NIH Public Access
Author Manuscript
Curr Opin Struct Biol. Author manuscript; available in PMC 2015 February 01.

Published in final edited form as:
Curr Opin Struct Biol. 2014 February ; 0: 106–114. doi:10.1016/j.sbi.2014.01.001.

Fitting CRISPR-associated Cas3 into the Helicase Family Tree

Ryan N. Jackson[1], **Matthew Lavin**[2], **Joshua Carter**[1], and **Blake Wiedenheft**[1]

[1]Department of Microbiology and Immunology, Montana State University, Bozeman MT 59718

[2]Department of Plant Sciences and Plant Pathology, Montana State University, Bozeman MT 59718

Abstract

Helicases utilize NTPs to modulate their binding to nucleic acids and many of these enzymes also unwind DNA or RNA duplexes in an NTP-dependent fashion. These proteins are phylogenetically related but functionally diverse, with essential roles in virtually all aspects of nucleic acid metabolism. A new class of helicases associated with RNA-guided adaptive immune systems in bacteria and archaea has recently been identified. Prokaryotes acquire resistance to invading genetic parasites by integrating short fragments of foreign nucleic acids into repetitive loci in the host chromosome known as CRISPRs (Clustered Regularly Interspaced Short Palindromic Repeats). CRISPR-associated gene 3 (*cas3*) encodes a conserved helicase protein that is essential for phage defense. Here we review recent advances in Cas3 biology, and provide a new phylogenetic framework that positions Cas3 in the helicase family tree. We anticipate that this Cas3 phylogeny will guide future biochemical and structural studies.

Keywords

Helicases; CRISPR; Cas3; adaptive immunity; bacteriophage; bacterial defense

Introduction

Helicases use nucleotide triphosphates (NTPs) to modulate binding to nucleic acids and in many cases these proteins couple the chemical energy of NTP hydrolysis with conformational changes that destabilize, modify and/or unwind nucleic acid structure [1-4]. These enzymes have evolved diverse functions essential for genome replication, repair, transcription, and translation. In humans, defects in helicases are associated with a wide range of diseases including cancer, neurodegenerative diseases, and developmental disorders [5,6]. Beyond their essential roles in genome maintenance and expression, helicases also perform non-canonical functions in RNA surveillance and decay, ribosome biogenesis, mRNA splicing, nuclear export of mRNA, and antiviral defense [7-11].

Cas3 proteins are a newly identified group of helicases that play a central role in many aspects of the CRISPR-mediated adaptive immune systems in bacteria and archaea (For recent reviews see [12-15]). CRISPR loci and their associated (*cas*) genes are remarkably diverse, and phylogenetic studies have identified three major Types (Type I, II and III) and

© 2014 The Authors. Published by Elsevier Ltd. All rights reserved.

correspondence: bwiedenheft@gmail.com.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

NIH-PA Author Manuscript
NIH-PA Author Manuscript
NIH-PA Author Manuscript

NIH-PA Author Manuscript NIH-PA Author Manuscript NIH-PA Author Manuscript

10 distinct subtypes (IA-F, IIA-B, IIIA-B) [16-18]. Cas3 is an essential component of all Type I systems (i.e., IA-F), but most biochemical and genetic studies have been performed with Cas3 proteins from only two subtypes (i.e., IE and IF), and it is unclear if these functional attributes can be generalized to all Cas3 helicases. Here we briefly review the phylogeny and function of helicases and present a new phylogenetic analysis that positions Cas3 proteins in the helicase family tree.

Helicase phylogeny, function, and structure

In the early 1990's, Gorbalenya and Koonin identified a series of conserved amino acid motifs that could be used to recognize and classify helicases [19]. These conserved motifs were used to establish a sequence-based classification system that delineated three main superfamilies (SF1-3) and two smaller helicase groups (DnaB-like and Rho-like). Later, Wigley and colleagues simplified the helicase classification system by renaming the DnaB-like, Rho-like, and AAA+ helicases as SF4, SF5, and SF6 (respectively), and established a naming system that categorized helicases according to their direction of unwinding and nucleic acid substrate preference [1] (**Figure 1A**). Sequence based classification of SF3, SF4, and SF5 provides direct mechanistic insight into the unwinding properties of these helicases. However, SF1 and SF2 helicases are functionally diverse, consisting of members that unwind duplexes with no defined polarity, or do not unwind duplexes at all. In an attempt to streamline the classification scheme for SF1 and SF2 helicases, Jankowsky and colleagues performed a comprehensive phylogenetic analysis of all SF1 and SF2 helicases from yeast, humans, and *E. coli.* They observed robust clustering of helicases into monophyletic clades called ‘families’ [3]. Importantly, each of these families shared mechanistic properties, demonstrating that helicase phylogeny could be used to predict function of SF1 and SF2 enzymes.

In spite of considerable amino acid sequence variability, all helicases share a conserved RecA-like fold that consists of a beta-sheet sandwiched by alpha helices [1]. SF1 and SF2 helicases contain tandem RecA-like folds on a single polypeptide [2,3], while SF3-6 helicases assemble into toroidal ring-like structures from six RecA-like monomers [4] (**Figure 1B and 1C**). In each case, the interface between the RecA-like domains form a cleft that is lined with conserved amino acids that coordinate NTP, Mg, and nucleic acid substrates (**Figure 1B and 1C**). NTP binding is coordinated by amino acids on two adjacent RecA-like domains. This bidentate binding site consists of the Walker A (motif I, phosphate binding-loop) and Walker B (motif II, Mg2+ binding aspartic acid) motifs on one side of the cleft and a conserved arginine (R) on the other [2,4] (**Figure 1A, B and C**). The guanidinium group of the arginine coordinates the gamma-phosphate of the NTP and NTP hydrolysis ‘unhitches’ the arginine from the nucleotide, thereby increasing conformational flexibility between the domains of the helicase core. This NTP-dependent conformational cycling between the closed (NTP bound) and open (unbound) state is the basis of locomotion and/or nucleic acid modification by helicases.

The monomeric helicases (SF1 and SF2) fall into three mechanistic categories: those that unwind duplex substrates with a preferred polarity (all SF1 families, DEAH/RHA, NS3/NPH-II, Ski2-like, RecQ), those that bind or translocate along duplex substrates without unwinding (Rig-I-like, Swi/Snf2, RecG, Type 1 and 3 restriction enzymes), and those that destabilize nucleic acid duplexes locally without directional unwinding (DEAD-box) [3]. Recent structures of DEAD-box and Rig-I like helicases bound to duplex substrates provide significant insight to the similarities and differences of these mechanisms [20-24]. These enzymes share common sequence signatures associated with converting NTP hydrolysis into molecular motion, but distinct structural features (i.e. loops and terminal accessory domains) explain their mechanistic differences.

NIH-PA Author Manuscript NIH-PA Author Manuscript NIH-PA Author Manuscript

Many SF1 and SF2 helicases separate nucleic acid duplexes by using the helicase core to drive a molecular wedge between the two strands of an oncoming duplex (**Figure 1D**). In these systems, the wedge (usually a β-hairpin) is positioned at the leading edge of the separation fork and locomotion is driven by conformational cycling between the NTP bound (closed) and unbound (open) states [3,25-27] (**Figure 1D**). The NTP is coordinated by amino acids located on both the N- and C-terminal RecA domains; thus, NTP serves as a structural staple that stabilizes the 'closed' state. NTP hydrolysis results in two coordinated transitions; the broken phosphate bond releases the conformational constraint holding the two RecA domains together, while the RecA domain at the leading edge (i.e. RecA2) releases its grip on the translocation strand (**Figure 1D**). RecA1 remains associated with the translocation strand, forcing RecA2 to move unidirectionally. The alternating action between open and closed states has been described as an inchworm that uses its hind legs to clasp the translocation strand while the front legs 'inch' forward, one base pair per cycle [1,25,26,28-31]. NTP binding locks the front legs onto the translocation strand and 'pulls' the rear domain into the conformationally closed state. Notably, helicases that move in the opposite direction share the same mechanistic principles of locomotion, but the nucleic acid binding affinity of the two domains are reversed [27,32]. In addition to the core helicase motor, accessory domains often participate in directional unwinding. Accessory domains that cap the RecA folds often contribute aromatic amino acids that pi-stack with nucleobases and provide a backstop for directional motion [2,3,25,26,28] (**Figure ID**).

The hexameric helicases (SF3-6) utilize at least two distinct ring-like architectures—either a flat-closed ring or a notched lock washer—to unwind nucleic acid (**Figure 1E and F**). SF3 and SF5 proteins assemble into flat closed-ring conformations, with paddle-like loops that extend from each subunit into the central channel [33,34]. Each loop subunit binds to sequential phosphates along the nucleic acid backbone in a spiral staircase orientation (**Figure 1E**). Loop height correlates with the catalytic state of the NTP binding pocket, suggesting that these paddles employ a coordinated escort mechanism in which the nucleic acid is pulled through the central channel during the NTPase cycle [4,33]. While the ring-like assemblies for these helicases are similar, SF3 helicases translocate in the 3' to 5' direction while SF5 enzymes translocate in 5' to 3' direction. Comparative structural studies suggest that the direction of these motors is defined by the order of NTP hydrolysis around the hexameric ring. Looking down the barrel of the ring from the 5' end of the translocation strand, the NTPase sites in SF3 helicase rings fire in a clockwise order, while SF5 firing is in a counter clockwise order [34] (**Figure 1E**).

SF4 helicases also assemble into hexameric ring-like structures, but these rings bind nucleic acid in a notched lock washer conformation [35]. Each subunit in the ring interacts with the translocation strand though loops that are consecutively positioned along the phosphate backbone. The structure of the DnaB helicase suggests a coordinated mechanism of locomotion with a power stroke that relies on binding of NTP at the leading edge, and NDP release at the trailing edge. Release of NDP coincides with release of the lagging end subunit, which preferentially reassociates at the leading edge upon NTP binding [35]. This model has been described as a hand-over-hand mechanism where the hand in back (lagging subunit) preferentially reassociates at the front in an NTP-dependent fashion (**Figure 1E**). Interestingly, recent cryo-EM reconstructions of the minichromosome maintenance (MCM) protein, a SF6 family helicase, reveal that this complex can adopt both the notched lock washer and open flat-ring conformations suggesting that at least some hexameric helicases are structurally dynamic [36].

NIH-PA Author Manuscript
NIH-PA Author Manuscript
NIH-PA Author Manuscript

Fitting Cas3 into the Helicase Family Tree

Phylogenetic analysis of helicases has been used to infer mechanistic features such as substrate preference, directionality, or NTP preference, but Cas3 helicases have not been included in these studies [1,3]. To determine the evolutionary history of Cas3 helicases and to gain mechanistic insight regarding Cas3 helicase function, we performed a phylogenetic analysis of Cas3 along with helicases from each of the major helicase superfamilies (i.e., SF1-SF6) (**Figure 2 and Supplemental Figure 1**). Cas3 sequences from each CRISPR subtype were selected from both archaeal and bacterial genomes. Cas3 sequences from some subtypes were only found in bacteria (i.e., Type IF), and others were mainly observed in archaea (i.e., Type IA). We also included two recently identified Cas3 sequences found in viral genomes [37]. Cas3 sequences are diverse, due in part to differences in N- or C-terminal accessory domains, but all Cas3 sequences contain a common core helicase domain. To compare only helicase-specific features, we restricted our phylogenetic analysis to amino acids associated with the core helicase domains. The Cas3 helicase sequences were aligned to SF1 and SF2 helicases from yeast, *E. coli*, humans [3], and select sequences from SF3-6. Sequences were initially aligned using Clustal Omega [38] and manually curated using conserved sequence motifs as structural benchmarks (**Figure 2A and Supplemental Figure 1**). Phylogenetic analysis results in trees with branching patterns that resolve each of the major helicase superfamilies (SF1-6) (**Figure 2B**). The Cas3 sequences form a new branch within the SF2 lineage that is most closely related to SF2 helicase families that unwind with defined polarity, such as the DEAH/RHA and NS3/NPH-II.

All Cas3 proteins have sequence signatures that are characteristic of SF2 helicases (**Figure 2A**). The twelve conserved motifs that unify SF2 helicases are similar in Cas3 proteins [3]. Beyond the highly conserved NTP binding site formed by motifs I, II and VI (Walker A, Walker B, and arginine finger), all Cas3 helicases contain motif IVa, which is a signature of all SF2 proteins (**Figure 2A**). Additionally, Cas3 sequences contain a SAT or TAT sequence at motif III, which clearly distinguishes SF2 helicases from SF1.

A feature that distinguishes the Cas3 helicase from other SF2 helicase motors is motif IV. In SF2 helicases, motif IV contains a conserved aromatic residue that participates in NTP binding, NTP hydrolysis, and helps hold the translocation strand in place as the RecA1 domain moves during NTP binding [25,28,39]. This conserved aromatic residue in motif IV is missing in Cas3 proteins and instead contains a conserved asparagine (**Figure 2A**). The functional significance of an asparagine at this position awaits clarification by structures and biochemical studies. Regardless of its function, this distinct sequence signature can be used to distinguish Cas3 helicases from other SF2 members.

We anticipate that Cas3 helicases unwind duplex nucleic acid with an inchworm-like mechanism that has been described for other SF2 helicases. Biochemical studies have shown that Type IE Cas3 proteins are ATP-dependent helicases that unwind dsDNA in a 3’ to 5’ direction [40-42]. The close phylogenetic relationship of Cas3 proteins to SF2 helicase families that unwind with defined polarity, suggests that Type I Cas3 helicases will unwind duplexes with 3’ to 5’ directionality [3]. However, strand separation by related SF2 helicases (NS3, DEAH/RHA, and Ski2) requires a β-hairpin located between motifs Va and VI on domain 2 [25,28]. Interestingly, Cas3 sequences do not appear to have a β-hairpin at this location. This suggests that Cas3 proteins use an alternative strand splitting feature, similar to what has been observed in some SF1 helicase families [26,27,29] (**Supplementary Figure 1**).

Translocating SF2 helicases often contain an accessory domain that is juxtaposed to the RecA core and interacts with nucleic acid bases [2,3]. In the NS3 helicase from HCV and

NIH-PA Author Manuscript NIH-PA Author Manuscript NIH-PA Author Manuscript

Ski2-like helicases, C-terminal accessory domains contribute to translocation by providing aromatic residues that pi-stack with nucleobases on the translocation strand [25,28]. Mutation of these residues or removal of the C-terminal domain in the Ski2-like helicase Hel308 results in impaired unwinding, while maintaining normal NTPase activity [25]. These data suggest that the C-terminal accessory domain increases nucleic acid affinity, and prevents the backward sliding of RecA-like domains during translocation. All Cas3 sequences contain a C-terminal accessory domain of unknown function. Based on the similar architecture of Cas3 to related NS3 and Ski2-like proteins we anticipate that at least one of the functions of the C-terminal domain is to enhance DNA binding and to orient the motor for directional unwinding.

Cas3 Phylogeny

Cas3 is a defining feature of all Type I CRISPR systems [16]. However, Type I systems are diverse and previous phylogenic studies performed using a complex multi-component approach have identified at least 6 distinct subtypes (IA-IF) [16-18]. We hypothesized that the helicase domain of Cas3 proteins evolved under subtype-specific selective pressures, and that the Cas3 phylogeny might reflect mechanistic distinctions between the different subtypes. To test this hypothesis we performed a phylogenetic analysis on Cas3 sequences from each of the six different subtypes (**Figure 3**). Consistent with our hypothesis, Cas3 proteins resolve into well-supported monophyletic clades that mirror their previous phylogenetic assignments. This suggests that the Cas3 helicase core domains can be used as a simple proxy for classifying CRISPR-system subtypes.

Each Cas3 protein displays sequence characteristics within the helicase core that can be used for subtype classification. Some of these distinctions are subtle, and distinctions between nearest neighbors (e.g. subtypes IB and IC) may require consideration of all twelve conserved amino-acid motifs. However, amino acid sequences in the Q motif and Walker B motif can be used to differentiate Cas3 subtypes (**Figure 3A**). The Walker B motif (DExx) coordinates a Mg2+ ion through outer sphere interactions with the carboxyl group of the aspartic acid (D), and the glutamic acid (E) is suggested to act as a catalytic base in NTP hydrolysis. Although not invariant, each Cas3 subtype contains a defining Walker B motif that can be used in conjunction with other sequence features, such as the Q motif to define the CRISPR system subtype. The glutamine of the Q motif provides specificity for ATP by coordinating the N6 and N7 positions of the adenosine ring. The DEAH/RHA and NS3 helicases do not contain this motif and promiscuously bind any NTP within a structurally different binding pocket [9]. Cas3 subtypes ID, IE and IF contain a Q motif and we anticipate that Cas3 proteins from these three subtypes will preferentially associate with ATP. This preference has recently been experimentally determined for Cas3 proteins from the Type IE system [41]. Cas3 proteins from the IA, IB and IC subtypes do not contain a Q motif and are expected to be more promiscuous with regard to their NTP preference.

Central role of Cas3 in CRISPR–mediated immunity

CRISPR-mediated adaptive immunity proceeds in three stages: new sequence acquisition, CRISPR RNA biogenesis, and target interference [12-15] (**Figure 4**). In the first stage of adaptive immunity, foreign DNA (viral or plasmid) is inserted into the CRISPR locus of the host. CRISPR loci are transcribed and processed into short CRISPR derived RNAs (crRNAs) that contain a sequence derived from a previously encountered foreign nucleic acid. Cas proteins bind the crRNAs and the resulting ribonucleoprotein complex patrols the intracellular environment for detection of invading DNA. However, since the CRISPR locus is the template for generating crRNAs, each crRNA is complementary to at least two distinct targets: an invading phage or plasmid sequence (called a protospacer) and the ‘spacer’

NIH-PA Author Manuscript NIH-PA Author Manuscript NIH-PA Author Manuscript

sequence in the CRISPR locus of the host. CRISPR RNA-guided surveillance complexes avoid ‘self’ (i.e. spacers in the CRISPR) and efficiently target ‘non-self’ (i.e. protospacers) through protein-mediated recognition of a short sequence motif called a protospacer adjacent motif (PAM). The surveillance complex in the Type IE system is called Cascade (CRISPR-associated complex for antiviral defense) [43]. Cascade is a 405-kDa ribonucleoprotein complex composed of 11 subunits of five functionally essential Cas proteins and a 61-nucleotide crRNA [43-46]. Cascade engages invading nucleic acids through recognition of a three nucleotide PAM motif, which is proposed to destabilize the local DNA duplex for complementary sequence sampling by the crRNA [47,48]. Base pairing between the crRNA and the complementary DNA target triggers a conformational change in Cascade that bends the target DNA and displaces the non-target strand (R-loop) [45,49]. Cas3 is recruited to the target bound Cascade complex [49]. In the Type IE CRISPR systems, ATP enhances Cas3 recruitment to R-loops and R-loop binding enhances ATPase activity of Cas3 by 44-fold [40]. These data demonstrate a significant functional interaction between Cas3, ATP, and R-loops presented by Cascade. The HD-nuclease domain of Cas3 nicks the R-loop, and then unidirectionally degrades duplex substrates [40,41,49]. Mutations of the Walker A and Walker B motifs in subfamilies IE and IF compromise CRISPR-mediated protection [47,48], suggesting that the enhanced Cas3 recruitment and/or unidirectional unwinding are essential for efficient elimination of invading DNA.

In addition to its essential role in interference, Cas3 has also been implicated in new sequence acquisition. In *E. coli*, Cas3 has been shown to promote the rapid acquisition of new spacer sequences [50,51]. The mechanism of this ‘priming’ phenomenon remains unclear, but these data suggest that Cas3-mediated unwinding and/or degradation of target DNA may act as a signal for recruiting the acquisition machinery to a DNA invader (**Figure 4**). A role for Cas3 in CRISPR adaptation is also supported by the fusion of Cas3 with genes involved in adaptation (i.e., Cas1 and Cas2) [17,52]. In Type IF CRISPR systems, the Cas3 protein is fused to Cas2, and Cas1 has been identified in Cas3 pull-down experiments [53].

Outlook

Cas3 helicases are diverse, and while some subtypes are more closely related than others (e.g., IE and IF), sequences from each Cas3 subtype maintain distinctions within the helicase core that may have functional implications (e.g., NTP preference). The enzymatic activities of Cas3 have only been clearly studied in the Type IE system. While all Cas3 helicases are anticipated to participate in the degradation of foreign DNA, we hypothesize that each subtype may have evolved unique properties that expand the functional role of these helicases. We anticipate that future biochemical and structural studies will clarify the role of these diverse enzymes and define the molecular signals on the target bound Cascade complex that are responsible for Cas3 recruitment.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

We are grateful to Eckhard Jankowsky, members of the Wiedenheft lab and to members of the MSU CRISPR supergroup for feedback and thoughtful discussion on this manuscript. J.C. is supported by a grant for undergraduate research from the Howard Hughes Medical Institute (#52006931). Research in the Wiedenheft lab is supported by the National Institutes of Health (P20GM103500 and R01GM108888), National Science Foundation EPSCoR (EPS-110134), the M.J. Murdock Charitable Trust, and the Montana State University Agricultural Experimental Station.

NIH-PA Author Manuscript NIH-PA Author Manuscript NIH-PA Author Manuscript

References

1*. Singleton MR, Dillingham MS, Wigley DB. Structure and mechanism of helicases and nucleic acid translocases. Annu Rev Biochem. 2007; 76:23–50. [PubMed: 17506634] [This review provides a comprehensive summary of the structure, function and nomenclature of helicases.]

2. Pyle AM. Translocation and unwinding mechanisms of RNA and DNA helicases. Annual Review of Biophysics. 2008; 37:317–336.

3*. Fairman-Williams ME, Guenther UP, Jankowsky E. SF1 and SF2 helicases: family matters. Current Opinion in Structural Biology. 2010; 20:313–324. [PubMed: 20456941] [This paper demonstrates that helicase phylogeny can be used to predict function of SF1 and SF2 family enzymes.]

4. Enemark EJ, Joshua-Tor L. On helicases and other motor proteins. Curr Opin Struct Biol. 2008; 18:243–257. [PubMed: 18329872]

5. Steimer L, Klostermeier D. RNA helicases in infection and disease. RNA Biology. 2012; 9:751–771. [PubMed: 22699555]

6. Suhasini AN, Brosh RM. Disease-causing missense mutations in human DNA helicase disorders. Mutation Research-Reviews in Mutation Research. 2013; 752:138–152. [PubMed: 23276657]

7. Garneau NL, Wlusz J, Wilusz CJ. The highways and byways of mRNA decay. Nature Reviews Molecular Cell Biology. 2007; 8:113–126.

8. Martin R, Straub AU, Doebele C, Bohnsack MT. DExD/H-box RNA helicases in ribosome biogenesis. RNA Biology. 2013; 10:4–18. [PubMed: 22922795]

9. Cordin O, Hahn D, Beggs JD. Structure, function and regulation of spliceosomal RNA helicases. Current Opinion in Cell Biology. 2012; 24:431–438. [PubMed: 22464735]

10. Linder P, Jankowsky E. From unwinding to clamping - the DEAD box RNA helicase family. Nature Reviews Molecular Cell Biology. 2011; 12:505–516.

11. Fullam A, Schroder M. DExD/H-box RNA helicases as mediators of anti-viral innate immunity and essential host factors for viral replication. Biochimica Et Biophysica Acta-Gene Regulatory Mechanisms. 2013; 1829:854–865.

12. Sorek R, Lawrence CM, Wiedenheft B. CRISPR-mediated Adaptive Immune Systems in Bacteria and Archaea. Annu Rev Biochem. 2013; 82

13. Reeks J, Naismith JH, White MF. CRISPR interference: a structural perspective. Biochemical Journal. 2013; 453:155–166. [PubMed: 23805973]

14. Westra ER, Swarts D, Staals R, Jore MM, Brouns SJJ, Oost Jvd. The CRISPRs, They Are A-Changin': How Prokaryotes Generate Adaptive Immunity. Annu. Rev. Genet. 2012; 46:311–339. [PubMed: 23145983]

15. Wiedenheft B, Sternberg SH, Doudna JA. RNA-guided genetic silencing systems in bacteria and archaea. Nature. 2012; 482:331–338. [PubMed: 22337052]

16*. Makarova KS, Haft DH, Barrangou R, Brouns SJ, Charpentier E, Horvath P, Moineau S, Mojica FJ, Wolf YI, Yakunin AF, et al. Evolution and classification of the CRISPR-Cas systems. Nat Rev Microbiol. 2011; 9:467–477. [PubMed: 21552286] [This paper describes the current classification scheme for CRISPR subtypes as well as the nomenclature of their associated (cas) genes.]

17. Makarova KS, Grishin NV, Shabalina SA, Wolf YI, Koonin EV. A putative RNA-interference-based immune system in prokaryotes: computational analysis of the predicted enzymatic machinery, functional analogies with eukaryotic RNAi, and hypothetical mechanisms of action. Biol Direct. 2006; 1:7. [PubMed: 16545108]

18. Haft DH, Selengut J, Mongodin EF, Nelson KE. A guild of 45 CRISPR-associated (Cas) protein families and multiple CRISPR/Cas subtypes exist in prokaryotic genomes. PLoS Comput Biol. 2005; 1:e60. [PubMed: 16292354]

19. Gorbalenya AE, Koonin EV. Helicases - Amino-Acid-Sequence Comparisons and Structure-Function-Relationships. Current Opinion in Structural Biology. 1993; 3:419–429.

20. Mallam AL, Del Campo M, Gilman B, Sidote DJ, Lambowitz AM. Structural basis for RNA-duplex recognition and unwinding by the DEAD-box helicase Mss116p. Nature. 2012; 490:121–125. [PubMed: 22940866]

NIH-PA Author Manuscript NIH-PA Author Manuscript NIH-PA Author Manuscript

21. Jiang FG, Ramanathan A, Miller MT, Tang GQ, Gale M, Patel SS, Marcotrigiano J. Structural basis of RNA recognition and activation by innate immune receptor RIG-I. Nature. 2011; 479:423–U184. [PubMed: 21947008]
22. Kowalinski E, Lunardi T, McCarthy AA, Louber J, Brunel J, Grigorov B, Gerlier D, Cusack S. Structural Basis for the Activation of Innate Immune Pattern-Recognition Receptor RIG-I by Viral RNA. Cell. 2011; 147:423–435. [PubMed: 22000019]
23. Luo DH, Ding SC, Vela A, Kohlway A, Lindenbach BD, Pyle AM. Structural Insights into RNA Recognition by RIG-I. Cell. 2011; 147:409–422. [PubMed: 22000018]
24. Wu B, Peisley A, Richards C, Yao H, Zeng XH, Lin C, Chu FX, Walz T, Hur S. Structural Basis for dsRNA Recognition, Filament Formation, and Antiviral Signal Activation by MDA5. Cell. 2013; 152:276–289. [PubMed: 23273991]
25. Buttner K, Nehring S, Hopfner KP. Structural basis for DNA duplex separation by a superfamily-2 helicase. Nature Structural & Molecular Biology. 2007; 14:647–652.
26. Velankar SS, Soultanas P, Dillingham MS, Subramanya HS, Wigley DB. Crystal structures of complexes of PcrA DNA helicase with a DNA substrate indicate an inchworm mechanism. Cell. 1999; 97:75–84. [PubMed: 10199404]
27. Saikrishnan K, Powell B, Cook NJ, Webb MR, Wigley DB. Mechanistic Basis of 5′-3′ Translocation in SF1B Helicases. Cell. 2009; 137:849–859. [PubMed: 19490894]
28. Gu MG, Rice CM. Three conformational snapshots of the hepatitis C virus NS3 helicase reveal a ratchet translocation mechanism. Proceedings of the National Academy of Sciences of the United States of America. 2010; 107:521–528. [PubMed: 20080715]
29. Lee JY, Yang W. UvrD helicase unwinds DNA one base pair at a time by a two-part power stroke. Cell. 2006; 127:1349–1360. [PubMed: 17190599]
30. Qi Z, Pugh RA, Spies M, Chemla YR. Sequence-dependent base pair stepping dynamics in XPD helicase unwinding. Elife. 2013; 2:e00334. [PubMed: 23741615]
31. Myong S, Ha T. Stepwise translocation of nucleic acid motors. Current Opinion in Structural Biology. 2010; 20:121–127. [PubMed: 20061135]
32. Kuper J, Wolski SC, Michels G, Kisker C. Functional and structural studies of the nucleotide excision repair helicase XPD suggest a polarity for DNA translocation. Embo Journal. 2012; 31:494–502. [PubMed: 22081108]
33. Enemark EJ, Joshua-Tor L. Mechanism of DNA translocation in a replicative hexameric helicase. Nature. 2006; 442:270–275. [PubMed: 16855583]
34*. Thomsen ND, Berger JM. Running in Reverse: The Structural Basis for Translocation Polarity in Hexameric Helicases. Cell. 2009; 139:523–534. [PubMed: 19879839] [This recent paper describes an alternative mechanism for hexameric strand unwinding based on the structure of DnaB complexed with nucleic acid and an NTP analog in a notched lock washer configuration.]
35. Itsathitphaisarn O, Wing RA, Eliason WK, Wang JM, Steitz TA. The Hexameric Helicase DnaB Adopts a Nonplanar Conformation during Translocation. Cell. 2012; 151:267–277. [PubMed: 23022319]
36. Costa A, Ilves I, Tamberg N, Petojevic T, Nogales E, Botchan MR, Berger JM. The structural basis for MCM2-7 helicase activation by GINS and Cdc45. Nature Structural & Molecular Biology. 2011; 18:471–U110.
37. Seed KD, Lazinski DW, Calderwood SB, Camilli A. A bacteriophage encodes its own CRISPR/Cas adaptive response to evade host innate immunity. Nature. 2013; 494:489–491. [PubMed: 23446421]
38. Sievers F, Wilm A, Dineen D, Gibson TJ, Karplus K, Li WZ, Lopez R, McWilliam H, Remmert M, Soding J, et al. Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. Molecular Systems Biology. 2011; 7
39. Banroques J, Cordin O, Doere M, Linder P, Tanner NK. A conserved phenylalanine of motif IV in superfamily 2 helicases is required for cooperative, ATP-dependent binding of RNA substrates in DEAD-box proteins. Molecular and Cellular Biology. 2008; 28:3359–3371. [PubMed: 18332124]
40**. Mulepati S, Bailey S. In Vitro Reconstitution of an Escherichia coli RNA-guided Immune System Reveals Unidirectional, ATP-dependent Degradation of DNA Target. J Biol Chem. 2013;

NIH-PA Author Manuscript

288:22184–22192. [PubMed: 23760266] [This study demonstrates that Cas3 recruitment to Cascade is enhanced by ATP.]

41. Sinkunas T, Gasiunas G, Fremaux C, Barrangou R, Horvath P, Siksnys V. Cas3 is a single-stranded DNA nuclease and ATP-dependent helicase in the CRISPR/Cas immune system. EMBO J. 2011; 30:1335–1342. [PubMed: 21343909]

42**. Sinkunas T, Gasiunas G, Waghmare SP, Dickman MJ, Barrangou R, Horvath P, Siksnys V. In vitro reconstitution of Cascade-mediated CRISPR immunity in Streptococcus thermophilus. EMBO J. 2013; 32:385–394. [PubMed: 23334296] [This study developed a reconstituted system to demonstrate that Cascade, Cas3 and ATP are essential to achieve target duplex DNA target degradation.]

43. Brouns SJ, Jore MM, Lundgren M, Westra ER, Slijkhuis RJ, Snijders AP, Dickman MJ, Makarova KS, Koonin EV, van der Oost J. Small CRISPR RNAs guide antiviral defense in prokaryotes. Science. 2008; 321:960–964. [PubMed: 18703739]

44. Jore MM, Lundgren M, van Duijn E, Bultema JB, Westra ER, Waghmare SP, Wiedenheft B, Pul U, Wurm R, Wagner R, et al. Structural basis for CRISPR RNA-guided DNA recognition by Cascade. Nat Struct Mol Biol. 2011; 18:529–536. [PubMed: 21460843]

45**. Wiedenheft B, Lander GC, Zhou K, Jore MM, Brouns SJJ, van der Oost J, Doudna JA, Nogales E. Structures of the RNA-guided surveillance complex from a bacterial immune system. Nature. 2011; 477:486–489. [PubMed: 21938068] [This paper presents structures of a CRISPR-RNA guided surveillance complex (Cascade) and demonstrated that target binding results in a conformational rearrangement.]

46. Duijn, Ev; Barbu, IM.; Barendregt, A.; Jore, MM.; Wiedenheft, B.; Lundgren, M.; Westra, ER.; Brouns, SJJ.; Doudna, JA.; Oost, Jvd, et al. Native Tandem and Ion Mobility Mass Spectrometry highlight Structural and Modular Similarities in CRISPR-associated protein complexes from Escherichia coli and Pseudomonas aeruginosa. Mol Cell Proteomics. 2012; 11:1430–41. [PubMed: 22918228]

47*. Sashital DG, Wiedenheft B, Doudna JA. Mechanism of foreign DNA selection in a bacterial adaptive immune system. Mol Cell. 2012; 48:606–615. [PubMed: 22521690] [This paper demonstrates that Cas3 is essential for target primed spacer acquisition of foreign DNA into the CRISPR locus.]

48. Westra ER, Semenova E, Datsenko KA, Jackson R, Wiedenheft B, Severinov K, Brouns SJ. Type I-E CRISPR-Cas Systems Discriminate Target from Non-Target DNA through Base Pairing-Independent PAM Recognition. PLoS Genet. 2013; 9:e1003742. [PubMed: 24039596]

49. Westra ER, van Erp PBG, Künne T, Wong SP, Staals RHJ, Seegers CLC, Bollen S, Jore MM, Semenova E, Severinov K, et al. CRISPR immunity relies on the consecutive binding and degradation of negatively supercoiled invader DNA by Cascade and Cas3. Mol Cell. 2012; 46:595–605. [PubMed: 22521689]

50. Datsenko KA, Pougach K, Tikhonov A, Wanner BL, Severinov K, Semenova E. Molecular memory of prior infections activates the CRISPR/Cas adaptive bacterial immunity system. Nat Commun. 2012; 3:945. [PubMed: 22781758]

51. Swarts DC, Mosterd C, van Passel MW, Brouns SJ. CRISPR Interference Directs Strand Specific Spacer Acquisition. PLoS One. 2012; 7:e35888. [PubMed: 22558257]

52. Plagens A, Tjaden B, Hagemann A, Randau L, Hensel R. Characterization of the CRISPR/Cas subtype I-A system of the hyperthermophilic crenarchaeon Thermoproteus tenax. J Bacteriol. 2012; 194:2491–2500. [PubMed: 22408157]

53*. Richter C, Gristwood T, Clulow JS, Fineran PC. In vivo protein interactions and complex formation in the Pectobacterium atrosepticum subtype I-F CRISPR/Cas System. PLoS One. 2012; 7:e49549. [PubMed: 23226499] [This study is the first to show a direction interaction between Cas3 and Cas1, suggesting a role for Cas3 in new sequence acquisition.]

54. Huelsenbeck JP, Ronquist F. MRBAYES: Bayesian inference of phylogenetic trees. Bioinformatics. 2001; 17:754–755. [PubMed: 11524383]

55. Abascal F, Zardoya R, Posada D. ProtTest: selection of best-fit models of protein evolution. Bioinformatics. 2005; 21:2104–2105. [PubMed: 15647292]

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

Highlights

- Cas3 helicases are a central component of most CRISPR-mediated adaptive immune systems.
- Cas3 enzymes are multi-domain proteins that unwind and degrade invading nucleic acids.
- Cas3 proteins are SF2 helicases, most closely related to the DEAH/RHA and NS3/NPH-II families.
- The Cas3 helicase domain evolves according to subtype specific selective pressures.
- The Cas3 helicase domain can be used as a proxy for classifying CRISPR-system subtypes.

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

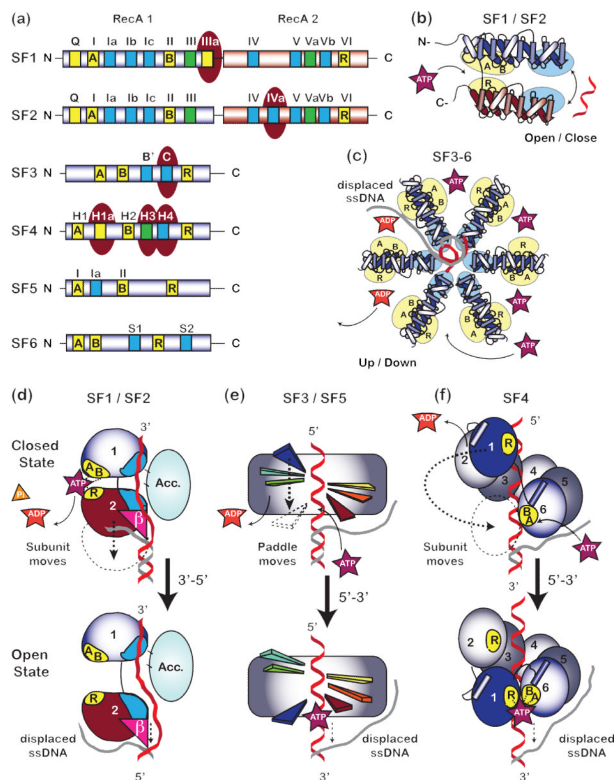

Figure 1. Structures and unwinding mechanisms of helicase superfamilies

(a) A schematic of the core helicase domain. The N-terminal RecA domain (RecA1) is represented by a blue cylinder and the C-terminal RecA domain (RecA2) is shown as a red cylinder. Conserved amino acid motifs are colored according to helicase function. Motifs in yellow are involved in NTP binding/hydrolysis, green are associated with translocation, and blue interact with nucleic acid. Motifs that are unique to specific superfamilies are highlighted with a red oval. The Walker A (A), Walker B (B) and arginine finger (R) motifs are conserved across all helicase superfamilies. **(b)** Topology diagrams depicting the secondary structure of the tandem RecA-like folds observed in SF1 and SF2 helicases. The RecA-like domains form a cleft that contains an NTP binding pocket (yellow) and a nucleic acid binding site (blue). NTP binding and hydrolysis causes the cleft to cycle between the closed and open states. **(c)** SF3-6 helicases assemble into toroidal hexamers that radially array the bipartite NTP binding sites. **(d)** Schematic of the unwinding mechanism for SF1 and SF2 helicases. The top and bottom panels represent closed (NTP-bound) and open (unbound) states, respectively. The RecA-like domains and conserved motifs are colored as in (a). NTP-dependent conformational changes drive a wedge (colored pink) between the oncoming strands of a duplex. **(e)** A schematic of the unwinding mechanism of the flat hexameric SF3 and SF5 helicases. The flat ring is depicted as a rectangle. The translocation strand threads through a central pore in the hexamer. The top panel shows the nucleic acid binding loops arranged in a spiral staircase configuration. The bottom panel depicts a downward motion of the top loop (blue wedge), during NTP binding and hydrolysis. **(f)** Schematic of the SF4 and SF6 hexamer bound to ssDNA and nucleotides before (top panel) and after (bottom panel) a NDP release. NTP binding at an empty site coupled with ADP release at an adjacent site moves the top domain of the lock washer in a 5’ to 3’ direction.

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

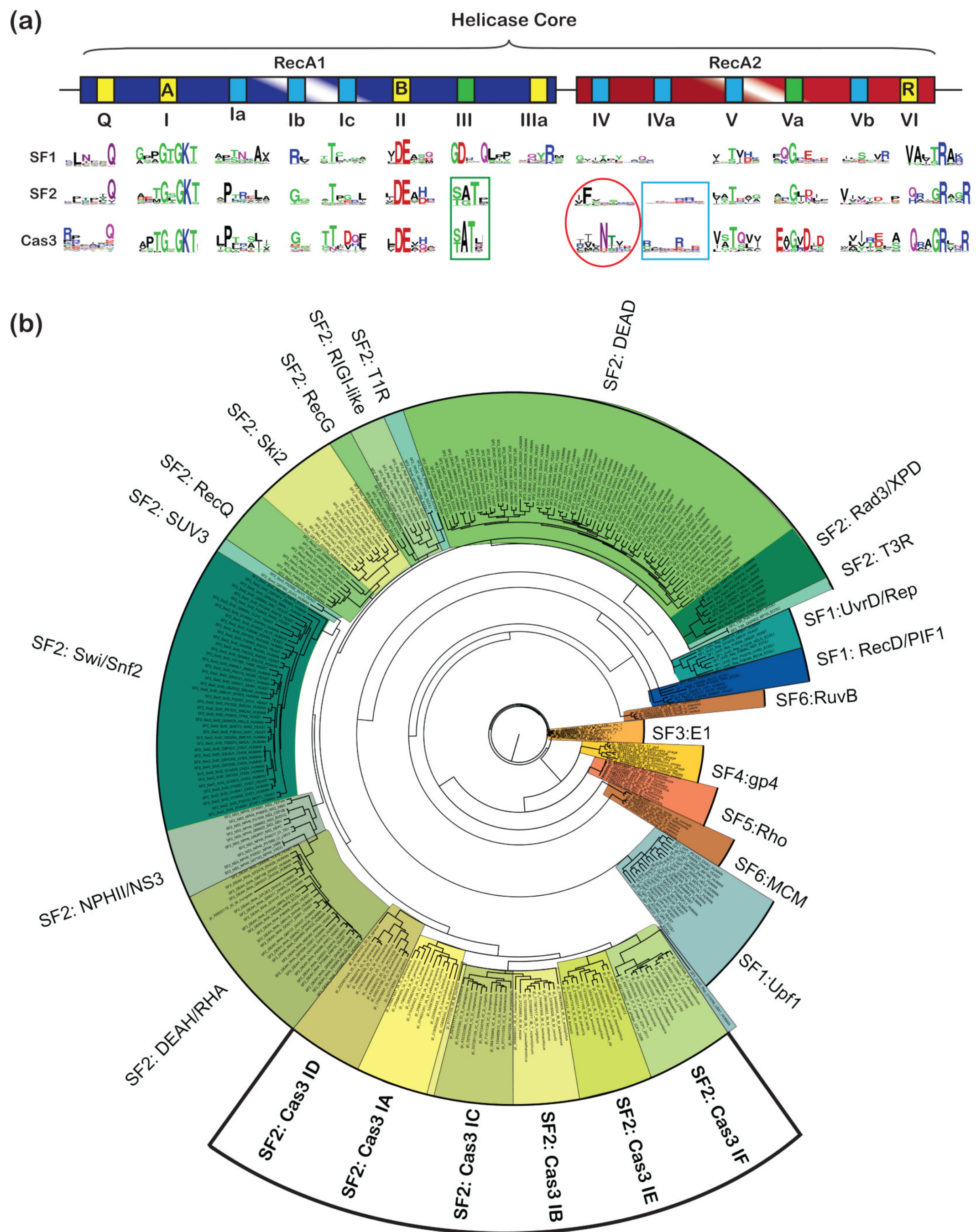

Figure 2. Fitting Cas3 into the helicase family tree

(a) Sequence logos of the conserved motifs in the core helicase domain of SF1, SF2 and Cas3 proteins. Green and blue boxes denote motifs that define Cas3 as SF2 helicases. Cas3 contains a unique motif (IV, red circle) not observed in other SF2 helicases. **(b)** Phylogenetic tree including 265 sequences representing SF1-6 helicases. Amino acid sequences from the core RecA helicase domains of 68 different Cas3 proteins were aligned to the helicase core domains of representative sequences from all superfamilies. Sequences were aligned with Clustal Omega and manually curated in Se-Al (see Supplementary Figure 1 for alignment). N- and C-terminal accessory domain sequences were not included. The alignment contained 878 amino acid positions, 572 of which are parsimony-informative (i.e.

the position had at least one alternative amino acid in more than one sequence). Phylogenetic analysis was carried out with a Bayesian approach using MrBayes [54]. Tree topologies were sampled every 250 generations for 10^6 generations using the WAG evolutionary model with fixed amino acid frequencies and gamma-shape rate variation with a proportion of invariable sites as recommended by ProtTest [55]. Posterior probabilities for all of the marked clades ranged 0.95-1.00.

NIH-PA Author Manuscript NIH-PA Author Manuscript NIH-PA Author Manuscript

NIH-PA Author Manuscript
NIH-PA Author Manuscript
NIH-PA Author Manuscript

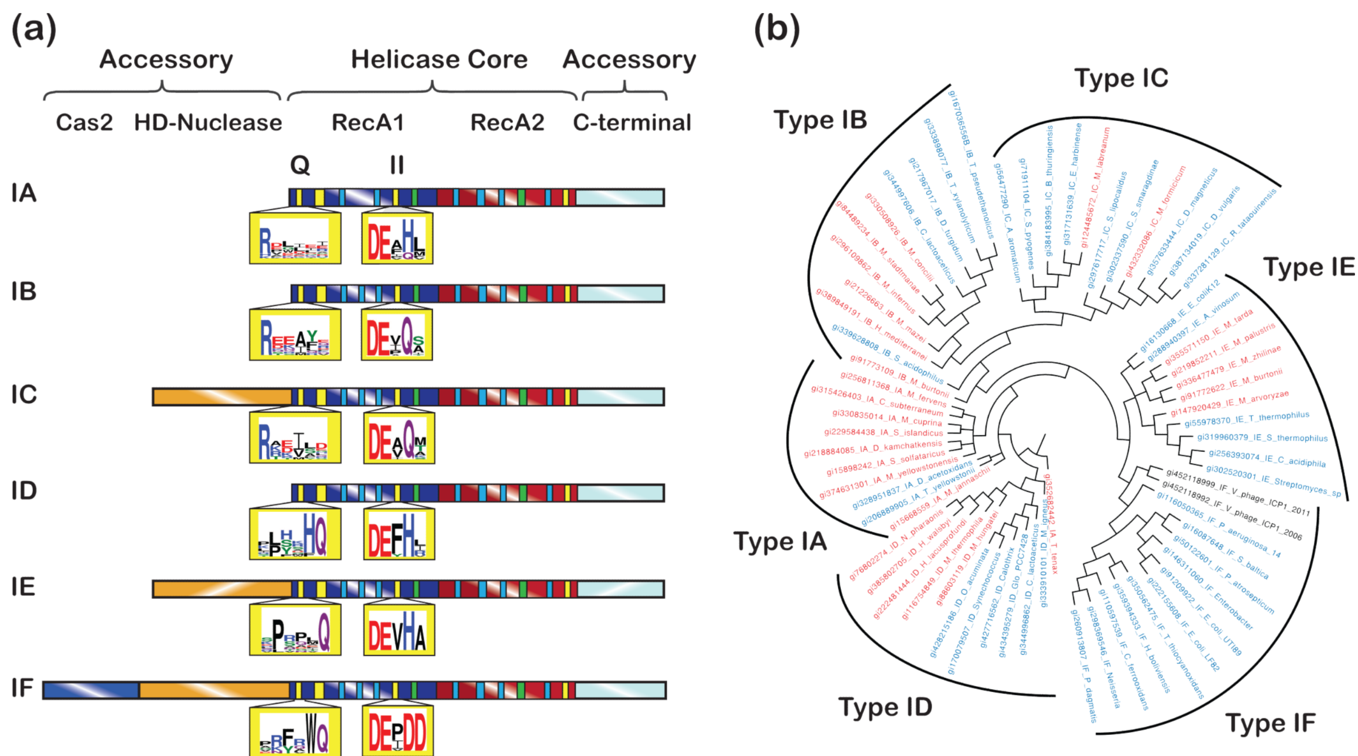

Figure 3. Cas3 proteins form well-supported clades that support previously delineated CRISPR subtypes

(a) Schematic of the helicase core and accessory domains commonly observed in each Cas3 subtype. Conserved helicase motifs are colored as in Figure 1. Cas3 sequences in the Q motif and motif II (i.e. Walker B) can be used to delineate Cas3 subtype association. **(b)** Phylogenetic tree of the core helicase domains from 68 different Cas3 proteins. Alignments and phylogenies were performed as described in Figure 2. Bacterial sequences are colored red, archaeal sequences blue and viral sequences black. Posterior probabilities for all of the marked clades ranged 0.95-1.00.

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

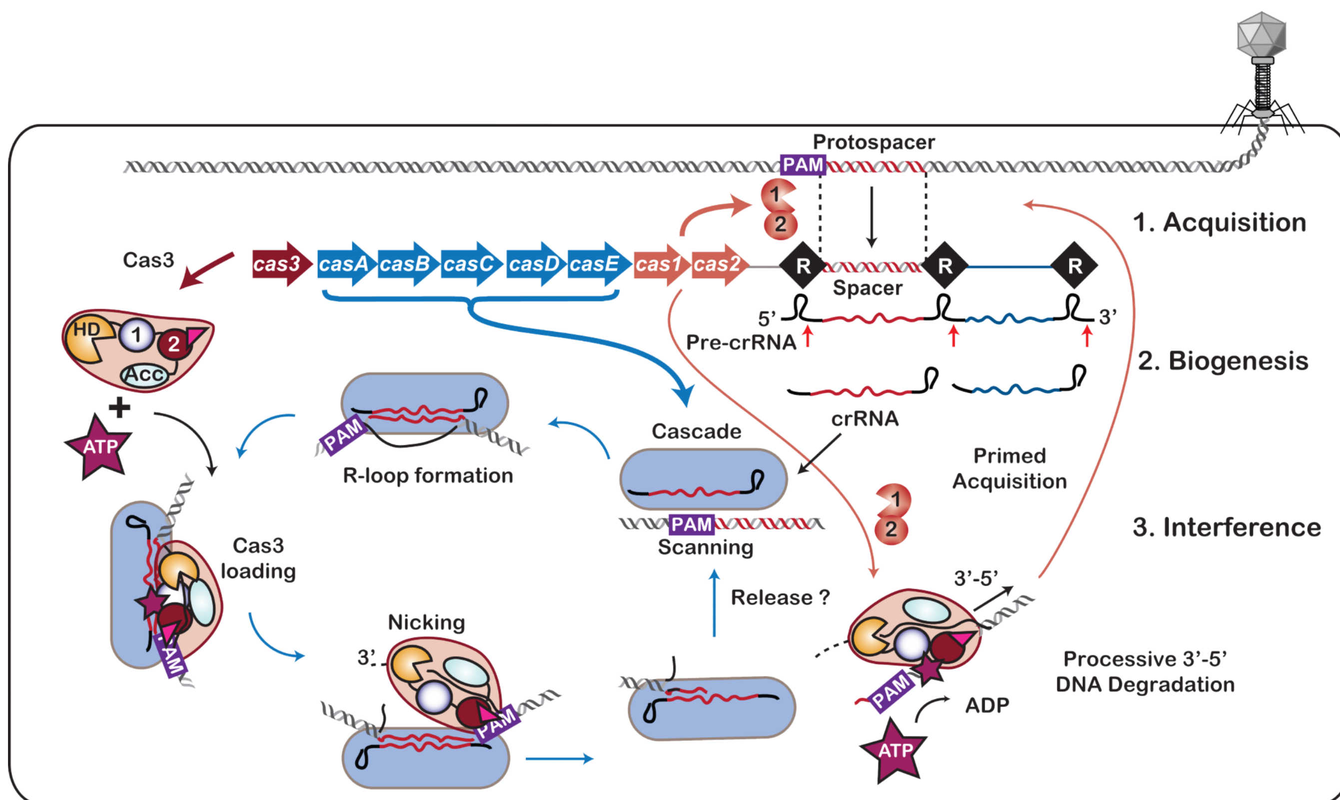

Figure 4. The central role of Cas3 in CRISPR-associated adaptive immunity

CRISPR-mediated immunity proceeds in three basic steps: acquisition, crRNA biogenesis, and interference. Fragments of foreign DNA (protospacers) are acquired from regions of the invading genome that are flanked by short sequence motifs called protospacer adjacent motifs (PAMs). Protospacers are inserted into the CRISPR locus between direct repeats (black squares) by a mechanism that involves Cas1 and Cas2 proteins. The CRISPR locus is transcribed (pre-crRNA) and processed (little red arrows) into small crRNAs that are loaded into a crRNA-guided surveillance complex called Cascade (blue oval). Cascade is anticipated to facilitate target detection by scanning, and target recognition results in R-loop formation. The target bound surveillance complex recruits the effector nuclease-helicase Cas3 through a mechanism that is enhanced by ATP. Cas3 binds the R-loop and nicks the displaced strand. In the presence of ATP, Cas3 unidirectionally degrades the DNA target in a 3' to 5' direction. Cas3-mediated degradation may serve as a signal that recruits Cas1 and Cas2, resulting in the rapid acquisition of new spacers derived from the target strand of the DNA invader. This phenomenon is called priming.