Spacer capture and integration by a type I-F Cas1–Cas2-3 CRISPR adaptation complex

Robert D. Fagerlund[a,1], Max E. Wilkinson[a,b,1,2], Oleg Klykov[c,d,1], Arjan Barendregt[c,d], F. Grant Pearce[e,f], Sebastian N. Kieper[a,3], Howard W. R. Maxwell[a], Angela Capolupo[c,d], Albert J. R. Heck[c,d], Kurt L. Krause[b], Mihnea Bostina[a,g], Richard A. Scheltema[c,d], Raymond H. J. Staals[a,4], and Peter C. Fineran[a,5]

[a]Department of Microbiology and Immunology, University of Otago, Dunedin 9054, New Zealand; [b]Department of Biochemistry, University of Otago, Dunedin 9054, New Zealand; [c]Biomolecular Mass Spectrometry and Proteomics, Bijvoet Center for Biomolecular Research and Utrecht Institute for Pharmaceutical Sciences, Utrecht University, 3584 CH Utrecht, The Netherlands; [d]Netherlands Proteomics Centre, Utrecht University, 3584 CH Utrecht, The Netherlands; [e]Biomolecular Interactions Centre, University of Canterbury, Christchurch 8020, New Zealand; [f]School of Biological Sciences, University of Canterbury, Christchurch 8020, New Zealand; and [g]Otago Centre for Electron Microscopy, University of Otago, Dunedin 9054, New Zealand

Edited by Jennifer A. Doudna, University of California, Berkeley, CA, and approved May 18, 2017 (received for review November 6, 2016)

CRISPR-Cas adaptive immune systems capture DNA fragments from invading bacteriophages and plasmids and integrate them as spacers into bacterial CRISPR arrays. In type I-E and II-A CRISPR-Cas systems, this adaptation process is driven by Cas1–Cas2 complexes. Type I-F systems, however, contain a unique fusion of Cas2, with the type I effector helicase and nuclease for invader destruction, Cas3. By using biochemical, structural, and biophysical methods, we present a structural model of the 400-kDa $Cas1_4$–$Cas2\text{-}3_2$ complex from *Pectobacterium atrosepticum* with bound protospacer substrate DNA. Two Cas1 dimers assemble on a Cas2 domain dimeric core, which is flanked by two Cas3 domains forming a groove where the protospacer binds to Cas1–Cas2. We developed a sensitive in vitro assay and demonstrated that Cas1–Cas2-3 catalyzed spacer integration into CRISPR arrays. The integrase domain of Cas1 was necessary, whereas integration was independent of the helicase or nuclease activities of Cas3. Integration required at least partially duplex protospacers with free 3′-OH groups, and leader-proximal integration was stimulated by integration host factor. In a coupled capture and integration assay, Cas1–Cas2-3 processed and integrated protospacers independent of Cas3 activity. These results provide insight into the structure of protospacer-bound type I Cas1–Cas2-3 adaptation complexes and their integration mechanism.

CRISPR-Cas | phage resistance | horizontal gene transfer | spacer acquisition | mass spectrometry

Clustered regularly interspaced short palindromic repeats (CRISPRs) and their Cas proteins are prokaryote adaptive immune systems that provide defense against invading elements, typically phages and plasmids (1, 2). The systems are evolutionarily diverse, organized into two major classes and multiple types and subtypes (3). CRISPR arrays consist of repeats separated by spacers that are usually derived from invaders (4). Arrays are transcribed and processed into CRISPR RNAs (crRNAs) containing a single spacer sequence (5, 6). In type I CRISPR-Cas systems, crRNAs and Cas proteins assemble into Cascade complexes (5, 7, 8) that bind complementary sequences (protospacers) to elicit invader DNA destruction by the nuclease-helicase protein Cas3 (9–11).

Upon encountering an invader with no matches to existing spacers, new invader-derived spacers can be selected and processed (captured) and integrated into the CRISPR array, which is called naïve adaptation (12–14). Spacer capture is biased to occur beside stalled replication forks and other DNA breaks, and the RecBCD complex is proposed to have a role in generating spacer precursors (15, 16). However, adaptation still occurs in the absence of RecBCD, albeit less efficiently, indicating that other pathways exist. Additionally, in type I systems, spacers are acquired next to protospacer adjacent motifs (PAMs), which are also required for interference (17). Spacer integration relies on Cas1 and Cas2, which are almost universal in CRISPR-Cas systems (3, 18). Cas1 is a nuclease/integrase and Cas2 is a small, apparently structural protein (19, 20). Cas1 and Cas2 domain proteins were first shown to interact in vivo in the type I-F system of *Pectobacterium atrosepticum* (21). More recently, structures of type I-E Cas1–Cas2 complexes have illuminated aspects of protospacer binding (19, 22, 23). Integration has been studied in vitro for the type I-E and II-A Cas1–Cas2 complexes (24, 25), and the reverse reaction (disintegration) only requires Cas1 (26). These studies provided valuable insight into integration in these specific systems. However, there is considerable CRISPR-Cas diversity, and adaptation mechanisms in other systems are unexplored.

Type I-F systems encode a Cascade (also known as Csy) complex, a Cas1 protein, and a unique fusion of Cas2 and Cas3 (Cas2-3), meaning this protein engages in both adaptation and interference (21, 27). Formation of type I-F Cas1–Cas2-3 complexes is likely to be important for rapid primed adaptation (28, 29). Priming can occur either during interference or in response to invaders that have

Significance

CRISPR-Cas systems provide prokaryotic adaptive immunity against invading genetic elements. For immunity, fragments of invader DNA are integrated into CRISPR arrays by Cas1 and Cas2 proteins. Type I-F systems contain a unique fusion of Cas2 to Cas3, the enzyme responsible for destruction of invading DNA. Structural, biophysical, and biochemical analyses of Cas1 and Cas2-3 from *Pectobacterium atrosepticum* demonstrated that they form a 400-kDa complex with a $Cas1_4$:$Cas2\text{-}3_2$ stoichiometry. Cas1–Cas2-3 binds, processes, and catalyzes the integration of DNA into CRISPR arrays independent of Cas3 activity. The arrangement of Cas3 in the complex, together with its redundant role in processing and integration, supports a scenario where Cas3 couples invader destruction with immunization—a process recently demonstrated in vivo.

Author contributions: R.H.J.S. and P.C.F. conceived the study; R.D.F., M.E.W., A.J.R.H., M.B., R.A.S., R.H.J.S., and P.C.F. designed research; R.D.F., M.E.W., O.K., A.B., F.G.P., S.N.K., H.W.R.M., A.C., M.B., and R.H.J.S. performed research; R.D.F., M.E.W., O.K., A.B., F.G.P., S.N.K., H.W.R.M., A.C., A.J.R.H., K.L.K., M.B., R.A.S., R.H.J.S., and P.C.F. analyzed data; and R.D.F., M.E.W., O.K., R.A.S., R.H.J.S., and P.C.F. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

Data deposition: The data reported in this paper have been deposited in the Electron Microscopy Database (accession no. EMD-8660).

[1]R.D.F., M.E.W., and O.K. contributed equally to this work.

[2]Present address: MRC Laboratory of Molecular Biology, Cambridge CB2 0QH, United Kingdom.

[3]Present address: Kavli Institute of Nanoscience and Department of BioNanoscience, Delft University of Technology, 2629 HZ Delft, The Netherlands.

[4]Present address: Laboratory of Microbiology, Department of Agrotechnology and Food Sciences, Wageningen University, 6708 WE Wageningen, The Netherlands.

[5]To whom correspondence should be addressed. Email: peter.fineran@otago.ac.nz.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1618421114/-/DCSupplemental.

escaped interference through protospacer or PAM mutations, triggering rapid acquisition from the foreign element to restore immunity (29–32). Multiple type I systems undergo priming, and for type I-E it requires Cascade, crRNA, Cas1, Cas2, and Cas3 (28, 30, 33, 34). During type I-F priming, new spacers are captured from regions adjacent to the escaped protospacer in a process that involves the 3′-to-5′ translocation of a Cas1–Cas2-3 complex (28, 29). The Cas2-3 fusion in I-F systems means that when Cas3 is recruited to targets identified by a Cascade–crRNA complex, Cas2 will always be in tow, which can result in Cas1 corecruitment. This may directly couple Cas3 helicase and nuclease functions with Cas1 and Cas2 adaptation activities. It is plausible that related Cas1–Cas2–Cas3 complexes occur in all type I systems to promote priming.

The structure, stoichiometry, and mechanism of adaptation complexes containing Cas1, Cas2, and Cas3 domains are unknown. Therefore, we characterized the type I-F Cas1–Cas2-3 complex biophysically and, by reconstituting adaptation in vitro, determined its role in spacer capture and integration. We propose that Cas1 and Cas2 alone are sufficient for PAM recognition, spacer capture, and integration, whereas Cas3 likely acts earlier to increase adaptation during priming and interference by acting as a helicase and nuclease in the generation of spacer precursors.

Results

Cas1 and Cas2-3 Form a 400-kDa Cas1$_4$–Cas2-3$_2$ Complex. To investigate the Cas1–Cas2-3 complex, we coexpressed StrepII-tagged Cas1 (37.6 kDa) and untagged Cas2-3 (124.9 kDa) from *P. atrosepticum* and purified the complex by affinity and size-exclusion chromatography (SEC) (Fig. 1*A* and Fig. S1*A*). The elution profile of Cas1–Cas2-3 gave a mass of ∼390 kDa (Fig. S1*A*), which was further supported by SEC–right-angle light scattering (RALS), which estimated 397 kDa (Fig. 1*B*). Sedimentation velocity by analytical ultracentrifugation (AUC) yielded a major species at 13.5S, which corresponded to ∼380 kDa, whereas a minor species at 4.5S (∼75 kDa) was consistent with some Cas1 dimer dissociating from Cas1–Cas2-3 (Fig. 1*C*).

To clarify the Cas1–to–Cas2-3 ratio, we estimated the absolute protein quantities via shotgun mass spectrometry (MS) using intensity-based absolute quantification (iBAQ) (35), which revealed a 2:1 ratio (Table S1). Top-down MS on a denatured Cas1–Cas2-3 complex enabled an accurate mass measurement of Cas1 as 37,557.62 ± 0.84 Da, whereas free Cas2-3 was not observed. To define the stoichiometry, we determined the accurate mass by native Orbitrap MS. The spectra showed four distinct charge distributions (Fig. 1*D*) originating from species with molecular masses (and relative abundance) of 363,983 ± 63 Da (7%), 400,471 ± 31 Da (16%), 409,413 ± 54 Da (47%), and 418,403 ± 42 Da (30%). The 400-kDa species is in complete agreement with the predicted stoichiometry of Cas1$_4$–Cas2-3$_2$. The larger species containing 9 and 18 kDa of extra mass are likely to be complexes copurified with captured DNA. This mass could correspond to 29 to 30 nt/bp of ssDNA (9 kDa), dsDNA (18 kDa), or variations thereof. The less abundant 364-kDa species was assigned as Cas1$_3$–Cas2-3$_2$. In conclusion, Cas1 and Cas2-3 form a 400-kDa complex with a stoichiometry of Cas1$_4$:Cas2-3$_2$, with a significant proportion bound to nucleic acids (see below).

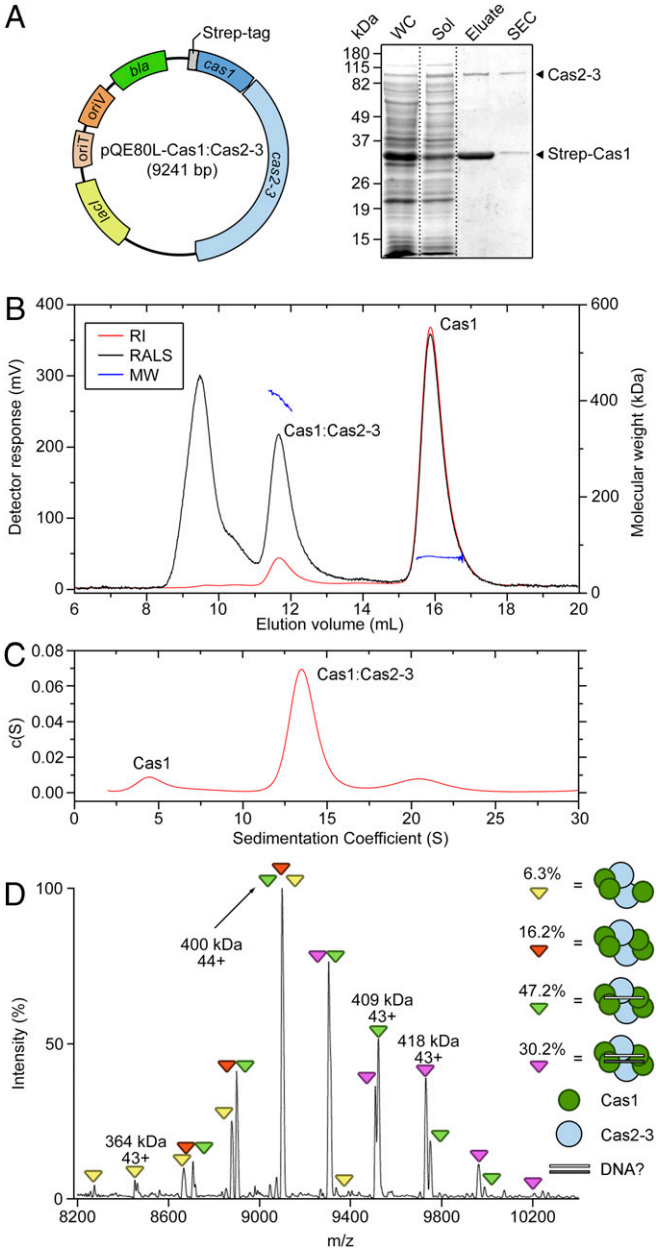

Fig. 1. Cas1 and Cas2-3 form a 400-kDa Cas1$_4$–Cas2-3$_2$ complex. (*A*) Expression and purification of StrepII-Cas1–Cas2-3. (*A*, *Left*) The coexpression plasmid is depicted. (*A*, *Right*) SDS/PAGE of whole-cell (WC), soluble (Sol), Strep-Tactin elution (Eluate), and SEC fractions. A SEC chromatogram with molecular weight (MW) standards is shown in Fig. S1*A*. (*B*) SEC-RALS. MW (blue) was calculated from the refractive index (RI; red) and RALS (black). (*C*) Sedimentation velocity. (*D*) Native MS. Predicted stoichiometries are shown, and iBAQ is shown in Table S1.

Molecular Architecture of Cas1–Cas2-3. Type I-F Cas2-3 has multiple domains: an N-terminal Cas2, followed by a Cas3 region containing an HD endonuclease and an SF2 helicase (two RecA domains), and an accessory C-terminal domain (36) (Fig. 2*A*). To gain structural insight into the full complex, we used modeling, cross-linking, and electron microscopy (EM). To create an initial model of *P. atrosepticum* Cas1–Cas2-3, we made a homology model from *Pseudomonas aeruginosa* Cas2-3 (37) and used the *Escherichia coli* type I-E Cas1–Cas2–protospacer structure (22, 23) and *P. atrosepticum* Cas1 dimer (38). To optimize the model, we performed cross-linking with MS to identify proximal regions within proteins and obtain spatial restraints. Due to lysine side-chain flexibility, we assumed a 40-Å maximum distance between C_α and C_α. There were 19 unique cross-links for Cas1–Cas2-3 (Fig. 2*A* and Table S2), including 13 intralinks (within a single protein) and 6 interlinks (between two proteins). Of the 13 identified intralinks, 2 belong to Cas1 and 11 to the Cas3 region of Cas2-3. Mapping the cross-links on our model placed two Cas1 intralinks and eight Cas2-3 intralinks within 40 Å, validating the Cas2-3 homology model. The three Cas2-3 outliers suggest that residues 500 and 823, in the RecA1 and RecA2 domains, reside in flexible regions.

Next, we examined the complex by EM and obtained a map of ∼25-Å resolution (Fig. 2 *B* and *C* and Fig. S1 *B*–*F*). To find the best fit, Cas1 and Cas2-3 were modeled into the density map by

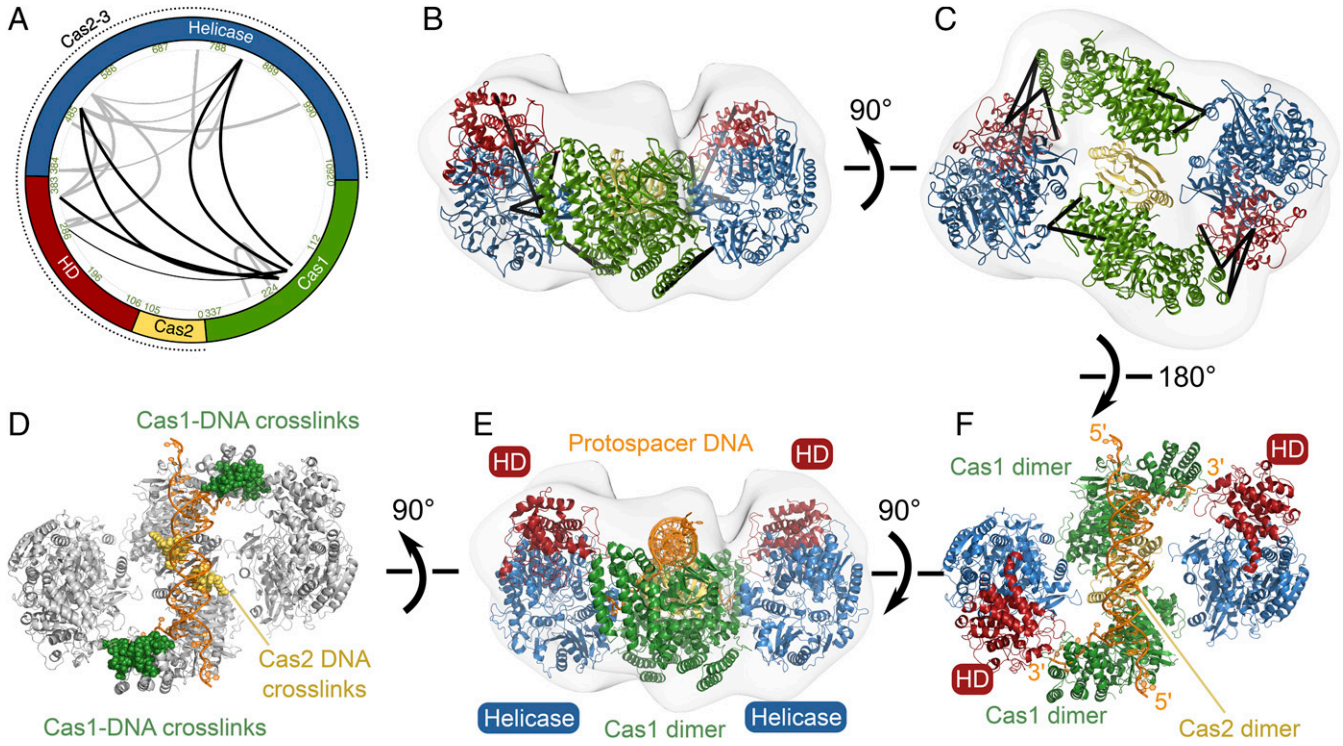

Fig. 2. Molecular architecture of $Cas1_4$–$Cas2\text{-}3_2$. (*A*) Schematic of intralinks (depicted as gray lines; thin lines are outliers) and interlinks (black) (Table S2). Cas1 is in green, the Cas2-3 domain organization is shown with Cas2 in yellow, the HD nuclease of Cas3 in red, and the remainder of Cas3 in blue (SF2 helicase, composed of RecA1 and RecA2, and a C-terminal domain). (*B* and *C*) Modeling of Cas1–Cas2-3 into EM density. Interprotein cross-links are shown (black lines) and EM data are shown in Fig. S1 *B*–*F*. (*D*) DNA-interacting peptides were mapped on Cas1 (positions 201 to 213 and 273 to 294; green) and the Cas2 domain of Cas2-3 (82 to 95; yellow). Details are in Dataset S1. (*E* and *F*) The final Cas1–Cas2-3 model with features labeled and protospacer DNA shown in orange.

taking into account the cross-linking data and a flexible linker between the Cas2 and Cas3 domains (residues 85 to 100). Supporting this flexibility, the *P. aeruginosa* Cas2-3 cryo-EM structure had no Cas2 density (39). Of the 12 intermolecular connections between Cas1 and Cas3 (Fig. 2 *A*–*C* and Table S2), 10 were within cross-linking range. The Cas3 helicase domain has conformational flexibility (9), which might contribute to the cross-link outliers. Overall, the EM map accommodated our model based on the stoichiometry and cross-linking analyses, but also revealed a region of density not accounted for by Cas1–Cas2-3 alone (Fig. 2*B*). Combined with the extra mass detected by native MS, we hypothesized this density was due to protospacer DNA.

To test whether Cas1–Cas2-3 contained DNA and to analyze DNA placement, we designed an assay involving formaldehyde cross-linking of proteins to DNA and LC-MS/MS detection of enriched DNA-bound peptides. Three peptides were highly enriched in the DNA–protein cross-linked samples, indicating their interaction with DNA (Dataset S1). These were Cas1 residues 201 to 213 and 273 to 294, and Cas2-3 peptide 82 to 95 within the Cas2 domain (Fig. 2*D*). These positions in the type I-F complex are consistent with the protospacer-binding regions in type I-E Cas1–Cas2 (22, 23), which enabled modeling of a captured protospacer on Cas1–Cas2-3 that correlated with the location of additional EM density (Fig. 2 *D*–*F*). We also directly assayed DNA binding by Cas1–Cas2-3 using electrophoretic mobility-shift assays and observed that Cas1–Cas2-3 can bind short linear ssDNA and dsDNA substrates of the same sequence with similar affinities, consistent with the masses observed in native MS (Fig. S2*A*). Interestingly, the type I-E Cas1 protein was recently shown to associate with non-dsDNA in vivo (40).

The resulting model (Fig. 2 *E* and *F*) shows two Cas1 dimers assembled on either end of a Cas2 domain dimeric core, similar to the *E. coli* I-E complex (19, 22, 23). The two Cas3 domains flank the Cas2 core and span the region between the Cas1 dimers, with the Cas3 HD region in proximity to the catalytic integrase lobe of one Cas1 dimer and the Cas3 C-terminal region near the inactive lobe of the other Cas1 dimer. The protospacer-binding surface is augmented by the HD nuclease subdomains of the Cas3 domains, forming a long groove. In *Thermobifida fusca* Cas3, ssDNA can be guided by the helicase domain toward the HD nuclease domain for fragmentation (9). Remarkably, in Cas1–Cas2-3 the DNA-binding site of each HD domain is adjacent to the catalytic integrase lobe of Cas1, suggesting a pipeline of DNA-processing active sites that might deliver DNA to Cas1.

Cas1–Cas2-3 Catalyzes Spacer Integration in Vitro. To reconstitute in vitro spacer integration by Cas1–Cas2-3, we established a novel assay. *P. atrosepticum* has three CRISPR arrays and all are active in vivo (28, 29). Because CRISPR1 is the most active (~70% of acquisitions), Cas1–Cas2-3 was incubated with a 32-bp dsDNA protospacer and a plasmid containing CRISPR1 with its leader (Fig. 3*A* and Fig. S2*B*). The protospacer was chosen due to its frequent naïve acquisition in vivo, and because ~90% of spacers are 32 nt long (29). Integration products could be half-sites, resulting from attack by one end of the protospacer, or full-sites, resulting from two half-site integrations by each end of the protospacer (25). Integration was detected by PCR using a forward protospacer primer and a reverse primer in spacer 6 (Fig. 3 *A* and *B*). Integration was detected as early as 5 min, was optimal in 1 to 2 h (Fig. 3*B*), and occurred in either orientation with a similar efficiency and into each repeat–spacer junction (Fig. 3 *B* and *C*). By sequencing, we confirmed that spacers integrated at repeat–spacer boundaries, showing that Cas1–Cas2-3 alone recognizes CRISPR repeats. Integrations occurred into the top DNA strand (i.e., the 5′ end of repeats; Fig. 3 *B* and *C*) or the bottom DNA strand (i.e., the 3′ end of repeats; Fig. S2*C*). Although integrations occurred into multiple repeat–spacer junctions, acquisition was slightly favored at the leader-proximal repeat

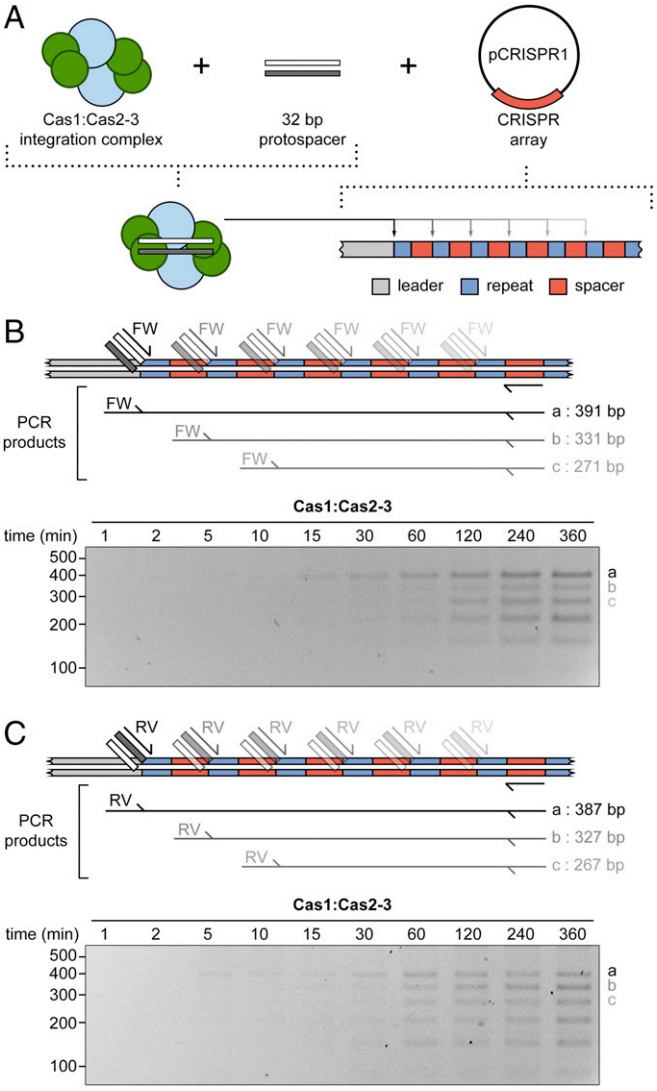

Fig. 3. Cas1–Cas2-3 catalyzes spacer integration into CRISPR arrays. (*A*) Schematic of the integration assay. Purified Cas1–Cas2-3 was incubated with a protospacer (32-nt dsDNA) and the CRISPR1 array on a plasmid (pCRISPR1), and integration was detected by PCR and gel electrophoresis. (*B*) Time course of protospacer integration into CRISPR1. A 32-bp protospacer (PF1647+PF1648) was incubated with Cas1–Cas2-3 and pCRISPR1 for 1 to 360 min and integration was detected by PCR [primers PF1649 (forward primer; FW) and PF1822]. (*C*) As in *B*, but reactions were performed with the reverse primer (PF1650; RV) in combination with PF1822. Similar reactions were performed with a leader-specific primer (PF1792) and PF1649 or PF1650 (Fig. S2*C*). Metal cofactor requirements are shown in Fig. S2*D*. The marker shown is in bp.

(Figs. 3 and 4). Acquisition was metal-dependent, because EDTA inhibited the reactions (Fig. S2*D*), presumably by chelating metal cofactors that were copurified with Cas1–Cas2-3. Mg^{2+} and Ca^{2+} enhanced the reaction, whereas Mn^{2+}, Fe^{2+}, Ni^{2+}, Co^{2+}, and Zn^{2+} decreased specificity. In summary, Cas1–Cas2-3 catalyzes spacer integration in either orientation into repeat–spacer junctions, with a slight preference for the leader-proximal repeat.

In Vitro Integration Requires Cas1–Cas2-3 but Not Cas3. To test whether the entire Cas1–Cas2-3 complex was required for in vitro integration, Cas1 and Cas2-3 were purified separately. Integration was not supported by Cas1 or Cas2-3 alone, but was robust with the Cas1–Cas2-3 complex (Fig. 4*A*). The role of the Cas3 part of the Cas2-3 fusion in adaptation is unknown, but might assist in primed spacer acquisition (28, 29, 41) and naïve adaptation (41). The Cas1–Cas2-3 architecture suggested that the Cas3 nuclease and helicase active sites are not directly involved in spacer integration, as they would not contact the protospacer (Fig. 2). To test this, the nuclease and helicase domains were inactivated with well-characterized site-directed mutations (D124A and D591A), which disrupt the HD and DExx helicase motif II, respectively (9, 10). As we predicted, spacer integration by Cas1–Cas2-3 (D124A) and Cas1–Cas2-3(D591A) mutant complexes was unaffected (Fig. 4*B*). Because we showed that the active sites of the Cas3 domain did not influence spacer integration, we tested whether the entire Cas3 domain was necessary for integration. Following deletion of the Cas3 domain, we performed an integration assay with Cas1–Cas2(ΔCas3), which showed that integration still occurred but was less specific (Fig. S3*A*). To examine the role of Cas1, we mutated a metal-coordinating active-site aspartate (D269A), which abolished primed acquisition in vivo (38). Assembly of the resulting Cas1(D269A)–Cas2-3 complex was unaffected (Fig. S3 *B* and *C*) yet integration was abrogated (Fig. 4*B*), demonstrating the key role of the integrase activity of Cas1 in adaptation. Some Cas2 proteins have nuclease activity (42, 43), and the *P. atrosepticum* Cas2 domain contains some conserved residues at potential catalytic sites (Fig. S3*D*). However, the purified I-F Cas2 domain lacked detectable nuclease activity against a range of single- and double-stranded DNA substrates (Fig. S3*E*). Therefore, Cas2 appears to play a structural and DNA-binding role that is consistent with an in vivo mutagenesis and adaptation study in the *E. coli* I-E Cas2 (19). In summary, the entire Cas1–Cas2-3 complex, but not the helicase or nuclease activities of the Cas3 domain, is required for spacer integration in vitro.

Substrate Requirements for Integration. The natural substrates used during integration by Cas1–Cas2-3 are unknown. Because multiple methods showed that Cas1–Cas2-3 bound dsDNA and ssDNA, we tested integration using ssDNA protospacers. Integration of 32-nt ssDNA protospacers was barely detectable, whereas the dsDNA protospacer integrated efficiently (Fig. 4*C*). Next, we tested the 3′ end-group requirement for the nucleophile to attack the CRISPR array and showed that phosphorylated 3′ ends (3′-Ps) blocked integration, demonstrating that 3′-OH groups were necessary (Fig. 4*D*), similar to I-E Cas1–Cas2 (24, 26). Based on our high-throughput in vivo acquisition data (29), we anticipated that Cas1–Cas2-3 would typically use 32-bp protospacers but tolerate other lengths. Therefore, we tested integration of different-length protospacers (Fig. 4*E*). The 32- and 27-bp protospacers were integrated but integration was severely reduced for substrates <27 bp, which correlated with their inability to outcompete binding of the 32-bp protospacer to Cas1–Cas2-3 (Fig. S3*F*). In addition, a 60-bp substrate was integrated, albeit less efficiently (Fig. S3*G*). Because the integration assay cannot discriminate full-site from half-site intermediates, we consider that integration of longer substrates represents half-site events, where Cas1–Cas2-3 binds to the ends of the dsDNA. Therefore, these would not form new spacers in vivo. Indeed, in vivo spacers 34 bp or longer constitute <0.5% of events, and very rarely are 40- to 50-bp spacers detected (29). In the protospacer-bound type I-E Cas1–Cas2 structures, the 23-bp duplex is "bracketed" by two Cas1 tyrosines that splay the DNA and position 5 nt of the 3′ end of the ssDNA in the active site (22, 23). *P. atrosepticum* Cas1 has a histidine at the equivalent position (His-26) that might mediate splaying of the protospacer and, indeed, Cas1–Cas2-3 can efficiently integrate splayed substrates (Fig. S3*H*). The I-F consensus spacer is 32 bp, suggesting that Cas1–Cas2-3 typically binds a 22-bp duplex, explaining the lack of integration with protospacers <27 bp that lack a minimum 22 bp of duplex DNA and 5 nt of the 3′ end required to reach the Cas1 active site.

Next, we tested the effects of CRISPR topology on Cas1–Cas2-3–mediated integration. Integration occurred into a supercoiled plasmid containing CRISPR1, but not when it was linearized (Fig. 4*F*).

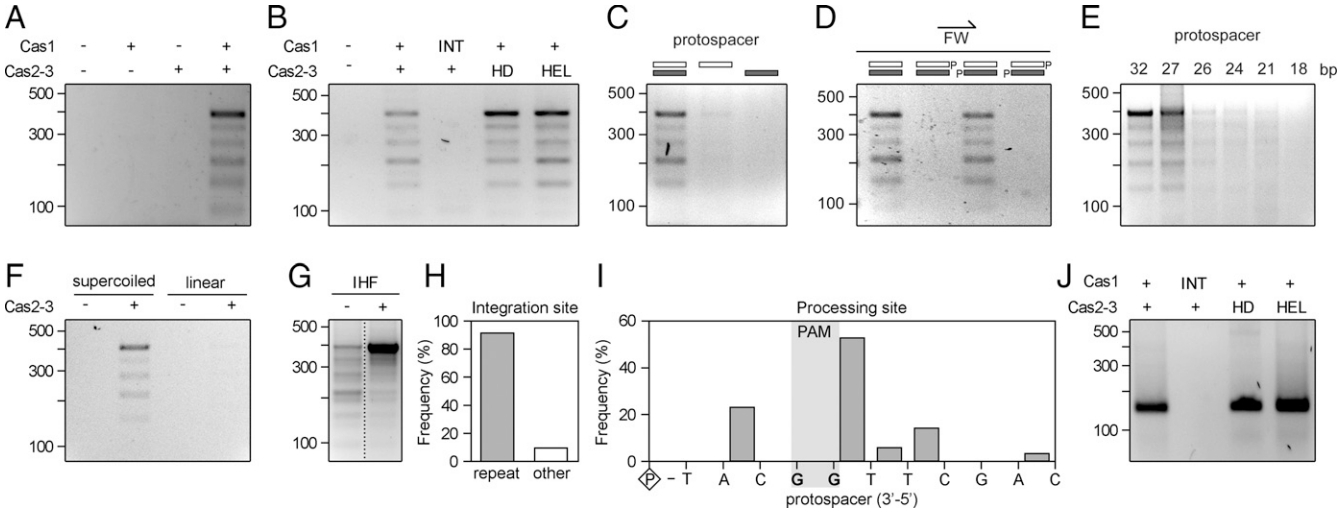

Fig. 4. In vitro integration and capture activity of Cas1–Cas2-3. (*A*–*G*) Integration assays using pCRISPR1 (*A*) with no protein, Cas1, Cas2-3, or Cas1–Cas2-3 and a 32-bp protospacer (PF1647+PF1648); (*B*) with no protein, WT Cas1–Cas2-3, or Cas1 integrase (D269A; INT), Cas3 nuclease (D124A; HD), or Cas3 helicase (D591A; HEL) mutant complexes. The protospacer was as per *A*. Cas1(D269A)–Cas2-3 was purified and assembled into a stable complex (Fig. S3); (*C*) with ssDNA or dsDNA 32-nt/bp substrates. Oligonucleotides were PF1647 and PF1648; (*D*) in the presence and absence of 3′-P–blocked oligonucleotides. Lanes show a 32-bp protospacer with free 3′-OH groups (PF1647+PF1648), an F 3′-P group (PF1882+PF1648), an R 3′-P group (PF1647+PF1883), or both 3′-P groups (PF1882+PF1883); (*E*) with dsDNA protospacers of decreasing length including 32 bp (PF1647+PF1648), 27 bp (PF1895+PF1896), 26 bp (PF2066+PF2067), 24 bp (PF1897+PF1898), 21 bp (PF1899+PF1900), and 18 bp (PF1901+PF1902); (*F*) using supercoiled pCRISPR1 or pCRISPR1 linearized by HindIII. The protospacer was as per *A*; and (*G*) with the 32-bp protospacer as per *A* incubated ± IHF. (*H*) Capture assay involving a plasmid with the CRISPR1 leader and one repeat (pPF1042). The bar graph depicts the proportion of integrated spacers found either directly adjacent to the repeat or elsewhere in the array (other) (Dataset S2). (*I*) Capture processing-site distribution following sequencing of captured substrates from *H*. (*J*) Capture assay as in *H* with WT Cas1–Cas2-3 or Cas1 integrase (D269A; INT), Cas3 nuclease (D124A; HD), or Cas3 helicase (D591A; HEL) mutant complexes. Integration was detected by PCR using primers PF1649+PF1822, except in *E* using primers PF1901+PF1822 and in *J* using primers PF1792+PF1997. The marker shown is in bp.

Our results, together with a similar supercoiled DNA preference exhibited by type I-E Cas1–Cas2 (24), suggest that DNA topology is important for integration by type I CRISPR-Cas systems. Furthermore, analysis of the CRISPR1 leader revealed three putative integration host factor (IHF)-binding sites (Fig. S2*B*). Therefore, we purified the DNA-binding protein, IHF, and tested its effect on integration. In support of recent work in the type I-E system (44), IHF promoted integration by type I-F Cas1–Cas2-3 and dramatically enhanced the specificity of leader-proximal integrations (Fig. 4*G*). IHF was also able to increase leader-proximal integrations for the Cas1–Cas2(ΔCas3) complex (Fig. S3*A*).

Spacer Capture and Integration by Cas1–Cas2-3. A critical question is how substrates are generated for integration by Cas1–Cas2-3. For type I-E systems, RecBCD may provide one route for the generation of precursor substrates for Cas1–Cas2 during naïve adaptation (16). Similarly, during priming, Cas3 has a role in substrate generation (45) and/or adaptation complex translocation (29). Therefore, we tested whether Cas1–Cas2-3 bound and processed (captured) substrates to generate protospacers proficient for integration. To identify processing sites, we blocked the ends with 3′-P and used splayed substrates to uniquely "barcode" both 5′ and 3′ ends of each DNA strand. We reconstituted a coupled capture–integration assay in vitro, showing that Cas1–Cas2-3 enabled processing and integration into a plasmid with the CRISPR1 leader and a single repeat. We cloned and sequenced these integration products (Dataset S2) and they were predominantly located precisely at the repeat–spacer junction (Fig. 4*H*, "repeat"), although some were inserted elsewhere in the array (Fig. 4*H*, "other"). Protospacer processing typically occurred adjacent to the GG PAM, although there was some off-site activity (Fig. 4*I*). Substrate capture and integration were independent of Cas3 activity, because complexes with helicase or nuclease mutations were proficient in this coupled assay (Fig. 4*J*). In contrast, the Cas1 integrase mutant complex was unable to acquire spacers, showing that integration, and perhaps capture, required the Cas1 and Cas2 parts of the complex and that Cas3 does not directly participate in these final steps of CRISPR adaptation.

Discussion

Adaptation in type I CRISPR-Cas systems involves three interrelated processes: (*i*) adaptation to threats not previously encountered (naïve); (*ii*) adaptation to those that have escaped direct interference (primed); and (*iii*) interference-driven adaptation that provides a positive feedback loop (14). Primed and interference-driven CRISPR adaptation requires coupling of both the interference (Cascade-crRNA and Cas3) and adaptation (Cas1 and Cas2) machinery, but how this is coordinated is unclear. Here we demonstrated the formation of a type I-F Cas1–Cas2-3 adaptation complex and characterized its integration activity. The complex contains two Cas1 dimers assembled onto the dimeric-scaffolding Cas2 domains of two Cas2-3 proteins in a $Cas1_4$:$Cas2\text{-}3_2$ stoichiometry. Modeling, cross-linking, and EM showed that the Cas3 domain occupies the region between the Cas1 dimers along the side of the Cas2 core, which positions the HD domain of Cas3 in proximity to the catalytic lobe of one Cas1 dimer and the C-terminal region of Cas3 next to the inactive lobe of the opposite Cas1 dimer. This Cas1 and Cas2 core is similar to the crystal structures of the I-E Cas1–Cas2 complex (19, 23, 24).

To study adaptation by Cas1–Cas2-3, we developed a sensitive PCR-based assay that detects protospacer integration without requiring radioactivity or high-throughput sequencing. Using this assay, we demonstrated that both Cas1 and Cas2-3 were needed for protospacer integration, without a requirement for the helicase and nuclease functions of the Cas3 domain, although we found a nonenzymatic role of the Cas3 domain in integration specificity (Fig. S3*A*). We observed no apparent orientation bias for which end of the protospacer integrated into the CRISPR repeat, whereas in vivo the rate of protospacer flipping is rare (29). In vivo, capture and integration might be coupled to ensure correct

protospacer orientation for nucleophilic attack. We tested this in vitro using a coupled capture–integration assay but observed no detectable bias in integration orientation, even in the presence of IHF (Fig. S4), which suggests that additional factors are involved in vivo. Interestingly, although the additional Cas3 domain in the Cas1–Cas2-3 complex is a clear structural difference from the I-E Cas1–Cas2 complex, the integration requirements were similar between the systems and Cas3 was not essential for these final steps. The Cas1 and Cas2 proteins of type I-E and I-F systems are structurally quite different. The I-F Cas1 has unique asymmetry and the sequence of the Cas2 domain of type I-F Cas2-3 is divergent (and directly fused to Cas3) (Fig. S2*C*). Because type I-F and I-E systems also differ in their PAM recognition and primed spacer acquisition mechanisms, their similarities in the integration mechanism are of considerable interest.

This type I-F Cas1–Cas2-3 complex raises the possibility of similar complexes in all type I systems. Priming occurs in type I-B, I-C, I-E, and I-F systems in multiple genera (28, 30, 33, 34, 46), suggesting that this efficient adaptation route is ubiquitous in type I systems. Therefore, the type I-F Cas1–Cas2-3 complex provides a view of adaptation complexes that are recruited to targets by the type I Cascade surveillance machineries. There are several nonmutually exclusive models for the role of Cas3 during priming. In type I-F systems, primed acquisition data are consistent with Cas2-3 helicase-dependent translocation being involved in new spacer selection, either via delivery of Cas1 (and the Cas2 domain) to the invader DNA and/or through substrate precursor generation via the HD nuclease (28, 29). A helicase-dependent model is supported by a type I-E single-molecule study, where nonnucleolytic Cas3 translocation from a primed target was stimulated by Cas1–Cas2 addition (47). However, during type I-E interference, Cas3 is recruited to Cascade-generated R-loops, unwinds DNA via its helicase activity, and feeds ssDNA to the HD nuclease for degradation (9, 10). Interference can also stimulate priming in both type I-E and I-F systems (29, 32) and, in type I-E systems, DNA degradation from Cas3 fuels priming by providing substrates to the adaptation complex (45). The Cas1–Cas2-3 structure reveals that the Cas3 HD nuclease active site is adjacent to the catalytic integrase lobe of Cas1. During priming, it is possible that ssDNA fragments liberated from the HD domain provide precursors to the Cas1–Cas2 portion of the complex either (*i*) directly or (*ii*) due to their high local concentration. Because Cas1–Cas2 can efficiently bind ssDNA but is a poor substrate for integration, it is likely that at least partially duplex DNA is required for integration. Cas3 HD activity on the opposite DNA strand might provide complementary fragments that reanneal and are then processed and integrated as new spacers. Interestingly, we detected DNA bound by Cas1–Cas2-3 by various techniques, and the native MS is suggestive of both ssDNA- and dsDNA-bound forms of the complex in vivo, suggesting reannealing could occur on the complex. Furthermore, cross-links suggested that Cas1–Cas2-3 binds protospacer DNA using a similar interface to the *E. coli* Cas1–Cas2–protospacer complex, although we cannot rule out the possibility that some of the ssDNA binds the Cas3 domain, as was observed in *T. fusca* Cas3 (9).

Importantly, the nuclease and helicase domains of Cas3 were not involved in the final capture and integration reactions catalyzed by Cas1–Cas2-3, whereas Cas1 activity was required. Therefore, we predict that naïve acquisition in type I-F systems can occur in the absence of the interference machinery when precursors are generated by other processes, such as via RecBCD (16). In contrast, naïve adaptation by I-F systems was proposed to involve all components of the interference machinery (41). Assuming the conservation of type I complexes composed of Cas1, Cas2, and Cas3, the proposed pipeline of DNA-processing active sites, starting from the helicase, to the nuclease, and then to the integrase, may account for the efficiency of primed adaptation and apply generally to type I systems. However, because the distribution of spacer selection differs between type I-E and type I-F systems, there are likely to be mechanistic distinctions.

Materials and Methods

Details of the materials and methods used in this study, including cloning and protein purification, SEC-RALS, AUC, mass spectroscopy, cross-linking analyses, electron microscopy, and structural modeling, are provided in *SI Materials and Methods*. Strains and plasmids are in Table S3, and Table S4 lists the oligonucleotides.

Integration and capture assays are described in detail in *SI Materials and Methods*. Briefly, reactions typically contained 10 nM protospacer and 70 nM Cas1–Cas2-3 and were incubated on ice for 15 min; then, 7.5 nM CRISPR plasmid was added and incubated at 25 °C for 1 h. Reactions were stopped at 65 °C for 20 min and integration was detected by PCR. Unless stated otherwise, primers were PF1649 and PF1822. Capture assays and PCR (primers PF1792+ PF1997) were performed as described above but with plasmid pPF1042, and the integration product was cloned into pGEM-T (Promega); plasmids were isolated from individual colonies and sequenced with primer PF861.

ACKNOWLEDGMENTS. We thank L. Burga for EM data collection, S. Jackson for reading the manuscript, and C. Richter and S. Luckner for purification trials. This research was funded by the Marsden Fund, Royal Society of New Zealand, a Rutherford Discovery Fellowship (to P.C.F.), and an Otago Research Grant (to M.B., P.C.F., and K.L.K.). R.H.J.S. was supported by an Otago Health Sciences Career Development Award. A.J.R.H. was financed by The Netherlands Organisation for Scientific Research (NWO) funded by the NWO Roadmap Initiative Proteins@Work (Project 184.032.201).

1. Barrangou R, Marraffini LA (2014) CRISPR-Cas systems: Prokaryotes upgrade to adaptive immunity. *Mol Cell* 54:234–244.
2. Mohanraju P, et al. (2016) Diverse evolutionary roots and mechanistic variations of the CRISPR-Cas systems. *Science* 353:aad5147.
3. Makarova KS, et al. (2015) An updated evolutionary classification of CRISPR-Cas systems. *Nat Rev Microbiol* 13:722–736.
4. Barrangou R, et al. (2007) CRISPR provides acquired resistance against viruses in prokaryotes. *Science* 315:1709–1712.
5. Brouns SJ, et al. (2008) Small CRISPR RNAs guide antiviral defense in prokaryotes. *Science* 321:960–964.
6. Hale C, Kleppe K, Terns RM, Terns MP (2008) Prokaryotic silencing (psi)RNAs in *Pyrococcus furiosus*. *RNA* 14:2572–2579.
7. Jackson RN, et al. (2014) Crystal structure of the CRISPR RNA-guided surveillance complex from *Escherichia coli*. *Science* 345:1473–1479.
8. Mulepati S, Héroux A, Bailey S (2014) Crystal structure of a CRISPR RNA-guided surveillance complex bound to a ssDNA target. *Science* 345:1479–1484.
9. Huo Y, et al. (2014) Structures of CRISPR Cas3 offer mechanistic insights into Cascade-activated DNA unwinding and degradation. *Nat Struct Mol Biol* 21:771–777.
10. Sinkunas T, et al. (2011) Cas3 is a single-stranded DNA nuclease and ATP-dependent helicase in the CRISPR/Cas immune system. *EMBO J* 30:1335–1342.
11. Westra ER, et al. (2012) CRISPR immunity relies on the consecutive binding and degradation of negatively supercoiled invader DNA by Cascade and Cas3. *Mol Cell* 46: 595–605.
12. Amitai G, Sorek R (2016) CRISPR-Cas adaptation: Insights into the mechanism of action. *Nat Rev Microbiol* 14:67–76.
13. Sternberg SH, Richter H, Charpentier E, Qimron U (2016) Adaptation in CRISPR-Cas systems. *Mol Cell* 61:797–808.
14. Jackson SA, et al. (2017) CRISPR-Cas: Adapting to change. *Science* 356:eaal5056.
15. Ivančić-Baće I, Cass SD, Wearne SJ, Bolt EL (2015) Different genome stability proteins underpin primed and naïve adaptation in *E. coli* CRISPR-Cas immunity. *Nucleic Acids Res* 43:10821–10830.
16. Levy A, et al. (2015) CRISPR adaptation biases explain preference for acquisition of foreign DNA. *Nature* 520:505–510.
17. Mojica FJ, Diez-Villasenor C, Garcia-Martinez J, Almendros C (2009) Short motif sequences determine the targets of the prokaryotic CRISPR defence system. *Microbiology* 155:733–740.
18. Yosef I, Goren MG, Qimron U (2012) Proteins and DNA elements essential for the CRISPR adaptation process in *Escherichia coli*. *Nucleic Acids Res* 40:5569–5576.
19. Nuñez JK, et al. (2014) Cas1-Cas2 complex formation mediates spacer acquisition during CRISPR-Cas adaptive immunity. *Nat Struct Mol Biol* 21:528–534.
20. Wiedenheft B, et al. (2009) Structural basis for DNase activity of a conserved protein implicated in CRISPR-mediated genome defense. *Structure* 17:904–912.
21. Richter C, Gristwood T, Clulow JS, Fineran PC (2012) In vivo protein interactions and complex formation in the *Pectobacterium atrosepticum* subtype I-F CRISPR/Cas system. *PLoS One* 7:e49549.
22. Nuñez JK, Harrington LB, Kranzusch PJ, Engelman AN, Doudna JA (2015) Foreign DNA capture during CRISPR-Cas adaptive immunity. *Nature* 527:535–538.
23. Wang J, et al. (2015) Structural and mechanistic basis of PAM-dependent spacer acquisition in CRISPR-Cas systems. *Cell* 163:840–853.

24. Nuñez JK, Lee AS, Engelman A, Doudna JA (2015) Integrase-mediated spacer acquisition during CRISPR-Cas adaptive immunity. *Nature* 519:193–198.
25. Wright AV, Doudna JA (2016) Protecting genome integrity during CRISPR immune adaptation. *Nat Struct Mol Biol* 23:876–883.
26. 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.
27. Cady KC, O'Toole GA (2011) Non-identity-mediated CRISPR-bacteriophage interaction mediated via the Csy and Cas3 proteins. *J Bacteriol* 193:3433–3445.
28. Richter C, et al. (2014) Priming in the type I-F CRISPR-Cas system triggers strand-independent spacer acquisition, bi-directionally from the primed protospacer. *Nucleic Acids Res* 42:8516–8526.
29. Staals RH, et al. (2016) Interference-driven spacer acquisition is dominant over naive and primed adaptation in a native CRISPR-Cas system. *Nat Commun* 7:12853.
30. Datsenko KA, et al. (2012) Molecular memory of prior infections activates the CRISPR/Cas adaptive bacterial immunity system. *Nat Commun* 3:945.
31. Fineran PC, et al. (2014) Degenerate target sites mediate rapid primed CRISPR adaptation. *Proc Natl Acad Sci USA* 111:E1629–E1638.
32. Semenova E, et al. (2016) Highly efficient primed spacer acquisition from targets destroyed by the *Escherichia coli* type I-E CRISPR-Cas interfering complex. *Proc Natl Acad Sci USA* 113:7626–7631.
33. Li M, Wang R, Zhao D, Xiang H (2014) Adaptation of the *Haloarcula hispanica* CRISPR-Cas system to a purified virus strictly requires a priming process. *Nucleic Acids Res* 42:2483–2492.
34. Rao C, et al. (2016) Active and adaptive *Legionella* CRISPR-Cas reveals a recurrent challenge to the pathogen. *Cell Microbiol* 18:1319–1338.
35. Schwanhäusser B, et al. (2011) Global quantification of mammalian gene expression control. *Nature* 473:337–342.
36. Jackson RN, Lavin M, Carter J, Wiedenheft B (2014) Fitting CRISPR-associated Cas3 into the helicase family tree. *Curr Opin Struct Biol* 24:106–114.
37. Wang X, et al. (2016) Structural basis of Cas3 inhibition by the bacteriophage protein AcrF3. *Nat Struct Mol Biol* 23:868–870.
38. Wilkinson ME, et al. (2016) Structural plasticity and in vivo activity of Cas1 from the type I-F CRISPR-Cas system. *Biochem J* 473:1063–1072.
39. Wang J, et al. (2016) A CRISPR evolutionary arms race: Structural insights into viral anti-CRISPR/Cas responses. *Cell Res* 26:1165–1168.
40. Musharova O, et al. (2017) Spacer-length DNA intermediates are associated with Cas1 in cells undergoing primed CRISPR adaptation. *Nucleic Acids Res* 45:3297–3307.
41. Vorontsova D, et al. (2015) Foreign DNA acquisition by the I-F CRISPR-Cas system requires all components of the interference machinery. *Nucleic Acids Res* 43:10848–10860.
42. Beloglazova N, et al. (2008) A novel family of sequence-specific endoribonucleases associated with the clustered regularly interspaced short palindromic repeats. *J Biol Chem* 283:20361–20371.
43. Nam KH, et al. (2012) Double-stranded endonuclease activity in *Bacillus halodurans* clustered regularly interspaced short palindromic repeats (CRISPR)-associated Cas2 protein. *J Biol Chem* 287:35943–35952.
44. Nuñez JK, Bai L, Harrington LB, Hinder TL, Doudna JA (2016) CRISPR immunological memory requires a host factor for specificity. *Mol Cell* 62:824–833.
45. Künne T, et al. (2016) Cas3-derived target DNA degradation fragments fuel primed CRISPR adaptation. *Mol Cell* 63:852–864.
46. Almendros C, Guzmán NM, García-Martínez J, Mojica FJ (2016) Anti-cas spacers in orphan CRISPR4 arrays prevent uptake of active CRISPR-Cas I-F systems. *Nat Microbiol* 1:16081.
47. Redding S, et al. (2015) Surveillance and processing of foreign DNA by the *Escherichia coli* CRISPR-Cas system. *Cell* 163:854–865.
48. Schuck P (2000) Size-distribution analysis of macromolecules by sedimentation velocity ultracentrifugation and Lamm equation modeling. *Biophys J* 78:1606–1619.
49. Rose RJ, Damoc E, Denisov E, Makarov A, Heck AJ (2012) High-sensitivity Orbitrap mass analysis of intact macromolecular assemblies. *Nat Methods* 9:1084–1086.
50. Kao A, et al. (2011) Development of a novel cross-linking strategy for fast and accurate identification of cross-linked peptides of protein complexes. *Mol Cell Proteomics* 10: M110.002212.
51. Liu F, Rijkers DT, Post H, Heck AJ (2015) Proteome-wide profiling of protein assemblies by cross-linking mass spectrometry. *Nat Methods* 12:1179–1184.
52. Cox J, Mann M (2008) MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. *Nat Biotechnol* 26:1367–1372.
53. Biasini M, et al. (2014) SWISS-MODEL: Modelling protein tertiary and quaternary structure using evolutionary information. *Nucleic Acids Res* 42:W252–W258.
54. Schrödinger L (2010) The PyMOL Molecular Graphics System. Version 1.7.6.2. Available at www.pymol.org.
55. Pettersen EF, et al. (2004) UCSF Chimera—A visualization system for exploratory research and analysis. *J Comput Chem* 25:1605–1612.
56. Rames M, Yu Y, Ren G (2014) Optimized negative staining: A high-throughput protocol for examining small and asymmetric protein structure by electron microscopy. *J Vis Exp* (90):e51087.
57. Mindell JA, Grigorieff N (2003) Accurate determination of local defocus and specimen tilt in electron microscopy. *J Struct Biol* 142:334–347.
58. Ludtke SJ, Baldwin PR, Chiu W (1999) EMAN: Semiautomated software for high-resolution single-particle reconstructions. *J Struct Biol* 128:82–97.
59. Scheres SH (2012) RELION: Implementation of a Bayesian approach to cryo-EM structure determination. *J Struct Biol* 180:519–530.
60. Mastronarde DN (1997) Dual-axis tomography: An approach with alignment methods that preserve resolution. *J Struct Biol* 120:343–352.
61. Di Tommaso, et al. (2011) T-Coffee: A web server for the multiple sequence alignment of protein and RNA sequences using structural information and homology extension. *Nuc acids Res* 39:W13-7.
62. Przybilski R, et al. (2011) Csy4 is responsible for CRISPR RNA processing in *Pectobacterium atrosepticum*. *RNA Biol* 8:517–528.