10
pyruvate/ketoisovalerate ferredoxin oxidoreductase subunit gamma	PF0971	19.99	23.24	UNIQUE	0	3.0	0	15
50S ribosomal protein L2	PF1822	26.00	22.59	UNIQUE	0	3.0	0	12
pyruvate ferredoxin oxidoreductase subunit beta	PF0965	36.22	15.11	UNIQUE	0	3.0	0	13
DNA/RNA-binding protein albA	PF1881	10.36	36.56	UNIQUE	0	3.0	0	17
thermosome, single subunit	PF1974	59.93	6.38	UNIQUE	0	3.0	0	12
ATPase, vanadate-sensitive	PF1399	65.66	6.45	UNIQUE	0	3.0	0	19
CasI	PF1118	37.44	10.56	UNIQUE	0	3.0	0	11
proteasome subunit alpha	PF1571	28.97	11.92	UNIQUE	0	4.0	0	17

hypothetical protein	PF1109	124.79	2.47	UNIQUE	0	2.0	0	9
30S ribosomal protein S24e	PF0253	11.65	39.39	UNIQUE	0	3.0	0	18
50S ribosomal protein L6	PF1808	20.83	14.13	UNIQUE	0	2.0	0	8
pyruvate ferredoxin oxidoreductase subunit alpha	PF0966	44.13	9.09	UNIQUE	0	3.0	0	8
glutamine synthetase	PF0450	50.13	13.21	UNIQUE	0	3.0	0	16
hypothetical protein	PF1191	38.76	7.25	UNIQUE	0	2.0	0	6
transcription factor	PF1295	21.28	7.85	UNIQUE	0	1.0	0	10
L-threonine 3-dehydrogenase	PF0991	37.78	10.06	UNIQUE	0	2.0	0	22
50S ribosomal protein L14	PF1814	15.21	10.64	UNIQUE	0	1.0	0	4
hypothetical protein	PF0496	30.57	9.85	UNIQUE	0	1.0	0	3
30S ribosomal protein S3	PF1819	23.40	6.67	UNIQUE	0	1.0	0	4
superoxide reductase	PF1281	14.30	15.32	UNIQUE	0	1.0	0	4
cell division protein CDC48	PF1882	89.13	1.26	UNIQUE	0	1.0	0	1
NDP-sugar synthase	PF0868	47.17	3.63	7.00	1	8.0	3	21
glutamate synthase subunit alpha	PF0205	54.40	4.38	6.67	2	7.0	6	40
50S ribosomal protein L12	PF1994	11.22	13.08	6.33	1	6.0	12	76
tryptophan synthase subunit beta	PF1706	42.48	5.41	2.56	2	5.0	9	23
hypothetical protein	PF0204	42.63	13.01	1.75	4	6.0	12	21
cell division protein CDC48	PF0963	94.04	18.64	1.58	18	12.0	126	199
s1p family ribosomal protein	PF0399	83.14	7.70	0.00	4	0.0	34	0
chromosome segregation ATPase	PF1843	134.91	4.67	0.00	4	0.0	7	0
glutamyl-tRNA ligase	PF1753	66.50	8.22	0.00	3	0.0	21	0
large helicase-like protein	PF1504	104.72	2.52	0.00	2	0.0	30	0
glutamyl-tRNA(Gln) amidotransferase subunit D	PF1461	48.59	4.79	0.00	1	0.0	1	0
Cas4-1 pAb Co-IP 300 mM NaCl Shotgun MS Analysis (in-gel)								
Annotation	Locus/ORF	MW (kDa)	Total % Seq Coverage	Fold Change Spectra	Total Peptides PI	Total Peptides IM	Total Spectra PI	Total Spectra IM
hypothetical protein	PF0206	30.52	6.55	UNIQUE	0	1	0	130
argininosuccinate synthase	PF0207	45.89	22.68	UNIQUE	0	7	0	25
elongation factor 1-alpha	PF1375	47.59	10.05	UNIQUE	0	4	0	12
fibrillarin	PF0059	25.73	16.30	UNIQUE	0	3	0	17
phosphoenolpyruvate synthase	PF0043	90.41	5.02	UNIQUE	0	4	0	6
methionine aminopeptidase	PF0541	32.80	10.85	UNIQUE	0	3	0	8
hypothetical protein	PF0797	40.81	6.63	UNIQUE	0	2	0	13
hypothetical protein	PF1918	43.87	8.51	UNIQUE	0	3	0	22
hypothetical protein	PF1183	14.39	13.28	UNIQUE	0	3	0	65
cell division protein CDC48	PF0963	94.04	3.82	26.75	3	21	8	214
glutamate synthase subunit alpha	PF0205	54.40	5.78	3.32	3	8	19	63
glutamine synthetase	PF0450	50.13	12.07	2.80	4	7	10	28
Cas4-1	PF1119	20.14	21.30	1.54	4	8	56	86
hypothetical protein	PF1142	29.95	10.11	0.76	4	3	21	16

Cas1	PF1118	37.44	32.30	0.66	10	10	166	110
tryptophan synthase subunit beta	PF1706	42.48	40.98	0.60	13	10	83	50
3-methyl-2-oxobutanoate hydroxymethyltransferase	PF1143	31.85	27.92	0.54	8	11	101	55
hypothetical protein	PF0204	42.63	14.63	0.23	5	5	26	6
DNA helicase	PF0085	156.88	5.47	0.00	3	0	13	0
cell division protein CDC48	PF1882	89.13	2.14	0.00	1	0	3	0
Cas4-1 pAb Co-IP 300 mM NaCl Shotgun MS Analysis (in-solution)								
Annotation	ORF	MW (kDa)	Total % Seq Coverage	Fold Change Spectra	Total Peptides PI	Total Peptides IM	Total Spectra PI	Total Spectra IM
argininosuccinate synthase	PF0207	45.89	24.39	UNIQUE	0	7	0	27
elongation factor 1-alpha	PF1375	47.59	13.55	UNIQUE	0	3	0	12
pyruvate/ketoisovalerate ferredoxin oxidoreductase subunit gamma	PF0971	19.99	17.84	UNIQUE	0	2	0	6
hypothetical protein	PF1587	32.17	8.45	UNIQUE	0	2	0	23
ATPase, vanadate-sensitive	PF1399	65.66	4.96	UNIQUE	0	2	0	7
transcription-associated protein TFIIIS	PF1855	48.70	6.37	UNIQUE	0	1	0	1
cell division protein CDC48	PF0963	94.04	3.82	3.63	3	13	40	145
hypothetical protein	PF1119	20.14	17.16	2.29	3	4	17	39
glutamate synthase subunit alpha	PF0205	54.40	6.97	2.27	3	7	15	34
50S ribosomal protein L12	PF1994	11.22	13.08	2.00	1	1	2	4
hypothetical protein	PF0204	42.63	19.78	1.64	6	8	25	41
ribonucleotide-diphosphate reductase subunit alpha	PF0440	200.79	3.05	0.60	4	2	20	12
ferredoxin	PF1909	7.28	88.06	0.00	3	0	21	0
hypothetical protein	PF0630	6.85	69.49	0.00	3	0	4	0
hypothetical protein	PF0537	90.32	4.68	0.00	3	0	6	0
cell division protein CDC48	PF1882	89.13	1.76	0.00	1	0	2	0
Cas4-1 pAb Co-IP 500 mM NaCl Shotgun MS Analysis (in-solution)								
Annotation	Locus/ORF	MW (kDa)	Total % Seq Coverage	Fold Change Spectra	Total Peptides NI	Total Peptides IM	Total Spectra NI	Total Spectra IM
cell division protein CDC48	PF0963	94.04	17.08	UNIQUE	0	14	0	140
Cas4-1	PF1119	20.14	54.44	UNIQUE	0	11	0	88
NDP-sugar synthase	PF0868	47.17	15.50	UNIQUE	0	5	0	17
hypothetical protein	PF1142	29.95	10.11	UNIQUE	0	4	0	54
elongation factor 1-alpha	PF1375	47.59	15.65	UNIQUE	0	4	0	10
prolyl endopeptidase	PF0825	70.76	10.39	UNIQUE	0	4	0	110
ski2-like helicase	PF0677	82.56	7.22	UNIQUE	0	4	0	12
formaldehyde:ferredoxin oxidoreductase wor5	PF1480	69.81	6.25	UNIQUE	0	3	0	109
tryptophan synthase subunit beta	PF1706	42.48	3.35	UNIQUE	0	1	0	3
pyruvate ferredoxin oxidoreductase subunit beta	PF0965	36.22	5.14	UNIQUE	0	1	0	3
ribonucleotide-diphosphate reductase subunit alpha	PF0440	200.79	5.06	0.00	10	0	59	0
hypothetical protein	PF0204	42.63	11.92	0.00	4	0	5	0

hypothetical protein	PF1915	45.44	5.82	0.00	3	0	19	0
aminotransferase	PF0293	36.87	8.18	0.00	4	0	40	0
ADP-dependent glucokinase	PF0312	51.22	7.03	0.00	3	0	3	0
hydrogenase maturation protein HypF	PF0559	87.34	2.72	0.00	2	0	13	0
iron(III) ABC transporter permease	PF1748	61.03	2.59	0.00	1	0	12	0
met-10+ protein	PF0500	32.39	4.68	0.00	1	0	1	0
cystathionine gamma-synthase	PF1266	41.16	2.19	0.00	1	0	13	0

Figure 2.8 Band Excision Mass Spectrometry of Cas4-1 pAb Immunoprecipitations

Immunoprecipitated material from Cas4-1 immunoprecipitations (500 mM NaCl from Figure 2.7 panel A) using lysates overexpressing Cas1, Cas2, and Cas4-1 (TPF41) was subject to mass spectrometry analyses. Band excision to identify Cas4-1 was performed at the expected molecular weight of Cas4-1 (~20 kDa). Pre-immune control antibody lanes (PI) and anti-Cas1 antibody lanes (I) are indicated. The area that was cut is shown (black box). Bands correlating to immune-specific Cas proteins are indicated by colored asterisks. The mass spectrometry results from the excised region are indicated in the next table (Table 2.7).

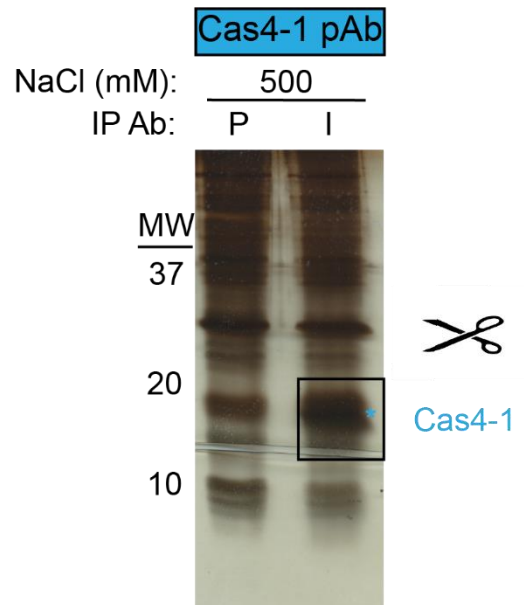


Table 2.8 Cas4-1 is Identified at the Expected Molecular Weight in Cas4-1

Immunoprecipitations

Band excision Co-IP mass spectrometry was performed as indicated in Figure 2.8 with band cutting at Cas4-1's expected molecular weight (20 kDa). Samples were generated from the same experiment as shown in Figure 2.7 panel A (500 mM NaCl). Proteins are listed as originally detailed in Table 2.2.

Cas4-1 pAb Co-IP 500 mM NaCl Band Cutting MS Analysis								
Annotation	ORF	MW (kDa)	Total % Seq Coverage	Fold Change Spectra	Total Peptides PI	Total Peptides IM	Total Spectra PI	Total Spectra IM
chromosome segregation protein	PF1167	103.76	6.92	UNIQUE	0	5	0	175
DNA-directed RNA polymerase subunit B	PF1564	126.91	8.06	UNIQUE	0	5	0	36
cell division protein CDC48	PF1882	89.13	3.27	UNIQUE	0	2	0	10
phosphoglycolate phosphatase	PF1419	25.47	6.49	UNIQUE	0	1	0	31
Cas4-1	PF1119	20.14	37.28	3.63	8	16	100	363
hypothetical protein	PF1438	69.86	3.79	2.75	2	4	4	11
CRISPR-associated protein Cas1	PF1118	37.44	22.98	1.86	7	8	35	65
hypothetical protein	PF1987	15.49	21.90	1.50	2	3	10	15
tryptophan synthase subunit alpha	PF1705	27.46	35.08	1.04	8	6	27	28
C/D box methylation guide ribonucleoprotein complex aNOP56 subunit	PF0060	46.74	14.93	1.00	6	6	21	21
hypothetical protein	PF0560	33.88	17.00	0.93	4	5	14	13
glutamine synthetase	PF0450	50.13	19.82	0.87	9	5	38	33
DNA/RNA-binding protein alba	PF1881	10.36	22.58	0.73	2	1	11	8
fibrillarin	PF0059	25.73	60.79	0.64	13	6	50	32
3-methyl-2-oxobutanoate hydroxymethyltransferase	PF1143	31.85	49.12	0.61	16	16	921	558
tryptophan synthase subunit beta	PF1706	42.48	39.43	0.58	14	12	182	105
thermosome, single subunit	PF1974	59.93	27.69	0.33	13	7	63	21
superoxide reductase	PF1281	14.30	33.87	0.29	2	1	14	4
elongation factor 1-alpha	PF1375	47.59	29.44	0.14	11	5	79	11
phosphoenolpyruvate synthase	PF0043	90.41	24.24	0.10	19	4	99	10
cell division protein CDC48	PF0963	94.04	51.61	0.10	44	10	229	23
hypothetical protein	PF1142	29.95	34.46	0	9	0	68	0
ATPase, vanadate-sensitive	PF1399	65.66	11.90	0	6	0	25	0
glutamate dehydrogenase	PF1602	47.07	15.00	0	6	0	20	0
hypothetical protein	PF0547	50.71	9.82	0	4	0	20	0
3-hydroxyisobutyrate dehydrogenase	PF0716	31.34	16.91	0	4	0	8	0
ATPase	PF1041	68.14	8.97	0	5	0	8	0
NDP-sugar synthase	PF0868	47.17	9.93	0	4	0	12	0
pyruvate ferredoxin oxidoreductase subunit alpha	PF0966	44.13	16.41	0	4	0	11	0
oligopeptide ABC transporter	PF1209	71.54	5.95	0	4	0	16	0
hypothetical protein	PF0797	40.81	7.49	0	2	0	12	0
alpha-amylase	PF0272	76.21	2.77	0	2	0	8	0
aldehyde:ferredoxin oxidoreductase	PF1961	69.31	2.41	0	1	0	128	0
cytochrome-c3 hydrogenase subunit gamma	PF0892	33.01	5.14	0	1	0	1	0

methionine aminopeptidase	PF0541	32.80	9.49	0	0	3	0	11
---------------------------	--------	-------	------	---	---	---	---	----

Figure 2.9 Cas2 Immunoprecipitations Co-Precipitate Cas1 in Co-IP/Western Blotting

A) Anti-Cas2 IgY Co-IP salt curve on a 15% SDS-PAGE. Co-immunoprecipitations were performed against Cas2 using the pAb IgY antibody from Figure 2.2. Each Co-IP was subject to three different salt wash conditions at 100, 300, and 500 mM salt using overexpressed lysates from TPF41. The background of the assay is controlled for via pre-immune antibodies (P) purified from hens before immunization compared to the anti-Cas2 immune antibody (I) made after immunization. A band enriched in the immune corresponding to Cas2's molecular weight of 10 kDa was observed in each condition (red asterisks). B) Cas2 Co-IP/western blot against Cas2 and Cas1. A separate Co-IP salt titration experiment was performed under similar conditions to demonstrate Cas2 and Cas1 isolation. Washes ranged from 100 mM to 1M NaCl. IP material was subject to western blotting against Cas2 and Cas1. Cas4-1 probing was not performed because previous data were inconclusive (Figure 2.7B). Cas2 was present in all immune conditions, while Cas1 was found enriched in immunes at 100 and 300 mM NaCl.

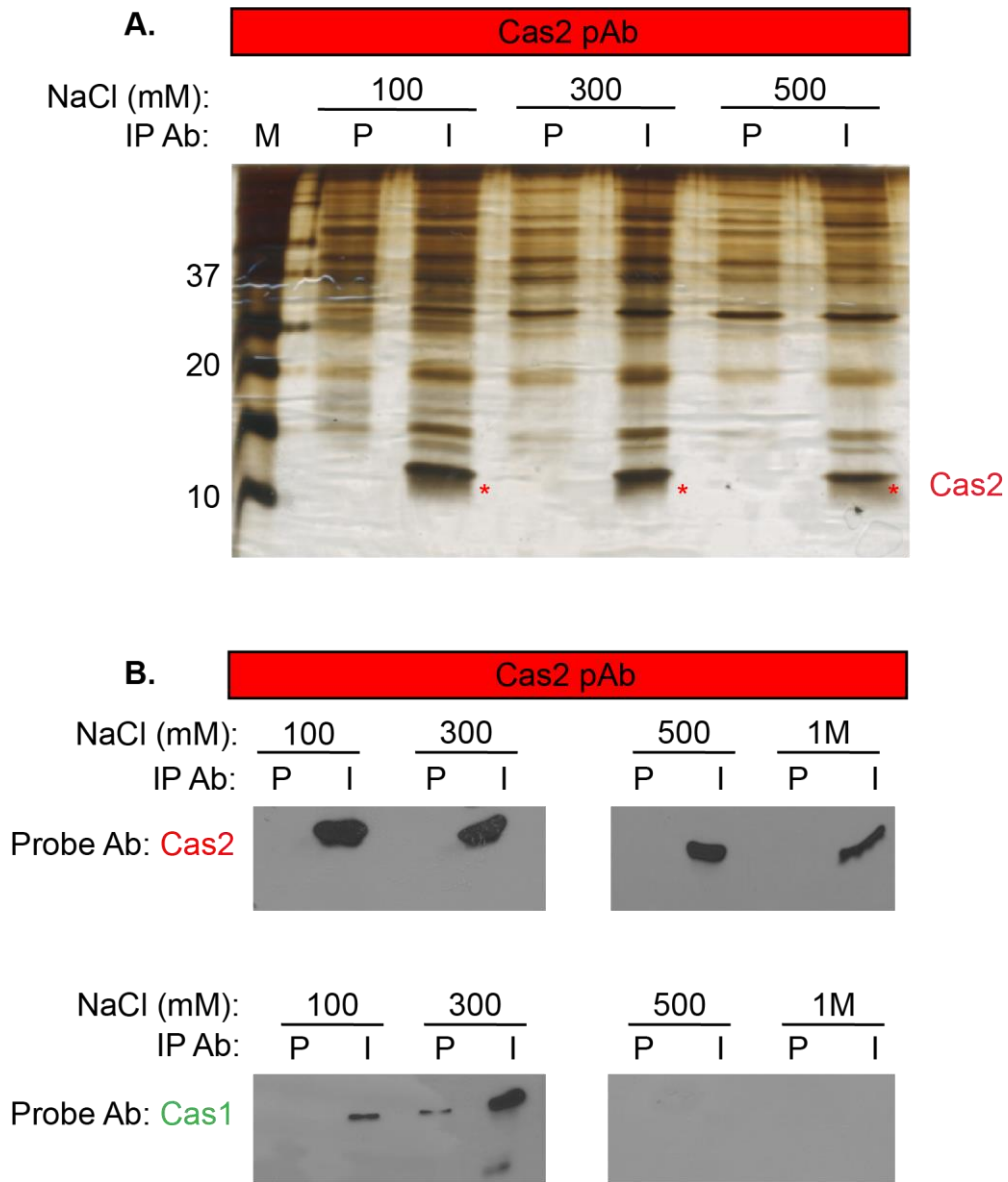


Table 2.9 Cas2 IP/shotgun Mass Spectrometry Illustrates the Immune Enrichment of Cas2, but Non-Specific Isolation of Cas1 and Cas4-1

The table below represents shotgun mass spectrometry results from Cas2 IgY IPs performed with in-gel sample preparation (see Figure 2.5). This analysis correlates with the salt analysis experiment that is shown in Figure 2.9 panel A performed at 100 and 300 mM NaCl with lysates overexpressing Cas1, Cas2, and Cas4-1 (TPF41). Table information is setup as indicated in Table 2.2 with the difference of a pre-immune control (PI) compared to a non-immune control antibody.

Cas2 pAb Co-IP 100 mM NaCl Shotgun MS Analysis (in-gel)								
Annotation	ORF/Locus	MW (kDa)	Total % Seq Coverage	Fold Change Spectra	Total Peptides PI	Total Peptides IM	Total Spectra PI	Total Spectra IM
phospho-sugar mutase	PF1729	26.46	33.33	UNIQUE	0	8	0	57
oxidoreductase	PF0276	53.71	16.98	UNIQUE	0	7	0	41
hypothetical protein	PF0204	42.63	15.72	UNIQUE	0	7	0	71
Cas2	PF1117	9.95	32.94	UNIQUE	0	3	0	59
acetylornithine/acetyl-lysine aminotransferase	PF1685	41.08	17.62	UNIQUE	0	6	0	73
methionine aminopeptidase	PF0541	32.80	12.20	UNIQUE	0	3	0	20
hypothetical protein	PF1142	29.95	10.86	UNIQUE	0	3	0	76
alpha-amylase	PF0272	76.24	1.54	UNIQUE	0	1	0	14
cytochrome-c3 hydrogenase subunit gamma	PF0892	33.01	5.14	6.80	1	5	5	34
N-ethylammeline chlorohydrolase	PF1538	46.66	4.30	6.75	1	4	8	54
aminohydrolase	PF0818	36.99	9.55	1.78	2	3	9	16
ribosomal protein s6 modification protein	PF1682	30.82	16.48	1.76	4	7	38	67
translation initiation factor IF-2	PF1717	44.99	15.57	1.72	5	8	25	43
Cas4-1	PF1119	20.14	27.81	1.70	6	8	98	167
CRISPR-associated protein Cas1	PF1118	37.44	24.22	1.57	9	12	125	196
3-methyl-2-oxobutanoate hydroxymethyltransferase	PF1143	31.85	25.44	1.37	8	8	101	138
thermosome, single subunit	PF1974	59.93	10.38	1.26	7	6	47	59
glutamate synthase subunit alpha	PF0205	54.40	15.94	1.16	8	10	75	87
hypothetical protein	PF1393	74.43	11.08	0.97	6	4	38	37
transcription initiation factor IIB	PF1377	34.07	17.00	0.81	4	5	180	146
tryptophan synthase subunit beta	PF1706	42.48	26.55	0.75	10	13	251	187
acetyl-lysine deacetylase	PF1686	35.89	31.60	0.71	8	6	93	66
cell division protein CDC48	PF0963	94.04	19.00	0.70	16	12	117	82
glutamine synthetase	PF0450	50.13	22.78	0.60	11	8	154	92
NDP-sugar synthase	PF0868	47.17	23.00	0.51	7	4	39	20
elongation factor 1-alpha	PF1375	47.59	25.70	0.21	12	4	126	26
tryptophan synthase subunit alpha	PF1705	27.46	32.66	0.00	9	0	103	0
DNA-directed RNA polymerase subunit B	PF1564	126.91	5.82	0.00	6	0	34	0
hypothetical protein	PF1615	108.72	7.17	0.00	6	0	27	0
asparaginyl-tRNA ligase	PF0155	50.06	12.44	0.00	5	0	44	0
DNA-directed RNA polymerase subunit A'	PF1563	103.38	8.13	0.00	5	0	23	0
hypothetical protein	PF1454	68.66	7.65	0.00	3	0	21	0
ornithine carbamoyltransferase	PF0594	35.14	9.21	0.00	1	0	22	0

Cas2 pAb Co-IP 300 mM NaCl Shotgun MS Analysis (in-gel)								
Annotation	Locus/ORF	MW (kDa)	Total % Seq Coverage	Fold Change Spectra	Total Peptides PI	Total Peptides IM	Total Spectra PI	Total Spectra IM
glutamate synthase subunit alpha	PF0205	54.40	32.67	UNIQUE	0	16	0	181
cell division protein CDC48	PF0963	94.04	14.93	UNIQUE	0	13	0	127
Cas2	PF1117	9.95	64.71	UNIQUE	0	5	0	118
methionine aminopeptidase	PF0541	32.80	22.03	UNIQUE	0	6	0	56
acetyl-lysine deacetylase	PF1686	35.89	19.33	UNIQUE	0	6	0	43
hypothetical protein	PF0797	40.81	18.16	UNIQUE	0	6	0	24
hypothetical protein	PF1142	29.95	13.48	UNIQUE	0	4	0	125
hypothetical protein	PF1326	18.54	17.18	UNIQUE	0	3	0	30
aminohydrolase	PF0818	36.99	9.55	UNIQUE	0	2	0	5
hypothetical protein	PF1438	69.86	6.47	UNIQUE	0	3	0	29
C/D box methylation guide ribonucleoprotein complex aNOP56 subunit	PF0060	46.74	5.72	UNIQUE	0	3	0	21
NDP-sugar synthase	PF0868	47.17	10.17	UNIQUE	0	3	0	11
alpha-amylase	PF0272	76.24	2.93	UNIQUE	0	2	0	32
hypothetical protein	PF1454	68.66	3.40	UNIQUE	0	1	0	9
hypothetical protein	PF1987	15.49	14.60	UNIQUE	0	1	0	2
oxidoreductase	PF0276	53.71	5.80	2.33	3	9	33	77
hypothetical protein	PF0204	42.63	10.84	2.22	5	8	37	82
elongation factor 1-alpha	PF1375	47.59	5.84	1.76	3	5	37	65
DNA helicase	PF0504	94.37	6.25	1.55	5	8	38	59
CRISPR-associated protein Cas1	PF1118	37.44	17.08	1.47	7	18	339	497
3-methyl-2-oxobutanoate hydroxymethyltransferase	PF1143	31.85	18.37	1.35	8	14	286	386
Cas4-1	PF1119	20.14	27.81	0.91	6	6	132	120
glutamine synthetase	PF0450	50.13	18.91	0.89	10	10	129	115
tryptophan synthase subunit beta	PF1706	42.48	28.61	0.84	13	15	320	270
phosphoenolpyruvate synthase	PF0043	90.41	6.36	0.53	5	3	47	25

tryptophan synthase subunit alpha	PF1705	27.46	20.56	0.52	6	6	86	45
thermosome, single subunit	PF1974	59.93	12.02	0.35	8	6	95	33
hypothetical protein	PF0547	50.71	8.71	0.00	4	0	52	0
hypothetical protein	PF1931	27.62	5.26	0.00	1	0	4	0
N-ethylmeline chlorohydrolase	PF1538	46.66	2.39	0.00	1	0	30	0
hypothetical protein	PF0737	5.26	25.00	0.00	1	0	29	0

Figure 2.10 Band Excision Mass Spectrometry of Cas2 pAb Immunoprecipitations

Immunoprecipitated material from the Cas2 polyclonal antibody immunoprecipitation (500 mM NaCl from Figure 2.9 panel A using lysates from TPF41 (Cas1, Cas2, Cas4-1 overexpression) was subject to mass spectrometry analyses. Band excision to identify Cas2 was performed at the protein's expected molecular weight (~10 kDa). Pre-immune control antibody lanes (PI) and anti-Cas1 antibody lanes (I) are indicated. The area that was cut is shown (black box). Bands correlating to immune-specific Cas proteins are indicated by colored asterisks. The mass spectrometry from the excised region is indicated in the next table (Table 2.10).

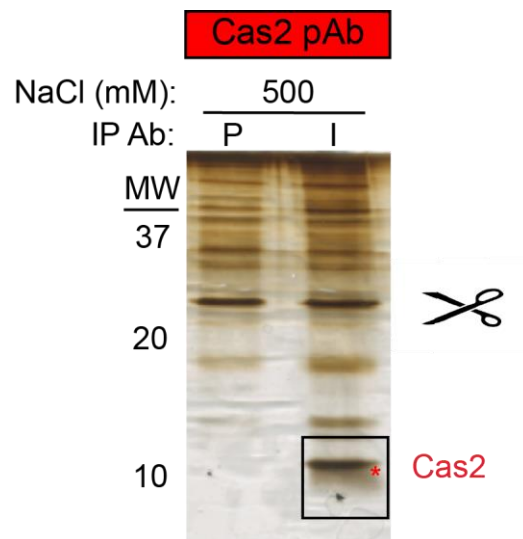


Table 2.10 Cas2 is Identified at the Expected Molecular Weight in Cas2 pAb IP/MS

Analyses

Band excision Co-IP/MS was performed as indicated in Figure 2.10 with band cutting at expected molecular weight of Cas2 (~10 kDa). Samples were generated from the same experiment as shown in Figure 2.9 panel A (500 mM NaCl, TPF41 lysates). The protein identification table is listed as originally detailed in Table 2.2.

Cas2 pAb Co-IP 500 mM NaCl Band Cutting MS Analysis								
Annotation	Locus/ORF	MW (kDa)	Total % Seq Coverage	Fold Change Spectra	Total Peptides PI	Total Peptides IM	Total Spectra PI	Total Spectra IM
HTH transcription regulator	PF0340	12.70	17.59	UNIQUE	0	2	0	23
Cas2	PF1117	9.95	8.24	UNIQUE	0	2	0	163
pyruvate ferredoxin oxidoreductase subunit delta	PF0967	11.99	11.43	UNIQUE	0	1	0	11
glutamine synthetase	PF0450	50.13	2.05	UNIQUE	0	1	0	12
histone a2	PF1722	7.26	13.43	UNIQUE	0	1	0	12
30S ribosomal protein S28e	PF1368	8.08	15.49	UNIQUE	0	1	0	9
30S ribosomal protein S17e	PF1491	7.90	14.93	UNIQUE	0	1	0	6
.1n hypothetical protein	PF0222	7.25	11.67	UNIQUE	0	1	0	19
cytochrome-c3 hydrogenase subunit gamma	PF0892	33.01	6.51	UNIQUE	0	1	0	2
hypothetical protein	PF1932	31.11	3.72	UNIQUE	0	1	0	6
glutamate synthase subunit alpha	PF0205	54.40	1.79	UNIQUE	0	1	0	5
3-methyl-2-oxobutanoate hydroxymethyltransferase	PF1143	31.85	14.13	0.05	6	1	315	17
thermosome, single subunit	PF1974	59.93	15.66	0	11	0	273	0
tryptophan synthase subunit beta	PF1706	42.48	16.24	0	8	0	216	0
cell division protein CDC48	PF0963	94.04	7.05	0	7	0	78	0
tryptophan synthase subunit alpha	PF1705	27.46	17.34	0	5	0	104	0
CRISPR-associated protein Cas1	PF1118	37.44	14.60	0	5	0	82	0
phosphoenolpyruvate synthase	PF0043	90.41	3.92	0	3	0	33	0
AsnC family transcriptional regulator	PF2053	17.19	12.84	0	2	0	19	0

Figure 2.11 Size Exclusion Chromatography of *P. furiosus* Cell Lysate Suggests Cas1 and Cas4-1 May Interact

TPF41 (overexpression of Cas1, Cas2, Cas4-1) cell lysates were subject to size separation via gel filtration on a Sephacryl, 16/60 S200 HR column. Prior to the separation of TPF41 lysates, a solution of five standard proteins (thyroglobulin 670 kDa, γ -globulin 158 kDa, ovalbumin 44 kDa, myoglobin 17 kDa, and vitamin B12 1.35 kDa) was run to standardize molecular weight. The elution peak fraction of each of these standard proteins as indicated by A680 UV plot are labeled (black arrows). After TPF 41 lysates containing adaptation Cas proteins had been run and eluted, fractions were subjected to western blotting probing for Cas1 (A), Cas2 (B), and Cas4-1 (C). Peak fractions for each protein are indicated by the corresponding colored arrows: Cas1 green, Cas2 red, Cas4-1 blue.

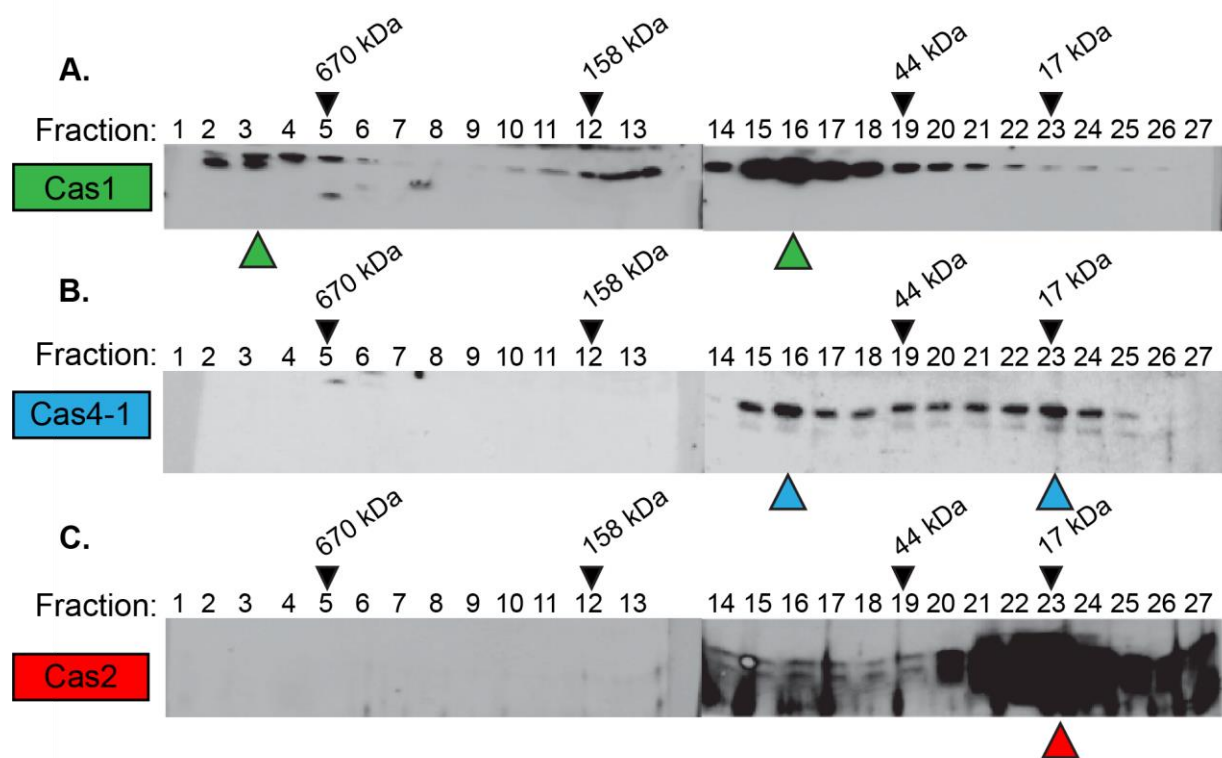


Figure 2.12 Experimental Design of Cas1 Nickel Chromatography

Pyrococcus furiosus lysates were subjected to anti-polyhistidine tag pulldowns with Cas1 polyhistidine-tagged at its C-terminus. Lysates overexpressing adaptation Cas proteins (Cas1, Cas2, Cas4-1, and Cas4-2) (TPF68) were incubated under anaerobic conditions with nickel resin in a batch-style purification. Post incubation, the resin was subjected to five one mL (10 column volume) washes with increasing amounts of imidazole to remove contaminant proteins. Three washes were repeated with 20mM imidazole, followed by two washes at 50 and 100 mM imidazole. Nickel affinity purified proteins were then eluted twice with high imidazole concentration at 50 mM (200 ul/ 2 column volumes). Next, purified proteins were subject to both SDS-PAGE and western blotting analysis to examine protein content.

Input: Pfu lysates, Cas1 10Xhis tagged + Cas overexpressed

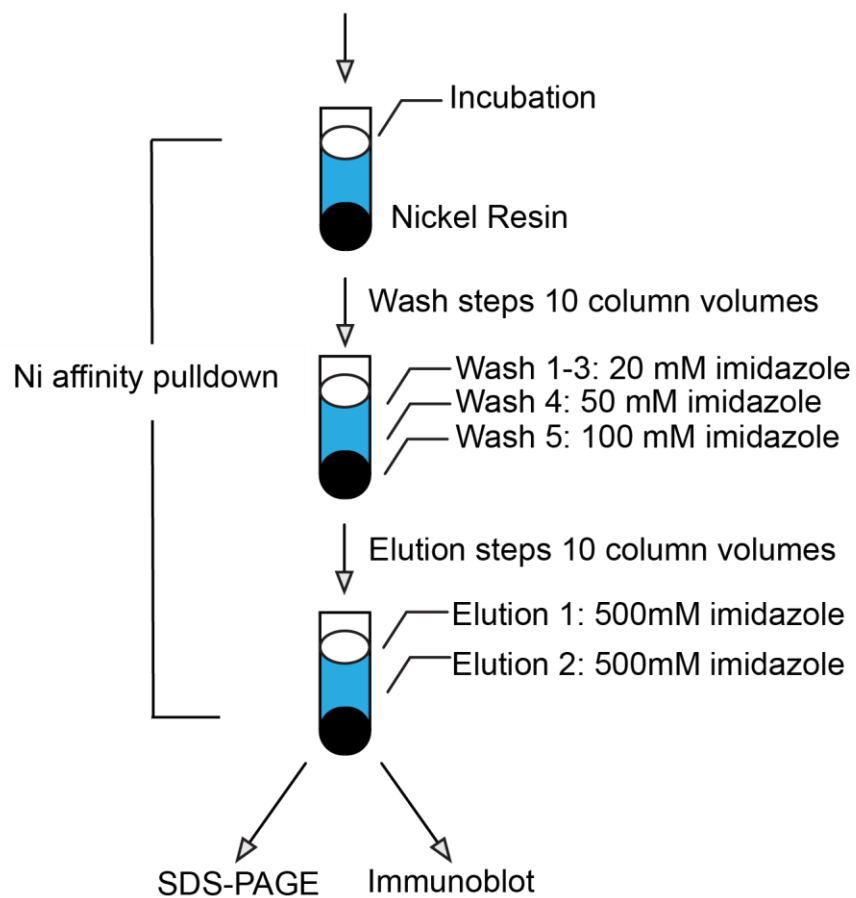
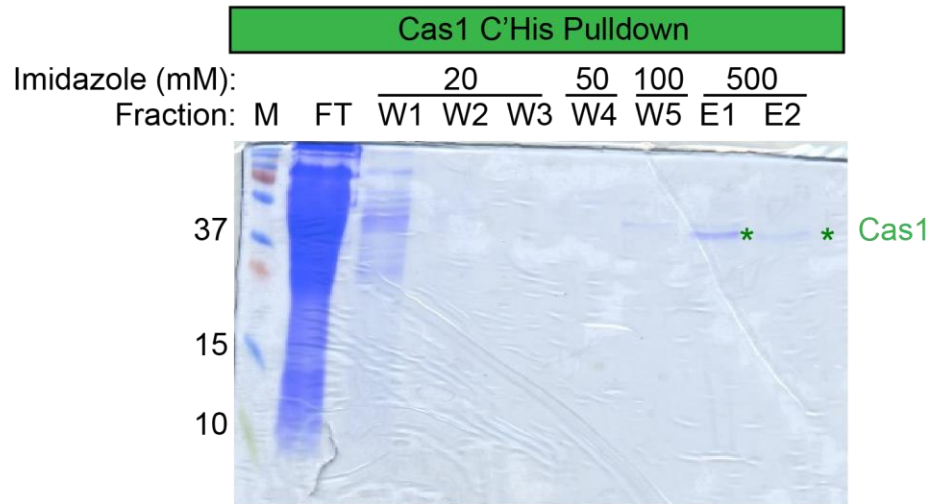


Figure 2.13 Nickel chromatography Demonstrates a Similar Pattern of Proteins as Cas1 Co-IP

Pyrococcus furiosus cell lysates with polyhistidine-tagged Cas1 (overexpression of Cas1, Cas2, Cas4-1, and Cas4-2) (TPF68 transformed with pEAW002) were subjected to nickel affinity purification. A) Coomassie-stained SDS-PAGE of Cas1 purification fractions. Fractions range from flowthrough (FT), washes (W1-W5), and elutions (E1/E2) according to Figure 2.12. The amount of imidazole used per wash or elution is indicated above each fraction. B) Silver-stained SDS-PAGE of Cas1 polyhistidine affinity purification. The same fractions as illustrated in panel A were further tested using silver staining in panel B. Additionally, a Cas1 tagless control lysate (TPF68) (-) was run side-by-side the previous fractions where Cas1 was tagged (+) in this assay to determine background proteins. Protein content present in wash fractions is present in the top gel while proteins present in elution conditions are found in the bottom gel. Colored asterisks detail expected Cas protein band presence enriched in elution fractions, while black asterisks illustrate major non-Cas protein bands. Cas protein identity was later tested in follow-up western blot experiments (Figure 2.14)

A.



B.

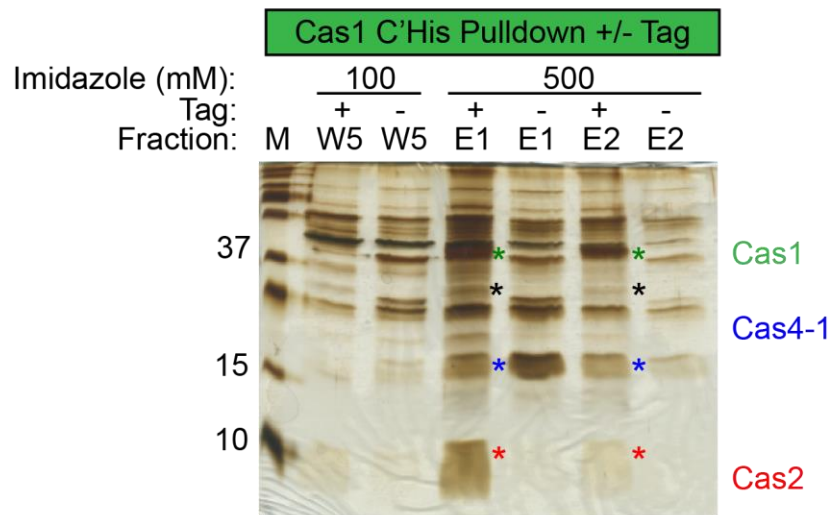
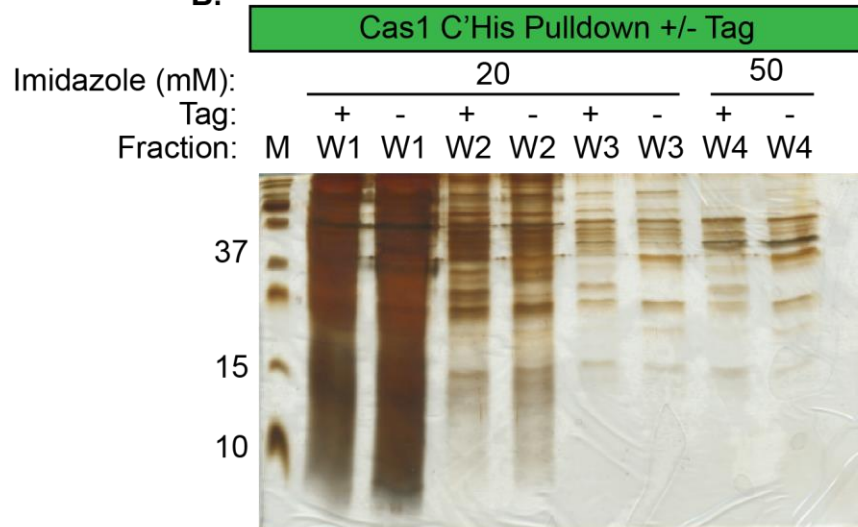


Figure 2.14 Affinity Chromatography/Western Blotting Demonstrates Cas1 and Cas4-1 Association

Western blotting experiment mirroring Figure 2.13. Cas1 with a C-terminal polyhistidine tag, overexpressed in *Pyrococcus furiosus* with Cas2, Cas4-1, and Cas4-2, was subject to nickel affinity purification and western blotting. A) total protein (T), soluble (S), flow-through (FT), wash (W1-W5), and elution (E1-E2) fractions were subject to Cas1 western blotting. The Cas1 target protein is found enriched in elution fractions. The amount of imidazole used to wash and elute fractions is indicated. B) The same fractions from panel A were subjected to western blot probing for Cas2 and Cas4-1 proteins for evidence of protein co-elution. Additionally, a tagless Cas1 lysate control (TPF68) was included to ascertain background of Cas2 and Cas4-1 proteins as in Figure 2.13B.


A. Cas1 C'His Pulldown

Imidazole (mM): 20 50 100 500
 Fraction: M T S FT W1 W2 W3 W4 W5 E1 E2
 Probe Ab: Cas1




B. Cas1 C'His Pulldown +/- Tag


Imidazole (mM): 20
 Tag: - + - + - + - +
 Fraction: M FT FT W1 W1 W2 W2 W3 W3
 Probe Ab: Cas4-1



Imidazole (mM): 50 100 500
 Tag: - + - + - + - +
 Fraction: W4 W4 W5 W5 E1 E1 E2 E2
 Probe Ab: Cas4-1



Imidazole (mM): 20
 Tag: - + - + - + - +
 Fraction: M FT FT W1 W1 W2 W2 W3 W3
 Probe Ab: Cas2



Imidazole (mM): 50 100 500
 Tag: - + - + - + - +
 Fraction: W4 W4 W5 W5 E1 E1 E2 E2
 Probe Ab: Cas2




Figure 2.15 Cas1 Associates with Nucleic Acids

Nucleic acid material eluted from Cas1 chromatin immunoprecipitation using A) A5D9 and B) A6.C11 anti-Cas1 mAbs run on 15% denaturing PAGE post-stained with SYBR Gold. ChIP was performed with TPF66 Pfu lysate (Cas1 and Cas2 overexpression, Δ Cas4-1) or TPF51 (Δ Cas1, Cas2, Cas4-1 overexpression) to give the background of the assay. Eluted nucleic acids were treated with RNase A to identify nucleic acid as DNA or RNA.

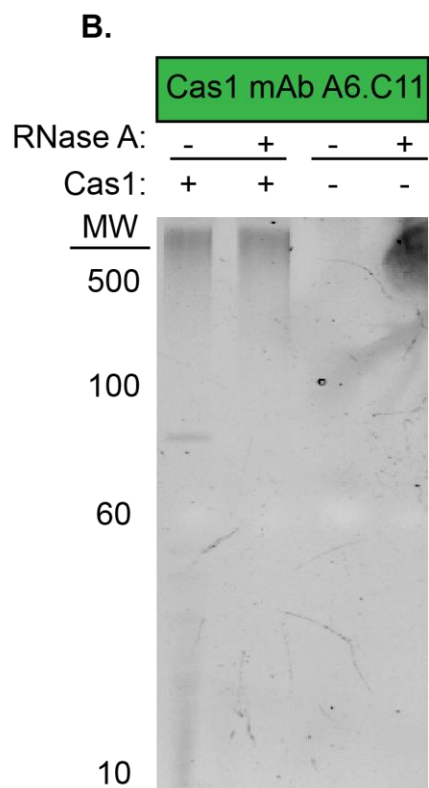
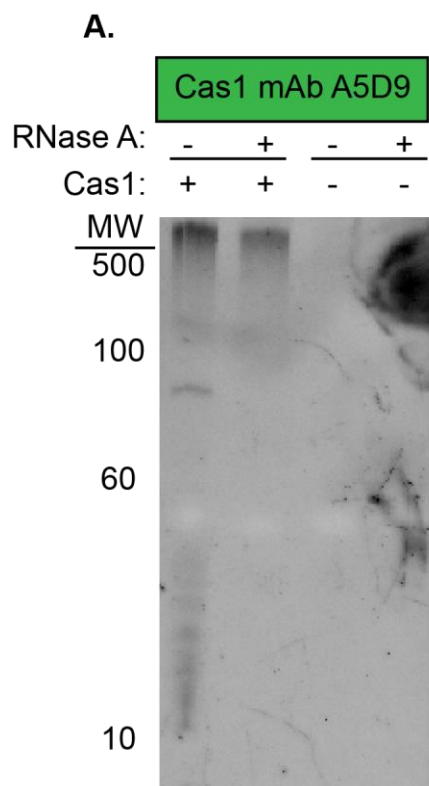


Table 2.11 Cas1 mAb Co-IP/MS summary results: Cas1 Associates with Cas, Ribosomal, and DNA-Binding Proteins

The table below shows summary data from proteins identified from all Cas1 mAb Co-IP experiments represented categorically. Proteins were grouped according to their function: CRISPR-Cas, DNA binding & topology, ribosomal and translation, other, and unannotated. All associated spectra and peptide counts for Cas1 mAb Co-IPs shown in Tables 2.2, 2.3, 2.4, and 2.6 were added together to give total counts. Only identifications that were immune-specific (minimum two-fold immune spectral count enrichment) are shown. Thus, specific proteins across ALL experiments are shown according to predicted protein function. As all proteins shown are specific, they are sorted by the number of immune spectra. Overall, Cas1 in this analysis was found to be associated with Cas2, Cas4-1, and a bevy of ribosomal, DNA binding, and other proteins of unknown function.

Cas Proteins Cas1 mAb Co-IP Summary								
Annotation	Locus/ORF	MW (kDa)	Fold Change Peptide	Fold Change Spectra	Total Peptides NI	Total Peptides IM	Total Spectra NI	Total Spectra IM
Cas1	PF1118	37.43	17.09	89.57	11	188	65	5822
Cas4-1	PF1119	20.14	62	540	2	124	5	2700
Cas2	PF1117	9.95	UNIQUE	UNIQUE	0	2	0	18
DNA Binding & Topology Proteins								
Annotation	Locus/ORF	MW (kDa)	Fold Change Peptide	Fold Change Spectra	Total Peptides NI	Total Peptides IM	Total Spectra NI	Total Spectra IM
DNA/RNA-binding protein Alba	PF1881	10.36	9.75	107.58	4	39	12	1291
DNA Helicase	PF0504	94.37	UNIQUE	UNIQUE	0	93	0	764
AsnC family transcriptional regulator	PF2053	17.19	8.00	51.00	3	24	3	153
DNA-directed RNA polymerase	PFC_07045	103.38	UNIQUE	UNIQUE	0	12	0	138
Histone a2	PF1722	7.26	UNIQUE	UNIQUE	0	7	0	119
Dna2-nam7 helicase family protein	PF0572	74.52	UNIQUE	UNIQUE	0	5	0	115
Histone a1	PF1831	7.38	UNIQUE	UNIQUE	0	6	0	78
DNA-directed RNA polymerase	PF1564	126.91	UNIQUE	UNIQUE	0	17	0	74
Replication factor A	PF2020	41.00	UNIQUE	UNIQUE	0	14	0	68
Hef nuclease	PFC_08485	86.74	UNIQUE	UNIQUE	0	9	0	63
DNA primase DnaG	PF1725	51.02	6.00	17.00	2	12	3	51
Type 2 DNA topoisomerase 6 subunit A	PF1578	44.03	UNIQUE	UNIQUE	0	14	0	48
DNA repair and recombination protein RadA	PF1926	38.36	UNIQUE	UNIQUE	0	12	0	46
DNA helicase	PF0085	156.90	UNIQUE	UNIQUE	0	7	0	46
DNA polymerase II large subunit	PF0019	143.06	UNIQUE	UNIQUE	0	4	0	30
DNA topoisomerase 1	PF0494	121.09	UNIQUE	UNIQUE	0	9	0	28
DNA topoisomerase VI subunit B	PF1579	64.37	UNIQUE	UNIQUE	0	4	0	15
AsnC family transcriptional regulator	PF0739	24.11	UNIQUE	UNIQUE	0	2	0	11

DNA-directed RNA polymerase subunit A"	PF1562	44.36	UNIQUE	UNIQUE	0	4	0	9
Ribosomal & Translation								
Annotation	Locus/ORF	MW (kDa)	Fold Change Peptide	Fold Change Spectra	Total Peptides NI	Total Peptides IM	Total Spectra NI	Total Spectra IM
C/D box methylation guide ribonucleoprotein complex aNOP56 subunit	PF0060	46.74	55.00	186.83	2	110	6	1121
Fibrillarin-like rRNA/tRNA 2'-O-methyltransferase	PF0059	25.73	UNIQUE	UNIQUE	0	80	0	825
50S ribosomal protein L4	PF1824	28.67	UNIQUE	UNIQUE	0	76	0	658
50S ribosomal protein L3	PF1825	41.35	UNIQUE	UNIQUE	0	79	0	599
Probable tRNA pseudouridine synthase B	PF1785	38.52	UNIQUE	UNIQUE	0	58	0	525
50S ribosomal protein L12	PF1994	11.22	UNIQUE	UNIQUE	0	30	0	481
30S ribosomal protein S7	PF1558	24.62	UNIQUE	UNIQUE	0	47	0	430
30S ribosomal protein S19e	PF1499	17.33	UNIQUE	UNIQUE	0	25	0	421
50S ribosomal protein L13P	PF1645	16.28	UNIQUE	UNIQUE	0	27	0	406
30S ribosomal protein S5	PF1804	26.52	UNIQUE	UNIQUE	0	53	0	364
50S ribosomal protein L18e	PF1646	13.71	UNIQUE	UNIQUE	0	19	0	344
50S ribosomal protein L18P	PF1805	23.15	UNIQUE	UNIQUE	0	29	0	332
30S ribosomal protein S17e	PF1491	7.90	UNIQUE	UNIQUE	0	26	0	306
30S ribosomal protein S17P	PF1815	13.11	UNIQUE	UNIQUE	0	16	0	299
30S ribosomal protein S15	PF2056	18.57	UNIQUE	UNIQUE	0	25	0	298
50S ribosomal protein L30P	PF1803	17.68	UNIQUE	UNIQUE	0	32	0	293

50S ribosomal protein L10	PF1993	37.09	UNIQUE	UNIQUE	0	36	0	268
50S ribosomal protein L1	PF1992	23.82	UNIQUE	UNIQUE	0	26	0	268
30S ribosomal protein S13	PF1650	16.89	UNIQUE	UNIQUE	0	27	0	253
Translation initiation factor 2 subunit gamma	PF1717	44.99	UNIQUE	UNIQUE	0	44	0	252
30S ribosomal protein S11	PF1648	14.70	UNIQUE	UNIQUE	0	18	0	248
50S ribosomal protein L31e	PF0378	11.03	UNIQUE	UNIQUE	0	17	0	240
50S ribosomal protein L2	PF1822	26.00	UNIQUE	UNIQUE	0	42	0	240
50S ribosomal protein L15e	PF0876	22.56	UNIQUE	UNIQUE	0	25	0	233
50S ribosomal protein L32e	PF1807	15.51	UNIQUE	75.33	0	17	3	226
30S ribosomal protein S6e	PF0488	13.95	UNIQUE	UNIQUE	0	23	0	217
50S ribosomal protein L7Ae	PF1367	13.37	UNIQUE	UNIQUE	0	11	0	210
30S ribosomal protein S2	PF1640	22.97	UNIQUE	UNIQUE	0	28	0	207
50S ribosomal protein L6	PF1808	20.83	UNIQUE	UNIQUE	0	26	0	205
30S ribosomal protein S3	PF1819	23.40	UNIQUE	UNIQUE	0	38	0	192
Probable Brix domain-containing ribosomal biogenesis protein	PF2010	26.04	UNIQUE	UNIQUE	0	22	0	176
50S ribosomal protein L15P	PF1802	16.37	UNIQUE	UNIQUE	0	26	0	174
50S ribosomal protein L22P	PF1820	17.61	UNIQUE	UNIQUE	0	19	0	173
50S ribosomal protein L29P	PF1818	8.51	UNIQUE	UNIQUE	0	17	0	167
30S ribosomal protein S4	PF1649	21.31	UNIQUE	UNIQUE	0	21	0	152
30S ribosomal protein S10	PF1376	11.72	UNIQUE	UNIQUE	0	14	0	144
50S ribosomal protein L10e	PF1279	20.90	UNIQUE	UNIQUE	0	22	0	142

30S ribosomal protein S4e	PF1812	28.11	UNIQUE	UNIQUE	0	24	0	132
30S ribosomal protein S3Ae	PF2054	22.91	UNIQUE	UNIQUE	0	29	0	131
30S ribosomal protein S8e	PF1069	14.24	UNIQUE	UNIQUE	0	25	0	131
50S ribosomal protein L11	PF1991	17.61	UNIQUE	UNIQUE	0	13	0	124
50S ribosomal protein L24P	PF1813	14.30	UNIQUE	UNIQUE	0	10	0	122
30S ribosomal protein S9	PF1644	15.26	UNIQUE	UNIQUE	0	8	0	119
30S ribosomal protein S19P	PF1821	15.30	UNIQUE	UNIQUE	0	15	0	114
30S ribosomal protein S12	PF1559	16.41	UNIQUE	UNIQUE	0	27	0	112
tRNA-splicing ligase RtcB	PFC_10530	108.72	UNIQUE	UNIQUE	0	20	0	106
30S ribosomal protein S8	PF1809	14.63	UNIQUE	UNIQUE	0	7	0	104
50S ribosomal protein L18Ae	PF0376	9.30	UNIQUE	UNIQUE	0	8	0	92
50S ribosomal protein L14e	PF0819	8.94	UNIQUE	UNIQUE	0	9	0	87
50S ribosomal protein L23P	PF1823	9.85	UNIQUE	UNIQUE	0	13	0	82
S1p family ribosomal protein	PF0399	83.14	UNIQUE	UNIQUE	0	12	0	82
50S ribosomal protein L14P	PF1814	15.21	UNIQUE	UNIQUE	0	12	0	79
50S ribosomal protein L39e	PF0379	6.28	UNIQUE	UNIQUE	0	7	0	78
Translation initiation factor 2 subunit alpha	PF1140	31.88	UNIQUE	UNIQUE	0	15	0	74
Ribosomal RNA small subunit methyltransferase Nep1	PF1262	25.92	UNIQUE	UNIQUE	0	10	0	72
50S ribosomal protein L21e	PF1035	11.32	UNIQUE	UNIQUE	0	4	0	72
30S ribosomal protein S24e	PF0253	11.65	UNIQUE	UNIQUE	0	8	0	67
50S ribosomal protein L35Ae	PF1872	9.71	UNIQUE	UNIQUE	0	9	0	66

H/ACA RNA-protein complex component Gar1	PFC_09635	11.45	UNIQUE	UNIQUE	0	5	0	65
Translation initiation factor 2 subunit beta	PF0481	16.21	UNIQUE	UNIQUE	0	11	0	65
Translation initiation factor 6	PF0377	24.56	UNIQUE	UNIQUE	0	12	0	60
50S ribosomal protein L19e	PF1806	17.84	UNIQUE	UNIQUE	0	7	0	59
50S ribosomal protein L5	PF1811	21.20	UNIQUE	UNIQUE	0	8	0	57
Elongation factor 1- alpha	PF1375	47.59	3.33	5.44	3	10	9	49
50S ribosomal protein L37e	PF1541	7.20	UNIQUE	UNIQUE	0	3	0	45
30S ribosomal protein S27e	PF0218	6.85	UNIQUE	UNIQUE	0	5	0	39
30S ribosomal protein S14 type Z	PF1810	6.59	UNIQUE	UNIQUE	0	4	0	28
tRNA/rRNA methyltransf erase	PF0251	44.99	UNIQUE	UNIQUE	0	10	0	25
Ribosomal protein s6 modification protein	PF1682	30.82	UNIQUE	UNIQUE	0	4	0	24
Probable transcription termination protein NusA	PF1560	16.43	UNIQUE	UNIQUE	0	5	0	24
H/ACA RNA-protein complex component Gar1	PF1791	11.45	UNIQUE	UNIQUE	0	4	0	24
50S ribosomal protein L30e	PF1561	10.69	UNIQUE	UNIQUE	0	3	0	23
Putative snRNP Sm- like protein	PF1542	8.42	UNIQUE	UNIQUE	0	3	0	19
nol1-nop2- sun family nucleolar protein IV	PF0666	52.16	UNIQUE	UNIQUE	0	5	0	18
nol1-nop2- sun family nucleolar protein IV	PF0666	52.16	UNIQUE	UNIQUE	0	5	0	18
30S ribosomal protein S28e	PF1368	8.08	UNIQUE	UNIQUE	0	2	0	16
H/ACA RNA-protein complex	PF1141	7.20	UNIQUE	UNIQUE	0	2	0	15

component Nop10p								
Translation initiation factor 1A	PFC_07140	13.05	UNIQUE	UNIQUE	0	3	0	14
Ribonuclease P protein component 2	PF1378	13.81	UNIQUE	UNIQUE	0	2	0	13
tRNA- splicing endonuclease	PF0266	19.89	UNIQUE	UNIQUE	0	2	0	12
DEXX-box atpase	PF0634	54.56	UNIQUE	UNIQUE	0	2	0	10
nol1-nop2- sun family nucleolar protein II	PF1265	50.39	UNIQUE	UNIQUE	0	3	0	9
Protein translation factor SUI1 homolog	PF1817	11.39	UNIQUE	UNIQUE	0	2	0	8
thiamine biosynthesis/ tRNA modification protein ThiI	PF1288	40.69	1.00	2.67	2	2	3	8
seryl-tRNA ligase	PF1204	53.15	UNIQUE	UNIQUE	0	2	0	6
Other Identifications: RNA binding, Cell division, Metabolism								
Annotation	Locus/ORF	MW (kDa)	Fold Change Peptide	Fold Change Spectra	Total Peptides NI	Total Peptides IM	Total Spectra NI	Total Spectra IM
Thermosome , single subunit	PF1974	59.93	2.16	8.49	19	41	51	433
Cell division protein CDC48	PF0963	94.04	2.00	25.31	7	14	13	329
Nucleoside- triphosphatas e PFC_01625	PF0501	20.83	17.50	124.00	2	35	2	248
Chromosome partition protein Smc	PF1843	134.91	0.00	9.19	4	0	16	147
Ornithine carbamoyltra nsferase	PF0594	35.14	UNIQUE	UNIQUE	0	7	0	125
Proteasome- activating nucleotidase	PF0115	44.76	UNIQUE	UNIQUE	0	27	0	123
ATPase	PF1041	68.14	UNIQUE	UNIQUE	0	28	0	88
Exosome complex component Rrp41	PF1568	27.81	UNIQUE	UNIQUE	0	19	0	83
Exosome complex component Rrp42	PFC_07065	30.47	UNIQUE	UNIQUE	0	9	0	80
Proteasome subunit alpha	PF1571	28.97	UNIQUE	UNIQUE	0	6	0	62
RNA- processing protein	PF1580	25.46	UNIQUE	UNIQUE	0	7	0	47
Exosome complex	PFC_07075	29.50	UNIQUE	UNIQUE	0	9	0	41

component Rrp4								
Tryptophan synthase beta chain	PF1706	42.48	UNIQUE	UNIQUE	0	14	0	38
Beta- glucosidase	PF0442	49.71	UNIQUE	UNIQUE	0	11	0	36
RNA 3'- terminal phosphate cyclase	PF1549	36.76	UNIQUE	UNIQUE	0	8	0	36
Acetyl-lysine deacetylase	PF1686	35.89	UNIQUE	UNIQUE	0	15	0	33
enolase	PF0215	46.79	UNIQUE	UNIQUE	0	6	0	30
Inosine-5'- monophosph ate dehydrogena se	PF1794	21.32	UNIQUE	UNIQUE	0	7	0	27
Type II secretion system protein	PFC_04200	129.90	3.00	4.50	1	3	6	27
bacteriochlor ophyll synthase, 43 kDa subunit	PF0454	40.76	1.67	6.50	3	5	4	26
RNA- binding protein FAU-1	PF0022	53.62	UNIQUE	UNIQUE	0	6	0	21
pyruvate formate- lyase activating enzyme-like protein	PF1397	40.36	UNIQUE	UNIQUE	0	5	0	16
Periplasmic sugar binding protein	PF0119	61.21	UNIQUE	UNIQUE	0	2	0	15
ADP- dependent glucokinase	PF0312	51.22	UNIQUE	UNIQUE	0	3	0	15
UPF0201 protein PFC_05700	PFC_05700	15.71	UNIQUE	UNIQUE	0	4	0	13
Exosome complex RNA- binding protein Csl4	PF0052	21.94	UNIQUE	UNIQUE	0	4	0	12
exosome complex RNA- binding protein Rrp4	PF1569	29.50	UNIQUE	UNIQUE	0	4	0	11
aspartate carbamoyltra nsferase catalytic subunit	PF0599	34.68	UNIQUE	UNIQUE	0	4	0	11
acetyl-CoA acetyltransfe rase	PF0973	40.94	UNIQUE	UNIQUE	0	3	0	11
glutamine synthetase	PF0450	50.13	UNIQUE	UNIQUE	0	3	0	10

NDP-sugar synthase	PF0868	47.17	UNIQUE	UNIQUE	0	3	0	10
exosome complex RNA-binding protein Rrp42	PF1567	30.47	UNIQUE	UNIQUE	0	2	0	10
BtpA family protein	PF0860	28.56	UNIQUE	UNIQUE	0	2	0	9
hydrogenase maturation protein HypF	PF0559	87.34	UNIQUE	UNIQUE	0	3	0	9
bifunctional phosphoman nomutase/phosphoglucomutase	PF0588	49.58	UNIQUE	UNIQUE	0	2	0	8
GTP-binding protein	PF0998	40.64	UNIQUE	UNIQUE	0	2	0	7
Phosphorylase	PF1535	97.63	UNIQUE	UNIQUE	0	2	0	5
aldehyde ferredoxin oxidoreductase	PF0464	73.88	UNIQUE	UNIQUE	0	2	0	5
ribulose bisphosphate carboxylase	PF1156	47.30	UNIQUE	UNIQUE	0	2	0	5
aldehyde ferredoxin oxidoreductase	PF1203	68.71	UNIQUE	UNIQUE	0	2	0	5
L-isoaspartyl protein carboxyl methyltransferase	PF1896	28.46	UNIQUE	UNIQUE	0	1	0	4
2-ketoisovalerate ferredoxin oxidoreductase subunit beta	PF0968	34.73	UNIQUE	UNIQUE	0	2	0	4
Arginase	PF0499	27.26	UNIQUE	UNIQUE	0	1	0	2
Uncharacterized/Unannotated Proteins								
Annotation	Locus/ORF	MW (kDa)	Fold Change Peptide	Fold Change Spectra	Total Peptides NI	Total Peptides IM	Total Spectra NI	Total Spectra IM
Uncharacterized protein	PF0490	27.82	UNIQUE	UNIQUE	0	49	0	469
Uncharacterized protein	PF0797	40.81	UNIQUE	UNIQUE	0	40	0	306
Uncharacterized protein	PF0206	30.52	1.33	5.83	3	4	41	239
Uncharacterized protein	PF0070	36.82	UNIQUE	UNIQUE	0	19	0	233
Uncharacterized protein	PF0496	30.57	UNIQUE	UNIQUE	0	18	0	180
Uncharacterized protein	PF1625	14.85	UNIQUE	UNIQUE	0	18	0	158
Uncharacterized protein	PF1927	23.42	UNIQUE	UNIQUE	0	23	0	157
Uncharacterized protein	PF1101	49.03	UNIQUE	UNIQUE	0	16	0	62

Uncharacterized protein	PF2018	31.12	UNIQUE	UNIQUE	0	14	0	57
hypothetical protein	PF1867	33.11	UNIQUE	UNIQUE	0	10	0	49
Uncharacterized protein	PF0213	30.26	UNIQUE	UNIQUE	0	11	0	47
hypothetical protein	PF0341	38.04	UNIQUE	UNIQUE	0	9	0	46
Uncharacterized protein	PFC_04875	30.65	UNIQUE	UNIQUE	0	11	0	42
Uncharacterized protein	PF0829	7.55	UNIQUE	UNIQUE	0	6	0	42
Uncharacterized protein	PFC_04545	9.03	UNIQUE	UNIQUE	0	5	0	39
Uncharacterized protein	PF1201	19.34	UNIQUE	UNIQUE	0	7	0	38
Uncharacterized protein	PF1826	29.23	UNIQUE	UNIQUE	0	9	0	35
hypothetical protein	PF1142	29.95	4.50	8.50	2	9	4	34
Uncharacterized protein	PFC_09020	18.20	UNIQUE	UNIQUE	0	8	0	31
Uncharacterized protein	PF1827	31.77	UNIQUE	UNIQUE	0	4	0	29
Uncharacterized protein	PF1393	74.43	UNIQUE	UNIQUE	0	7	0	22
hypothetical protein	PF1615	108.72	UNIQUE	UNIQUE	0	5	0	22
Uncharacterized protein	PF1498	16.49	UNIQUE	UNIQUE	0	3	0	19
Uncharacterized protein	PF1027	43.48	UNIQUE	UNIQUE	0	7	0	17
Uncharacterized protein	PF1038	23.55	UNIQUE	UNIQUE	0	5	0	16
Uncharacterized protein	PF0124	39.38	3.50	7.50	2	7	2	15
hypothetical protein	PF1059	45.21	UNIQUE	UNIQUE	0	5	0	14
hypothetical protein	PF1191	38.76	1.00	7.00	2	2	2	14
Uncharacterized protein	PF0099	36.66	UNIQUE	UNIQUE	0	5	0	12
Uncharacterized protein	PF1641	37.68	UNIQUE	UNIQUE	0	4	0	11
Uncharacterized protein	PF0051	25.68	UNIQUE	UNIQUE	0	2	0	5
Uncharacterized protein	PF1184	28.40	UNIQUE	UNIQUE	0	2	0	3
Uncharacterized protein	PFC_04395	45.23	UNIQUE	UNIQUE	0	1	0	3
Uncharacterized protein	PFC_04935	29.95	UNIQUE	UNIQUE	0	1	0	2
Uncharacterized protein	PFC_05155	38.96	UNIQUE	UNIQUE	0	1	0	0
Uncharacterized protein	PF0537	90.32	UNIQUE	UNIQUE	0	3	0	0

APPENDIX

Strains and Plasmid Inventory

<i>P. furiosus</i> Strains	Relevant Characteristics	Source or Reference
COM1	Base of each strain used in this study	Farakas et.al 2011
JFW02 (WT)	Δ pyrF Δ trpAB	Farakas et.al 2011
TPF41	Cas1, Cas2, Cas4-1, overexpressed	This Study
TPF51	Δ Cas1, Cas2, Cas 4-1 overexpressed	This Study
TPF66	Cas1, Cas2 overexpressed, Δ Cas4-1	This Study
TPF68	Cas1, Cas2, Cas4-1, Cas4-2 overexpressed	This Study
TPF40	Cas1, Cas2 overexpressed, Δ Cas4-1	This Study
Plasmids	Relevant Characteristics	Source or Reference
pEAW002	Cas1 10X polyhis C'tag overexpressed	This Study

Commonly Used Abbreviations

pAb: polyclonal antibody

mAb: monoclonal antibody

IP: immunoprecipitation

Co-IP: co-immunoprecipitation

MS: tandem mass spectrometry

SDS-PAGE: sodium dodecyl sulfate polyacrylamide gel electrophoresis

IP/Western: western blotting performed on immunoprecipitated material

IP/MS: mass spectrometry performed on immunoprecipitated material