RNA-guided complex from a bacterial immune system enhances target recognition through seed sequence interactions

Blake Wiedenheft[a,b], Esther van Duijn[c,1], Jelle B. Bultema[d,1], Sakharam P. Waghmare[e,1], Kaihong Zhou[a,1], Arjan Barendregt[c], Wiebke Westphal[b], Albert J. R. Heck[c], Egbert J. Boekema[d], Mark J. Dickman[e], and Jennifer A. Doudna[a,b,f,g,2]

[a]Howard Hughes Medical Institute; [b]Department of Molecular and Cell Biology, University of California, Berkeley, CA 94720; [f]Physical Biosciences Division, Lawrence Berkeley National Laboratory, Berkeley, CA 94720; Department of Chemistry, University of California, Berkeley, CA 94720; [g]Department of Chemistry, University of California, Berkeley, CA 94720; [c]Biomolecular Mass Spectrometry and Proteomics Group, Bijvoet Center for Biomolecular Research, Utrecht Institute for Pharmaceutical Sciences, and The Netherlands Proteomics Center, Utrecht University, Padualaan 8, 3584 CH Utrecht, The Netherlands; [d]Electron Microscopy Group, Groningen Biomolecular Sciences and Biotechnology Institute, University of Groningen, Nijenborgh 4, 9747 AG, Groningen, The Netherlands; and [e]Department of Chemical and Biological Engineering, ChELSI Institute, University of Sheffield, Mappin Street, Sheffield S1 3JD, United Kingdom

Prokaryotes have evolved multiple versions of an RNA-guided adaptive immune system that targets foreign nucleic acids. In each case, transcripts derived from clustered regularly interspaced short palindromic repeats (CRISPRs) are thought to selectively target invading phage and plasmids in a sequence-specific process involving a variable cassette of CRISPR-associated (*cas*) genes. The CRISPR locus in *Pseudomonas aeruginosa* (PA14) includes four *cas* genes that are unique to and conserved in microorganisms harboring the Csy-type (CRISPR system yersinia) immune system. Here we show that the Csy proteins (Csy1–4) assemble into a 350 kDa ribonucleoprotein complex that facilitates target recognition by enhancing sequence-specific hybridization between the CRISPR RNA and complementary target sequences. Target recognition is enthalpically driven and localized to a "seed sequence" at the 5′ end of the CRISPR RNA spacer. Structural analysis of the complex by small-angle X-ray scattering and single particle electron microscopy reveals a crescent-shaped particle that bears striking resemblance to the architecture of a large CRISPR-associated complex from *Escherichia coli*, termed Cascade. Although similarity between these two complexes is not evident at the sequence level, their unequal subunit stoichiometry and quaternary architecture reveal conserved structural features that may be common among diverse CRISPR-mediated defense systems.

Cmr | RNA interference | RNA silencing | Argonaute | surveillance system

Clustered regularly interspaced short palindromic repeats (CRISPRs) are the genetic record of an RNA-based adaptive immune system that is prevalent among prokaryotes. Each CRISPR locus consists of a series of short repeats that are separated by nonrepetitive spacer sequences derived from foreign genetic elements (1, 2). These repetitive elements rapidly expand in response to phage challenge by site-specifically integrating short fragments of the foreign DNA at one end of the evolving CRISPR (3–5). CRISPR adaptation results in sequence-specific resistance to genetic parasites containing a complementary sequence (4, 5).

The genes flanking CRISPRs encode proteins that have been implicated as mediators of these diverse immune systems. Genetic experiments in *Streptococcus thermophilus* provided initial evidence for the role of CRISPR-associated (Cas) proteins in adaptive immunity, but assigning function to *cas* genes in other organisms has been challenging due to a lack of primary sequence conservation (4). Phylogenetic analyses have identified distinct subfamilies of the CRISPR system, which are named using three letter abbreviations (reviewed in ref. 6). Each immune system includes a distinct set of 4–10 *cas* genes that are associated with a particular CRISPR repeat sequence type (7–9).

CRISPR loci are transcribed as long precursor RNAs that are recognized and processed into short CRISPR RNAs (crRNAs) by CRISPR-specific endoribonucleases. The high-resolution crystal structure of a crRNA bound to Csy4, the CRISPR-specific endoribonuclease from *Pseudomonas aeruginosa* (PA14), revealed a unique combination of sequence- and structure-specific interactions that explain this highly selective association (10). A CRISPR-specific endoribonuclease in *Escherichia coli* (Cse3, also known as CasE) performs an analogous function, although this enzyme is structurally distinct from Csy4 (11, 12).

In the *E. coli* system, CasE and its associated crRNA are essential components of a multisubunit macromolecular complex termed Cascade (CRISPR-associated complex for antiviral defense). Cascade is composed of an unequal stoichiometry of five subfamily-specific Cas proteins (Cse-type) that have been designated CasA-CasE (11). Protection against phage challenge requires both the Cascade complex and an additional protein (Cas3), which is predicted to function as a helicase and nuclease (11, 13). Although the mechanism of CRISPR-mediated phage interference in *E. coli* is not currently known, phage challenge experiments have shown that crRNAs complementary to phage DNA are significantly more effective at reducing phage titers than are crRNAs that target the corresponding RNA sequence (11). CRISPR-mediated DNA targeting also occurs in *Staphylococcus epidermidis* (Csm-type) and *S. thermophilus* (Csn-type) (5, 14). By contrast, biochemical evidence in *Pyrococcus furiosus* (Cmr-type) indicates that the Cmr proteins form a complex that specifically cleaves RNA targets at a fixed distance from the 3′-end of the crRNA (15).

Here we report the discovery of a CRISPR-associated complex from the PA14 strain of *P. aeruginosa*. The complex is composed of a unique set of proteins, which have previously been shown to be exclusive to and conserved in the Csy subfamily (CRISPR system yersinia) of CRISPR-mediated immune systems (7, 9). We show that this complex participates in target recognition by facilitating sequence-specific hybridization between the crRNA and complementary targets. Similar to mRNA recognition by Argonaute proteins during RNA interference (RNAi) in eukar-

yotes, CRISPR target selection is governed by a "seed sequence" at the 5′ end of the crRNA spacer. Although comprised of distinct proteins, the stoichiometry and the morphology of the Csy complex resemble the architecture of the Cascade complex from *E. coli*. These findings suggest that large CRISPR-associated ribonucleoproteins mediate surveillance and target recognition in diverse CRISPR-mediated immune systems.

Results

Csy Complex Assembly. The CRISPR-mediated adaptive immune system in *P. aeruginosa* (PA14) consists of two CRISPRs that flank six *cas* genes (Fig. 1*A*); crystal structures for two of the proteins associated with this immune system have been published (10, 16). Cas1 and Cas3 are frequently associated with diverse immune system subtypes, whereas the four *csy* genes (*csy*1–4) are exclusive to and conserved in microorganisms that harbor the Csy yersinia type immune system (7, 9). Partial dyad symmetry in the PA14 repeat sequence results in long precursor CRISPR transcripts consisting of a series of 28-nucleotide (nt) repeats with stable stem-loop structures that are separated by 32-nt spacer sequences (Fig. 1*A*). Csy4 selectively binds and cleaves at the 3′ end of each stem-loop structure in these pre-CRISPR RNAs, producing a library of 60-nt crRNAs that each include a unique 32-nt spacer sequence (10). Previous work has shown that Csy4 remains stably associated with the 3′ stem-loop of the crRNA after cleavage, but a role for the other Csy proteins has not been determined.
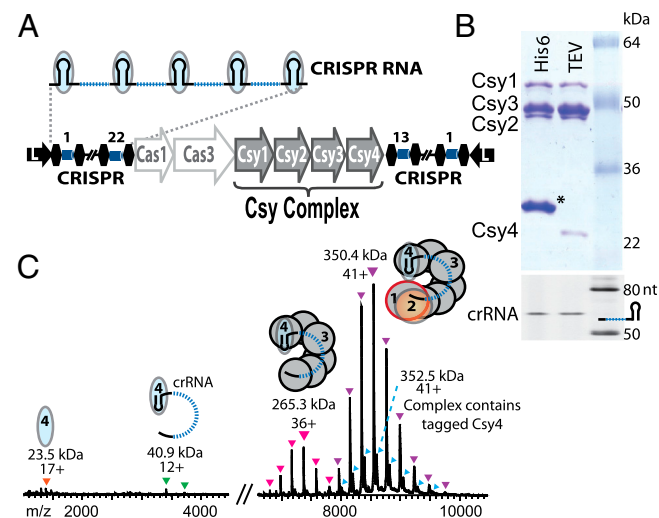

Fig. 1. The Csy proteins assemble into a large ribonucleoprotein complex. (*A*) Two CRISPR loci flank a set of *cas* genes in the *P. aeruginosa* (PA14) genome. Each CRISPR consists of a series of direct repeats (black hexagons) that are separated by unique spacer sequences (blue cylinders). Both CRISPRs are flanked by a leader sequence (black arrow). Dyad symmetry within each direct repeat results in a CRISPR RNA transcript consisting of a series of hairpins (black) that are recognized by Csy4 (cyan oval). Each repeat is separated by unique spacer sequences (dashed blue line). (*B*) Coomassie-blue stained SDS-polyacrylamide gel of the affinity purified Csy complexes (*Upper*). An N-terminal His-tags (*) on Csy4 can be used as bait to pull down the other untagged Csy proteins. The His-tag is removable by treating with TEV protease (right lane). Denaturing polyacrylamide gel of phenol extracted crRNAs isolated from the Csy complexes (*Lower*). The N-terminal His-tag does not interfere with particle assembly or CRISPR RNA processing. (*C*) Native mass spectrum of the Csy complex. The intact Csy complex has a total molecular weight of 350.4 kDa (purple triangles) with a subunit stoichiometry corresponding to $Csy1_1 : Csy2_1 : Csy3_6 : Csy4_1 : crRNA_1$. A complex with a slightly higher mass (352.5 kDa) is also observed. The additional mass in this complex is due to incomplete removal of the His-tag from the Csy4 subunit following digestion with the TEV protease (blue triangles). A complex lacking Csy1 and Csy2 was also identified (pink triangles). At the low m/z region of the spectrum free Csy4 (orange triangle) and a complex of Csy4 and crRNA (green triangles) are observed. For each distribution the charge state of the major peak is given.

To analyze the function of the Csy proteins, we cloned each of the *csy* genes from *P. aeruginosa* (PA14) with an N-terminal His6-tag and coexpressed each of these proteins with untagged versions of the other Csy proteins. Although a stable heterodimeric complex consisting of Csy1 and Csy2 could be purified independent of other proteins or RNA, Csy3, and Csy4 form only transient interactions in the absence of a CRISPR RNA (Fig. S1 *A and B*). In contrast, coexpression of the Csy proteins together with a cognate CRISPR RNA results in formation of a stable complex that could be purified by nickel affinity chromatography using any one of the His6-tagged Csy proteins (Fig. 1*B* and Fig. S1). The Csy complex migrates as a single species with a retention volume consistent with a molecular mass of approximately 350 kDa on a size-exclusion column and SDS-polyacrylamide gels of the purified complex indicate that the Csy proteins are not stoichiometric (Fig. 1*B* and Fig. S1).

Stoichiometry and Architecture of the Csy Ribonucleoprotein Complex. The Cascade complex from *E. coli* consists of five Cse proteins that assemble with an unequal stoichiometry that includes six copies of Cse4 (CasC) (13). Although the Csy proteins do not share significant sequence similarity with any of the Cse proteins in Cascade, Coomassie-stained SDS-polyacrylamide gels indicate that the Csy3 protein is overrepresented in the Csy complex (Fig. 1*B* and Fig. S1). We used a combination of mass spectrometric techniques to determine the composition and structural architecture of the Csy complex. The Csy complex was first analyzed by denaturing tandem mass spectrometry, resulting in accurate mass measurements for each Csy subunit that are consistent with their theoretical masses (Table S1). The mass of each subunit was used to interpret the mass of the intact complex as determined by native mass spectrometry. A complex composed of one copy of each Csy protein and a single crRNA would have a mass of approximately 164 kDa. However, analysis of the intact assembly by native mass spectrometry showed three major charge state distributions, corresponding to masses of 352.5 kDa ± 16 Da, 350.4 kDa ± 22 Da, and 265.3 kDa ± 9 Da (Fig. 1*C* and Table S1). The two larger masses are consistent with a Csy complex with a subunit stoichiometry corresponding to one Csy1, one Csy2, six Csy3s, one Csy4, and one crRNA ($Csy1_1 : Csy2_1 : Csy3_6 : Csy4_1 : crRNA_1$). The mass difference (2,052 Da) between these two complexes is due to incomplete removal of the tag from Csy4 (mass of the tag: 2,051 Da). The third charge state centered around an m/z ratio of approximately 7,400 corresponds to a subcomplex that is missing Csy1 and Csy2. This suggests that Csy1 and Csy2 are positioned on the periphery of the complex and also reflects the stable interaction between these two subunits that was observed in coexpression experiments. At the low m/z region of the spectrum, free Csy4 and a Csy4/crRNA complex are observed.

The generation of mature crRNAs from long precursor CRISPR transcripts is thought to be an essential step in all CRISPR-mediated immune systems. CRISPR RNA processing in *P. furiosus* and *E. coli* results in mature crRNAs that have a 2′,3′-cyclic phosphate on the 3′ end (13, 17). The crRNAs from *P. furiosus* undergo additional 3′-end processing by an unknown mechanism that results in two predominant crRNA species (15, 18). To determine the chemical nature of *P. aeruginosa* crRNAs, we used denaturing RNA chromatography and electrospray ionization mass spectrometry (ESI-MS) to analyze crRNAs isolated directly from the Csy complex. To simplify the analysis, crRNAs were isolated following coexpression of the Csy proteins with a synthetic CRISPR containing eight repeats and seven identical spacers (Fig. S2*A*). Chromatography of the crRNAs isolated from this complex reveals a single RNA species with a retention time consistent with a mature crRNA that is 60-nt in length (Fig. S2*B*). ESI-MS analysis of the intact crRNA and oligoribonucleotide fragments generated from RNase T1 and RNase A (Fig. S2 *C and D*) indicate that pre-CRISPR RNAs are cleaved by Csy4 on

the 3′ side of the CRISPR RNA hairpin, generating a 60-nt crRNA with 5′-hydroxyl and 3′-phosphate termini (MW 19,328.5) (Fig. S2E). Further verification of the 3′ phosphate was obtained upon acid treatment of the crRNA, after which no change in mass was observed using ESI-MS (Fig. S2E). We found no evidence for 3′-end trimming or any nucleoside modifications to the mature crRNA.

Target Recognition Through Seed Sequence Interactions. Target recognition is essential for identifying and silencing foreign nucleic acids in all CRISPR systems. To test the target recognition properties of the Csy complex, we performed electrophoretic mobility shift assays (EMSA) using substrates that contain either a sequence complementary to the crRNA spacer or a noncomplementary oligonucleotide of the same length to control for nonspecific interactions (Fig. 2A and Fig. S3). The Csy complex bound to single-stranded DNA (ssDNA) containing a target sequence complementary to the crRNA spacer sequence with subnanomolar affinity (K_d = 0.5 nM), whereas there was no detectable association with ssDNA substrates that do not contain a target sequence (Fig. 2A).

Target recognition is defined by complementary base pairing between the spacer sequence of the crRNA and the target sequence, but a specific role for the Cas proteins in the process of target recognition has not been established. To determine the influence of the Csy proteins on target recognition we removed the proteins by phenol extraction and repeated the binding assay with the crRNA and the target DNA in isolation. Removal of the Csy proteins results in a significant decrease in crRNA binding affinity (Fig. 2B). Low nanomolar binding affinities between the crRNA and the target could be restored in the absence of the Csy complex only after heat annealing the isolated nucleic acids (Fig. S3C).

The PA14 genome contains two CRISPR loci, both of which contain spacer sequences that are identical to sequences in double-stranded DNA (dsDNA) phage that are known to infect *Pseudomonas* (Fig. S4). To test the Csy complex for its ability to recognize a target sequence within a DNA duplex, we generated a dsDNA substrate by annealing the target and nontarget oligos that were used in Fig. 2A. In contrast to the high-affinity interactions measured for ssDNA substrates, we observe significantly weaker binding affinities for dsDNA substrates (Fig. 2C). Similar results are observed when the target strand of the duplex is labeled (Fig. S3 D and E).

The crRNA is an essential structural subunit that is required for stable assembly of the Csy complex. Whereas this high-affinity RNA–protein interaction protects the sugar-phosphate backbone from degradation by cellular nucleases, it may also limit accessibility of the crRNA for base pairing with target sequences. This problem is not unique to CRISPR RNA-guided silencing. In fact, small RNAs bound to the eukaryotic RNAi machinery face a

similar problem. Argonaute proteins from prokaryotes have been important models for understanding target recognition in eukaryotic RNAi. Structural and biochemical studies have shown that Argonaute proteins facilitate target recognition by preordering the first 8-nt of the guide in a helical configuration (19). This region has been termed the seed sequence and is critical for target recognition in eukaryotes (20).

To test the relative importance of sequences within the crRNA spacer region for target recognition, we used isothermal titration calorimetry (ITC) to measure the thermodynamic parameters for guide–target interactions in 8-nt steps across the length of the crRNA (Fig. 3A). Strikingly, high-affinity interactions between the crRNA and targets are localized to the 5′ end of the spacer sequence. An 8-nt ssDNA oligo complementary to the first 8-nt of the spacer (nts 1–8) binds with a K_d of approximately 100 nM, whereas an 8-nt bridging oligo corresponding to nts 5–12 of the spacer sequence binds with a K_d that is approximately fourfold weaker. Oligos complementary to regions outside the seed (nts 1–8) do not bind with measurable affinity.

In addition to measuring binding affinities, ITC also provides an accurate measure of the enthalpy (ΔH) and entropy (ΔS) of any bimolecular interaction. The higher affinity interactions measured for nts 1–8 coincides with a larger enthalpy as compared to the bridging oligo interaction (Fig. 3B). This difference is even more significant when the G + C content of these two oligos is taken into consideration and suggests that nucleotides 3′ of the seed are not available for binding and thus do not contribute to the binding enthalpy. Nucleic acid hybridization is generally an exothermic process driven by enthalpy with an entropic penalty. The entropic (ΔS) penalty for hybridization at the seed (-106 cal.mol^{-1} K^{-1}) is greater than that for the bridging oligo (-81 cal.mol^{-1} K^{-1}), consistent with the idea that nucleotides 3′ of the seed are not accessible for binding.

Structure of the Csy Complex. Affinity tagged purification and mass spectrometry analysis indicate that the Csy proteins assemble into a large stable ribonucleoprotein complex in which Csy3 forms a hexameric structure. The size and stability of this complex suggested that the complex would be amenable to structure determination by transmission electron microscopy (TEM) followed by single particle classification. Classification of TEM projection images of the Csy complex resulted in a limited number of distinct classes, because the particles have a preferred orientation on grids. Two-dimensional class averages reveal a crescent-shaped particle with approximate dimensions of 120 × 150 Å (Fig. 4A). A regularly shaped and evenly spaced feature that runs along the backbone of the crescent-shaped architecture is clearly visible in class averages (Fig. 4A and Fig. S5). This repeating structural unit is consistent with the hexameric stoichiometry of Csy3 and suggests that Csy3 forms the backbone of this complex. In addition to the backbone, we also observe additional density at one
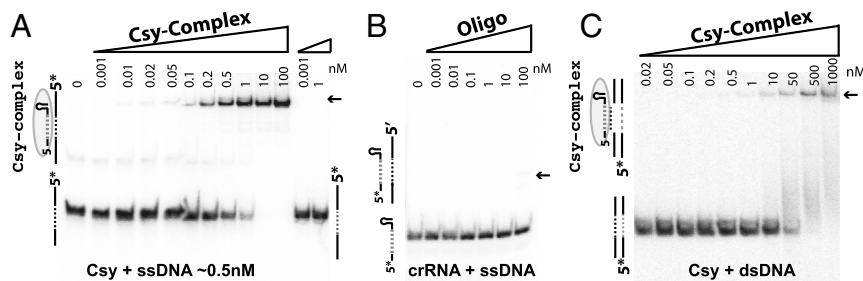

Fig. 2. The Csy complex facilitates sequence-specific hybridization. (*A*) A 32-nt target sequence (dashed line) located in the center of the of a 100-nt ssDNA is recognized and bound by the Csy complex (cyan oval) with a K_d of 0.5 nM (left lanes), whereas the Csy complex does not interact with nontarget DNA (far right two lanes). Nucleic acids are 5′ end labeled with ^{32}P (5*). (*B*) Hybridization between the crRNA and the ssDNA target are nearly undetectable in the absence of the Csy proteins (faint band). (*C*) The Csy complex binds complementary sequences in preformed dsDNA duplexes, however binding affinities are significantly weaker than that observed to ssDNA targets. The arrow indicates the location of the shifted product.
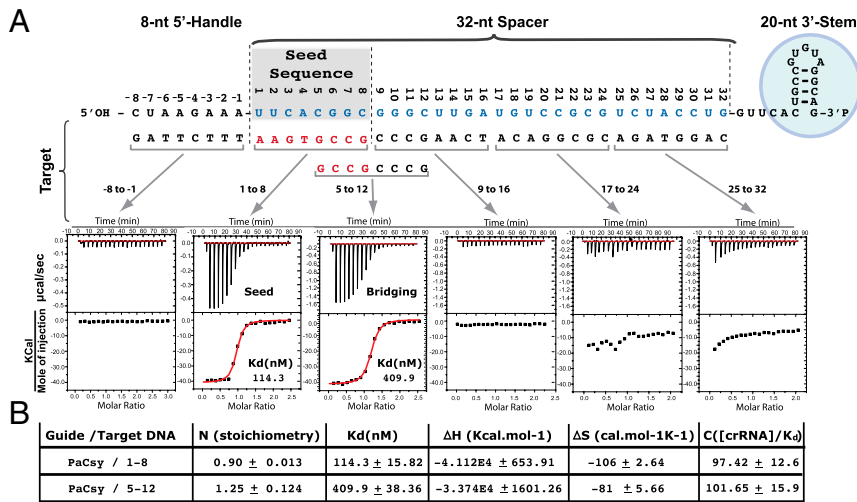
Fig. 3. The Csy complex enhances target recognition through seed-sequence interactions. (*A*) Schematic representation of the mature crRNA showing the 5′ hydroxyl and the 3′ phosphate. Eight-nucleotide oligos complementary to discrete regions of the crRNA were titrated into a 200 ul sample cell containing the Csy complex. The raw ITC data for each titration is shown for each oligo. Titrations performed using oligos outside the seed were originally done using 0.5 mM Csy complex and 0.5 mM oligo. These titrations were repeated at higher concentrations (4–10×) with similar results. Titrations preformed using oligos complementary to the seed were done using 0.01 mM Csy complex and 0.1 mM oligo. (*B*) Average parameters from triplicate experiments.

Guide /Target DNA	N (stoichiometry)	Kd(nM)	ΔH (Kcal.mol-1)	ΔS (cal.mol-1K-1)	C([crRNA]/Kd)
PaCsy / 1–8	0.90 ± 0.013	114.3 ± 15.82	-4.112E4 ± 653.91	-106 ± 2.64	97.42 ± 12.6
PaCsy / 5–12	1.25 ± 0.124	409.9 ± 38.36	-3.374E4 ± 1601.26	-81 ± 5.66	101.65 ± 15.9

end of the Csy3 crescent. The stable association between Csy1 and Csy2 (established by coexpression and mass spectrometry) is consistent with the additional density on one end being composed of the Csy1-2 heterodimer and tandem mass spectrometry experiments suggest that these two subunits are localized at the periphery of this complex.

To complement the two-dimensional structural analysis by TEM, we used small-angle X-ray scattering for three-dimensional shape determination of the Csy complex (Fig. 4*B*). Scattering curves generated from thirty exposures were superimposable and linear in the Guinier region, suggesting that there was no radiation-induced change in the X-ray scattering properties (Fig. S6*A*). These data were radially integrated, averaged, and background subtracted, resulting in a final scatter plot that includes scattering vectors (*q*) ranging from 0.0141 to 0.1369 Å⁻¹ (Fig. S6*B*). The paired-set of distances (P_r) between scattering electrons indicates that the most frequently sampled interatomic distance in the Csy complex is 52.0 Å (Fig. S6*C*). This real-space distribution of scattering pairs indicates a monodispersed sample with real-space dimensions consistent with reciprocal-space Guinier approxima-

tions and agrees with the dimensions observed by TEM. Twenty independent bead models were generated by simulated annealing and these models were aligned, filtered, and averaged based on occupancy. Nineteen of the original bead models are included in the final volumetric reconstruction, suggesting that this is a reliable and stable three-dimensional model of the Csy complex (Fig. S6*D*). The overall crescent-shaped morphology of the complex is similar to that observed by TEM and the additional asymmetric density is clearly evident in the small-angle X-ray scattering (SAXS) model. The arch of the Csy complex spans approximately 200 Å when measured from tip to tail. Although we do not directly observe the crRNA in either structure, the dimensions and the dependence of complex formation on the crRNA are consistent with an RNA that extends along the arch of the structure.

Discussion

Constant selective pressure from invasive genetic elements has driven the diversification of both RNA and protein components of CRISPR-mediated adaptive immune systems. Despite this sequence diversity, we find that the Csy proteins of *P. aeruginosa* assemble into a 350 kDa complex with some structural and functional similarities to the Cascade complex formed by *E. coli* Cas proteins (Fig. S7). The Csy complex includes a crRNA, produced by the CRISPR-specific endoribonuclease (Csy4) that is required for complex assembly (10). Unlike CRISPR RNA processing in *E. coli* (by CasE or Cse3) and *P. furiosus* (by Cas6), which produce crRNAs with a 2′,3′-cyclic phosphate, the crRNAs generated by Csy4 have a terminal 3′ phosphate (Fig. S2 *E and F*). These differences in the chemical nature of mature crRNAs between CRISPR subfamilies highlight possible mechanistic differences in both CRISPR RNA biogenesis and RNA–protein complex assembly.

The mature crRNA isolated from the Csy complex can be divided into three discrete segments (8-nt 5′ handle/32-nt spacer/20-nt 3′ hairpin) that have distinct functions (Fig. 3*A*). The 3′ hairpin provides a recognition site for endonuclease binding and cleavage. The 5′-handle, whose length is conserved between distinct immune system types, may play a role in target discrimination similar to that observed for crRNAs in *S. epidermidis* (21). Spacer sequences are thought to have a universal function in target recognition. Phage challenge experiments in *S. thermophilus* and in natural microbial communities have shown that spacers
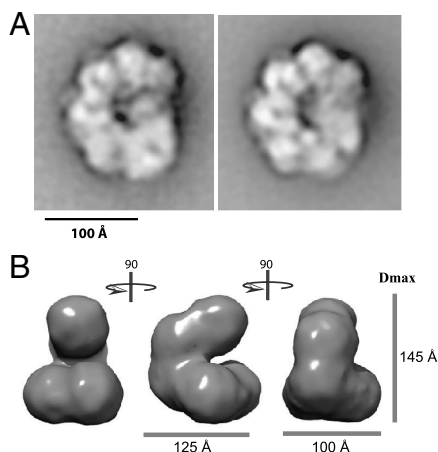

Fig. 4. EM and SAXS reconstructions of the Csy complex reveal a crescent-shaped particle. (*A*) Averaged two-dimensional EM projections of the Csy complex reveal a 120 × 150 Å crescent-shaped particle. (*B*) Ab initio SAXS reconstruction of the Csy complex reveals a crescent-shaped particle reminiscent of the Cascade complex from *E. coli* (13).

are acquired from foreign genetic elements (3–5). The length of the spacer sequence varies between immune system subtypes, but is generally conserved within each subtype. The two CRISPRs in *P. aeruginosa* (PA14) contain a total of 35 spacers that are each 32-nt in length. Nine of these spacers are complementary to sequences found in nine different dsDNA phage that are known to infect *Pseudomonas* (Fig. S4). Spacers complementary to both the coding and template strands of the target phage have been identified, suggesting that this immune system may target DNA.

In this study we used a reconstituted in vitro system to show that the Csy proteins participate in target recognition by facilitating sequence-specific hybridization between the crRNA and a complementary target sequence. Binding assays indicate that the Csy complex can bind target sequences within a dsDNA duplex, but the binding affinities are significantly weaker than binding to ssDNA targets (Fig. 2). This suggests that additional cofactors may be required to facilitate unwinding of the dsDNA for efficient presentation of ssDNA targets, or that the Csy complex is recruited to single-stranded regions of the phage chromosome that are presented over the course of the viral life cycle.

The possible functional similarity between Argonaute and the Csy complex in enhancing guide RNA base-pairing to a target sequence prompted investigation of the thermodynamics of crRNA binding to a complementary target sequence. Thermodynamic parameters of crRNA-target sequence hybridization identified a high-affinity binding site located at the 5′ end of the crRNA spacer sequence. Remarkably, oligonucleotides complementary to regions outside this "seed" region do not bind with measureable affinity. Although the experiments presented here do not precisely define the boundaries of this high-affinity seed-like binding site, our data suggest that nucleotides 3′ of the seed are not available for base pairing prior to target matching at the seed. We do not expect that the seed-binding properties shared by the Csy complex and Argonautes reflects a common evolutionary ancestry, but rather that the length of the seed may be a more universal property that is topologically defined by the number of nucleotides available for interacting with a target substrate in the span of a single helical turn.

A short, high-affinity target recognition site may be advantageous for rapid surveillance and efficient detection of foreign nucleic acid invaders. We propose a seed-nucleation mechanism, where the initial homology search is reduced from 32-nt to an 8-nt seed match. Reduced binding requirements at the seed may significantly enhance the rate of scanning and thus helps explain the efficiency of this surveillance system. Detection of a seed match could induce a conformational change that exposes additional binding sites and may recruit other silencing factors like Cas3. Biochemical and genetic data support base pairing downstream of the seed. A target complementary to the entire 32-nt spacer binds with a K_d of 0.5 nM as compared to a K_d of 100 nM for the 8-nt seed oligo (Figs. 2 and 3). Chemical probing experiments done on the Cascade complex bound to target DNA also indicate that the 5′ end of the spacer binds to targets with high affinity and genetic experiments in *S. thermophilus* have identified single nucleotide mutations in both phage and plasmids that escape immune system suppression (5, 13, 22).

Structural analysis of the Csy complex by native mass spectrometry, EM and SAXS reveal a crescent-shaped particle with an unequal subunit stoichiometry of $Csy1_1 : Csy2_1 : Csy3_6 : Csy4_1 : crRNA_1$ (Figs. 1 and 4). The shape and stoichiometry of this complex is reminiscent of the Cascade complex from *E. coli*, which is a multisubunit ribonucleoprotein complex consisting of Cse-type Cas proteins (CasA-E) (Fig. S7). The Cascade complex also contains proteins present in unequal stoichiometry ($CasA_1 : CasB_2 : CasC_6 : CasD_1 : CasE_1 : crRNA_1$) and forms a seahorse-shaped structure (13). These two CRISPR-associated complexes share a central crescent-shaped architecture with a regularly shaped and evenly spaced feature that runs along the spine of each complex (Fig. 4*A*) (13). This feature is consistent with the hexameric assembly of Csy3 in the Csy complex and CasC (Cse4) in Cascade (Fig. S7).

Although the Csy complex and Cascade share a central crescent shape, the large tail on Cascade, which has been identified as the CasA subunit, is a major distinction between these two architectures (Fig. S7). This structural difference has functional significance in target recognition. CasA is a 50 kDa protein that mediates nonsequence-specific interactions with DNA (13). The Csy complex is approximately 50 kDa lighter than Cascade (Fig. 2), only interacts with DNA in a sequence-specific fashion (Fig. 3), does not contain a CasA homolog and is missing a CasA-like tail (Fig. 4, Fig. S7, and ref 13).

The EM and SAXS models of this complex provide a unique opportunity for structural comparison between distinct CRISPR-associated complexes and explain differences between how these complexes interact with target nucleic acids. The discovery of the Csy complex reveals an unanticipated architectural similarity between distinct CRISPR systems and shared seed-like binding region with Argonaute proteins. In addition, these results provide a structural basis for investigating how the crRNA, and the spacer sequence in particular, is presented for base pairing with cognate nucleic acids. Finally, the finding that the 5′ end of the spacer sequence is uniquely important for target binding uncovers a remarkable parallel with Argonaute-mediated target recognition by micro-RNAs in eukaryotes. Future experiments will be required to determine and compare the underlying molecular mechanisms involved in these distinct pathways for genetic silencing.

Material and Methods

Primers used for cloning are listed in Table S2 and methods for expression, purification, mass spectrometry, binding assays, electron microscopy, and small angle X-ray scattering are available in *SI Text*.

1. Bolotin A, Quinquis B, Sorokin A, Ehrlich SD (2005) Clustered regularly interspaced short palindrome repeats (CRISPRs) have spacers of extrachromosomal origin. *Microbiology* 151:2551–2561.

2. Mojica FJ, Diez-Villasenor C, Garcia-Martinez J, Soria E (2005) Intervening sequences of regularly spaced prokaryotic repeats derive from foreign genetic elements. *J Mol Evol* 60:174–182.

3. Andersson AF, Banfield JF (2008) Virus population dynamics and acquired virus resistance in natural microbial communities. *Science* 320:1047–1050.

4. Barrangou R, et al. (2007) CRISPR provides acquired resistance against viruses in prokaryotes. *Science* 315:1709–1712.

5. Garneau JE, et al. (2010) The CRISPR/Cas bacterial immune system cleaves bacteriophage and plasmid DNA. *Nature* 468:67–71.

6. van der Oost J, Jore MM, Westra ER, Lundgren M, Brouns SJ (2009) CRISPR-based adaptive and heritable immunity in prokaryotes. *Trends Biochem Sci* 34:401–407.

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

8. Kunin V, Sorek R, Hugenholtz P (2007) Evolutionary conservation of sequence and secondary structures in CRISPR repeats. *Genome Biol* 8:R61.

9. Makarova KS, Grishin NV, Shabalina SA, Wolf YI, Koonin EV (2006) A putative RNA-interference-based immune system in prokaryotes: computational analysis of

the predicted enzymatic machinery, functional analogies with eukaryotic RNAi, and hypothetical mechanisms of action. *Biol Direct* 1:7.

10. Haurwitz RE, Jinek M, Wiedenheft B, Zhou K, Doudna JA (2010) Sequence- and structure-specific RNA processing by a CRISPR endonuclease. *Science* 329: 1355–1358.

11. Brouns SJ, et al. (2008) Small CRISPR RNAs guide antiviral defense in prokaryotes. *Science* 321:960–964.

12. Ebihara A, et al. (2006) Crystal structure of hypothetical protein TTHB192 from Thermus thermophilus HB8 reveals a new protein family with an RNA recognition motif-like domain. *Protein Sci* 15:1494–1499.

13. Jore MM, et al. (2010) Structural basis for CRISPR RNA-guided DNA recognition by Cascade. *Nat Struct Mol Biol*, in press.

14. Marraffini LA, Sontheimer EJ (2008) CRISPR interference limits horizontal gene transfer in staphylococci by targeting DNA. *Science* 322:1843–1845.

15. Hale CR, et al. (2009) RNA-guided RNA cleavage by a CRISPR RNA–Cas protein complex. *Cell* 139:945–956.

16. Wiedenheft B, et al. (2009) Structural basis for DNase activity of a conserved protein implicated in CRISPR-mediated antiviral defense. *Structure* 17:904–912.

17. Carte J, Wang R, Li H, Terns RM, Terns MP (2008) Cas6 is an endoribonuclease that generates guide RNAs for invader defense in prokaryotes. *Gene Dev* 22:3489–3496.

18. Hale C, Kleppe K, Terns RM, Terns MP (2008) Prokaryotic silencing (psi)RNAs in Pyrococcus furiosus. *RNA* 14:2572–2579.

19. Parker JS, Parizotto EA, Wang M, Roe SM, Barford D (2009) Enhancement of the seed-target recognition step in RNA silencing by a PIWI/MID domain protein. *Mol Cell* 33:204–214.

20. Bartel DP (2009) MicroRNAs: Target recognition and regulatory functions. *Cell* 136: 215–233.

21. Marraffini LA, Sontheimer EJ (2010) Self versus non-self discrimination during CRISPR RNA-directed immunity. *Nature* 463:568–571.

22. Deveau H, et al. (2008) Phage response to CRISPR-encoded resistance in Streptococcus thermophilus. *J Bacteriol* 190:1390–1400.

BIOCHEMISTRY