Advances in Experimental Medicine and Biology 1414
Protein Reviews

M. Zouhair Atassi *Editor*

Protein Reviews

Volume 23

MOREMEDIA

Springer

Advances in Experimental Medicine and Biology

Protein Reviews

Volume 1414

Series Editor

M. Zouhair Atassi
Biochemistry and Molecular Biology, Baylor College of Medicine, Houston, TX, USA

The aim of the Protein Reviews is to serve as a publication vehicle for articles that focus on important current and fundamental aspects of protein structure, function, evolution and genetics. Publications will be selected based on their significance to the study of biological systems, their relevance to the study of health and disease or their contribution to technological developments. Proteins associated with diseases or the appearance and progression of diseases are also important subjects that may be covered in this series. Additionally, proteins that could function as potential biomarkers for prediction and analysis, and also as an object for treatment and for the design of unique, novel and target-specific therapeutics are also of interest to this book series. The issues may include biochemistry, biophysics, immunology, molecular biology, genetics, molecular and cellular mechanisms of action, clinical studies and new pioneering therapies. A given volume may be focused, or may be a selected assortment of different current topics. This book series publishes accepted articles online upon acceptance. It will afterwards assemble relevant articles as single volumes. Invited authors are nominated by the Editorial Board or by other experts. However, interested individuals may suggest a topic or may propose a person to review a current important topic. Such interested authors should contact the editor before submitting a manuscript. Authors selected by the Editorial Board will be invited by the editor. The authors of the articles are selected from leading basic or medical scientists in academic and industrial organizations. The manuscripts are reviewed and evaluated in the usual manner by experts in the topic of the manuscript. The articles will be published online no later than two months after the editorial review and acceptance.

M. Zouhair Atassi
Editor

Protein Reviews

Volume 23

Springer

Editor
M. Zouhair Atassi
Biochemistry and Molecular Biology
Baylor College of Medicine
Houston, TX, USA

ISSN 0065-2598 ISSN 2214-8019 (electronic)
Advances in Experimental Medicine and Biology
ISSN 2520-1891 ISSN 2520-1905 (electronic)
Protein Reviews
ISBN 978-3-031-28669-8 ISBN 978-3-031-28670-4 (eBook)
https://doi.org/10.1007/978-3-031-28670-4

© Springer Nature Switzerland AG 2023
This work is subject to copyright. All rights are reserved by the Publisher, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, reuse of illustrations, recitation, broadcasting, reproduction on microfilms or in any other physical way, and transmission or information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed.
The use of general descriptive names, registered names, trademarks, service marks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.
The publisher, the authors, and the editors are safe to assume that the advice and information in this book are believed to be true and accurate at the date of publication. Neither the publisher nor the authors or the editors give a warranty, expressed or implied, with respect to the material contained herein or for any errors or omissions that may have been made. The publisher remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

This Springer imprint is published by the registered company Springer Nature Switzerland AG
The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

Preface

Protein Reviews is a book series published by Springer Nature, as a subseries of Advances in Experimental Medicine and Biology (Impact Factor: 2.45). Twenty three volumes have appeared. To see previous volumes, please go to: www.springer.com/series/14330.

The review articles had focused on important current and fundamental aspects of protein structure, function, evolution, genetics, significance to the biological system, and/or relevance to a health and disease. Proteins associated with a certain disease or the appearance and progression of a disease are also covered in this series. Before publishing accepted reviews in print, they will be published online. All chapters are indexed in PubMed and the Web of Science. Review articles had been by invitation, but some were also selected from proposed topics. An author who wished to write a manuscript could submit his/her proposal to the editor.

Here, in the first chapter "Structure and Function of SNM1 Family Nucleases," Hsuan-Yi Wu, Yuanzhang Zheng, Adrian R. Laciak, Nian N. Huang, Mary Koszelak-Rosenblum, Andrew J. Flint, Grant Carr, and Guangyu Zhu discuss three human nucleases, SNM1A, SNM1B/Apollo, and SNM1C/Artemis, that belong to the SNM1 gene family. These nucleases are involved in various cellular functions and share a similar catalytic domain which is characterized as a fused metallo-β-lactamase and a CPSF-Artemis-SNM1-PSO2 domain. SNM1A and SNM1B/Apollo are exonucleases, whereas SNM1C/Artemis is an endonuclease. The review contains a summary of recent findings on SNM1's cellular and biochemical functions, and also structural biology studies. In addition, protein structure prediction by an artificial intelligence program provides a different view of the non-catalytic domain, used with current X-ray results.

The second chapter by Tamara Flusche and Rakhi Rajan, entitled "Molecular Details of DNA Integration by CRISPR-Associated Proteins During Adaptation in Bacteria and Archaea," deals with clustered regularly interspaced short palindromic repeats (CRISPR) and CRISPR-associated proteins that constitute an adaptive immune system in bacteria and archaea, where immunological memory is retained in the CRISPR locus as short pieces of the intruding nucleic acid, termed spacers.

The third chapter by Paul R. Gardner, entitled "Ordered Motions in the Nitric-Oxide Dioxygenase Mechanism of Flavohemoglobin and Assorted Globins with Tightly Coupled Reductases," discusses how nitric oxide

dioxygenases (NODs) activate and combine O_2 with NO to form nitrate. A variety of oxygen-carrying hemoglobins with associated partner reductases or electron donors function as enzymatic NODs. Analyzing kinetic and structural studies of the archetypal two-domain microbial flavohemoglobin-NOD, Dr. Gardner presents an allosteric mechanism that employs selective tunnels for O_2 and NO, gates for NO, nitrate, transient O_2 association with ferric heme, and an O_2 and NO-initiated ferric heme spin crossover-driven, motion-controlled, and dipoleregulated electron-transfer switch.

The last chapter entitled "Structural Analyses of the Multicopper Site of CopG Support a Role as a Redox Enzyme" by Andrew C. Hausrath and Megan M. McEvoy, examines how metal ions, including copper, are used as broad-spectrum biocides in a variety of clinical and environmental settings. The copG gene is a common component of such copper resistance protein clusters, but its contribution to copper resistance is not well understood. In this chapter, the authors summarize the available information about the CopG protein encoded by this gene. The recent structure is compared to diverse copper-containing metallochaperones, metalloenzymes, and electron transfer proteins.

I was the Founding Editor in 1999 and continued as the Editor for 23 years. Now, because of health reasons, I really must retire and regrettably the book review series will also be retired and cease publication.

I would like to thank the authors in volume 23 and all previous volumes for their outstanding contributions, and the readers for their continued support. I also want to pay special thanks to my wife Lena MousaPasha Atassi, MD, for her patient help and support during my work on these books.

Houston, TX, USA M. Zouhair Atassi

Contents

About the Editor

M. Zouhair Atassi is the emeritus Robert A. Welch chair of chemistry, professor of biochemistry and molecular biology, and professor of pathology and immunology at Baylor College of Medicine, Houston, Texas. Previously, he was professor of biochemistry and immunology at Mayo Medical School and Mayo Clinic, Rochester, Minnesota. He was president of the International Symposium of the Immunobiology of Proteins and Peptides and the Institute of Immunology; founding editor and editor-in-chief of *Critical Reviews in Immunology*, *Protein Journal*, and *Protein Reviews*; and served on the editorial boards of nine other peer-reviewed journals. Professor Atassi is the editor or co-editor of 37 books; author or co-author of over 500 articles in peer-reviewed journals on protein structure, biological function, and molecular and cellular immune recognition; and has received several distinguished awards, medals, and prizes in these fields.

Adv Exp Med Biol - Protein Reviews (2023) 23: 27–43
https://doi.org/10.1007/5584_2022_730
© Springer Nature Switzerland AG 2022
Published online: 20 July 2022

Molecular Details of DNA Integration by CRISPR-Associated Proteins During Adaptation in Bacteria and Archaea

Tamara Flusche and Rakhi Rajan

Abstract

Clustered regularly interspaced short palindromic repeats (CRISPR) and CRISPR-associated (Cas) proteins constitute an adaptive immune system in bacteria and archaea, where immunological memory is retained in the CRISPR locus as short pieces of the intruding nucleic acid, termed spacers. The adaptation to new infections occurs through the integration of a new spacer into the CRISPR array. For immune protection, spacers are transcribed into CRISPR RNAs (crRNA) that are used to guide the effector nuclease of the system in sequence-dependent target cleavage. Spacers originate as a prespacer from either DNA or RNA depending on the CRISPR-Cas system being observed, and the nearly universal Cas proteins, Cas1 and Cas2, insert the prespacer into the CRISPR locus during adaptation in all systems that contain them. The mechanism of site-specific prespacer integration varies across CRISPR classes and types, and distinct differences can even be found within the same subtype. In this review, the current knowledge on the mechanisms of prespacer integration in type II-A CRISPR-Cas systems will be described. Comparisons of the currently characterized type II-A systems show that distinct mechanisms exist within different members of this subtype and are correlated to sequence-specific interactions of Cas proteins and the DNA elements present in the CRISPR array. These observations indicate that nature has fine-tuned the mechanistic details while performing the basic step of DNA integration by Cas proteins, which offers unique advantages to develop Cas1-Cas2-based biotechnology.

Keywords

Cas1 · Cas2 · Cas9 · CRISPR adaptation · CRISPR-Cas · Csn2 · DNA integration · Integrase · Prespacer · Type II CRISPR systems

Abbreviations

CRISPR	Clustered regularly interspaced short palindromic repeats
Cas	CRISPR-associated
crRNA	CRISPR RNA
pre-crRNA	pre-CRISPR RNA
nt	nucleotide
nts	nucleotides
PAM	protospacer adjacent motif

T. Flusche and R. Rajan (✉)
Department of Chemistry and Biochemistry, Price Family Foundation Institute of Structural Biology, Stephenson Life Sciences Research Center, University of Oklahoma, Norman, OK, USA
e-mail: r-rajan@ou.edu

PFS	protospacer flanking site
ssDNA	single-stranded DNA
dsDNA	double-stranded DNA
tracrRNA	*trans*-activating-crRNA
ssRNA	single-stranded RNA
E. coli	*Escherichia coli*
IHF	integration host factor

1 Introduction

Clustered regularly interspaced short palindromic repeats (CRISPR) were discovered as repetitive sequences in the genomes of bacteria and archaea (Ishino et al. 1987; Mojica et al. 1993, 2000). These repetitive sequences, termed "repeats," were separated by short, fixed length regions that originated from extracellular DNA, termed "spacers" (Bolotin et al. 2005; Mojica et al. 2005). Soon after, CRISPR-associated (Cas) proteins were found to have roles in integrating new spacers into the CRISPR array and targeting extracellular DNA for cleavage in a sequence-dependent manner, establishing the CRISPR-Cas operon as one functional system (Jansen et al. 2002; Haft et al. 2005; Barrangou et al. 2007). CRISPR-Cas systems have been established as adaptive immune systems found in archaea and bacteria, and some systems have been shown to target intruding RNA, while others have been identified as essential for gene regulation (Barrangou et al. 2007; Hale et al. 2009; Silas et al. 2016; Mohr et al. 2018; Faure et al. 2019; Newsom et al. 2021).

1.1 CRISPR-Cas Systems as Adaptive Immune Systems

Even though CRISPR systems vary in their compositions and mechanisms, the overall steps leading to a successful immune defense, adaptation, expression, and interference are the same in all organisms that have been characterized so far (Fig. 1a). When foreign genetic elements, such as plasmids or bacteriophage, enter the cell, a short piece of the invading DNA, or RNA in the case of some systems, will be acquired and integrated into the prokaryotic genome as a part of the repeat-spacer array (Barrangou et al. 2007; Nuñez et al. 2014, 2015; Silas et al. 2016; Mohr et al. 2018). This process is termed "adaptation" (Fig. 1a). The new spacers are preferentially integrated upstream of the first repeat, though integration can occur ectopically at other positions in the array (Fig. 1a) (Erdmann and Garrett 2012; Nuñez et al. 2015; McGinn and Marraffini 2016). In the second step of "expression," the CRISPR array is transcribed into pre-CRISPR RNA (pre-crRNA), which is then processed into mature crRNAs containing specific regions of one repeat-spacer unit (Fig. 1a). In the final step of "interference," the crRNA binds a single or multiple Cas proteins to form an effector complex to find the target site on the invading nucleic acid that is complementary to the guide region of the crRNA (~20-32 nucleotide (nt) long region) (Jore et al. 2011; Sternberg et al. 2014). Cas nucleases either site-specifically introduce double-stranded breaks or degrade long regions of the intruder genome, inactivating the infection (Fig. 1a) (Brouns et al. 2008; Garneau et al. 2010). To distinguish native nucleic acids from the intruder during the interference step, different CRISPR systems employ distinct mechanisms, such as a protospacer adjacent motif (PAM) or protospacer flanking site (PFS), to prevent self-DNA/RNA cleavage (Deveau et al. 2008; Mojica et al. 2009; Marraffini and Sontheimer 2010; Abudayyeh et al. 2016). The PAM and PFS sequence requirements depend on the organism and CRISPR system with strand preferences in some of the CRISPR types (Mojica et al. 2009; Vink et al. 2021).

1.2 Classification of CRISPR-Cas Systems

The different CRISPR-Cas systems have been categorized into 2 different classes, 6 types, and 33 subtypes (Makarova et al. 2018, 2020a). Each CRISPR-Cas system contains an operon of several Cas proteins and a CRISPR array consisting of multiple repeats and spacers (Makarova et al. 2002, 2006; Haft et al. 2005). In certain CRISPR

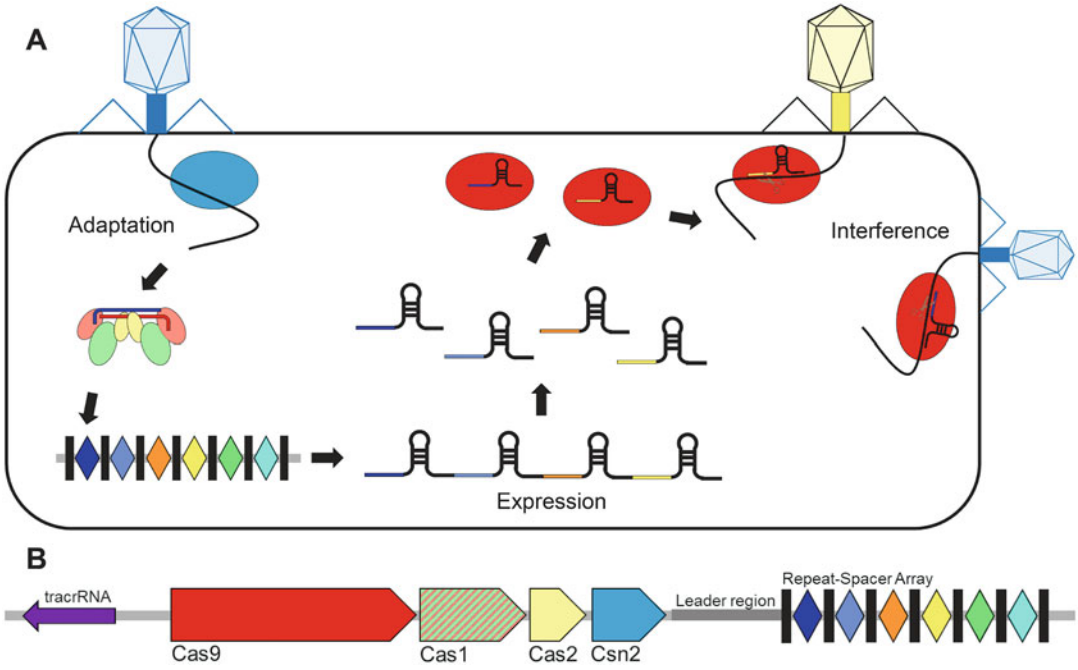

Fig. 1 CRISPR-Cas systems as an adaptive immune system. (**a**) Cartoon schematic of the different steps involved in CRISPR-Cas immune response. There are three distinct steps in CRISPR immunity. Step 1 is adaptation, which enables insertion of spacers acquired from invading nucleic acids. The excised nucleic acid is processed into a prespacer by protein complexes composed of Cas proteins as well as yet uncharacterized cellular proteins (represented by the blue circle). The processed prespacer (red and blue strands) is integrated into the repeat-spacer array (rectangle-diamond array) of the CRISPR locus by Cas1 (green and salmon circles) and Cas2 (yellow circles). Step 2 is expression, where the repeat-spacer array is transcribed into pre-crRNA and then processed into individual crRNAs (hairpin structures). The crRNAs are taken up by Cas proteins (red circle) forming an effector complex that surveys the host cell for the presence of intruder genome. In step 3, the ribonuclear complex locates the intruder by sequence specific interactions of crRNA with the intruder genome, followed by cleavage and inactivation of the intruder by signature Cas proteins, offering protection from invasion. (**b**) Cartoon schematic of a type II-A CRISPR locus

systems, the DNA region immediately upstream of the first repeat of the CRISPR array, known as the leader region, holds promoters for transcription of the locus as well as motifs essential for adaptation (Jansen et al. 2002; Pourcel et al. 2005; Pougach et al. 2010; Yosef et al. 2012; Wei et al. 2015a).

Class 1 is defined by multi-protein effector complexes. Class 1 consists of types I, III, and IV. Type I has an effector complex known as the CRISPR-associated complex for antiviral defense or cascade, which is composed of multiple Cas proteins that bind the crRNA (Brouns et al. 2008; Jore et al. 2011; Wiedenheft et al. 2011). The cascade complex targets DNA based on complementarity with the crRNA, followed by recruitment of the signature protein Cas3, which uses an HD endonuclease for DNA degradation (Brouns et al. 2008; Sinkunas et al. 2011; Jore et al. 2011; Huo et al. 2014). Type III systems are unique due to their ability to perform crRNA-dependent RNA cleavage as well as transcription-coupled single-stranded DNA (ssDNA) cleavage, both of which are essential for complete protection against the intruder (Hale et al. 2009; Staals et al. 2013; Samai et al. 2015). The RNA and DNA cleavages also trigger the synthesis of cyclic oligoadenylates by Cas10, which activates nonspecific cellular RNA cleavage by Csm6/Csx1, arresting cell growth (Niewoehner et al. 2017;

Kazlauskiene et al. 2017; Koonin and Makarova 2018). Type IV systems are highly derived CRISPR systems that generally do not have components such as adaptation modules or Cas interference nucleases (Makarova et al. 2020a; Moya-Beltrán et al. 2021). While some type IV subtypes have been shown to cleave plasmids by association of CRISPR-Cas components with accessory proteins, some subtypes have been predicted to be involved in non-CRISPR-mediated immune defense and gene regulation (Taylor et al. 2021; Moya-Beltrán et al. 2021). Future research is essential to characterize type IV systems that are very distinct from the currently well-characterized CRISPR-Cas systems.

Class 2 systems are defined by their use of specific multi-domain single proteins as the effector complex. Class 2 consists of types II, V, and VI. Type II contains the effector nuclease Cas9, which possesses two separate endonuclease domains (HNH and RuvC) for double-stranded DNA (dsDNA) cleavage (Haft et al. 2005; Makarova et al. 2006; Jinek et al. 2012; Gasiunas et al. 2012; Sternberg et al. 2014). Cas9 requires a *trans*-activating-crRNA (tracrRNA) along with the crRNA to perform DNA interrogation and cleavage (Deltcheva et al. 2011). Type V contains the effector nuclease Cas12, which uses one nuclease domain (RuvC) for successive cleavages of both strands of DNA (Zetsche et al. 2015; Makarova et al. 2015). Type VI contains the effector nuclease Cas13 which targets RNA and uses two HEPN nuclease domains to inflict *cis* and *trans*-RNA cleavages to induce dormancy in phage-infected bacteria (Abudayyeh et al. 2016; East-Seletsky et al. 2016).

1.3 CRISPR Adaptation Has a Conserved Mechanism Across the Different CRISPR Types

Even though the exact protein composition of the different CRISPR-Cas systems vary widely as mentioned above, most of them possess Cas1 and Cas2 proteins (Makarova et al. 2020a). Though they vary in their sequence similarities, these homologous proteins are the driving force of the adaptation module in CRISPR immunity (Makarova et al. 2020a). Cas1 was first identified as a divalent metal-dependent DNA endonuclease (Wiedenheft et al. 2009). Metal binding pocket residues, including metal coordinating residues, have been found to be acidic (D/E) (Wiedenheft et al. 2009; Rollie et al. 2015). It houses the catalytic site for the integration reaction, and mutations of the key active site residues abolish integration entirely (Rollie et al. 2015). Cas2 was identified as a divalent metal-dependent single-stranded RNA (ssRNA) endoribonuclease with a ferredoxin-like fold (Beloglazova et al. 2008). However, Cas2 has not been observed to have any catalytic function during prespacer integration since mutation of the key active site residues does not inhibit integration (Rollie et al. 2015; Nuñez et al. 2014). Cas1 and Cas2 form a heterohexameric structure with the stoichiometry of $Cas1_4$-$Cas2_2$ to promote integration, which is critical for the integration reaction to occur (Fig. 2a) (Wright and Doudna 2016; Xiao et al. 2017).

Type II CRISPR-Cas systems have been the focus of many studies due to the presence of the signature nuclease Cas9, which is heavily used in many biotechnological and genome editing applications. All type II systems have Cas9, Cas1, and Cas2 in the CRISPR locus and are further categorized into three subtypes based on a secondary signature protein (Chylinski et al. 2013, 2014; Makarova et al. 2020a). Type II-A contains the protein Csn2 (Fig. 1b), which has been found to have dsDNA binding activity, with no demonstrated catalytic activity (Nam et al. 2011; Ellinger et al. 2012; Arslan et al. 2013). Type II-B contains the protein Cas4, which has been shown to have divalent metal-dependent endonuclease activity as well as 5′ to 3′ exonuclease activity against ssDNA (Lemak et al. 2013). While the role of Cas4 in type II-B adaptation is not currently known, Cas4s in other CRISPR types have been demonstrated to have a role in adaptation (Rollie et al. 2018; Kieper et al. 2018; Lee et al. 2018, 2019; Makarova et al. 2020a). Type II-C contains only Cas9, Cas1, and Cas2 in its *cas* operon.

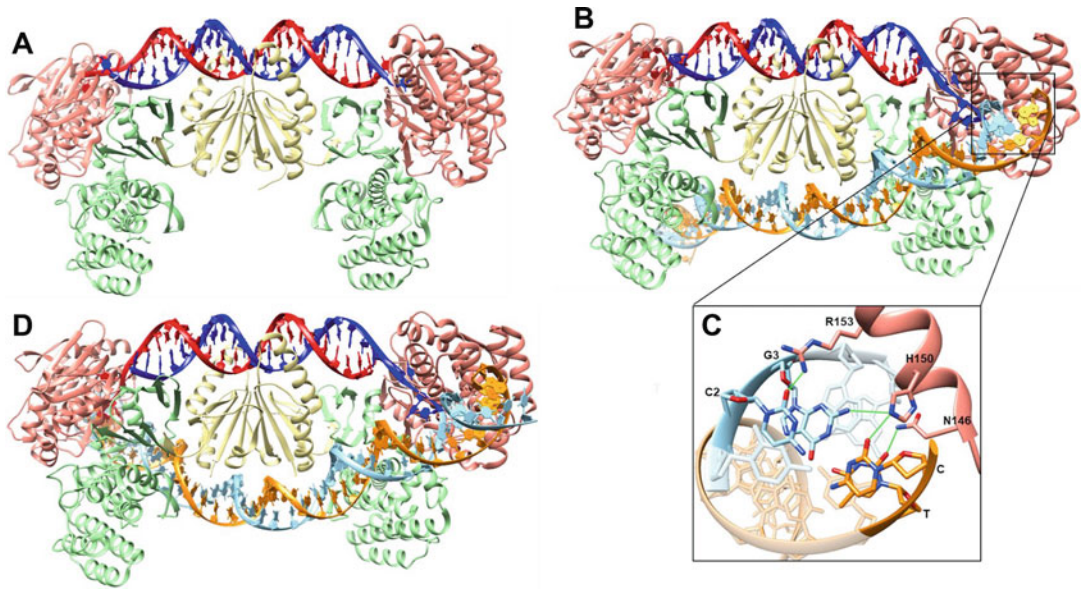

Fig. 2 Components of prespacer integration in type II-A systems. (**a**) Crystal structure of the $Cas1_4$-$Cas2_2$-prespacer integration complex; PDB ID: 5XVN (Xiao et al. 2017). (**b**) Crystal structure of the integration complex where the prespacer has been integrated into the leader-repeat junction of the target DNA; PDB ID: 5XVO (Xiao et al. 2017). (**c**) Close-up of the leader recognition helix of Cas1. Residues of the helix make sequence-specific contacts with nucleotides of the leader (interactions are shown in lime green); PDB ID: 5XVO (Xiao et al. 2017). (**d**) Crystal structure of the integration complex where the prespacer has been integrated into both sides of the target DNA; PDB ID: 5XVP (Xiao et al. 2017). Color scheme: proximal Cas1, green; distal Cas1, salmon; Cas2, yellow; prespacer, red and blue strands; target DNA, cyan and orange strands.

In this review, we will be describing the mechanisms of prespacer acquisition and integration in CRISPR systems, with more in-depth mechanistic details being presented for the type II-A systems. A prespacer is a processed section of an intruder DNA or RNA that will be integrated into a repeat-spacer array. The prespacer integration mechanism has been well characterized in type II-A and type I-E; however, type II-A is distinct from type I-E in that accessory protein factors are not needed to facilitate sequence-specific prespacer integration into the CRISPR array. This independent mechanism of type II-A Cas1 and Cas2 to integrate DNA site specifically can potentially be advantageous for Cas1-Cas2-based biotechnology (Shipman et al. 2016, 2017; Schmidt et al. 2018).

2 Integration in Type II-A Systems

2.1 Prespacer Acquisition Precedes Integration

Prespacer acquisition is the process of acquiring a piece of invading nucleic acid so that it can be integrated into the repeat-spacer array. A complete mechanism for prespacer acquisition in type II-A systems has yet to be elucidated due to several outstanding questions. Current knowledge shows that all four Cas proteins in the type II-A locus as well as tracrRNA are essential for prespacer acquisition under in vivo conditions and that all four Cas proteins are able to form a stable complex and co-purify through a size exclusion column (Wei et al. 2015b; Heler et al.

2015; Wilkinson et al. 2019). Prespacer acquisition preferentially occurs around regions of double-strand breaks in DNA such as those created through AddAB-mediated repair, restriction enzyme digestion, or preexisting *cos* sites in viral DNA. As such, prespacer acquisition is thought to begin with Cas9 either identifying a PAM sequence near a dsDNA break or generating a dsDNA break through crRNA-mediated cleavage (Modell et al. 2017; Nussenzweig et al. 2019; Maguin et al. 2022). The flanking region of DNA is then excised and processed by Cas9 or other undefined nuclease(s) to ensure that the prespacer DNA contains features essential for Cas1-Cas2 to perform the integration reaction, such as the presence of an accessible 3′-OH on both ends of the prespacer and removal of the PAM sequence (Nuñez et al. 2015; Jakhanwal et al. 2021).

Even though Cas9 is essential for adaptation, its catalytic activity is dispensable. It has been observed that spacer uptake occurs from both the bacterial genome and intruding DNA when Cas9 is catalytically inactive (Wei et al. 2015b; Heler et al. 2015). Mutational studies of Cas9 showed that the substitution of I473F increased the rate of spacer acquisition at the expense of increased spacer uptake from the bacterial genome (Heler et al. 2017). This mutation has not been observed in any Cas9 protein so far and thus indicates the importance of avoiding autoimmune spacer sequences and establishes Cas9's role in propagating daughter cells with only functionally relevant spacers (Wei et al. 2015b; Heler et al. 2015, 2017).

Another interesting feature of adaptation is "primed adaptation," a process by which preexisting spacers enable faster acquisition of new spacers in a CRISPR system (Swarts et al. 2012; Datsenko et al. 2012). This mechanism has been well established in type I systems, where escaper phages cause stalled interference that triggers new spacer acquisition through a variety of pathways (Xue et al. 2015, 2016; Künne et al. 2016). Primed adaptation occurs at a higher frequency and with higher efficiency than naïve adaptation in type I systems (Datsenko et al. 2012). Recent studies are pointing to features in type II-A adaptation that resemble primed adaptation (Nicholson et al. 2018; Nussenzweig et al. 2019). Bioinformatic analysis of spacer distribution showed evidence of priming in several CRISPR systems, including types II-A and II-C, with the existence of a directionality in spacer selection in certain subsets (Nicholson et al. 2018). Experimental analysis testing two different type II-A systems, from *Streptococcus pyogenes* and *Streptococcus thermophilus*, showed an increased uptake of new spacers when an active Cas9 was able to induce targeted DNA cleavage with a preexisting spacer (Nussenzweig et al. 2019). A difference noted in type II systems is that primed adaptation occurs with a completely complementary crRNA, rather than only with escaper phages, implicating that type II-A systems proactively acquire new spacers for protection against future escaper phages (Nussenzweig et al. 2019).

It is unknown how Cas9 specifically interacts with the other Cas proteins during spacer acquisition, but binding assays have shown that Cas9 interacts and forms a stable complex with Csn2 in vitro but not with Cas1 or Cas2 (Ka et al. 2018). Csn2 has been proposed to protect the prespacer from degradation during the adaptation process as it has been shown to bind dsDNA ends, but the mechanism by which it assists prespacer integration is unclear (Arslan et al. 2013; Wilkinson et al. 2019). Cryo-electron microscopy structures of Cas proteins from the type II-A CRISPR3 locus of *Streptococcus thermophilus* revealed a $Csn2_4$-$Cas1_8$-$Cas2_4$ prespacer capture complex protecting ~30 base pairs of dsDNA (Wilkinson et al. 2019). It is still unclear how this observed complex functions in in vivo settings. Based on the current literature, it is speculated that the $Cas1_4$-$Cas2_2$ heterohexameric integration complex acquires the prespacer in a form that is ready to proceed to the chemical step of integration from a complex composed of all four type II-A Cas proteins (Jakhanwal et al. 2021).

2.2 Cas1-Cas2 Is the Minimal Subunit for Site-Specific Prespacer Integration

Prespacers are integrated into the repeat-spacer array by the integration complex through a transesterification reaction (Nuñez et al. 2014). The Cas1s distal to the Cas2 dimer catalyze the transesterification between the 3′-hydroxyl of the prespacer strand and the phosphate backbone of the genomic DNA (Fig. 2b) (Nuñez et al. 2015). The Cas2 dimer holds the prespacer DNA through interactions with the phosphate backbone of the DNA (Nuñez et al. 2014; Xiao et al. 2017). The optimal length of prespacer is variable depending on the organism and has been linked to the size and orientation of the Cas2 dimer in the integration complex, which in turn affects the distance between the Cas1 active sites for catalysis (Xiao et al. 2017). The integration complex is the minimal subunit for site-specific prespacer integration in type II-A systems when the correctly processed prespacer is available for integration (Wright and Doudna 2016).

Prespacer integration is dependent on sequence-specific interactions between the integration complex and the leader region of DNA, the region immediately upstream of the first repeat of the CRISPR array (Jansen et al. 2002; Pourcel et al. 2005; Yosef et al. 2012; Wei et al. 2015a). Integration of a new prespacer occurs preferentially at the first repeat of the repeat-spacer array, a position referred to as the leader-repeat junction (Fig. 3) (Barrangou et al. 2007; Yosef et al. 2012). The sequence immediately upstream of the first repeat was found to be crucial for on-target integration, and crystal structures of the integration complex bound to a DNA representing the type II-A leader-repeat junction showed that Cas1 makes sequence specific contacts with the bases of the genomic DNA through a conserved alpha helix called the leader-recognition helix (Fig. 2b, c) (Wei et al. 2015a; McGinn and Marraffini 2016; Xiao et al. 2017; Budhathoki et al. 2020). If the sequence of the leader-repeat junction is mutated, ectopic spacer integration occurs, mainly at other repeat-spacer junctions within the repeat-spacer array (McGinn and Marraffini 2016).

Two transesterification reactions at both ends of the first repeat, coinciding with the leader-repeat and the repeat-spacer junctions, are essential for the complete integration of a prespacer into a CRISPR array. In type II-A systems, integration preferentially occurs at guanine nucleotides in both these locations (Kim et al. 2019). Once Cas1 has made sequence-specific contacts with the leader-repeat junction, a molecular-ruler mechanism defines the second integration site (Kim et al. 2019). The two integration reactions are not always sequential (Van Orden et al. 2020). The transesterification reaction is reversible at either the leader-repeat junction or the repeat-spacer junction, which has biological implications (Ma et al. 2021).

2.3 Mechanisms to Ensure Fidelity and Repair During CRISPR Adaptation

Integration at the leader-repeat junction followed by the repeat-spacer junction is essential for maintaining fidelity of prespacer integration and avoid insertions into sites within the genome that will disrupt the functionality of genes. While certain type II-A systems inherently possess a sequential order, leader-repeat junction followed by repeat-spacer integration, some others can integrate at either of these sites equally well under in vitro conditions (Kim et al. 2019; Budhathoki et al. 2020; Van Orden et al. 2020). Several mechanisms that are not fully characterized are proposed to ensure the fidelity of site-specific spacer insertion. One such mechanism to establish the fidelity is the ability of Cas1-Cas2 proteins to promote disintegration of the prespacer if the first integration occurs at a site other than the leader-repeat junction (Nuñez et al. 2015; Wright and Doudna 2016). Recently, disintegration of prespacers is being proposed as a prerequisite for the second site insertion to relieve topological stress on the genomic DNA during the prespacer integration as well as to initiate DNA repair mechanisms to fill in and seal the gaps following a new prespacer integration (Ma et al. 2021).

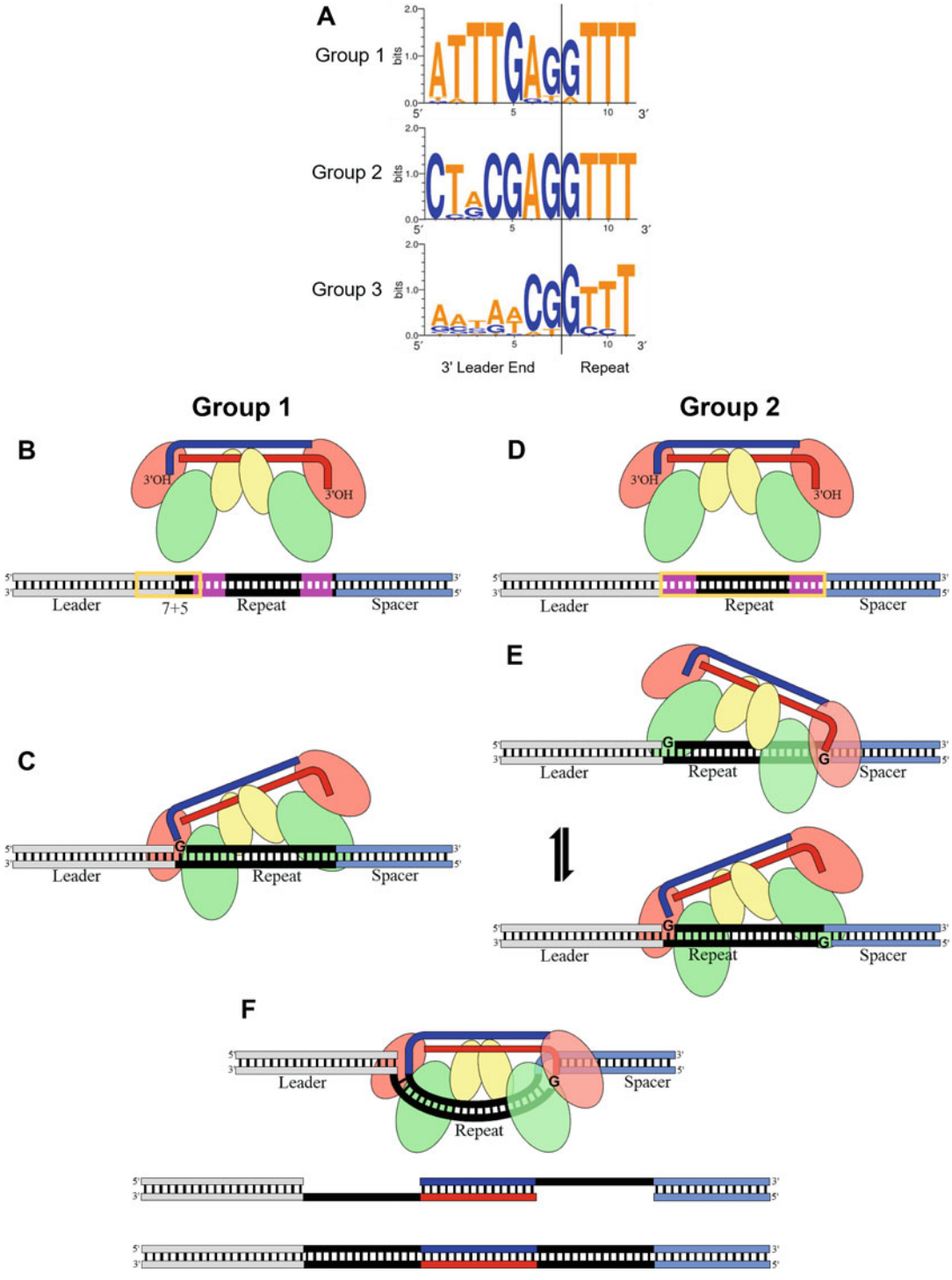

Fig. 3 Type II-A can be further divided into three distinct subgroups with differing integration mechanisms. (**a**) The three groups within type II-A are divided based on the sequence conservation at the 3′ end of the leader region. Group 1 has the consensus sequence of ATTTGAG. Group 2 has the consensus sequence of CTRCGAG, where R is any purine. Group 3 has the consensus sequence of NNNNNCG, where N is any nucleotide. (**b**) Group 1 target recognition occurs with a minimal unit of 7 nt of the leader and 5 nt of the repeat. (**c**) Group 1 prespacer integration ensues with the first of two sequential transesterification reactions occurring at the leader-

Integration of the prespacer leaves ssDNA gaps composed of the repeat region on either side of the newly integrated spacer (Fig. 3f). These gaps are believed to be repaired by cellular polymerases and ligases, leading to the addition of a new repeat sequence upstream of the newly integrated spacer. Recent evidence has shown that resolution of the post-synaptic complex created during prespacer integration and subsequent gap repair can occur in tandem through transcription-coupled DNA repair (Budhathoki et al. 2020). smFRET experiments observed that the Cas1-Cas2 heterohexameric complex remained bound to the fully integrated prespacer in a post-synaptic complex. It is postulated that transcription-coupled DNA repair can play a significant role in the resolution of the post-synaptic complex as transcription with RNA polymerase occurs more frequently than replication with DNA polymerase in a bacterial cell (Budhathoki et al. 2020).

2.4 Leader-Repeat as a Motif for Site-Specific Spacer Insertion in Type II-A Systems

As mentioned previously, the leader-repeat junction is crucial for providing site-specificity and independency of prespacer integration by the type II-A Cas1-Cas2 proteins. Recent studies have found that based on the sequence conservation of the last seven nucleotides at the 3′ end of the leader region, the type II-A CRISPR systems can be divided into three distinct groups, each with a distinct mechanism for prespacer insertion (Fig. 3a) (Van Orden et al. 2017, 2020). Interestingly, phylogenetic analyses of the leader-repeat junction and each of the four Cas proteins from a comprehensive list of type II-A systems segregated them into the same three groups based on the leader-repeat conservation (Van Orden et al. 2017). This indicates coevolution of the leader-repeat sequence and the Cas proteins in the three groups of type II-A systems and the sequence-specific protein-DNA interactions within each group for efficient prespacer integration (Van Orden et al. 2017, 2020). Despite all proteins being of the type II-A loci, Cas1-Cas2 proteins are not cross-compatible between the type II-A subgroups. Cas1 from one type II-A group will not form an integration complex with Cas2 from another group under in vitro conditions (Van Orden et al. 2020).

When investigating the differences in integration mechanism between these three groups, it was found that each group possesses a distinct biochemical mechanism for prespacer insertion (Van Orden et al. 2020). Groups 1 and 2, which both manifested strong leader-repeat conservation, can perform efficient prespacer integration under in vitro conditions using the isolated Cas1-Cas2 proteins from each group (Wright and Doudna 2016; Kim et al. 2019; Van Orden et al. 2020). Group 3 systems have the least conserved 3′ leader sequence, and biochemical analysis showed little to no activity by this Cas1-Cas2 complex in integrating into its cognate leader-repeat sequence under in vitro conditions (Van Orden et al. 2020). Group 2 Cas1-Cas2 was found to be able to perform the two prespacer integration steps, at the leader-repeat junction and at the repeat-spacer junction, independent of the other reaction (Fig. 3e) (Budhathoki et al. 2020; Van Orden et al. 2020). Group 1 Cas1-Cas2, however,

Fig. 3 (continued) repeat junction. (**d**) Group 2 target recognition occurs with the whole repeat region. (**e**) Both leader-repeat and repeat-spacer transesterification reactions can occur independent of the other reaction. Disintegration may occur if the repeat-spacer reaction occurs first or integration happens at an off-target site. (**f**) Full site integration is achieved when the prespacer is integrated at both the leader-repeat and repeat-spacer junctions. In group 1, the second of two sequential transesterification reactions occur at the repeat-spacer junction and requires the whole repeat sequence for this reaction to proceed. Group 2 which does not follow a sequential pattern completes the second reaction to attach both the ends of the prespacer. The integration complex disengages and leaves single-stranded gaps on either side of the newly integrated spacer. These gaps are proposed to be filled and ligated by host polymerases and repair machinery

follows a sequential order in the process with the leader-repeat integration preceding the repeat-spacer integration under in vitro conditions (Fig. 3c) (Kim et al. 2019; Van Orden et al. 2020).

The sequence requirements of the 3′ leader end in integration varies between the type II-A subgroups (Van Orden et al. 2020). Under in vitro conditions, while group 1 Cas1-Cas2 strictly requires a 12-nt-long leader-repeat junction (7 nt of the leader and 5 nt of the repeat) to catalyze the prespacer insertion at the leader-repeat junction (Fig. 3b), group 2 Cas1-Cas2 can perform independent leader-repeat and repeat-spacer insertions with the minimum requirement of a whole repeat region (36 nt long) without the need for the conserved leader end (Fig. 3c) (Wei et al. 2015a; Budhathoki et al. 2020; Van Orden et al. 2020). Addition of extra nucleotides at the 3′ leader end, along with the full repeat, drastically improves the efficiency of integration by the group 2 Cas1-Cas2 complex (Van Orden et al. 2020). This is distinct compared to the complete absence of integration in the group 1 system without the conserved leader-repeat junction sequence (Van Orden et al. 2020). Even though leader-repeat insertion requires only a 12-nt leader-repeat conservation in the group 1 system, the repeat-spacer insertion needs the full 36-nt repeat region along with the 3′ leader end (Van Orden et al. 2020). These observations point to the availability of different natural mechanisms to perform prespacer integration, while maintaining sequence specificity, which will enable discrete biotechnological applications.

Repeat sequences themselves can contribute differently to prespacer insertion in the different type II-A groups. The repeat regions are unique due to their palindromic sequence patterns. In several systems, the repeat sequence possesses inner or terminal inverted repeats (usually one set of two inverted sequences per repeat). The inverted repeat regions are crucial for the group 2 integration complex to recognize the genomic DNA, and integration is completely abolished when both inverted repeat sequences are altered (Fig. 3c) (Wright and Doudna 2016; Xiao et al. 2017; Budhathoki et al. 2020). Group 1 inverted repeats do not impact integration specificity; however, it does affect the efficiency of the repeat-spacer insertion (Fig. 3b) (Kim et al. 2019). This agrees with previous observations of target sequence requirements for each of these subgroups with group 2 needing the full repeat for insertion, while group 1 requires only the conserved leader-repeat junction (Kim et al. 2019; Budhathoki et al. 2020; Van Orden et al. 2020).

Significant differences are also seen in the prespacer forms that could be integrated by isolated Cas1-Cas2 proteins belonging to each type II-A group. Group 1 can efficiently integrate prespacers containing either 4- or 5-nt symmetrical overhangs at the 3′ ends (Van Orden et al. 2020). Group 2 is more robust as it can efficiently integrate prespacers with splayed, symmetrical, and nonsymmetrical overhangs at the 3′ end (Van Orden et al. 2020). This ability to integrate multiple forms of prespacer makes group 2 more autonomous than group 1.

3 Integration in Other CRISPR Systems

Of the other CRISPR types, integration has been most studied in type I-E from *Escherichia coli* K12. In type I-E systems, a $Cas1_4$-$Cas2_2$ heterohexameric complex performs spacer integration with important differences. Spacer acquisition begins with the Cas1-Cas2 complex recognizing the PAM sequence (Wang et al. 2015; Yoganand et al. 2019). Cas1 is able to recognize the cognate PAM sequence through sequence specific interactions with the PAM-complementary sequence and does not depend on other Cas proteins for this function (Wang et al. 2015; Yoganand et al. 2019). This independent function allows for the overexpression of type I-E Cas1 and Cas2 alone, with no other Cas proteins, to facilitate the insertion of functionally relevant spacers in a bacterial system devoid of any CRISPR types (Yosef et al. 2012; Nuñez et al. 2014).

Another important distinction in type I-E systems is the dependence of the integration complex on another cellular protein, integration host

factor (IHF), to enable sequence-specific prespacer insertion in vivo (Nuñez et al. 2016; Yoganand et al. 2017). IHF binds to a conserved motif ahead of the leader-repeat junction and bends the DNA, inducing a 180° turn (Nuñez et al. 2016; Yoganand et al. 2017). This large bend also allows an upstream DNA motif to interact with a secondary site on Cas1 and enables protein-protein interactions between IHF and Cas1 essential for sequence-specific spacer integration (Wright et al. 2017). Due to this requirement of an accessory protein, 60 nts upstream of the first repeat are required for site-specific prespacer integration in type I-E systems, as opposed to 7 nts required for some type II-A systems (Yosef et al. 2012; Van Orden et al. 2020). Interestingly, a recent bioinformatic analysis showed that only 25% of type I-E leaders have an IHF binding site, indicating IHF is not a universal mechanism in type I-E systems for establishing site specificity during spacer insertion (Santiago-Frangos et al. 2021). Among the systems lacking IHF motifs, as many as 20% of type I-E system possess a conserved leader-anchoring motif, a feature similar to the conserved 3′ leader end present in type II-A systems, though the sequences vary between the two (Santiago-Frangos et al. 2021). This work again shows that there are distinct differences in spacer integration specificity mechanisms even within one CRISPR subtype.

The roles of other Cas proteins in CRISPR adaptation have been established in some CRISPR systems. Cas4 is found in certain type I and type V subtypes as well as in type II-B, but its activity has only been characterized in a few type I systems (Makarova et al. 2020a). Cas4 is a nuclease that participates in PAM-dependent prespacer processing during adaptation in types I-A, I-C, and I-D, and it forms a complex with Cas1 and Cas2 (Rollie et al. 2018; Kieper et al. 2018; Lee et al. 2018, 2019). In certain archaea, Cas4 has been shown to define orientation of the spacer during integration (Shiimori et al. 2018). While Cas9's role in PAM selection is established in type II-A systems, the mechanisms defining spacer orientation is still unknown (Wei et al. 2015b; Heler et al. 2015).

Similar to the association of type II-A system with AddAB repair proteins for spacer sources, genome maintenance components, such as RecBCD and SbcCD, have been linked to naïve vs primed adaptation in type I-E systems (Radovčić et al. 2018; Kurilovich et al. 2019; Levy et al. 2015). It was shown that during naïve adaptation in *E. coli*, spacers are preferentially extracted from replication forks that are repaired by the RecBCD complex and that the presence of high-density *chi* sites in the host DNA prevents self-spacer insertion (Levy et al. 2015). It was shown that double mutants lacking RecB-SbcD or RecB-RecJ, but not single mutants, significantly inhibit primed adaptation, indicating redundancy in the mechanism of RecBCD and SbcCD in spacer acquisition (Kurilovich et al. 2019). As mentioned previously, primed adaption occurs in type I systems where spacers are acquired at a faster rate from the stalled interference complex due to mutations in the escaper phages that prevents cleavage of the intruder DNA (Xue et al. 2015, 2016; Künne et al. 2016).

While most focus on the characterization of CRISPR adaptation has been in type I-E and type II-A systems found in bacteria, recent studies have started to reveal the molecular mechanisms of adaptation in CRISPR-Cas systems found in archaea (Rollie et al. 2018; Shiimori et al. 2018; Makarova et al. 2020b; Stachler et al. 2020; Kolesnik et al. 2021). Although the catalytic insertion of prespacers by Cas1-Cas2 is believed to be universally conserved, other features such as the requirement of an accessory, non-Cas protein for prespacer processing, and/or site-specific insertion, the role of Cas4 in PAM recognition, prespacer trimming and ensuring directionality of prespacer insertion, and the ability of certain systems to derive prespacers from RNA targets are different from CRISPR adaptation in bacteria (Hale et al. 2009; Rollie et al. 2018; Shiimori et al. 2018; Artamonova et al. 2020; Stachler et al. 2020; Li et al. 2021).

4 Conclusion

Prespacer integration has been well characterized in type II-A CRISPR-Cas systems; however, there is still much to uncover about the rest of the adaptation process in type II-A as well as in types II-B and II-C.

Adaptation in type II-B systems has not been demonstrated so far. While type II-B systems contain Cas4 in the *cas* operon, similar to type I systems, the roles of Cas4 and Cas1-Cas2 as well as the DNA sequence requirements in type II-B systems are currently unknown (Chylinski et al. 2014; Shiimori et al. 2018).

Type II-C is unique among type II systems since it possesses only Cas9, Cas1, and Cas2, though there is a secondary subtype, type II-C2, found in archaea that contains Cas4 (Makarova et al. 2020a). A few studies have investigated the nature of adaptation in type II-C bacterial systems, examples being the observed adaptation in *Campylobacter jejuni* when infected with Class III *Campylobacter* phages that contain a phage-encoded Cas4-like protein and naïve adaptation observed in *Riemerella anatipestifer* (Hooton and Connerton 2015; He et al. 2018). Further studies are needed to determine the differences in activity of Cas1 and Cas2 between the different type II subtypes.

Available mechanisms of prespacer integration show similarities and differences between the different CRISPR types. Uncovering such mechanisms is important for biotechnological advancements in cell barcoding and cell-based information storage. A large gap still exists in terms of deciphering all the distinct steps essential for adaptation, especially regarding the roles played by cellular proteins and preparation of prespacers, and future research will be needed to develop a comprehensive picture of CRISPR adaptation.

Acknowledgments Tamara Flusche was supported partly by grants from the National Science Foundation (MCB-1716423, awarded to Dr. Rakhi Rajan) and by an Institutional Development Award (IDeA) from the National Institute of General Medical Sciences of the National Institutes of Health (grant number P20GM103640 to Dr. Ann West).

Conflicts of Interest The authors declare no competing commercial or financial interest.

References

Abudayyeh OO, Gootenberg JS, Konermann S, Joung J, Slaymaker IM, Cox DBT, Shmakov S, Makarova KS, Semenova E, Minakhin L, Severinov K, Regev A, Lander ES, Koonin EV, Zhang F (2016) C2c2 is a single-component programmable RNA-guided RNA-targeting CRISPR effector. Science 353(6299): aaf5573. https://doi.org/10.1126/science.aaf5573

Arslan Z, Wurm R, Brener O, Ellinger P, Nagel-Steger L, Oesterhelt F, Schmitt L, Willbold D, Wagner R, Gohlke H, Smits SHJ, Pul Ü (2013) Double-strand DNA end-binding and sliding of the toroidal CRISPR-associated protein Csn2. Nucleic Acids Res 41(12):6347–6359. https://doi.org/10.1093/nar/gkt315

Artamonova D, Karneyeva K, Medvedeva S, Klimuk E, Kolesnik M, Yasinskaya A, Samolygo A, Severinov K (2020) Spacer acquisition by Type III CRISPR–Cas system during bacteriophage infection of Thermus thermophilus. Nucleic Acids Res 48(17):9787–9803. https://doi.org/10.1093/nar/gkaa685

Barrangou R, Fremaux C, Deveau H, Richards M, Boyaval P, Moineau S, Romero DA, Horvath P (2007) CRISPR provides acquired resistance against viruses in prokaryotes. Science 315(5819):1709–1712. https://doi.org/10.1126/science.1138140

Beloglazova N, Brown G, Zimmerman MD, Proudfoot M, Makarova KS, Kudritska M, Kochinyan S, Wang S, Chruszcz M, Minor W, Koonin EV, Edwards AM, Savchenko A, Yakunin AF (2008) A novel family of sequence-specific endoribonucleases associated with the clustered regularly interspaced short palindromic repeats*. J Biol Chem 283(29):20361–20371. https://doi.org/10.1074/jbc.M803225200

Bolotin A, Quinquis B, Sorokin A, Ehrlich SD (2005) Clustered regularly interspaced short palindrome repeats (CRISPRs) have spacers of extrachromosomal origin. Microbiology 151(8):2551–2561. https://doi.org/10.1099/mic.0.28048-0

Brouns SJJ, Jore MM, Lundgren M, Westra ER, Slijkhuis RJH, Snijders APL, Dickman MJ, Makarova KS, Koonin EV, van der Oost J (2008) Small CRISPR RNAs guide antiviral defense in prokaryotes. Science 321(5891):960–964. https://doi.org/10.1126/science.1159689

Budhathoki JB, Xiao Y, Schuler G, Hu C, Cheng A, Ding F, Ke A (2020) Real-time observation of CRISPR spacer acquisition by Cas1–Cas2 integrase. Nat Struct Mol Biol 27(5):489–499. https://doi.org/10.1038/s41594-020-0415-7

Chylinski K, Le Rhun A, Charpentier E (2013) The tracrRNA and Cas9 families of Type II CRISPR-Cas immunity systems. RNA Biol 10(5):726–737. https://doi.org/10.4161/rna.24321

Chylinski K, Makarova KS, Charpentier E, Koonin EV (2014) Classification and evolution of Type II CRISPR-Cas systems. Nucleic Acids Res 42(10): 6091–6105. https://doi.org/10.1093/nar/gku241

Datsenko KA, Pougach K, Tikhonov A, Wanner BL, Severinov K, Semenova E (2012) Molecular memory of prior infections activates the CRISPR/Cas adaptive bacterial immunity system. Nat Commun 3(1):945. https://doi.org/10.1038/ncomms1937

Deltcheva E, Chylinski K, Sharma CM, Gonzales K, Chao Y, Pirzada ZA, Eckert MR, Vogel J, Charpentier E (2011) CRISPR RNA maturation by trans-encoded small RNA and host factor RNase III. Nature 471(7340):602–607. https://doi.org/10.1038/nature09886

Deveau H, Barrangou R, Garneau JE, Labonté J, Fremaux C, Boyaval P, Romero DA, Horvath P, Moineau S (2008) Phage response to CRISPR-encoded resistance in streptococcus thermophilus. J Bacteriol 190(4):1390–1400. https://doi.org/10.1128/JB.01412-07

East-Seletsky A, O'Connell MR, Knight SC, Burstein D, Cate JHD, Tjian R, Doudna JA (2016) Two distinct RNase activities of CRISPR-C2c2 enable guide-RNA processing and RNA detection. Nature 538(7624):270–273. https://doi.org/10.1038/nature19802

Ellinger P, Arslan Z, Wurm R, Tschapek B, MacKenzie C, Pfeffer K, Panjikar S, Wagner R, Schmitt L, Gohlke H, Pul Ü, Smits SHJ (2012) The crystal structure of the CRISPR-associated protein Csn2 from Streptococcus agalactiae. J Struct Biol 178(3):350–362. https://doi.org/10.1016/j.jsb.2012.04.006

Erdmann S, Garrett RA (2012) Selective and hyperactive uptake of foreign DNA by adaptive immune systems of an archaeon via two distinct mechanisms. Mol Microbiol 85(6):1044–1056. https://doi.org/10.1111/j.1365-2958.2012.08171.x

Faure G, Makarova KS, Koonin EV (2019) CRISPR–Cas: complex functional networks and multiple roles beyond adaptive immunity. J Mol Biol 431(1):3–20. https://doi.org/10.1016/j.jmb.2018.08.030

Garneau JE, Dupuis M-È, Villion M, Romero DA, Barrangou R, Boyaval P, Fremaux C, Horvath P, Magadán AH, Moineau S (2010) The CRISPR/Cas bacterial immune system cleaves bacteriophage and plasmid DNA. Nature 468(7320):67–71. https://doi.org/10.1038/nature09523

Gasiunas G, Barrangou R, Horvath P, Siksnys V (2012) Cas9–crRNA ribonucleoprotein complex mediates specific DNA cleavage for adaptive immunity in bacteria. Proc Natl Acad Sci 109(39):E2579–E2586. https://doi.org/10.1073/pnas.1208507109

Haft DH, Selengut J, Mongodin EF, Nelson KE (2005) A Guild of 45 CRISPR-Associated (Cas) protein families and multiple CRISPR/Cas subtypes exist in prokaryotic genomes. PLOS Comput Biol 1(6):e60. https://doi.org/10.1371/journal.pcbi.0010060

Hale CR, Zhao P, Olson S, Duff MO, Graveley BR, Wells L, Terns RM, Terns MP (2009) RNA-guided RNA cleavage by a CRISPR RNA-Cas protein complex. Cell 139(5):945–956. https://doi.org/10.1016/j.cell.2009.07.040

He Y, Wang M, Liu M, Huang L, Liu C, Zhang X, Yi H, Cheng A, Zhu D, Yang Q, Wu Y, Zhao X, Chen S, Jia R, Zhang S, Liu Y, Yu Y, Zhang L (2018) Cas1 and Cas2 from the Type II-C CRISPR-Cas system of Riemerella anatipestifer are required for spacer acquisition. Front Cell Infect Microbiol 8. https://doi.org/10.3389/fcimb.2018.00195

Heler R, Samai P, Modell JW, Weiner C, Goldberg GW, Bikard D, Marraffini LA (2015) Cas9 specifies functional viral targets during CRISPR–Cas adaptation. Nature 519(7542):199–202. https://doi.org/10.1038/nature14245

Heler R, Wright AV, Vucelja M, Bikard D, Doudna JA, Marraffini LA (2017) Mutations in Cas9 enhance the rate of acquisition of viral spacer sequences during the CRISPR-Cas immune response. Mol Cell 65(1): 168–175. https://doi.org/10.1016/j.molcel.2016.11.031

Hooton SPT, Connerton IF (2015) Campylobacter jejuni acquire new host-derived CRISPR spacers when in association with bacteriophages harboring a CRISPR-like Cas4 protein. Front Microbiol 5:744. https://doi.org/10.3389/fmicb.2014.00744

Huo Y, Nam KH, Ding F, Lee H, Wu L, Xiao Y, Farchione MD, Zhou S, Rajashankar K, Kurinov I, Zhang R, Ke A (2014) Structures of CRISPR Cas3 offer mechanistic insights into Cascade-activated DNA unwinding and degradation. Nat Struct Mol Biol 21(9):771–777. https://doi.org/10.1038/nsmb.2875

Ishino Y, Shinagawa H, Makino K, Amemura M, Nakata A (1987) Nucleotide sequence of the iap gene, responsible for alkaline phosphatase isozyme conversion in Escherichia coli, and identification of the gene product. J Bacteriol 169(12):5429–5433

Jakhanwal S, Cress BF, Maguin P, Lobba MJ, Marraffini LA, Doudna JA (2021) A CRISPR-Cas9–integrase complex generates precise DNA fragments for genome integration. Nucleic Acids Res 49(6):3546–3556. https://doi.org/10.1093/nar/gkab123

Jansen R, van Embden JDA, Gaastra W, Schouls LM (2002) Identification of genes that are associated with DNA repeats in prokaryotes. Mol Microbiol 43(6): 1565–1575. https://doi.org/10.1046/j.1365-2958.2002.02839.x

Jinek M, Chylinski K, Fonfara I, Hauer M, Doudna JA, Charpentier E (2012) A programmable dual RNA-guided DNA endonuclease in adaptive bacterial immunity. Science 337(6096):816–821. https://doi.org/10.1126/science.1225829

Jore MM, Lundgren M, van Duijn E, Bultema JB, Westra ER, Waghmare SP, Wiedenheft B, Pul Ü, Wurm R, Wagner R, Beijer MR, Barendregt A, Zhou K, Snijders APL, Dickman MJ, Doudna JA, Boekema EJ, Heck AJR, van der Oost J, Brouns SJJ (2011) Structural basis for CRISPR RNA-guided DNA recognition by Cascade. Nat Struct Mol Biol 18(5):529–536. https://doi.org/10.1038/nsmb.2019

Ka D, Jang DM, Han BW, Bae E (2018) Molecular organization of the Type II-A CRISPR adaptation module and its interaction with Cas9 via Csn2. Nucleic Acids

Res 46(18):9805–9815. https://doi.org/10.1093/nar/gky702

Kazlauskiene M, Kostiuk G, Venclovas Č, Tamulaitis G, Siksnys V (2017) A cyclic oligonucleotide signaling pathway in Type III CRISPR-Cas systems. Science 357(6351):605–609. https://doi.org/10.1126/science.aao0100

Kieper SN, Almendros C, Behler J, McKenzie RE, Nobrega FL, Haagsma AC, Vink JNA, Hess WR, Brouns SJJ (2018) Cas4 Facilitates PAM-Compatible Spacer Selection during CRISPR Adaptation. Cell Rep 22(13):3377–3384. https://doi.org/10.1016/j.celrep.2018.02.103

Kim JG, Garrett S, Wei Y, Graveley BR, Terns MP (2019) CRISPR DNA elements controlling site-specific spacer integration and proper repeat length by a Type II CRISPR–Cas system. Nucleic Acids Res 47(16):8632–8648. https://doi.org/10.1093/nar/gkz677

Kolesnik MV, Fedorova I, Karneyeva KA, Artamonova DN, Severinov KV (2021) Type III CRISPR-Cas systems: deciphering the most complex prokaryotic immune system. Biochem Mosc 86(10):1301–1314. https://doi.org/10.1134/S0006297921100114

Koonin EV, Makarova KS (2018) Discovery of oligonucleotide signaling mediated by CRISPR-associated polymerases solves two puzzles but leaves an enigma. ACS Chem Biol 13(2):309–312. https://doi.org/10.1021/acschembio.7b00713

Künne T, Kieper SN, Bannenberg JW, Vogel AIM, Miellet WR, Klein M, Depken M, Suarez-Diez M, Brouns SJJ (2016) Cas3-derived target DNA degradation fragments fuel primed CRISPR adaptation. Mol Cell 63(5):852–864. https://doi.org/10.1016/j.molcel.2016.07.011

Kurilovich E, Shiriaeva A, Metlitskaya A, Morozova N, Ivancic-Bace I, Severinov K, Savitskaya E (2019) Genome maintenance proteins modulate autoimmunity mediated primed adaptation by the Escherichia coli Type I-E CRISPR-Cas system. Genes 10(11):872. https://doi.org/10.3390/genes10110872

Lee H, Zhou Y, Taylor DW, Sashital DG (2018) Cas4-dependent prespacer processing ensures high-fidelity programming of CRISPR arrays. Mol Cell 70(1):48–59.e5. https://doi.org/10.1016/j.molcel.2018.03.003

Lee H, Dhingra Y, Sashital DG (2019) The Cas4-Cas1-Cas2 complex mediates precise prespacer processing during CRISPR adaptation. eLife 8:e44248. https://doi.org/10.7554/eLife.44248

Lemak S, Beloglazova N, Nocek B, Skarina T, Flick R, Brown G, Popovic A, Joachimiak A, Savchenko A, Yakunin AF (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(46):17476–17487. https://doi.org/10.1021/ja408729b

Levy A, Goren MG, Yosef I, Auster O, Manor M, Amitai G, Edgar R, Qimron U, Sorek R (2015) CRISPR adaptation biases explain preference for acquisition of foreign DNA. Nature 520(7548):505–510. https://doi.org/10.1038/nature14302

Li M, Gong L, Cheng F, Yu H, Zhao D, Wang R, Wang T, Zhang S, Zhou J, Shmakov SA, Koonin EV, Xiang H (2021) Toxin-antitoxin RNA pairs safeguard CRISPR-Cas systems. Science 372(6541):eabe5601. https://doi.org/10.1126/science.abe5601

Ma C-H, Javanmardi K, Finkelstein IJ, Jayaram M (2021) Disintegration promotes protospacer integration by the Cas1-Cas2 complex. eLife 10:e65763. https://doi.org/10.7554/eLife.65763

Maguin P, Varble A, Modell JW, Marraffini LA (2022) Cleavage of viral DNA by restriction endonucleases stimulates the Type II CRISPR-Cas immune response. Mol Cell 82(5):907–919.e7. https://doi.org/10.1016/j.molcel.2022.01.012

Makarova KS, Aravind L, Grishin NV, Rogozin IB, Koonin EV (2002) A DNA repair system specific for thermophilic Archaea and bacteria predicted by genomic context analysis. Nucleic Acids Res 30(2):482–496. https://doi.org/10.1093/nar/30.2.482

Makarova KS, Grishin NV, Shabalina SA, Wolf YI, Koonin EV (2006) A putative RNA-interference-based immune system in prokaryotes: computational analysis of the predicted enzymatic machinery, functional analogies with eukaryotic RNAi, and hypothetical mechanisms of action. Biol Direct 1:7. https://doi.org/10.1186/1745-6150-1-7

Makarova KS, Wolf YI, Alkhnbashi OS, Costa F, Shah SA, Saunders SJ, Barrangou R, Brouns SJJ, Charpentier E, Haft DH, Horvath P, Moineau S, Mojica FJM, Terns RM, Terns MP, White MF, Yakunin AF, Garrett RA, van der Oost J, Backofen R, Koonin EV (2015) An updated evolutionary classification of CRISPR–Cas systems. Nat Rev Microbiol 13(11):722–736. https://doi.org/10.1038/nrmicro3569

Makarova KS, Wolf YI, Koonin EV (2018) Classification and nomenclature of CRISPR-Cas systems: where from here? CRISPR J 1(5):325–336. https://doi.org/10.1089/crispr.2018.0033

Makarova KS, Wolf YI, Iranzo J, Shmakov SA, Alkhnbashi OS, Brouns SJJ, Charpentier E, Cheng D, Haft DH, Horvath P, Moineau S, Mojica FJM, Scott D, Shah SA, Siksnys V, Terns MP, Venclovas Č, White MF, Yakunin AF, Yan W, Zhang F, Garrett RA, Backofen R, van der Oost J, Barrangou R, Koonin EV (2020a) Evolutionary classification of CRISPR–Cas systems: a burst of class 2 and derived variants. Nat Rev Microbiol 18(2):67–83. https://doi.org/10.1038/s41579-019-0299-x

Makarova KS, Wolf YI, Shmakov SA, Liu Y, Li M, Koonin EV (2020b) Unprecedented diversity of unique CRISPR-Cas-related systems and Cas1 homologs in Asgard Archaea. CRISPR J 3(3):156–163. https://doi.org/10.1089/crispr.2020.0012

Marraffini LA, Sontheimer EJ (2010) Self versus non-self discrimination during CRISPR RNA-directed immunity. Nature 463(7280):568–571. https://doi.org/10.1038/nature08703

McGinn J, Marraffini LA (2016) CRISPR-Cas systems optimize their immune response by specifying the site of spacer integration. Mol Cell 64(3):616–623. https://doi.org/10.1016/j.molcel.2016.08.038

Modell JW, Jiang W, Marraffini LA (2017) CRISPR-Cas systems exploit viral DNA injection to establish and maintain adaptive immunity. Nature 544(7648): 101–104. https://doi.org/10.1038/nature21719

Mohr G, Silas S, Stamos JL, Makarova KS, Markham LM, Yao J, Lucas-Elío P, Sanchez-Amat A, Fire AZ, Koonin EV, Lambowitz AM (2018) A reverse transcriptase-Cas1 fusion protein contains a Cas6 domain required for both CRISPR RNA biogenesis and RNA spacer acquisition. Mol Cell 72(4):700–714.e8. https://doi.org/10.1016/j.molcel.2018.09.013

Mojica FJM, Juez G, Rodriguez-Valera F (1993) Transcription at different salinities of Haloferax mediterranei sequences adjacent to partially modified PstI sites. Mol Microbiol 9(3):613–621. https://doi.org/10.1111/j.1365-2958.1993.tb01721.x

Mojica FJM, Díez-Villaseñor C, Soria E, Juez G (2000) Biological significance of a family of regularly spaced repeats in the genomes of archaea, bacteria and mitochondria. Mol Microbiol 36(1):244–246. https://doi.org/10.1046/j.1365-2958.2000.01838.x

Mojica FJM, Díez-Villaseñor C, García-Martínez J, Soria E (2005) Intervening sequences of regularly spaced prokaryotic repeats derive from foreign genetic elements. J Mol Evol 60(2):174–182. https://doi.org/10.1007/s00239-004-0046-3

Mojica FJM, Díez-Villaseñor C, García-Martínez J, Almendros CY (2009) (2009) Short motif sequences determine the targets of the prokaryotic CRISPR defence system. Microbiology 155(3):733–740. https://doi.org/10.1099/mic.0.023960-0

Moya-Beltrán A, Makarova KS, Acuña LG, Wolf YI, Covarrubias PC, Shmakov SA, Silva C, Tolstoy I, Johnson DB, Koonin EV, Quatrini R (2021) Evolution of Type IV CRISPR-Cas systems: insights from CRISPR loci in integrative conjugative elements of Acidithiobacillia. CRISPR J 4(5):656–672. https://doi.org/10.1089/crispr.2021.0051

Nam KH, Kurinov I, 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(35):30759–30768. https://doi.org/10.1074/jbc.M111.256263

Newsom S, Parameshwaran HP, Martin L, Rajan R (2021) The CRISPR-Cas mechanism for adaptive immunity and alternate bacterial functions fuels diverse biotechnologies. Front Cell Infect Microbiol 10: 619763. https://doi.org/10.3389/fcimb.2020.619763

Nicholson TJ, Jackson SA, Croft BI, Staals RHJ, Fineran PC, Brown CM (2018) Bioinformatic evidence of widespread priming in Type I and II CRISPR-Cas systems. RNA Biol 16(4):566–576. https://doi.org/10.1080/15476286.2018.1509662

Niewoehner O, Garcia-Doval C, Rostøl JT, Berk C, Schwede F, Bigler L, Hall J, Marraffini LA, Jinek M (2017) Type III CRISPR–Cas systems produce cyclic oligoadenylate second messengers. Nature 548(7669): 543–548. https://doi.org/10.1038/nature23467

Nuñez JK, Kranzusch PJ, Noeske J, Wright AV, Davies CW, Doudna JA (2014) Cas1–Cas2 complex formation mediates spacer acquisition during CRISPR–Cas adaptive immunity. Nat Struct Mol Biol 21(6): 528–534. https://doi.org/10.1038/nsmb.2820

Nuñez JK, Lee ASY, Engelman A, Doudna JA (2015) Integrase-mediated spacer acquisition during CRISPR–Cas adaptive immunity. Nature 519(7542): 193–198. https://doi.org/10.1038/nature14237

Nuñez JK, Bai L, Harrington LB, Hinder TL, Doudna JA (2016) CRISPR immunological memory requires a host factor for specificity. Mol Cell 62(6):824–833. https://doi.org/10.1016/j.molcel.2016.04.027

Nussenzweig PM, McGinn J, Marraffini LA (2019) Cas9 cleavage of viral genomes primes the acquisition of new immunological memories. Cell Host Microbe 26(4):515–526.e6. https://doi.org/10.1016/j.chom.2019.09.002

Pougach K, Semenova E, Bogdanova E, Datsenko KA, Djordjevic M, Wanner BL, Severinov K (2010) Transcription, processing, and function of CRISPR cassettes in Escherichia coli. Mol Microbiol 77(6): 1367–1379. https://doi.org/10.1111/j.1365-2958.2010.07265.x

Pourcel C, Salvignol G, Vergnaud GY (2005) CRISPR elements in Yersinia pestis acquire new repeats by preferential uptake of bacteriophage DNA, and provide additional tools for evolutionary studies. Microbiology 151(3):653–663. https://doi.org/10.1099/mic.0.27437-0

Radovčić M, Killelea T, Savitskaya E, Wettstein L, Bolt EL, Ivančić-Baće I (2018) CRISPR–Cas adaptation in Escherichia coli requires RecBCD helicase but not nuclease activity, is independent of homologous recombination, and is antagonized by 5′ ssDNA exonucleases. Nucleic Acids Res 46(19): 10173–10183. https://doi.org/10.1093/nar/gky799

Rollie C, Schneider S, Brinkmann AS, Bolt EL, White MF (2015) Intrinsic sequence specificity of the Cas1 integrase directs new spacer acquisition. eLife 4: e08716. https://doi.org/10.7554/eLife.08716

Rollie C, Graham S, Rouillon C, White MF (2018) Prespacer processing and specific integration in a Type I-A CRISPR system. Nucleic Acids Res 46(3): 1007–1020. https://doi.org/10.1093/nar/gkx1232

Samai P, Pyenson N, Jiang W, Goldberg GW, Hatoum-Aslan A, Marraffini LA (2015) Co-transcriptional DNA and RNA Cleavage during Type III CRISPR-Cas Immunity. Cell 161(5):1164–1174. https://doi.org/10.1016/j.cell.2015.04.027

Santiago-Frangos A, Buyukyoruk M, Wiegand T, Krishna P, Wiedenheft B (2021) Distribution and phasing of sequence motifs that facilitate CRISPR adaptation. Curr Biol 31(16):3515–3524.e6. https://doi.org/10.1016/j.cub.2021.05.068

Schmidt F, Cherepkova MY, Platt RJ (2018) Transcriptional recording by CRISPR spacer acquisition from RNA. Nature 562(7727):380–385. https://doi.org/10.1038/s41586-018-0569-1

Shiimori M, Garrett SC, Graveley BR, Terns MP (2018) Cas4 nucleases define the PAM, length, and orientation of DNA fragments integrated at CRISPR loci. Mol Cell 70(5):814–824.e6. https://doi.org/10.1016/j.molcel.2018.05.002

Shipman SL, Nivala J, Macklis JD, Church GM (2016) Molecular recordings by directed CRISPR spacer acquisition. Science 353(6298):aaf1175. https://doi.org/10.1126/science.aaf1175

Shipman SL, Nivala J, Macklis JD, Church GM (2017) CRISPR–Cas encoding of a digital movie into the genomes of a population of living bacteria. Nature 547(7663):345–349. https://doi.org/10.1038/nature23017

Silas S, Mohr G, Sidote DJ, Markham LM, Sanchez-Amat A, Bhaya D, Lambowitz AM, Fire AZ (2016) Direct CRISPR spacer acquisition from RNA by a natural reverse transcriptase–Cas1 fusion protein. Science 351(6276):aad4234. https://doi.org/10.1126/science.aad4234

Sinkunas T, Gasiunas G, Fremaux C, Barrangou R, Horvath P, Siksnys V (2011) Cas3 is a single-stranded DNA nuclease and ATP-dependent helicase in the CRISPR/Cas immune system. EMBO J 30(7):1335–1342. https://doi.org/10.1038/emboj.2011.41

Staals RHJ, Agari Y, Maki-Yonekura S, Zhu Y, Taylor DW, van Duijn E, Barendregt A, Vlot M, Koehorst JJ, Sakamoto K, Masuda A, Dohmae N, Schaap PJ, Doudna JA, Heck AJR, Yonekura K, van der Oost J, Shinkai A (2013) Structure and activity of the RNA-targeting Type III-B CRISPR-Cas complex of Thermus thermophilus. Mol Cell 52(1):135–145. https://doi.org/10.1016/j.molcel.2013.09.013

Stachler A-E, Wörtz J, Alkhnbashi OS, Turgeman-Grott I, Smith R, Allers T, Backofen R, Gophna U, Marchfelder A (2020) Adaptation induced by self-targeting in a Type I-B CRISPR-Cas system. J Biol Chem 295(39):13502–13515. https://doi.org/10.1074/jbc.RA120.014030

Sternberg SH, Redding S, Jinek M, Greene EC, Doudna JA (2014) DNA interrogation by the CRISPR RNA-guided endonuclease Cas9. Nature 507(7490):62–67. https://doi.org/10.1038/nature13011

Swarts DC, Mosterd C, van Passel MWJ, Brouns SJJ (2012) CRISPR interference directs strand specific spacer acquisition. PLoS One 7(4):e35888. https://doi.org/10.1371/journal.pone.0035888

Taylor HN, Laderman E, Armbrust M, Hallmark T, Keiser D, Bondy-Denomy J, Jackson RN (2021) Positioning diverse Type IV structures and functions within Class 1 CRISPR-Cas systems. Front Microbiol 12:671522. https://doi.org/10.3389/fmicb.2021.671522

Van Orden MJ, Klein P, Babu K, Najar FZ, Rajan R (2017) Conserved DNA motifs in the Type II-A CRISPR leader region. PeerJ 5:e3161. https://doi.org/10.7717/peerj.3161

Van Orden MJ, Newsom S, Rajan R (2020) CRISPR Type II-A subgroups exhibit phylogenetically distinct mechanisms for prespacer insertion. J Biol Chem 295(32):10956–10968. https://doi.org/10.1074/jbc.RA120.013554

Vink JNA, Baijens JHL, Brouns SJJ (2021) PAM-repeat associations and spacer selection preferences in single and co-occurring CRISPR-Cas systems. Genome Biol 22(1):281. https://doi.org/10.1186/s13059-021-02495-9

Wang J, Li J, Zhao H, Sheng G, Wang M, Yin M, Wang Y (2015) Structural and mechanistic basis of PAM-dependent spacer acquisition in CRISPR-Cas systems. Cell 163(4):840–853. https://doi.org/10.1016/j.cell.2015.10.008

Wei Y, Chesne MT, Terns RM, Terns MP (2015a) Sequences spanning the leader-repeat junction mediate CRISPR adaptation to phage in Streptococcus thermophilus. Nucleic Acids Res 43(3):1749–1758. https://doi.org/10.1093/nar/gku1407

Wei Y, Terns RM, Terns MP (2015b) Cas9 function and host genome sampling in Type II-A CRISPR–Cas adaptation. Genes Dev 29(4):356–361. https://doi.org/10.1101/gad.257550.114

Wiedenheft B, Zhou K, Jinek M, Coyle SM, Ma W, Doudna JA (2009) Structural basis for DNase activity of a conserved protein implicated in CRISPR-mediated genome defense. Structure 17(6):904–912. https://doi.org/10.1016/j.str.2009.03.019

Wiedenheft B, Lander GC, Zhou K, Jore MM, Brouns SJJ, van der Oost J, Doudna JA, Nogales E (2011) Structures of the RNA-guided surveillance complex from a bacterial immune system. Nature 477(7365):486–489. https://doi.org/10.1038/nature10402

Wilkinson M, Drabavicius G, Silanskas A, Gasiunas G, Siksnys V, Wigley DB (2019) Structure of the DNA-bound spacer capture complex of a Type II CRISPR-Cas system. Mol Cell 75(1):90–101.e5. https://doi.org/10.1016/j.molcel.2019.04.020

Wright AV, Doudna JA (2016) Protecting genome integrity during CRISPR immune adaptation. Nat Struct Mol Biol 23(10):876–883. https://doi.org/10.1038/nsmb.3289

Wright AV, Liu J-J, Knott GJ, Doxzen KW, Nogales E, Doudna JA (2017) Structures of the CRISPR genome integration complex. Science 357(6356):1113–1118. https://doi.org/10.1126/science.aao0679

Xiao Y, Ng S, Nam KH, Ke A (2017) How Type II CRISPR–Cas establish immunity through Cas1–Cas2-mediated spacer integration. Nature 550(7674):137–141. https://doi.org/10.1038/nature24020

Xue C, Seetharam AS, Musharova O, Severinov K, Brouns SJ, Severin AJ, Sashital DG (2015) CRISPR interference and priming varies with individual spacer sequences. Nucleic Acids Res 43(22):10831–10847. https://doi.org/10.1093/nar/gkv1259

Xue C, Whitis NR, Sashital DG (2016) Conformational control of Cascade interference and priming activities in CRISPR immunity. Mol Cell 64(4):826–834. https://doi.org/10.1016/j.molcel.2016.09.033

Yoganand KNR, Sivathanu R, Nimkar S, Anand B (2017) Asymmetric positioning of Cas1–2 complex and

Integration Host Factor induced DNA bending guide the unidirectional homing of protospacer in CRISPR-Cas Type I-E system. Nucleic Acids Res 45(1): 367–381. https://doi.org/10.1093/nar/gkw1151

Yoganand KN, Muralidharan M, Nimkar S, Anand B (2019) Fidelity of prespacer capture and processing is governed by the PAM-mediated interactions of Cas1-2 adaptation complex in CRISPR-Cas Type I-E system. J Biol Chem 294(52):20039–20053. https://doi.org/10.1074/jbc.RA119.009438

Yosef I, Goren MG, Qimron U (2012) Proteins and DNA elements essential for the CRISPR adaptation process in Escherichia coli. Nucleic Acids Res 40(12): 5569–5576. https://doi.org/10.1093/nar/gks216

Zetsche B, Gootenberg JS, Abudayyeh OO, Slaymaker IM, Makarova KS, Essletzbichler P, Volz S, Joung J, van der Oost J, Regev A, Koonin EV, Zhang F (2015) Cpf1 is a single RNA-guided endonuclease of a Class 2 CRISPR-Cas system. Cell 163(3):759–771. https://doi.org/10.1016/j.cell.2015.09.038