Cas1–Cas2 complex formation mediates spacer acquisition during CRISPR–Cas adaptive immunity

James K Nuñez, Philip J Kranzusch, Jonas Noeske, Addison V Wright, Christopher W Davies & Jennifer A Doudna

The initial stage of CRISPR–Cas immunity involves the integration of foreign DNA spacer segments into the host genomic CRISPR locus. The nucleases Cas1 and Cas2 are the only proteins conserved among all CRISPR–Cas systems, yet the molecular functions of these proteins during immunity are unknown. Here we show that Cas1 and Cas2 from Escherichia coli form a stable complex that is essential for spacer acquisition and determine the 2.3-Å-resolution crystal structure of the Cas1–Cas2 complex. Mutations that perturb Cas1–Cas2 complex formation disrupt CRISPR DNA recognition and spacer acquisition in vivo. Active site mutants of Cas2, unlike those of Cas1, can still acquire new spacers, thus indicating a nonenzymatic role of Cas2 during immunity. These results reveal the universal roles of Cas1 and Cas2 and suggest a mechanism by which Cas1–Cas2 complexes specify sites of CRISPR spacer integration.

RESULTS

Cas1 and Cas2 form a specific complex in vitro and in vivo

The E. coli K12 (MG1655) strain has two endogenous CRISPR loci, one of which is flanked by eight Cas-encoding genes (Fig. 1a). In agreement with a previously developed assay, when Cas1 and Cas2 from K12 were overexpressed in E. coli BL21-AI cells, which lack all Cas-encoding genes, new spacer acquisition was detectable.
Figure 1 Cas1 and Cas2 associate to form a complex. (a) Representation of the CRISPR–Cas locus of *E. coli* K12. The 33-bp spacers (orange and purple squares) are separated by 28-bp repeats (black diamonds). The black half arrows flanking the leader and repeat-spacer arrays represent the positions of the primers used for PCR amplification in the spacer acquisition assays in BL21-AI cells. (b) Agarose gel of the PCR-amplified CRISPR-I locus of BL21-AI cells after induced expression of empty vector, Cas1, Cas2 or both Cas1 and Cas2. Distinct bands represent the number of repeat spacer–array additions into the genomic parental CRISPR locus. (c) Flag and HA immunoprecipitations (IP) in lysates overexpressing Cas1 only, Cas2 only or both. Uncropped image of gel is shown in Supplementary Figure 5. (d) ITC trace of Cas1 injection into a Cas2-containing cell. The reported *N* (stoichiometry) and *Kd* values are averages from three independent experiments.

By PCR amplification of the CRISPR locus (Fig. 1b). We sequenced newly acquired spacers and verified that spacer acquisition in this model system retained accurate insertion of 33-bp spacers mostly derived from the foreign plasmid used for protein overexpression (Supplementary Table 1). In addition to the insertion of the 33-bp spacer, each acquisition event duplicates the first repeat (28 bp), thereby expanding the parental locus by 61 bp (refs. 5,28). Although these results demonstrate that spacer acquisition requires only the proteins Cas1 and Cas2, we observed variable PAM sequences adjacent to the protospacer in the foreign DNA. These results support the conclusion that the *E. coli* CRISPR interference machinery, the Cascade complex and Cas3 nuclease, are required for an accurate ‘priming’ process in which the interference stage is coupled to spacer acquisition to yield strict AAG PAM selection6,7,18,19.

With the finding that Cas1 and Cas2 are the only Cas proteins required for spacer acquisition, we tested whether Cas3 and Cas2 form a stable complex in vivo. We overexpressed Cas1-Flag and Cas2-hemagglutinin (HA) fusion proteins in BL21-AI cells and conducted immunoprecipitation experiments in cell lysates. We confirmed that the epitope-tagged proteins are active in acquiring new spacers (Supplementary Fig. 1a). Selective elution from Flag or HA affinity beads with either a 3×-Flag or an HA peptide resulted in the co-elution (Fig. 1c). To verify that this interaction was direct, we separately purified the untagged construct of each protein and determined the *Kd* of the interaction to be ~290 nM, as measured by isothermal titration calorimetry (ITC; Fig. 1d). The calculated stoichiometry of Cas1 to Cas2 from the ITC experiments was ~1.5. To further probe for the stoichiometry of the complex, we conducted sedimentation velocity analytical ultracentrifugation (AUC) experiments and detected a strong peak at 5.2 S with an apparent molecular weight of ~78.1 kDa (Supplementary Fig. 1b,c). This is consistent with a complex composed of one Cas1 dimer (66 kDa) and one Cas2 dimer (22 kDa). The retention time of the complex on a gel-filtration column was also consistent with the AUC experiments (Supplementary Fig. 1d). Thus, we conclude that one dimer of Cas1 and one dimer of Cas2 interact to form a heterotetramer in solution.

Crystal structure of the Cas1–Cas2 complex

To gain insights into the structural organization of the Cas1–Cas2 complex, we determined the crystal structure of the complex. Crystal structures of Cas1 and Cas2 alone from various organisms, including *E. coli* K12, have been reported (refs. 22–26,29 and PDB 4MAK). Cas1 proteins are asymmetrical homodimers with each monomer having an N-terminal β-sheet domain and C-terminal α-helical domain23,24,26. Cas2 proteins are symmetrical homodimers with a core ferredoxin fold (refs. 22,25,29 and PDB 4MAK). We purified each protein and reconstituted the complex in vitro. Gel-filtration chromatography showed the copurification of both proteins as one peak; SDS-PAGE analysis of the peak fractions also confirmed this (Supplementary Fig. 1d,e). The Cas1–Cas2 complex yielded crystals that diffracted X-rays to 2.3-Å resolution. We determined the structure by single-wavelength anomalous dispersion (SAD) on selenomethionine-derivatized crystals and refined the resulting model to an *R*_work/*R*_free of 22%/24% (Table 1).

The overall architecture of the asymmetric unit is a heterohexameric complex consisting of two Cas1 dimers (Cas1a-b and Cas1c-d) and one Cas2 dimer (Cas2a-b). The overall architecture of the asymmetric unit is a heterohexameric complex consisting of two Cas1 dimers (Cas1a-b and Cas1c-d) and one Cas2 dimer (Cas2a-b). The overall architecture of the asymmetric unit is a heterohexameric complex consisting of two Cas1 dimers (Cas1a-b and Cas1c-d) and one Cas2 dimer (Cas2a-b). The overall architecture of the asymmetric unit is a heterohexameric complex consisting of two Cas1 dimers (Cas1a-b and Cas1c-d) and one Cas2 dimer (Cas2a-b). The overall architecture of the asymmetric unit is a heterohexameric complex consisting of two Cas1 dimers (Cas1a-b and Cas1c-d) and one Cas2 dimer (Cas2a-b).

![Crystal structure of the Cas1–Cas2 complex](image-url)

**Figure 2** Crystal structure of the Cas1–Cas2 complex. (a) Ribbon diagram of the Cas1–Cas2 complex. (b) Stereo view of the Cas1–Cas2 complex superimposed on the crystal structure of Cas1 from *S. thermophilus* (PDB 4MAK). (c) Electron density map of the Cas1–Cas2 complex. The electron density map is contoured at 1σ and 2σ levels and colored blue and red, respectively.

**Table 1** Data collection and refinement statistics

| Data collection | Native | Se derivative |
|-----------------|--------|---------------|
| Space group     | *P*2₁  | *P*2₁         |
| Cell dimensions | a, b, c (Å) | 94.875, 125.70, 99.31 | 93.70, 127.70, 99.32 |
| α, β, γ (°)     | 90, 102.74, 90 | 90, 102.32, 90 |
| Resolution (Å)  | 62.85–2.3 (2.383–2.3) | 48.63–2.89 |
| *R*_merge       | 1.0301 (1.88) | 0.2045 (1.57) |
| t/α             | 9.19 (1.00) | 10.32 (1.37) |
| ComPLETENESS (%) | 97.22 (91.21) | 91.87 |
| REDUNDANCY      | 3.1 (2.9) | 7.8 (7.2) |
| Refinement      | Resolution (Å) | 62.85–2.3 (2.383–2.3) |
|                 | *R*_work/*R*_free | 0.225 / 0.245 |
|                 | No. reflections | 97,929 (9,271) |
|                 | No. atoms | 10,451 |
|                 | Protein | 9,926 |
|                 | Water | 525 |
|                 | B factors | 55.60 |
|                 | r.m.s. deviations | 53.90 |
|                 | Bond lengths (Å) | 0.003 |
|                 | Bond angles (°) | 0.62 |

One crystal was used per data set. Values in parentheses denote highest-resolution shell.
that sandwich one Cas2 dimer (Fig. 2). Cas1a and Cas1c make contacts with the Cas2 dimer, and we observed no contacts between Cas1b or Cas1d and the Cas2 dimer. The Cas1c-Cas2 protein-protein interface buries a large surface area of ~3,100 Å^2, whereas the Cas1a-Cas2 interface buries an additional 800 Å^2 contributed by the C terminus of Cas1a, as described further below. Superposition of the two Cas1 dimers (a-b dimer with c-d dimer) shows high structural similarity, with an r.m.s. deviation (r.m.s.d.) of 0.394 Å for the Ca atoms (Supplementary Fig. 2a,b). Similar contacts are present between Cas1a and Cas1c with Cas2 on opposite sides, thus creating a symmetrical complex. Although Cas1a and Cas2 predominantly form a heterotetrameric complex in solution, our crystal structure suggests that the complex may also be capable of accessing a hexameric state during acquisition.

Conformational changes and contacts within the complex

The interface between Cas1 and Cas2 consists of hydrogen-bonding, electrostatic and hydrophobic interactions. We observed extensive electrostatic contacts between three arginine residues (R245, R252 and R256) in α8 of Cas1 with two acidic residues (E65 and D84) of Cas2 (Fig. 3a and Supplementary Fig. 2c). The R252 residue is positioned between E65 and D84 and may sample salt bridges between the two acidic residues, although we observed continuous density between R252 and E65 at the Cas1a–Cas2 interface. In the same region, backbone hydrogen-bond contacts are present between the newly resolved Cas2 β7 C terminus and β4 of Cas1, as discussed further below.

To identify Cas1 and Cas2 conformational changes that occur upon complex formation, we superimposed previously determined structures of apo Cas1 (PDB 3NKD)\(^2\) and Cas2 (PDB 4MAK) from *E. coli* with the Cas1–Cas2 complex structure (Fig. 2b,c). In addition to minor conformational changes present in the canonical βαββαββαββ ferredoxin fold of Cas2, the C terminus forms two antiparallel β-sheet (β6–β7) that contact β4 of Cas1 (Figs. 2c and 3a). This region is unresolved in the apo-Cas2 structure, which terminates at the C terminus of β5. Presumably, the β6–β7 region is flexible before complex formation with Cas1.

Although Cas1 does not undergo major conformational changes upon Cas2 binding (0.69-Å backbone r.m.s.d.), the proline-rich C-terminal tail of Cas1a is distinctively ordered in only the bound state and is stabilized by hydrophobic and electrostatic contacts (Fig. 3b). At the middle of the tail, I291 from Cas1 is positioned in a hydrophobic pocket of Cas2 that includes W44 and W60 (Fig. 3b). The C terminus of Cas1c is likely to span the opposite face of Cas2 to fully encapsulate the dimer, although we did not observe electron density for this tail, owing to crystal packing of an adjacent complex on this face. We also observed structural rearrangements in the C-terminal α-helical domains of Cas1b and Cas1d as compared to apo Cas1 (Fig. 2b); however, it is unclear whether this conformational change results from complex formation or crystallographic packing of another complex next to these monomers.

**Cas1–Cas2 complex formation is required in vivo**

To determine the function of Cas1–Cas2 complex formation, we conducted spacer acquisition assays in cells expressing Cas1 and Cas2 bearing mutations at the interprotein interfaces. We found that a structured Cas2 C terminus is critical for function, because its deletion (∆β6–β7) prevented detectable spacer acquisition (Fig. 3c). This deletion removed the backbone interaction of Cas2 with β4 of Cas1 as well as with the DB8 residue at the electrostatic interface. In contrast, deletion of the Cas1 tail (AP282–S305) did not abolish spacer acquisition. Furthermore, mutation of Cas1 I291, which binds in a hydrophobic pocket with W44 and W60 of Cas2, had no effect on spacer acquisition (Fig. 3b,c). Thus, the C-terminal tail of Cas1 is not essential for

---

**Figure 2** Crystal structure of the Cas1–Cas2 complex. (a) Overall structure, consisting of a Cas2 dimer (yellow and orange) and two Cas1 dimers (denoted with suffixes a–d; blue and teal). (b) Superposition of the Cas1a–Cas1b dimer with the previously determined *E. coli* Cas1 structure (gray, PDB 3NKD\(^2\)). The blue arrows point to the last resolved residue in the 4MAK structure. The N and C indicate the termini of each monomer; r.m.s.d. values of the superpositions are indicated.
Figure 3 Disruption of complex formation affects spacer acquisition in vivo. (a) Close-up view of the Cas1a-Cas2 protein-protein interface, with annotations for the residues involved in electrostatic interactions. (b) View of the ordered C-terminal tail of Cas1a with the electron density mesh contoured at 1.0 σ. (c–e) Agarose gels of in vivo acquisition assays with mutations of Cas1 and Cas2 at the C termini (c) and the electrostatic interface for Cas1 (d) and Cas2 (e). WT, wild type. (f) Western blot of Flag immunoprecipitations in BL21-AI cells expressing Cas1-Flag and Cas2-HA or various mutations of Cas1 and Cas2. Despite the low expression of Cas2 E65R, its co-elution with Cas1 is detectable. Uncropped images of gels are shown in Supplementary Figure 5.

spacer acquisition, although it may supplement the critical interactions at the interface described below.

Mutations of residues involved in the electrostatic interactions between subunits have drastic effects on spacer acquisition. Cas1 constructs with mutations at the arginines of α8 to alanine (R245A, R252A or R256A) could still acquire spacers. However, constructs with mutations of the same residues to the opposite charge (R245D, R252E or R256E) supported little or no detectable spacer acquisition compared to that in wild-type Cas1 (Fig. 3d). To show that these mutations have little or no effect on Cas1 stability, we purified the R252E mutant to homogeneity. The mutant eluted at the expected retention time for wild-type Cas1 dimer (Supplementary Fig. 2d). In comparison, single mutations of either of the two acidic Cas2 interface residues E65 or D84 to alanine or arginine had little or no effect on spacer acquisition compared to that in wild-type Cas2. A double mutation of both residues to arginine (E65R and D84R) abolished spacer acquisition in vivo (Fig. 3e). Thus, although mutations in Cas1 at the electrostatic interface are more deleterious than those in Cas2, complementary charges in this interface permit acquisition.

To confirm that the observed in vivo effects are due to disruption of Cas1–Cas2 complex formation, we conducted Flag immunoprecipitation experiments in lysates of BL21-AI cells overexpressing Cas1-Flag and Cas2-HA mutants. Mutations that had little effect on spacer acquisition (Cas1 ∆tail and Cas2 E65R) did not perturb coprecipitation of Cas1 and Cas2, thus indicating that the mutants are still able to form the complex (Fig. 3f). In contrast, acquisition-defective mutants (Cas1 R252E and Cas2 ∆β6–β7) could no longer form a stable Cas1–Cas2 complex. This result highlights the importance of the structured Cas2 C terminus and the positioning of R252 between the two Cas2 acidic residues in the electrostatic interaction interface. Together, these findings support the conclusion that Cas1–Cas2 complex formation is required for spacer acquisition in vivo.

The catalytic activity of Cas2 is dispensable in vivo

Despite the available literature on the biochemical activities of Cas1 and Cas2, the functional roles of these proteins during spacer acquisition are unknown. Cas1 is reported to be a sequence-independent, metal-dependent nuclease that can cleave single-stranded (ss) DNA, linear and plasmid double-stranded (ds) DNA, ssRNA and various DNA-repair intermediates such as Holliday junctions. Three different Cas2 homologs have been found to have metal-dependent nuclease activity with a preference for ssRNA or dsDNA or to lack any detectable nuclease activity. The heterohexameric Cas1–Cas2 complex has five potential active sites: one for each Cas1 monomer and one at the Cas2 homodimer interface (Fig. 4a,b). We conducted spacer acquisition assays with active site residue mutations in Cas1 and Cas2 to determine whether the nuclease activities of both proteins are required.

Alanine substitution of the conserved Cas1 active site residues abolished spacer acquisition, thus demonstrating the critical role of Cas1 in metal-dependent DNA cleavage during the adaptation stage (Fig. 4c). Despite the low protein sequence conservation of Cas2 proteins, a conserved acidic residue from each monomer is positioned in the active site to coordinate a metal ion during catalysis in vitro. (Fig. 4b). Surprisingly, Cas2 mutated in the signature catalytic E9 residue to alanine or arginine supported spacer acquisition at frequencies similar to those observed in the presence of wild-type Cas2 (Fig. 4d). A mutation of this acidic residue has previously been shown to have drastic effects on nucleic acid substrate cleavage in vitro. Cas2 with an R14A mutation, which has also been shown to be catalytically inactive in the Sulfolobus solfataricus Cas2 in vitro, was still active in acquiring spacers. Of the nearby arginine residues, only the R18A construct had low spacer-acquisition levels. This residue interacts with Cas1 at the interprotein interface, as supported by its continuous electron density with the Cas1 backbone. These results support the notion that Cas1 is the likely nuclease that catalyzes the integration reaction, whereas the function of Cas2 during CRISPR–Cas immunity may not be nucleic acid cleavage.

The Cas1–Cas2 complex is essential for CRISPR-locus binding

The molecular basis for new CRISPR spacer acquisition at the leader-proximal end of the CRISPR locus has been unknown. Although it has been hypothesized that Cas1 or Cas2 might provide such spacer-acquisition
Figure 4 Complex formation is required for CRISPR DNA recognition. (a,b) Close-up views of the active sites of Cas1 (a) and Cas2 (b) with stick representations for the conserved residues mutated in vivo. In the middle is a general view of the active sites in the complex, highlighted in red. (c,d) Acquisition assays of active site residue mutations of Cas1 (c) and Cas2 (d). (e) Western blot of fractions in the biotinylated DNA affinity precipitations. The cartoon representations at top are the DNA constructs used with the 5′-biotin labels (stars). W.C.L., whole cell lysate. (f) Western blot of the DNA affinity precipitations in BL21-AI lysates from overexpression of Cas1-Flag only, Cas2-HA only or both. (g) DNA affinity precipitations in the same lysates as in e, with the same DNA constructs as in e. Uncropped images of gels are shown in Supplementary Figure 5.

specifcity, previous studies have focused on the individual activities of these two proteins. These studies reported sequence-nonspecific DNA-binding properties of purified Cas1 and Cas2 (refs. 23–26,30). Our discovery that Cas1 and Cas2 form an essential complex that is required for CRISPR spacer acquisition in vivo led us to test for CRISPR DNA binding by Cas1 and Cas2. We initially conducted electrophoretic mobility shift assays of purified Cas1 and Cas2, either alone or as a complex, with various DNA substrates. In agreement with previous findings for either protein alone, purified Cas1–Cas2 complex had no sequence-specific DNA binding activity. These results suggest that other host factors may be required to stimulate loading of Cas1 and/or Cas2 on the CRISPR locus.

To alternatively probe for CRISPR-locus binding specificity, we conducted biotinylated DNA–affinity precipitation assays in lysates of BL21-AI cells overexpressing Cas1-Flag and Cas2-HA (Supplementary Fig. 3a). We first tested the ability of Cas1 and/or Cas2 to bind a 186-bp 5′-biotinylated dsDNA containing two CRISPR repeats, two spacers and the minimal 60-bp leader sequence shown to be required for spacer acquisition in vivo. For a control, we used a DNA of similar length with no CRISPR sequence. After a series of washes to remove nonspecific binders, western blot analysis of the elution samples confirmed the preferential binding of Cas1 to the CRISPR DNA compared to the control DNA (Fig. 4e and Supplementary Fig. 3b). Surprisingly, we did not detect Cas2 in the elution samples; this could be because of the washing conditions removing the weakly bound Cas2. To determine whether Cas2 is required for the preferential binding of Cas1 to the CRISPR DNA, we conducted the affinity precipitation experiment in BL21-AI cell lysates containing overexpressed Cas1 only, Cas2 only or Cas1 and Cas2. Although we did not detect the presence of Cas2 in any of the elution samples, we found that Cas1 lost preference for CRISPR DNA binding in the absence of Cas2 (Fig. 4f). Cas1 was no longer able to recognize a DNA substrate when the conserved CRISPR leader sequence was replaced with random DNA (Supplementary Fig. 3c,d), thus indicating, in agreement with previous in vivo results, that sequence- or structure-specific interactions with the leader DNA may be required to direct spacer acquisition. To determine whether a linear motif accounts for sequence-specific recognition of CRISPR DNA, we conducted DNA affinity precipitation experiments, using dsDNA substrates of equal length that contained scrambled portions of the CRISPR leader sequence (Supplementary Fig. 3c,d). In contrast to the severe binding defect resulting from complete removal of the CRISPR leader sequence, shorter scrambled stretches have a much less pronounced effect on the ability of Cas1 to recognize the DNA substrate. These results suggest that Cas1 recognition of the CRISPR leader sequence occurs through a yet-unknown nonlinear sequence or structural basis.

Upon finding that disruption of Cas1–Cas2 complex formation negatively affects spacer acquisition, we tested whether this defect is due in part to the inability of the complex to recognize the CRISPR locus. We conducted the DNA affinity purifications in lysates of cells expressing Cas1 and Cas2 mutants that we tested previously for in vitro complex formation (Fig. 3). Mutants that supported spacer acquisition and formed a complex (Cas1 Δtail and Cas2 ΔE65R) retained the ability to bind the CRISPR DNA (Supplementary Fig. 3e). In contrast, mutants that did not support spacer acquisition (Cas1 R252E and Cas2 Δβ6–B7) lost the preference for CRISPR DNA recognition. A mutation of the active site E9 residue of Cas2 had no effect on complex formation or CRISPR DNA binding (Fig. 4g,h). Thus, mutations that disrupt complex formation may have lost the ability to support spacer acquisition, because of the inability to recognize the leader-repeat sequence of the CRISPR locus.

DISCUSSION

The acquisition of new spacer sequences into the CRISPR locus as part of the adaptive immune response in bacteria requires the two conserved CRISPR-associated proteins Cas1 and Cas2. Our findings show that Cas1 and Cas2 assemble into a stable complex whose formation
is essential for the incorporation of foreign DNA spacers into the host CRISPR locus in vivo. The 2.3-Å crystal structure of the Cas1–Cas2 complex reveals a 2:1 stoichiometry in which a Cas2 dimer binds two Cas1 dimers to form a crablike architecture that specifies the site of integration at the leader end of the CRISPR locus. In solution, the complex is stable as a heterotetramer containing one dimer each of Cas1 and Cas2, thus leaving open the possibility that the tetrameric form is a functional unit during integration.

Our findings point to likely biochemical functions of Cas1 and Cas2 within the complex. Both proteins have been investigated independently and shown to possess nonspecific nuclease activity in vitro. On the basis of our active site mutational studies in vivo (Fig. 4), in which catalytically defective Cas1 mutants were incapable of supporting spacer acquisition, Cas1 functions as a bona fide nuclease involved in the adaptation stage of CRISPR–Cas immunity. In contrast, the catalytic activity of Cas2 is unnecessary for integration of sequences into the CRISPR locus in vivo. Furthermore, the observation that Cas2 does not coprecipitate with the biotinylated DNA probes suggests that Cas2 may bind weakly to the Cas1–DNA complex or may not bind directly to DNA within the Cas1–Cas2 complex. Together with the finding that Cas1–Cas2 complexes have a marked preference for binding to the CRISPR locus, which serves as the target site for spacer integration (Fig. 4), these results suggest that Cas2 recruits Cas1 to the leader sequence through an indirect mechanism. It remains possible that the nuclease activity ofCas2 contributes to a CRISPR-independent process, as suggested by the structural homology between Cas2 and the VapDHi toxin of the VapDHi-VapX toxin-antitoxin system in Haemophilus influenzae.

Cas1’s ability to assemble with Cas2, in addition to its catalytic function, is essential for spacer acquisition. Mutations in either Cas1 or Cas2 that disrupt Cas1–Cas2 complex formation in vitro also interfere with spacer acquisition in vivo. Furthermore, this functionally critical interaction is conserved across divergent CRISPR systems. Recent experiments have provided evidence for Cas1–Cas2–containing complexes in the type I-A CRISPR system in the crenarchaeon Thermoproteus tenax—in which Cas1 and Cas2 exist as a fusion protein—and in the type I-F system in the plant pathogen Pectobacterium atrosepticum. Despite the essential nature of the Cas1–Cas2 interaction in E. coli, we note that the observed interprotein interface contacts may not be conserved in other CRISPR–Cas systems. Structural alignment of available Cas1 and Cas2 crystal structures shows poor conservation of the three critical arginines that form salt bridges with E65 and D84 of Cas2 (Supplementary Fig. 4a,b). These residues may vary in other Cas1–Cas2 protein complexes, or they may be replaced by different interactions that ensure Cas1–Cas2 assembly in divergent CRISPR–Cas systems.

Alignment of Cas2 crystal structures also reveals substantial structural flexibility outside of the core ββαβαβä ferredoxin fold (Supplementary Fig. 4c). In particular, the C terminus of the E. coli Cas2 in the Cas1–Cas2 complex is positioned ~90° from its position in the other Cas2 structures. The structural changes in the Cas2 C-terminal β6–β7 strands that we observe in the Cas1–Cas2 complex, and their requirement for both complex stability and in vivo spacer acquisition (Fig. 3c), underscore the role of Cas2 as a central structural component of the Cas1–Cas2 integration complex.

Together with previous work, our findings establish that at least two multiprotein complexes are fundamental for a fully functioning type I CRISPR–Cas system: a Cas1–Cas2 spacer-acquisition complex and an RNA-guided DNA-interference complex. Whether these complexes interact to form a multifunctional supercomplex is not yet known. However, an interesting hint about this possibility comes from P. atrosepticum, in which Cas2 exists as an N-terminal fusion with Cas3, the foreign DNA–targeting nuclease recruited by the DNA-interference complex. The Cas2-Cas3 fusion protein in this organism has also been shown to associate with Cas1 (ref. 33). It is thus possible that at least some type I CRISPR–Cas systems use the Cas1–Cas2 complex not only for new spacer acquisition but also for coupling this process to target recognition and destruction.

METHODS
Methods and any associated references are available in the online version of the paper.

Accession codes. Coordinates and structure factors for the Cas1–Cas2 complex have been deposited in the Protein Data Bank under accession code 4P61.

Note: Any Supplementary Information and Source Data files are available in the online version of the paper.

ACKNOWLEDGMENTS
We are grateful for the input on this work provided by members of the Doudna laboratory. We thank S. Floor, A.S. Lee, H.Y. Lee, R. Wilson, R. Wu and K. Zhou for technical assistance, the 8.3.1 beamline staff at the Advanced Light Source and A. Iavarone (University of California, Berkeley) for MS. We thank D. King (Howard Hughes Medical Institute, University of California, Berkeley) for Flag and HA peptides. This project was funded by a US National Science Foundation grant to J.A.D. (no. 1244557). J.K.N. and A.V.W. are supported by US National Science Foundation Graduate Research Fellowships and J.K.N. by a University of California, Berkeley Chancellor’s Fellowship. P.J.K. is supported as a Howard Hughes Medical Institute Fellow of the Life Sciences Research Foundation. J.N. is supported by a Long-Term Postdoctoral Fellowship from the Human Frontier Science Program Organization. J.A.D. is supported as an Investigator of the Howard Hughes Medical Institute.

AUTHOR CONTRIBUTIONS
J.K.N. performed the protein purification, biochemical and crystallography experiments. X-ray diffraction data were collected by J.K.N., P.J.K. and J.N., and structure determination was performed by J.K.N. and P.J.K. A.V.W. assisted J.K.N. with in vivo acquisition and immunoprecipitation assays. C.W.D. performed and analyzed analytical ultracentrifugation experiments. J.N. and J.A.D. designed the study, analyzed all data and wrote the manuscript.

COMPETING FINANCIAL INTERESTS
The authors declare no competing financial interests.

Reprints and permissions information is available online at http://www.nature.com/reprints/index.html.

1. Sorek, R., Lawrence, C.M. & Wiedenheft, B. CRISPR-mediated adaptive immune systems in bacteria and archaea. Annu. Rev. Biochem. 82, 237–266 (2013).
2. Mojica, F.J., Diez-Villasenor, C., Garcia-Martinez, J. & Soria, E. Intervening sequences of regularly spaced prokaryotic repeats derive from foreign genetic elements. J. Mol. Evol. 60, 174–182 (2005).
3. Bolotin, A., Quinquis, B., Sorkin, A. & Ehrlich, S.D. Clustered regularly interspaced short palindromic repeats (CRISPRs) have spacers of extrachromosomal origin. Microbiology 151, 2551–2561 (2005).
4. Barrangou, R. et al. CRISPR provides acquired resistance against viruses in prokaryotes. Science 315, 1709–1712 (2007).
5. Yusof, I., Goren, M.G. & Qimron, U. Proteins and DNA elements essential for the CRISPR adaptation process in Escherichia coli. Nucleic Acids Res. 40, 5569–5576 (2012).
6. Swarts, D.C., Mosterd, C., van Passe, M.W. & Brouns, S.J. CRISPR interference directly controls specific spacer acquisition. PLoS ONE 7, e35888 (2012).
7. Datsenko, K.A. et al. Molecular memory of prior infections activates the CRISPR/Cas adaptive bacterial immunity system. Nat. Commun. 3, 945 (2012).
8. Brouns, S.J. et al. Small CRISPR RNAs guide antiviral defense in prokaryotes. Science 321, 960–964 (2008).
9. Carle, J., Wang, R., Li, H., Tens, R.M. & Tens, P.M. Cas6 is an endoribonuclease that generates guide RNAs for invader defense in prokaryotes. Genes Dev. 22, 3489–3496 (2008).
10. Haurwitz, R.E., Jin, M., Wiedenheft, B., Zhou, K. & Doudna, J.A. Sequence- and structure-specific RNA processing by a CRISPR endonuclease. Science 329, 1355–1358 (2010).
11. Deltcheva, E. et al. CRISPR RNA maturation by trans-encoded small RNA and host factor RNase III. Nature 471, 602–607 (2011).
12. Sashital, D.G., Jinek, M. & Doudna, J.A. An RNA-induced conformational change required for CRISPR RNA cleavage by the endoribonuclease Cse3. Nat. Struct. Mol. Biol. 18, 680–687 (2011).
13. Wiedenheft, B. et al. Structures of the RNA-guided surveillance complex from a bacterial immune system. Nature 477, 486–489 (2011).
14. Jinek, M. et al. CRISPR RNA maturation by trans-encoded small RNA and host factor RNase III. Nature 471, 602–607 (2011).
15. Garneau, J.E. et al. The CRISPR/Cas bacterial immune system cleaves bacteriophage and plasmid DNA. Nature 468, 67–71 (2010).
16. Jore, M.M. et al. Structural basis for CRISPR RNA-guided DNA recognition by Cascade. Nat. Struct. Mol. Biol. 18, 529–536 (2011).
17. Makarova, K.S. et al. Evolution and classification of the CRISPR-Cas systems. Nat. Rev. Microbiol. 9, 467–477 (2011).
18. Savitskaya, E., Semenova, E., Dedkov, V., Metlitskaya, A. & Severinov, K. High-throughput analysis of type I-E CRISPR/Cas spacer acquisition in E. coli. RNA Biol. 10, 716–725 (2013).
19. Diez-Villaseñor, C., Guzman, N.M., Almendros, C., Garcia-Martinez, J. & Mojica, F.J. CRISPR-spacer integration reporter plasmids reveal distinct genuine acquisition specificities among CRISPR-Cas I-E variants of Escherichia coli. RNA Biol. 10, 792–802 (2013).
20. Mojica, F.J., Diez-Villasenor, C., Garcia-Martinez, J. & Almendros, C. Short motif sequences determine the targets of the prokaryotic CRISPR defence system. Microbiology 155, 733–740 (2009).
21. Sashital, D.G., Wiedenheft, B. & Doudna, J.A. Mechanism of foreign DNA selection in a bacterial adaptive immune system. Mol. Cell 46, 606–615 (2012).
22. Beloglavova, N. et al. A novel family of sequence-specific endonucleases associated with the clustered regularly interspaced short palindromic repeats. J. Biol. Chem. 283, 20361–20371 (2008).
Sedimentation velocity experiments were previously described and subsequently purified with the same purification protocol used for the native proteins. The selenomethionine proteins were overexpressed in minimal medium as described. The sample was loaded on a Superdex 75 (16/60) size-exclusion column, and with buffer B in the absence of imidazole. The selenomethionine-derivatized complex was concentrated to ~4 mg mL\(^{-1}\) and crystallized under cryogenic conditions at the Lawrence Berkeley National Laboratory Advanced Light Source (beamline 8.3.1). The diffraction data were collected in the presence of TEV protease to remove the affinity tags. The protein was concentrated and further purified on an MBPTrap HP (GE Healthcare) column and a Superdex 75 (16/60) size exclusion chromatography column in tandem, with buffer B in the absence of imidazole. The eluted protein was dialyzed against buffer B at 4 °C in the presence of TEV protease to obtain clear separation between the Cas1–Cas2 complex and Cas1-only peaks on gel filtration. There was no difference in the retention time of the complex when the proteins were preincubated at a 1:1 ratio.

In vitro complex formation, crystallization and structure determination. Purified Cas1 and Cas2 were separately dialyzed against 150 mM KCl, 20 mM HEPES-KOH, pH 7.4, 5% glycerol and 1 mM TCEP at 4 °C overnight. The proteins were incubated together at a 1:3 Cas1/Cas2 molar ratio for 1 h on ice. The sample was loaded on a Superdex 75 (16/60) size-exclusion column, and the peak fractions corresponding to the complex were pooled and concentrated for crystallization. We note that the molar ratio of Cas1/Cas2 for preincubation of an overnight culture in LB medium containing 50 µg mL\(^{-1}\) streptomycin, a sample was transferred (1:300) into a 10–ml culture containing 0.2% l-arabinose, 0.1 mM IPTG and 50 µg µL\(^{-1}\) streptomycin to induce protein expression. After 20–24 h, a sample of the culture was diluted in water, boiled at 95 °C for 5 min and centrifuged (16,100g). A sample of the supernatant was used as template for PCR amplification of the CRISPR locus with the same primers as previously described. The PCR reactions were analyzed on 1.5% agarose gels. For comparison in spacer-acquisition efficacy of mutant proteins, the OD\(_{600}\) of each culture was measured, and the amount of culture obtained for PCR amplification was normalized accordingly. The newly acquired spacers reported in Supplementary Table 1 were obtained by plating a sample of the culture on LB agar plates and amplifying the CRISPR-I locus of single clones to detect locus expansion. The products of clones with expanded loci were submitted for sequencing. The gene annotations were obtained from the NCBI Basic Local Alignment Search Tool (BLAST) with the E. coli BL21 (taxid: 469008) genome sequence. All of the acquisition assays reported in this study have been replicated at least three times.

Immunoprecipitation assays. The Cas1-Flag and Cas2-HA constructs were both cloned into pCDB-1b (Novagen) and transformed into E. coli BL21-AI (Invitrogen). After preparation of an overnight culture in LB medium containing 50 µg mL\(^{-1}\) streptomycin, the cells were pelleted and resuspended in lysis buffer (150 mM KCl, 50 mM Tris, pH 7.5, 1 mM TCEP, 1% Triton X-100, 0.5 mM PMSF and protease inhibitors). After sonication on ice, the lysates were cleared and rocked for 1.5–2 h at 4 °C with either anti-Flag M2 or anti-HA affinity resin (Sigma-Alrich). The resin was washed five times with 400 mM KCl, 50 mM Tris, pH 7.5, 1 mM TCEP and 1% Triton X-100. The proteins were eluted with either 100 ng µL\(^{-1}\) Flag peptide (DYKDDDDK) or HA peptide (YPYDVPDYA), synthesized by David King (HHMI, UC Berkeley). The epitope-tagged proteins were detected with monoclonal anti-Flag M2 or anti-HA mouse antibody (Cell Signaling 2999S, 1:10,000) or HRP-conjugated anti-HA mouse antibody (Cell Signaling 2999S, 1:10,000). Validation information for the antibodies is available on the manufacturers’ websites. All of the Flag immunoprecipitation experiments reported in this study have been replicated at least three times.

DNA affinity precipitation assays. The 186-bp CRISPR DNA bait was generated by PCR amplification of the BL21-AI CRISPR-I locus with 5′-biotin–conjugated forward and reverse primers (synthesized by Integrated Device Technology). The 186-bp control DNA was PCR amplified from the ori sequence of the pUC19 vector. The input lysates were prepared as described above for IP assays. The amount of biotinylated DNA probe was normalized to 100 nM, and the probe was rocked with the lysate at 4 °C for 30 min. Avidin agarose (Pierce; Fig. 4c) or streptavidin magnetic beads (NEB; Supplementary Fig. 4b) were added to the reaction and rocked for an additional 1.5–2 h. The samples were washed five times with lysis buffer, and the proteins were eluted with Laemmli buffer by boiling at 95 °C for 5 min. Western blotting was conducted to detect Cas1-Flag and Cas2-HA in the samples as described in the IP procedure.

Structure-based sequence alignments. The amino acid sequences of the Cas1 and Cas2 proteins were obtained from the RCSB Protein Data Bank. The alignment was generated by PROMALS3D, and the output was analyzed on Jalview. The BLOSUM62 score threshold on Jalview was set to 50% to generate the conservation colors. The PDB IDs of the Cas1 structures are 3LFX.
(Thermotoga maritima), 3PV9 (Pyrococcus horikoshii), 2Y2S (Aquifex aeolicus) and 3GOD (Pseudomonas aeruginosa). The PDB IDs of the Cas2 structures are 3OQ2 (Desulfovibrio vulgaris), 4ES2 (Bacillus halodurans), 1ZPW (Thermus thermophilus), 2IOX (Pyrococcus furiosus) and 2I8E (Sulfolobus solfataricus).

34. Kranzusch, P.J., Lee, A.S., Berger, J.M. & Doudna, J.A. Structure of human cGAS reveals a conserved family of second-messenger enzymes in innate immunity. Cell Reports 3, 1362–1368 (2013).

35. Van Duyne, G.D., Standaert, R.F., Karplus, P.A., Schreiber, S.L. & Clardy, J. Atomic structures of the human immunophilin FKBP-12 complexes with FK506 and rapamycin. J. Mol. Biol. 229, 105–124 (1993).

36. Kabsch, W. Xds. Acta Crystallogr. D Biol. Crystallogr. 66, 125–132 (2010).

37. Evans, P. Scaling and assessment of data quality. Acta Crystallogr. D Biol. Crystallogr. 62, 72–82 (2006).

38. Adams, P.D. et al. PHENIX: a comprehensive Python-based system for macromolecular structure solution. Acta Crystallogr. D Biol. Crystallogr. 66, 213–221 (2010).

39. Emsley, P. & Cowtan, K. Coot: model-building tools for molecular graphics. Acta Crystallogr. D Biol. Crystallogr. 60, 2126–2132 (2004).

40. Laue, T.M., Shah, B.D., Ridgeway, T.M. & Pelleiter, S.L. Analytical Ultracentrifugation in Biochemistry and Polymer Science 90–125 (Royal Society of Chemistry, 1992).

41. Brown, P.H. & Schuck, P. Macromolecular size-and-shape distributions by sedimentation velocity analytical ultracentrifugation. Biophys. J. 90, 4651–4661 (2006).

42. Schuck, P. Size-distribution analysis of macromolecules by sedimentation velocity ultracentrifugation and LamM equation modeling. Biophys. J. 78, 1606–1619 (2000).

43. Altschul, S.F., Gish, W., Miller, W., Myers, E.W. & Lipman, D.J. Basic local alignment search tool. J. Mol. Biol. 215, 403–410 (1990).

44. Pei, J., Kim, B.H. & Grishin, N.V. PROMALS3D: a tool for multiple protein sequence and structure alignments. Nucleic Acids Res. 36, 2295–2300 (2008).

45. Waterhouse, A.M., Procter, J.B., Martin, D.M., Clamp, M. & Barton, G.J. Jalview Version 2: a multiple sequence alignment editor and analysis workbench. Bioinformatics 25, 1189–1191 (2009).