nature

https://doi.org/10.1038/s41586-021-03206-x

Accelerated Article Preview

The Card1 nuclease provides defence during type-III CRISPR immunity

Received: 5 July 2020

Accepted: 7 January 2021

Accelerated Article Preview Published online 18 January 2021

Cite this article as: Rostøl, J. T. et al. The Card1 nuclease provides defence during type-III CRISPR immunity. *Nature* https://doi.org/10.1038/s41586-021-03206-x (2021).

Jakob T. Rostøl, Wei Xie, Vitaly Kuryavyi, Pascal Maguin, Kevin Kao, Ruby Froom, Dinshaw J. Patel & Luciano A. Marraffini

This is a PDF file of a peer-reviewed paper that has been accepted for publication. Although unedited, the content has been subjected to preliminary formatting. Nature is providing this early version of the typeset paper as a service to our authors and readers. The text and figures will undergo copyediting and a proof review before the paper is published in its final form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers apply.

ACCELERATED ARTICLE PREVIEW

The Card1 nuclease provides defence during type-III CRISPR immunity

https://doi.org/10.1038/s41586-021-03206-x

Received: 5 July 2020

Accepted: 7 January 2021

Published online: 18 January 2021

Jakob T. Rostøl[1,4], Wei Xie[2,4], Vitaly Kuryavyi[2], Pascal Maguin[1], Kevin Kao[1], Ruby Froom[1], Dinshaw J. Patel[2]✉ & Luciano A. Marraffini[1,3]✉

During the prokaryotic type III CRISPR-Cas immune response, infection triggers the production of cyclic oligoadenylates, which bind and activate CARF domain-containing proteins[1,2]. Many type III loci are associated with proteins in which the CARF domain is fused to an endonuclease-like domain[3,4]; however, with the exception of the well-characterized Csm6/Csx1 RNases[5,6], whether and how these inducible effectors provide defense is not known. Here we investigated one of such type III CRISPR accessory proteins, Card1. Card1 forms a symmetrical dimer with a large central cavity between its CARF and restriction endonuclease (REase) domains that binds cA_4. Ligand binding results in a conformational change where individual monomers rotate relative to each other to form a more compact dimeric scaffold wherein a Mn cation coordinates to the catalytic residues and activates the cleavage of single, but not double, stranded nucleic acids (DNA and RNA). *In vivo*, Card1 activation induces dormancy of the infected hosts to provide immunity against phage infection and plasmids. Our results highlight the diversity of strategies used by CRISPR systems to provide immunity.

Clustered, regularly interspaced short palindromic repeat (CRISPR) loci consist of repetitive DNA sequences intercalated with "spacer" sequences that match the genomes of viruses and plasmids that infect bacteria and archaea[7,8]. These loci are transcribed and processed to generate short RNA guides that contain the spacer sequence, known as the CRISPR RNA (crRNA)[9]. Different effector complexes, encoded by the CRISPR associated (*cas*) genes, use the crRNA guides to find and destroy the nucleic acids of the invader[10]. Depending on their *cas* gene content, CRISPR-Cas systems can be classified into six different types[11]. Of these, type III systems display the most complex targeting mechanism. The crRNA in the type III Cas10 effector complex recognizes complementary invader's transcripts[12,13], resulting in the activation of two catalytic domains within Cas10. The HD domain initiates single-stranded DNA (ssDNA) cleavage near the target transcription site[13,14]; i.e. within the genome of the invader. At the same time the Palm domain converts ATP into 3′-5′ cyclic oligoadenylate (cA) of various sizes, commonly cA_4 and cA_6[1,2]. These molecules function as secondary messengers that bind the CRISPR-Cas Associated Rossmann Fold (CARF) domain of Csm6[15] or Csx1[16], accessory RNases most commonly found in type III-A or III-B loci, respectively. Binding of cA to the CARF domain activates an RNase HEPN domain, through which Csm6 degrades host and invader transcripts non-specifically[5], inducing a growth arrest essential for the type III-A CRISPR-Cas immune response against targets that are transcribed either at low levels[5] or late in the viral infection cycle[17].

Recent bioinformatics studies revealed the existence of a great diversity of genes associated with type III CRISPR-*cas* loci[18,19]. Many of them contain CARF domains fused to different effector domains with predicted catalytic or regulatory functions[4]. One of the most abundant (found in 929/6665 CARF-containing proteins) is the PD-D/ExK domain[20]. Biochemical and structural analysis determined that one such protein, *Thermus thermophilus* Can1, is activated by cA_4 binding to introduce nicks in supercoiled DNA at 60 °C[3]. However, whether and how these type III-associated, CARF:PD-D/ExK proteins, can be activated to provide immunity to prokaryotes remains to be demonstrated. To investigate this, we characterized Tresu_2185, found in the type III-A CRISPR-*cas* locus of the mesophilic gram-negative spirochete *Treponema succinifaciens*[4] (Extended Data Fig. 1a). Tresu_2185 contains 373 amino acids (43.9 kDa), is a member of the Pfam family pfam09002 (domain of unknown function 1887, DUF1887), and is composed of an N-terminal CARF domain and a C-terminal restriction endonuclease-like (PD-D/ExK) domain, typically found in type II restriction endonucleases or Holliday junction resolvases[21], where the two acidic residues coordinate a divalent cation important for catalysis.

Card1-cA_4 degrades ssDNA and ssRNA

To evaluate the biochemical activity of Tresu_2185, we expressed and purified it from *Escherichia coli*, and incubated with different nucleic acids and cAs. We found that the addition of cA_4, but not cA_6, resulted in the degradation of circular ΦX174 (Fig. 1a, and Supplementary Figure 1, which provides the unedited pictures for all gels presented in this study) and M13 ssDNA (Extended Data Fig. 1b), but neither cA_4 nor cA_6 promoted cleavage of supercoiled PUT-7 plasmid dsDNA (Extended Data Fig. 1c), nor supercoiled or linearized ΦX174 dsDNA (Extended Data Figs. 1d, e), at least at 37 °C. ssDNA degradation required the addition of Mn, but not Mg, Ca or Zn, divalent cation (Extended Data Fig. 1f) and resulted in a smear of products, suggesting that ssDNA cleavage is non-specific. These results demonstrate that Tresu_2185 is a cA_4-activated, non-specific ssDNA nuclease. Next generation sequencing of the ΦX174 and M13 ssDNA degradation products showed cleavage

[1]Laboratory of Bacteriology, The Rockefeller University, New York, NY, USA. [2]Structural Biology Program, Memorial Sloan Kettering Cancer Center, New York, NY, USA. [3]Howard Hughes Medical Institute, The Rockefeller University, New York, NY, USA. [4]These authors contributed equally: Jakob T. Rostøl, Wei Xie. ✉e-mail: pateld@mskcc.org; marraffini@rockefeller.edu

ACCELERATED ARTICLE PREVIEW

across both genomes (Extended Data Figs. 1g, h), with an average product size of ~150 nucleotides (Extended Data Figs. 1i, j), preferentially upstream of T(A/G) sites (Figs. 1b and Extended data Fig. 1k).

We also tested activity against RNA substrates. Surprisingly, Tresu_2185 robustly degraded ssRNA oligonucleotides (Fig. 1c), but not dsRNA (Extended Data Fig. 2a). Both Mn and Mg cations supported the ssRNase activity (Extended Data Fig. 2b). We explored the sequence specificity using polyA, polyC and polyU oligonucleotides harboring fluorescent-quencher pairs, and found that all were equally degraded (Extended Data Fig. 2c), similarly to the non-specific RNase I control. Neither Tresu_2185 nor RNase I cleaved polyG, most likely due to the formation of higher order structures by G quartets[22]. Therefore, given the cleavage of both ssRNA and ssDNA, we renamed Tresu_2185 **cA**-activated ss**R**Nase and ss**D**Nase 1, or Card1. Comparison of the efficiencies of both activities showed that ssRNA oligos were cleaved at lower nuclease concentrations than those required to degrade ssDNA oligos of the same sequence (Extended Data Fig. 2d). Finally, we performed calorimetric studies to measure the affinity of Card1 for each cA. We calculated K_D values of 15 and 78 nM for cA_4 and cA_6, respectively (Extended Data Fig. 3a, b), with this modest 5-fold difference in ligand affinity not being sufficient to explain the drastic contrast in nuclease activity.

cA_4 induces structural changes in Card1

To address the effect of cA on the activity of Card1 at the atomic level, we solved the crystal structures of the apo-, cA_4-, and cA_6-bound protein. The 2.3 Å crystal structure of apo-Card1 showed a symmetric dimeric topology (statistics in the Methods section) with a sizeable central hole between the CARF and REase domains (ribbon and electrostatic views in Fig. 1d and Extended Data Fig. 3e, respectively). SEC-MALS studies confirmed that apo-Card1 retains a dimer topology in solution (Extended data Fig. 4a). The crystal structure of the co-crystallized cA_4-Card1 complex at 3.0 Å (dimer of dimers topology, Extended Data Fig. 4b) and cA_4-Card1 D294N mutant complex (dimer topology) at 1.95 Å resolution (statistics in the Methods section), shows the second messenger bound in the central cavity, with a Mn cation bound to the individual catalytic pocket of each monomer (ribbon and electrostatic views in Fig. 1e and Extended Data Fig. 3f, respectively). We readily placed the bound cA_4 (Extended Data Figs. 4d, f) and the bound Mn cation (Extended Data Figs. 4e, g) in the observed electron density for the complexes. The bound cA_4 within the central cavity adopts an approximately square planar alignment and is positioned towards the CARF domain end of the dimeric topology. SEC-MALS measurements on the cA_4-Card1 complex are consistent with dimer formation for the complex in solution (Extended Data Fig. 4c).

Pronounced conformational changes are observed following superposition of the apo- and cA_4-bound dimeric structures of Card1 with an r.m.s.d. = 5.2 Å (over 709 residues, Fig. 1f, Extended Data Fig. 5a). However, superposition of individual monomers of apo- and cA_4-bound Card1 yields a much smaller r.m.s.d. of 1.0 Å over 301 residues (Fig. 2a), suggesting that the transition to the active conformation involves a rotation of the monomers relative to each other after cA_4 binding (see Supplementary Information Video 1). In the context of monomer structural comparison, localized conformational transitions are observed for loop segments 122-129 and 336-341 in the proximity of bound cA_4 (Fig. 2b), and a selective shift observed for catalytic residue E308 associated with Mn binding in the catalytic pocket on complex formation (Fig. 2c). The structural transitions involving breaking and making of key interactions between monomers of the dimeric alignment of apo-Card1 on complex formation with cA_4 are shown in movie format for the cA_4-binding pocket (Supplementary Information Video 2), the loop 336-341 segments (Supplementary Information Video 3) and the top half of the REase domains (Supplementary Information Video 4) with relevant text attached to the movie captions.

A comparative DALI search for folds related to the cA_4-bound Card1 REase domain dimer identified DNA duplex (with looped out bases)-containing type II restriction endonucleases EndoMS (PDB: 5GKE)[23] with a Z score of 9.6 and PspGI (PDB: 3BM3)[24] with a Z score of 6.8 (Extended Data Fig. 5c-e). Yet, we were unable to detect cA_4-bound Card1-mediated cleavage of mismatch-containing DNA duplexes (Extended Data Fig. 5f-h).

Next, we defined the hydrogen bonding (Fig. 2d) and hydrophobic (Fig. 2e and Extended Data Fig. 6a) interactions between bound cA_4 and Card1. Notably, bases A1 and A3 adopt *anti* while A2 and A4 adopt *syn* glycosidic bond alignments for cA_4 in the bound state (Fig. 2d). The side chains of T39 and E41 recognize adenosines A1 and A3 through hydrogen bonding interactions, accounting for cA recognition specificity. Notably, L339 and Y340 sense bound cA_4 and are also adjacent to the REase active site (Fig. 2e). Within the active site, the Mn cation is bound to the E259, D294, E308, C309 and K310 residues (Extended data Fig. 6b). To experimentally validate these interactions, we monitored cA_4-mediated ssDNase cleavage activity for alanine mutants of residues lining the cA_4-binding (majority completely abolished cleavage activity) and catalytic (completely abolished cleavage activity) pockets of Card1 (Fig. 2f). With the exception of the Q13A change, we obtained similar trends when we analyzed the effect of these same mutations on the degradation of ssRNA (Fig. 2g). One notable difference, however, was the rapid RNA cleavage by all the mutants, probably due to the much lower substrate complexity (60-nt RNA oligonucleotide vs. 5,386-nt ΦX174 ssDNA), as well as the more efficient degradation of ssRNA compared to ssDNA (Extended Data Fig. 2d). Isothermal titration calorimetry (ITC)-based binding affinities between cA_4 and Card1 mutants (Extended Data Fig. 3f, g) yielded dissociation constants consistent with the loss of ssDNA degradation activity. Collectively, these data identify the structural changes and critical residues responsible for the activation of Card1 ssDNA cleavage upon binding of cA_4. Of note, the CARF domain of Card1, unlike Csm6[15], is not a ring nuclease, since conversion of cA_4 to ApA>p was not detected in the cA_4-Card1 complex (Extended Data Fig. 6c, d).

We also solved the structure of co-crystallized cA_6-Card1 complex at 2.1 Å resolution (statistics in the Methods section) with ribbon and electrostatic views shown in Extended Data Figs. 6e and 3g, respectively. Bound cA_6 is positioned in the same pocket as bound cA_4, with four of its bases (A1, A2, A4 and A5) adopting an approximate square planar alignment, while two poorly defined bases (A3 and A6) are directed downwards from the plane towards the REase domain (Extended Data Figs. 4h, 4i, 6f). Importantly, superposition of apo- and cA_6-bound dimeric structures exhibited minimal conformational changes (Fig. 2h; Extended Data Fig. 5b; r.m.s.d. = 0.36 Å over 669 residues). The absence of a conformational transition likely reflects a steric clash between the side chain of L339 with the downward pointing A3 and A6 bases of bound cA_6 (Extended Data Fig. 6g).

Card1 activation leads to growth arrest

Next, we investigated the function of Card1 during the type III-A CRISPR-Cas immune response. To do this, we constructed pCRISPR(+Card1) by cloning, into the staphylococcal plasmid pC194, the *Staphylococcus epidermidis* RP62 type III-A locus[8], which Cas10 complex genes display a high similarity (30-58%) to those of the *T. succinifaciens* type III-A locus, carrying the *card1* open reading frame instead of that of the cA-activated accessory protein of staphylococci, Csm6 (Extended Data Fig. 1a). As a control we introduced mutations that inactivate Card1 (E308A, K310A) in pCRISPR(dCard1), that inactivate Csm6 in pCRISPR(-Card1) or that lack a targeting spacer in pCRISPR(Δ*spc*). Each of these plasmids were transformed into *Staphylococcus aureus* RN4220 cells containing pTarget, a second plasmid producing a target transcript, i.e. complementary to the crRNA expressed by the pCRISPR plasmids, that is under the control of an anhydrotetracyclin

ACCELERATED ARTICLE PREVIEW

e(aTc)-inducible promoter[5]. In this experimental system, the addition of the inducer triggers the type III-A response, with the production of cA by the Palm domain of Cas10 and the subsequent activation of Card1. First, we looked at the effect of Card1 activation on the transcriptome. In striking contrast to previous results obtained for the type III-associated RNase Csm6, whose activation leads to significant RNA degradation[5], we were not able to detect significant differences between the presence or absence of Card1 nucleases activities by RNA-seq analysis (Extended Data Fig. 7a-c), nor by Northern blot of select host and target plasmid transcripts (Extended Data Fig. 7d).

Next, we explored whether Card1 attacks the host DNA. We hypothesized that chromosomal DNA lesions caused by the ssDNA activity of Card1 would result in cell toxicity. To test this, we monitored the growth of the different cultures after the addition of aTc, in the absence of antibiotic selection of pTarget (Fig. 3a). +Card1, but not dCard1, Cas10 with an inactive Palm mutant (D586A, D587A; Cas10^{Palm}), nor Δ*spc*, cultures displayed a severe growth defect after the induction of target transcription (Fig. 3a). We observed an increase in OD_{600} at around 10 hours after addition of aTc, which was a result of the propagation of "escaper" cells carrying re-arranged, non-functional pCRISPR or pTarget plasmids within +Card1 cultures (Extended Data Figs. 7e, f). The lack of growth induced by Card1 could be due to either the arrest or the death of individual cells within the culture. To distinguish between these possibilities, we enumerated viable staphylococci after Card1 induction, plating culture aliquots taken at different times after addition of aTc on solid media lacking the inducer (Fig. 3b). +Card1 cultures showed an initial decrease in colony counts of an order of magnitude that slowed down after one hour, demonstrating the presence of a population of viable cells that cannot grow, but do not quickly die, upon activation of the nuclease. Half of the cells in the culture are escapers that can form colonies even in the presence of aTc; the other half are *bona fide* dormant cells (Fig. 3b). Next, we looked at the effects of Card1 activation on pTarget by agarose gel electrophoresis of plasmid DNA (Fig. 3c). +Card1, but not dCard1, cells cleared pTarget, but not pCRISPR, 20 minutes after addition of the inducer. However, in cells lacking the ssDNase activity of Cas10 (H14A, D15A; Cas10^{HD}), but that are still able to produce cA[1], pTarget remained intact (Fig. 3d) and the Card1-mediated growth arrest was maintained (Extended Data Figs. 7g-j), a result that highlights the importance of Cas10 for target DNA destruction. Together with the data showing the absence of detectable transcript degradation, these results suggest that the ssDNase, but not the ssRNase, activity of Card1, can produce a growth arrest of the host cell, presumably by introducing DNA lesions in the host chromosome, and act synergistically with Cas10 to specifically eliminate the target DNA (but not pCRISPR, which has the same replication mechanism as pTarget).

Card1 provides anti-phage immunity

To test the importance of Card1 during immunity against phage infection we programmed the different pCRISPR plasmids with spacers targeting either early- or late-expressed viral genes, infected the cultures with staphylococcal virulent phages at a multiplicity of infection (MOI) of 2-8 and followed their growth. Similarly to previous results for the Csm6 accessory RNase[17], when the early *ORF9* transcript of the lambda-like, dsDNA phage Φ12γ3[25] was targeted, the presence of Card1 activity was not required for immunity (Fig. 4a, Extended Data Fig. 8a). In contrast, when immunity was activated by the late *ORF27* transcript, +Card1 but not dCard1 cultures were able to survive infection (Fig. 4b and Extended Data Fig. 8b). Mutations in the cA_4 binding pocket of Card1 also prevented immunity (Fig. 4c). Similar results were obtained after targeting of the staphylococcal phage ΦNM1γ6[26] (Extended Data Figs. 8a, c-d). We also measured phage propagation in cells programmed to target the ORF27 transcript of Φ12γ3 by counting plaque forming units (pfu) at different times after infection (Fig. 4d). We found that while the phage propagated to high titers in both -Card1 and Δ*spc* cultures, +Card1 cells effectively suppressed Φ12γ3 from the culture. Finally, in cells that express Cas10^{HD}, Card1 was able to provide immunity to growing cells, both when the *ORF9* (Fig. 4a) and *ORF27* (Fig. 4b) transcripts were targeted, as well as reduce the phage titer in the cultures (Extended Data Fig. 8e). At a higher MOI (~25), however, when the great majority of cells are infected, Card1 failed to provide immunity (Extended Data Figs. 8f-g), suggesting that in the absence of Cas10 nuclease activity and low MOIs, defense is achieved through the growth of the cells that are not infected and not arrested by Card1 activity. Altogether, these results demonstrate that Card1 is sufficient to provide anti-phage defense in staphylococci at low MOIs and also required for an efficient type III-A CRISPR-Cas immune response when the target is expressed late during infection.

Discussion

To date, two prokaryotic defense systems have been described that use cyclic oligonucleotide second messengers to activate auxiliary proteins needed for immunity, type III CRISPR-Cas[1,2] and CBASS (cyclic-oligonucleotide-based anti-phage signaling system)[27–29]. While the mechanisms of invader recognition and cyclic oligonucleotide synthesis are very different, both pathways lead to the activation of non-specific nucleases that affect host viability[5,28]. For type III CRISPR, the most common effector proteins are the non-specific ssRNases Csm6 and Csx1[18,19]. In contrast, for CBASS, the cyclic oligonucleotides have so far mainly been shown to activate a majority of non-specific dsDNases such as NucC[27] and Cap4[28]. Based on our findings, we propose that, at least in staphylococci, Card1 protective function is achieved by two separate but overlapping mechanisms. First, Card1 toxicity can provide an abortive infection mechanism of defense in which compromised cells stop growing and prevent the exponential replication of the phage. This activity is similar to both the function of Csm6 during type III-A immunity against plasmid-borne, weakly transcribed targets[5] and viral threats recognized late in the infection cycle[17], as well as to the defense provided by type VI systems[30]. The CBASS effector dsDNases NucC and Cap4 also provide defense causing abortive infection[27,28]. These nucleases, however, cause the irreversible death of the cell, most likely due to the introduction of more severe lesions to the host genome in the form of dsDNA cuts. The great diversity of non-specific nucleases associated with CRISPR and CBASS that mediate abortive infection mechanisms could reflect the adaptation to the many different viral threats and infectious cycles to which prokaryotes are exposed. Second, in contrast to Csm6 which lacks ssDNase activity, Card1 could directly destroy the phage genome. Many phages and plasmids copy their DNA through rolling-circle replication, which involves the formation of ssDNA intermediates[31], likely making them sensitive to Card1 digestion. Moreover, since Cas10 also cuts ssDNA, possibly at the transcription fork of the target[14], it could generate more ssDNA intermediates that are sensitive to Card1 cutting, a hypothesis that explains our result showing synergy between both nucleases to specifically destroy pTarget (Figs. 3e, f).

Card1 (Fig. 1e), ssRNA-specific Csm6[15] (Extended Data Fig. 9a) Csx1[16] (Extended Fig. 9b), and supercoiled dsDNA-specific nickase Can1[3] (Extended data Fig. 9c) all bind cA_4, while dsDNA-specific Cap4[28] binds cA_3 (Extended data Fig. 9d), in each case within a dimeric pocket formed by a pair of CARF domains. There is little sequence conservation amongst the CARF domains of Card1 (classified as CARF4 family[20]), Csm6 (CARF1), Can1 (CARF4) and Cap4 (SAVED3) proteins as shown for example in a comparative analysis of the sequence and topology of the ssDNA/ssRNA-specific Card1 (Extended Data Fig. 9e) and the ssRNA-specific Csm6 (Extended Data Fig. 9f), which both bind cA_4 using distinct topologies and intermolecular interactions (compare Fig. 2d for Card1 with Extended Data Fig. 9a for Csm6), highlighting the unique features of cA_4 recognition by the CARF domains of Card1 and activation of the ssDNase and ssRNase activities of its REase domains.

ACCELERATED ARTICLE PREVIEW

At the atomic level, there are also many differences between Card1 (Fig. 1 and 2) and the other effector nucleases activated by cyclic-oligonucleotides (Extended Data Fig. 9a-d). When compared to Csm6, where cA_4 is bound to the CARF domains in the dimeric Csm6 complex[15] (Extended Data Fig. 9a), cA_4 is located between the CARF and REase domains in the dimeric Card1 complex (Fig. 1h). In addition, unlike Csm6[15], degradation of cA_4 was not observed for Card1 (Extended Data Fig. 6c, d), raising the possibility that its ssDNase activity could be controlled by *trans*-acting ring nucleases[32]. We can also compare Card1 with Can1, a type III-associated, CARF-containing nickase activated by cA_4 whose structure was recently solved, but its role during type III CRISPR-Cas immunity is unknown[3]. Can1 contains a pair of CARF domains interspaced between nuclease and nuclease-like domains and binds cA_4 as a monomer (Extended Data Fig. 9c). Both CARF domains are involved in cA_4 recognition with the nuclease and nuclease-like domains brought together to form a postulated nucleic acid binding site. It is unclear whether these structural differences account for the distinct substrate specificity of each nuclease. Finally, Card1 has many structural differences with CBASS effectors. For example, NucC adopts a trimeric scaffold, with bound cA_3 ligands promoting the formation of a dimer of trimers with endonuclease activity[27]. Cap4[28], which is composed of an endonuclease followed by a SAVED domain, recognizes diverse cA_3 with mixed 2′,5′ and 3′,5 linkages through a pair of tandem CARF domains within the SAVED module (Extended Data Fig. 9d). Further, Cap4 proteins are activated through ligand-dependent oligomerization, with this higher order state proposed to mediate cleavage of target dsDNA.

Our inability to date at growing crystals of cA_4-bound Card1 in the presence of bound ssDNA, led us to undertake docking studies (using HDOCK program[33]) towards positioning ssDNAs (there are severe steric clashes on positioning dsDNA) in the individual catalytic pockets of the REase domains guided by available information from published structures of type II restriction enzyme-DNA complexes[34,35]. Our modeling effort is premised on stacked bases of B-DNA flanking the cleavage site and a two divalent cation-mediated catalytic mechanism of phosphodiester bond cleavage[36]. Energetically favored structural models (computational protocols outlined in the Methods section) of B-form ssDNA (ApC**p**T1pG2pA3, with the cleavable phosphate in bold) positioned with alternate strand directionalities in the REase pocket of one monomer are shown in Extended Data Fig. 10a, b; see Figure caption for additional alignment details). Further, there is ample room to position the modeled individual ssDNAs in the pair of REase catalytic pockets within the structure of cA_4-Card1 dimer (Extended Data Fig. 10c, d). Future structural studies will directly address the remaining mechanistic issues.

We wonder whether the ssRNase activity, which we failed to detect in staphylococci, is more relevant in Card1's natural host, *T. succinifaciens*. Card1-mediated RNA degradation is robust *in vitro*, and there is the possibility that the *S. aureus* cellular environment is inhibitory of this activity. On the other hand, a recent neighborhood analysis[20] showed an enrichment of the *csx1* RNase (cd09741, cd09732 and cd09747) near *card1* (pfam09002). The absence of significant *in vivo* RNase degradation for Card1 would assign these two genes with complementary, rather than redundant, functions, and perhaps influence the observed co-localization. Our study highlights the variety of defense systems and mechanisms that prokaryotic organisms have evolved to counteract the diversity and rapid evolution of their genetic parasites.

Online content

Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at https://doi.org/10.1038/s41586-021-03206-x.

1. Kazlauskiene, M., Kostiuk, G., Venclovas, C., Tamulaitis, G. & Siksnys, V. A cyclic oligonucleotide signaling pathway in type III CRISPR-Cas systems. *Science* **357**, 605-609 (2017).
2. Niewoehner, O. *et al*. Type III CRISPR-Cas systems produce cyclic oligoadenylate second messengers. *Nature* **548**, 543-548 (2017).
3. McMahon, S. A. *et al*. Structure and mechanism of a Type III CRISPR defence DNA nuclease activated by cyclic oligoadenylate. *Nat Commun* **11**, 500 (2020).
4. Makarova, K. S., Anantharaman, V., Grishin, N. V., Koonin, E. V. & Aravind, L. CARF and WYL domains: ligand-binding regulators of prokaryotic defense systems. *Front Genet* **5**, 102 (2014).
5. Rostol, J. T. & Marraffini, L. A. Non-specific degradation of transcripts promotes plasmid clearance during type III-A CRISPR-Cas immunity. *Nat Microbiol* **4**, 656-662 (2019).
6. Sheppard, N. F., Glover, C. V., 3rd, Terns, R. M. & Terns, M. P. The CRISPR-associated Csx1 protein of Pyrococcus furiosus is an adenosine-specific endoribonuclease. *RNA* **22**, 216-224 (2016).
7. Barrangou, R. *et al*. CRISPR provides acquired resistance against viruses in prokaryotes. *Science* **315**, 1709-1712 (2007).
8. Marraffini, L. A. & Sontheimer, E. J. CRISPR interference limits horizontal gene transfer in staphylococci by targeting DNA. *Science* **322**, 1843-1845 (2008).
9. Brouns, S. J. *et al*. Small CRISPR RNAs guide antiviral defense in prokaryotes. *Science* **321**, 960-964 (2008).
10. Jackson, R. N., van Erp, P. B., Sternberg, S. H. & Wiedenheft, B. Conformational regulation of CRISPR-associated nucleases. *Curr. Opin. Microbiol.* **37**, 110-119 (2017).
11. Makarova, K. S. *et al*. Evolutionary classification of CRISPR-Cas systems: a burst of class 2 and derived variants. *Nat. Rev. Microbiol.* **18**, 67-83 (2020).
12. Hale, C. R. *et al*. RNA-guided RNA cleavage by a CRISPR RNA-Cas protein complex. *Cell* **139**, 945-956 (2009).
13. Kazlauskiene, M., Tamulaitis, G., Kostiuk, G., Venclovas, C. & Siksnys, V. Spatiotemporal Control of Type III-A CRISPR-Cas Immunity: Coupling DNA Degradation with the Target RNA Recognition. *Mol. Cell* **62**, 295-306 (2016).
14. Samai, P. *et al*. Co-transcriptional DNA and RNA Cleavage during Type III CRISPR-Cas Immunity. *Cell* **161**, 1164-1174 (2015).
15. Jia, N., Jones, R., Yang, G., Ouerfelli, O. & Patel, D. J. CRISPR-Cas III-A Csm6 CARF Domain Is a Ring Nuclease Triggering Stepwise cA4 Cleavage with ApA>p Formation Terminating RNase Activity. *Mol. Cell* **75**, 944-956 (2019).
16. Molina, R. *et al*. Structure of Csx1-cOA4 complex reveals the basis of RNA decay in Type III-B CRISPR-Cas. *Nat Commun* **10**, 4302 (2019).
17. Jiang, W., Samai, P. & Marraffini, L. A. Degradation of phage transcripts by CRISPR-associated RNases enables type III CRISPR-Cas immunity. *Cell* **164**, 710-721 (2016).
18. Shmakov, S. A., Makarova, K. S., Wolf, Y. I., Severinov, K. V. & Koonin, E. V. Systematic prediction of genes functionally linked to CRISPR-Cas systems by gene neighborhood analysis. *Proc Natl Acad Sci U S A* **115**, E5307-E5316 (2018).
19. Shah, S. A. *et al*. Comprehensive search for accessory proteins encoded with archaeal and bacterial type III CRISPR-cas gene cassettes reveals 39 new cas gene families. *RNA Biol*, 1-13 (2018).
20. Makarova, K. S. *et al*. Evolutionary and functional classification of the CARF domain superfamily, key sensors in prokaryotic antivirus defense. *Nucleic Acids Res.* **48**, 8828-8847 (2020).
21. Kosinski, J., Feder, M. & Bujnicki, J. M. The PD-(D/E)XK superfamily revisited: identification of new members among proteins involved in DNA metabolism and functional predictions for domains of (hitherto) unknown function. *BMC Bioinformatics* **6**, 172 (2005).
22. Balaratnam, S. & Basu, S. Divalent cation-aided identification of physico-chemical properties of metal ions that stabilize RNA G-quadruplexes. *Biopolymers* **103**, 376-386 (2015).
23. Nakae, S. *et al*. Structure of the EndoMS-DNA Complex as Mismatch Restriction Endonuclease. *Structure* **24**, 1960-1971 (2016).
24. Szczepanowski, R. H. *et al*. Central base pair flipping and discrimination by PspGI. *Nucleic Acids Res.* **36**, 6109-6117 (2008).
25. Modell, J. W., Jiang, W. & Marraffini, L. A. CRISPR-Cas systems exploit viral DNA injection to establish and maintain adaptive immunity. *Nature* **544**, 101-104 (2017).
26. Goldberg, G. W., Jiang, W., Bikard, D. & Marraffini, L. A. Conditional tolerance of temperate phages via transcription-dependent CRISPR-Cas targeting. *Nature* **514**, 633-637 (2014).
27. Lau, R. K. *et al*. Structure and Mechanism of a Cyclic Trinucleotide-Activated Bacterial Endonuclease Mediating Bacteriophage Immunity. *Mol. Cell* **77**, 723-733 (2020).
28. Lowey, B. *et al*. CBASS Immunity Uses CARF-Related Effectors to Sense 3′-5′- and 2′-5′-Linked Cyclic Oligonucleotide Signals and Protect Bacteria from Phage Infection. *Cell* **182**, 38-49 (2020).
29. Millman, A., Melamed, S., Amitai, G. & Sorek, R. Diversity and classification of cyclic-oligonucleotide-based anti-phage signalling systems. *Nat Microbiol* **5**, 1608-1615 (2020).
30. Meeske, A. J., Nakandakari-Higa, S. & Marraffini, L. A. Cas13-induced cellular dormancy prevents the rise of CRISPR-resistant bacteriophage. *Nature* **570**, 241-245 (2019).
31. Wawrzyniak, P., Plucienniczak, G. & Bartosik, D. The Different Faces of Rolling-Circle Replication and Its Multifunctional Initiator Proteins. *Front Microbiol* **8**, 2353 (2017).
32. Athukoralage, J. S., Rouillon, C., Graham, S., Gruschow, S. & White, M. F. Ring nucleases deactivate type III CRISPR ribonucleases by degrading cyclic oligoadenylate. *Nature* **562**, 277-280 (2018).
33. Yan, Y., Tao, H., He, J. & Huang, S. Y. The HDOCK server for integrated protein-protein docking. *Nat Protoc* **15**, 1829-1852 (2020).
34. Deibert, M., Grazulis, S., Sasnauskas, G., Siksnys, V. & Huber, R. Structure of the tetrameric restriction endonuclease NgoMIV in complex with cleaved DNA. *Nat. Struct. Biol.* **7**, 792-799 (2000).
35. Pingoud, A., Wilson, G. G. & Wende, W. Type II restriction endonucleases - a historical perspective and more. *Nucleic Acids Res.* **44**, 8011 (2016).
36. Yang, W. Nucleases: diversity of structure, function and mechanism. *Q. Rev. Biophys.* **44**, 1-93 (2011).

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

© The Author(s), under exclusive licence to Springer Nature Limited 2021

ACCEPTED ARTICLE PROOF PREVIEW

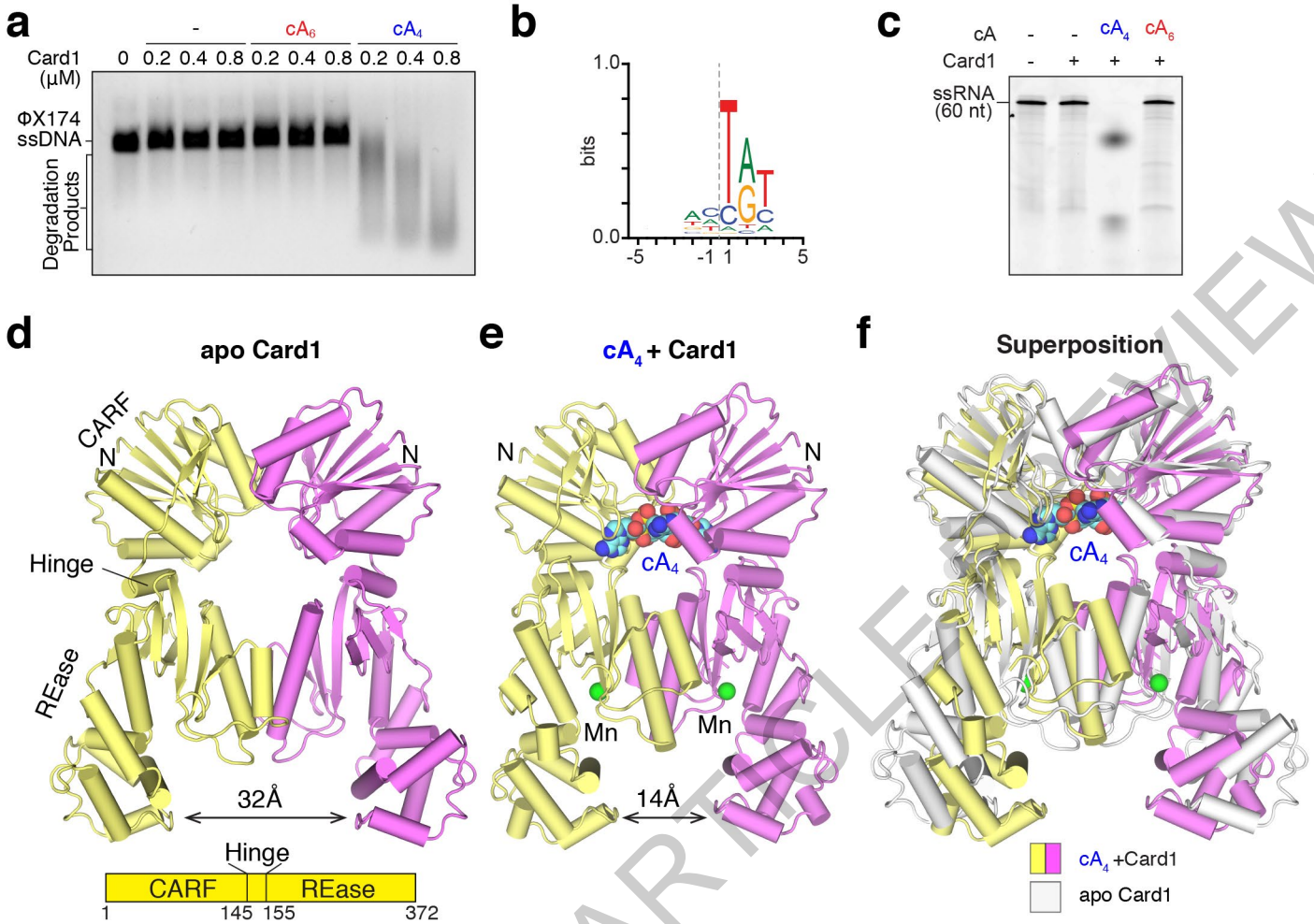


Fig. 1 | Card1 is a cA_4-activated single-stranded DNase and RNase. (a) Card1 cleavage of ΦX174 ssDNA for 30 minutes in the presence of cA_4 and cA_6; performed two times. **(b)** Cleavage preference of Card1, represented as a WebLogo, determined after next generation sequencing of ΦX174 degradation products. Five nucleotide positions upstream (-5 to -1) and downstream (1 to 5) of the detected cleavage sites are shown. **(c)** Card1 cleavage of a 60-nt ssRNA oligonucleotide for 15 minutes in the presence of cA_4 and cA_6; performed three times. **(d)** Crystal structure of dimeric apo-Card1 at 2.3 Å resolution. The monomers are colored yellow and magenta in the symmetrical dimer, with labeling of the CARF, hinge and REase domains, and highlighting the central hole between domains. **(e)** Crystal structure of co-crystallized cA_4-bound to dimeric Card1 at 3.0 Å resolution, with one cA_4 (in space-filling representation) bound per dimer and positioned within the periphery of the central hole. In addition, one Mn (in green) per monomer is bound in each REase catalytic pocket. **(f)** Conformational transitions following superposition of apo- (in silver) and cA_4-bound (in color) states of dimeric Card1.

ACCELERATED ARTICLE PREVIEW

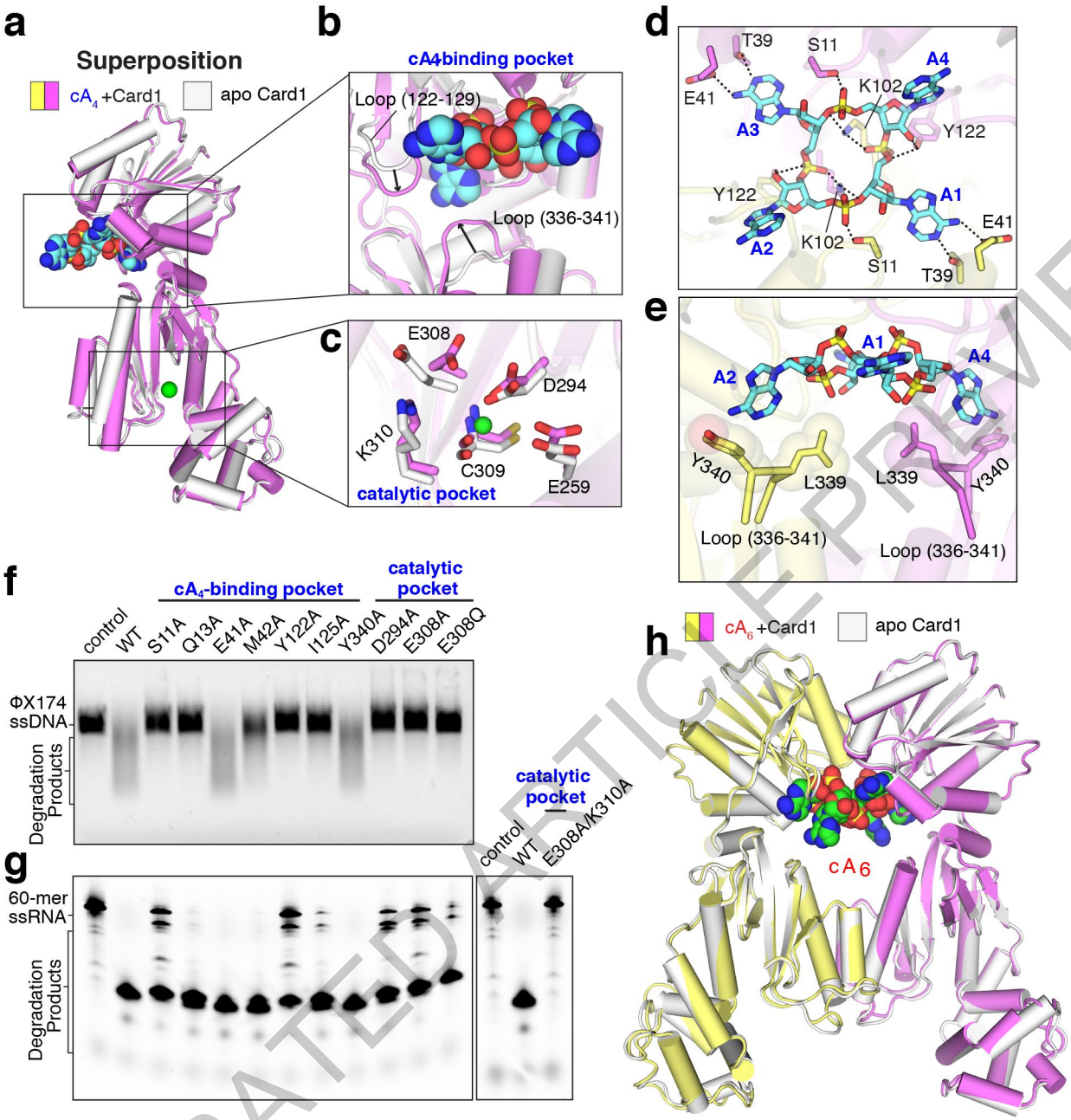


Fig. 2 | Identification of Card1 residues essential for cA4 activation and nucleic acid degradation. (a) Superposition of one monomer module of apo- (in silver) and cA_4-bound (in color) Card1 states. **(b)** Highlight of the changes in loop segments in the vicinity of bound cA_4 (see arrows). **(c)** Highlight of the changes in the Card1 catalytic pocket with an emphasis on the side chain of E308 following Mn coordination on complex formation. **(d, e)** Intermolecular hydrogen bonds **(d)** and hydrophobic interactions **(e)** between bound cA_4 and amino acids of dimeric Card1. **(f)** Impact of mutations of the residues identified in **(d, e)** on *in vitro* ΦX174 ssDNA cleavage activity; performed three times. **(g)** Same as **(f)** but using a fluorescently labeled 60-nt ssRNA substrate; performed two times. **(h)** Lack of conformational transitions following superposition of apo- (in silver) and cA_6-bound (in color) states of dimeric Card1.

ACCELERATED ARTICLE PREVIEW

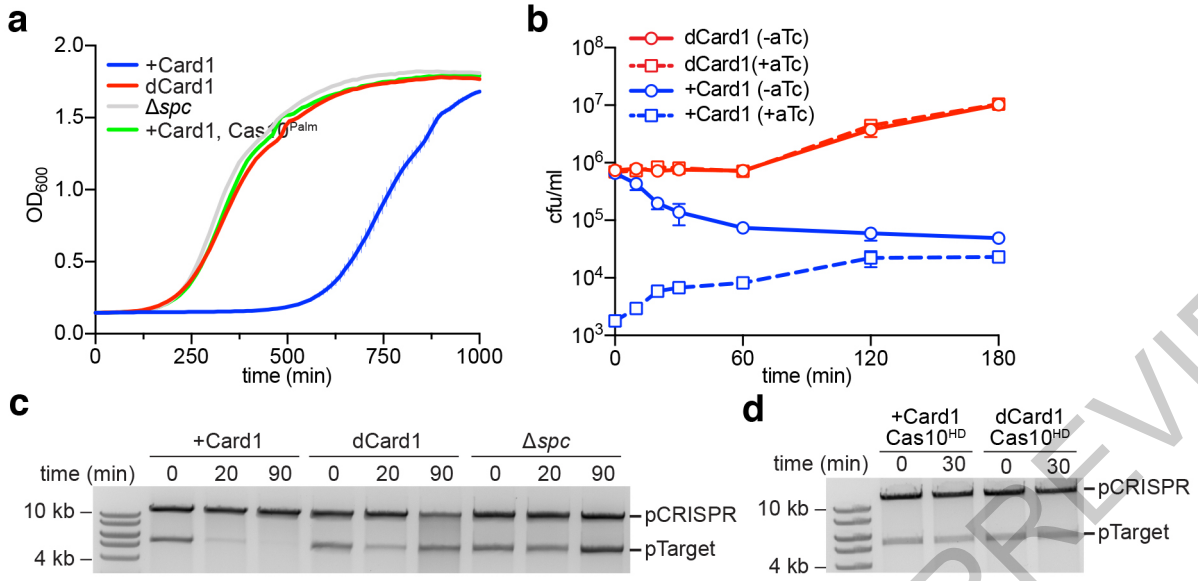


Fig. 3 | Card1 activation leads to a growth arrest of the cell and promotes destruction of target plasmids. (a) Growth of staphylococci carrying pTarget and different pCRISPR variants, measured as OD_{600} after the addition of aTc, in the absence of antibiotic selection for pTarget. Mean of three biological triplicates ± s.e.m. are reported. **(b)** Enumeration of colony-forming units (cfu) from staphylococcal cultures carrying different pCRISPR variants after the addition of aTc. At the indicated times after induction, aliquots were removed and plated on solid media with or without aTc to count the remaining viable cells. Mean of three biological replicates ± s.e.m are reported. **(c-d)** pTarget plasmid curing assay, where plasmid DNA was extracted from cells containing pTarget and different pCRISPR plasmids after addition of aTc. Plasmids were linearized and visualized by gel electrophoresis; performed three times.

ACCELERATED ARTICLE PREVIEW

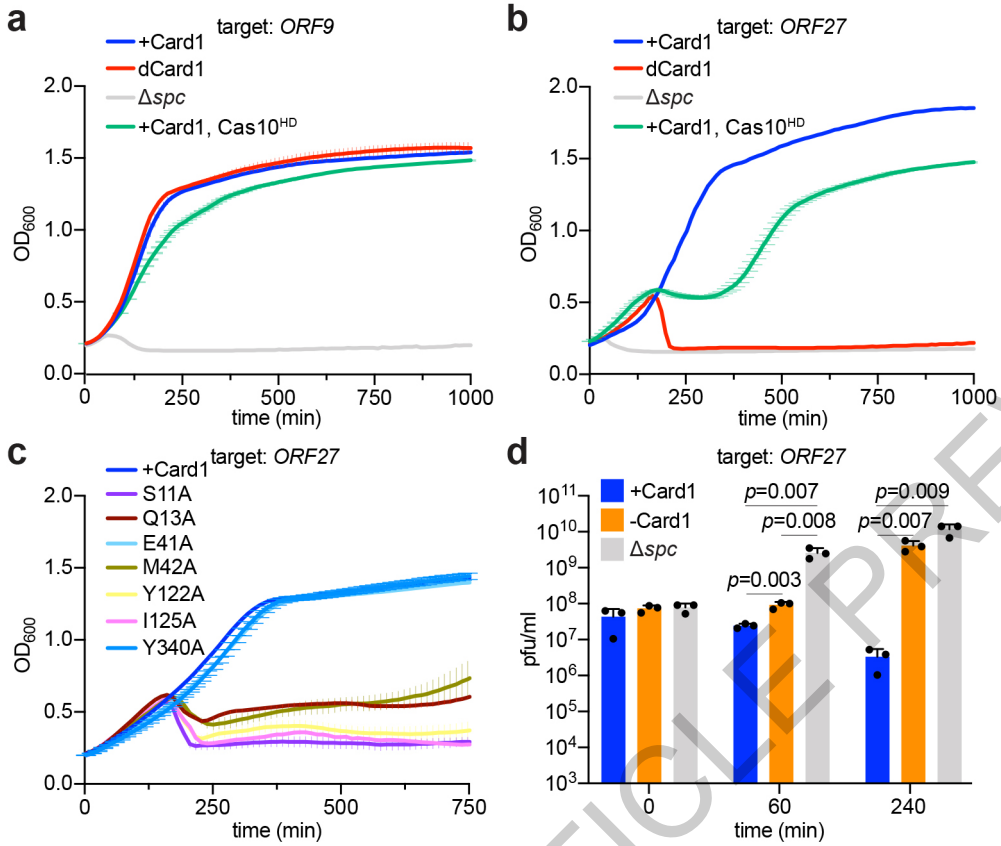


Fig. 4 | Card1 protects staphylococci from phage infection. (a) Growth of staphylococci carrying different pCRISPR variants programmed to target the *ORF9* transcript of Φ12γ3, measured as OD_{600} at different times after infection at a multiplicity of infection (MOI) between 2 and 8. Mean of three biological triplicates ± s.e.m. are reported. **(b)** Same as in **(a)** but targeting the *ORF27* transcript at an MOI ~8. Mean of three biological triplicates ± s.e.m. are reported. **(c)** Same as in **(b)** but following cultures carrying different mutations in the cA_4 binding pocket of Card1, at an MOI ~15. **(d)** Enumeration of plaque-forming units (pfu) within staphylococcal cultures carrying different pCRISPR variants, at the indicated times after infection with Φ12γ3 at an MOI ~10. Mean of three biological replicates ± s.e.m are reported. Significant *p* values ($p<0.05$), obtained with two-sided *t*-test, are shown.

ACCELERATED ARTICLE PREVIEW

Methods

Sequence alignments

Alignments and calculation of sequence identity and similarity (Extended Data Figure 1a) were determined using EMBOSS Needle pairwise sequence alignment[37]

Protein expression and purification

The corresponding sequence of full-length Card1 (1-372) was cloned to plasmid pJTR330 with a C-terminal hexahistidine (His6)-tag. The protein was overexpressed in *E. coli* strain BL21-CodonPlus(DE3)-RIL (Stratagene). Bacteria were grown at 37 °C to OD600 of 0.8 and induced by 0.5 mM isopropyl β-D-1-thiogalactopyranoside (IPTG) at 18 °C overnight. Bacteria cells were lysed by sonication in buffer A (20 mM Tris-HCl, 500 mM NaCl, pH 8.0) supplemented with 20 mM imidazole and 1 mM phenylmethylsulfonyl fluoride (PMSF). Cell lysates were centrifuged, and the supernatant was loaded onto a 5 mL HisTrap FF column (GE Healthcare) with extensive washing by buffer A supplemented with 50 mM imidazole. The target protein was eluted with buffer A supplemented with 300 mM imidazole. The eluate was further purified on 5 mL HiTrap Heparin column (GE Healthcare) by a linear gradient from 100 mM to 1 M NaCl, and then on Superdex 200 16/60 column pre-equilibrated in buffer B (20 mM Tris-HCl, pH 7.5, 150 mM NaCl, 1mM DTT). The high purity eluting fractions were detected by SDS-PAGE and collected. The protein was flash-frozen in liquid nitrogen and stored at −80 °C.

Selenomethionine (Se-Met) proteins were expressed in M9 medium (Sigma) supplied with L-selenomethionine. All Card1 mutants were generated by site-directed mutagenesis and confirmed by sequencing. The purifications of Se-Met and mutant proteins were the same as that for native proteins.

Crystallization and structure determination

Crystallization conditions were determined with crystal screens (Qiagen) by sitting-drop vapor diffusion. As for apo Card1, the protein was prepared at 15 mg/mL in buffer B. As for cA_4-Card1 or cA_6-Card1 complex, cA_4 or cA_6 at the final concentration of 1 mM was added to the 15 mg/mL Card1 in buffer B and incubated on ice for 1 hr before crystallization. Apo Card1 crystals were grown from drops with 1.5 µL protein solution and 1.5 µL reservoir solution (0.2 M K_2HPO_4, and 20% PEG3350 (w/v)). cA_4-Card1 crystals were grown from drops with 1.5 µL protein solution and 1.5 µL reservoir solution (0.1 M citric acid, 10% PEG6000 (w/v), final pH 5.0). cA_6-Card1 crystals were grown from drops with 1.5 µL protein solution and 1.5 µL reservoir solution (0.2 M NaCl, 0.1 M Na/K phosphate, pH 6.2, and 20% PEG1000 (w/v)). cA_4-Card1(D294N) crystals were grown from drops with 1.5 µL protein solution and 1.5 µL reservoir solution (0.1 M HEPES, pH 7, and 30% Jeffamine M-600 (v/v)). Crystals were cryoprotected by mother liquor containing 25% glycerol and flash-frozen in liquid nitrogen.

All the diffraction data sets were collected on the 24-ID beamline at the Advanced Photon Source (APS) at the Argonne National Laboratory, and auto-processed by XDS package[38] in the NE-CAT RAPD online server. The structure of apo- Card1 was solved by single wavelength anomalous diffraction method with the AutoSol and AutoBuild programs and followed by molecular replacement with PHASER program[39] in the PHENIX package[40]. The structure of cA_4-Card1 complex was solved by molecular replacement using one monomer of apo Card1 structure as the search model. The structure of cA_6-Card1 complex was solved by molecular replacement using the dimeric apo-Card1 structure as the search model. The structure of cA_4-Card1(D294N) complex was solved by molecular replacement using the dimeric cA_4-Card1 structure as the search model. Iterative manual model building was performed using the program COOT[41], and refinement with phenix.refine[42] to produce the final models. The statistics of the data collection and refinement are shown in Supplementary Methods Table 1. Figures were generated using PyMOL (http://www.pymol.org). Videos were generated by Chimera[43].

SEC-MALS experiments

For protein molar mass determination, purified Card1 proteins in buffer B with 1 mM cA_4 were analyzed using an ÄKTA-MALS system. A miniDAWN TREOS multi-angle light scattering detector (Wyatt Technology) and an Optilab T-rEX refractometer (Wyatt Technology) were used in-line with Superdex200 10/300 gel filtration column (GE Healthcare) pre-equilibrated in buffer B at a flow rate of 0.2 mL/min. Separation and ultraviolet detection were performed by ÄKTA Pure system (GE Healthcare), light scattering was monitored by miniDAWN TREOS system, and concentration was measured by the Optilab T-rEX differential refractometer. Molar masses of proteins were calculated using the Astra 6.1 program (Wyatt Technology) with a dn/dc value of 0.185 mL/g.

In vitro DNA/RNA cleavage assays

All reactions were performed at 37 °C in a reaction buffer containing 20 mM Tris-HCl, pH 7.5, 150 mM NaCl, 1mM DTT, and 5 mM $MnCl_2$, unless otherwise stated. *In vitro* DNA cleavage assays for circular ssDNA and supercoiled dsDNA were performed by incubating 10 µg ΦX174 ssDNA (New England Biolabs) or PUT-7 plasmid dsDNA (laboratory stock) substrates with different concentrations (200, 400 and 800 nM) of recombinant wild-type Card1 in the presence of 10 µM cA_4 or cA_6 in 20 µl reaction buffer. After 0.5 hr reaction at 37 °C, 6 X Orange DNA Loading Buffer (Thermofisher) was added. cA4 and cA6 was obtained from Biolog Life Science Institute GmbH & Co. KG (Bremen, Germany)

For the reactions in Extended Data figure 1b, d-e, the reactions were done as above, but with 250 nM Card1 and 2.5 uM of cA4, with 2 µg M13 ssDNA (NEB), 500 ng of non-linearized and linearized ΦX174 dsDNA (NEB), or 2 µg ΦX174 ssDNA (NEB), for 15 minutes, before being quenched by 25 mM EDTA. For testing the metal dependence of Card1, the reaction was performed as above, but with 5 mM of $MgCl_2$, $CaCl_2$, or $ZnSO_4$, in the presence of 1 mM EDTA.

For the RNA oligo cleavage assay, 250 nM of a Cy3-labelled RNA oligo was added to the reaction. The reaction products were run on Mini-PROTEAN TBE-Urea precast gel with 15% acrylamide (Bio-Rad), and visualised on an Amersham Typhoon 5 Biomolecular Imager. For degradation of ssRNA or dsRNA ladders, 1 ug of either ssRNA ladder (NEB) or dsRNA ladder (NEB) was digested. After 15 minutes, the reactions were stopped by the addition of 25 mM of EDTA. The reaction products were visualised by agarose gel electrophoresis.

For determining the nucleotide cleavage preference of the RNase activity of Card1, the reaction was performed as above, with 1 µM of each RNA oligo (IDT). The RNA oligos had a 5' end fluorophore (FAM) and a 3' end quencher (Iowa black), generating a fluorescent signal upon cleavage of the linker RNA. Fluorescent measurements were taken in a microplate reader (TECAN Infinite 200 PRO), using values from when the reaction was complete. 0.5 ul of RNase I (Thermo Fisher Scientific), which cuts next to all four RNA nucleotides, was used as a positive control. Guanine polynucleotides could not be tested due to their propensity to form degradation-resistant secondary structures.

For comparing relative catalytic rates of the DNase and RNase activities of Card1, Card1 was incubated with the reaction buffer described before, in the presence of one pair of DNA/RNA fluorescently labelled oligonucleotide of the same sequence (IDT). The increasing Card1 concentrations are 50 nM, 250 nM, 1.25 µM, and 6.25 µM respectively, and the cA4 concentrations are 250 nM, 1.25 µM, 6.25 µM, and 31.25 µM, respectively. The 30 nucleotide oligonucleotides (with the sequence of the top ΦX174 cut site as determined by NGS) had the DNA oligonucleotide labelled with Cy3 and the RNA oligonucleotide labelled with Cy5, and for the 50 nucleotide oligonucleotides (with the sequence of the top M13 cut site as determined by NGS) the DNA species was labelled with Cy5, and the RNA species labelled with Cy3. This allowed

ACCELERATED ARTICLE PREVIEW

the RNA and DNA oligonucleotides to be visualised on the same gel using different fluorescent filters. Each fluorophore was at 250 nM, and both reactions were done for 15 minutes before being quenched by 25 mM EDTA. The products were separated on a 15% acrylamide TBE-urea gels, and visualised on an Amersham Typhoon 5 Biomolecular Imager.

In vitro cA_4 cleavage assays

In *vitro* cA_4 cleavage assays were performed by incubating 2 μM Card1 or Csm6 with 100 μM cA4 in 150 μL reaction buffer containing 20 mM Tris–HCl, pH 7.5, 50 mM KCl, 50 mM NaCl, at 37 °C (for Card1) or 55 °C (for Csm6) in a time course. At desired time points, the protein was separated by using an Amicon Ultra 0.5-ml 10-kDa filter (Millipore) at 10,000 × g for 20 min at 4 °C. The filtered 100 μL reactions were analyzed using a 1 mL monoQ column (GE Healthcare) on an ÄKTA Pure system (GE Healthcare). Elution was performed with a linear gradient from 0.1 M to 1 M NaCl at a flow rate of 0.5 ml/min. The UV absorbance data recorded at 260 nm were employed, and final graphs were represented by GraphPad Prism v8.0.

Isothermal titration calorimetry

Wild-type and mutant Card1 proteins were diluted to the final concentrations of 11 to 41 μM in the ITC buffer composed (20 mM HEPES, pH 7.5, 150 mM NaCl, 0.5% β-mercaptoethanol) and were titrated against 150 to 500 μM cA_4 or cA_6 in the same buffer at 20 °C by using MicroCal PEAQ-ITC range (Malvern Panalytical, HTRSC, Rockefeller University). The titration sequences included a single 0.5 μL injection, followed by 18 injections of 2 μL each, with 2 min interval between injections and a stirring rate of 1000 rpm. Calorimetric data were analyzed using OriginLab software (GE Healthcare) and AFFINmeter web-based software (www.affinimeter.com), and final graphs were represented by GraphPad Prism v8.0.

Modeling of ssDNA into cA_4-Card1 complex using HDOCK

Molecular modeling was performed using the Maestro module from the Schrodinger suite of programs, version 2019-3. Since catalysis most likely involves a two divalent cation (metal) mechanism, we positioned a second metal within the catalytic pocket of the monomeric cA_4-Card 1 complex following comparison with the two metal structure of the catalytic pocket of the type II restriction enzyme NgoMIV-dsDNA complex that contains a similar positioning of catalytic residues. Superposition of the catalytic sites from the structure of the NgoMIV-dsDNA complex (1FIU.pdb)[34] and cA_4-Card1 complex has been performed by direct manipulation in three dimensions of the molecular images of the two structures on the graphical user interface of the Maestro program. The second divalent cation (metal 2) from 1FIU.pdb has been transferred into the file with a superpositioned cA_4-Card1 complex. The side chains of E259, D294 and K310 of cA_4-Card1 complex and E70, D140, and K187 of NgoMIV-dsDNA complex are fully superimposed. The carboxylate group of E308 of cA_4-Card1 coincided with the carboxylate of the co-crystallized acetic acid molecule, which coordinated two catalytic metal ions with its COO- group in the NgoMIV-dsDNA complex. The position of E201 in the NgoMIV-dsDNA complex was very close to that of K238 in the cA_4-Card1 complex. There is no analog for K238 in the NgoMIV-dsDNA complex. The substrate of NgoMIV is a dsDNA and hence the base-pairing provides substantial rigidity for coordinating its substrate. The residues E292 of cA_4-Card1 and E201 of NgoMIV-DNA complex are thus non-equivalent, but equidistant in their positions relative to the second metal ion in the NgoMIV-dsDNA complex. Due to the presence of K238, the COO- group of E292 in cA_4-Card1 occupies another position in the metal coordination sphere.

The monomeric cA_4-Card1 complex with two metals has been processed with protein preparation wizard of the Schrodinger suite and subjected to HDOCK docking protocol on addition of ssDNA. The protocol did not include any definition of the sites from the monomeric cA_4-Card1 receptor to interact with the ssDNA ligand. The ligand for docking computations had the sequence ApCpT1pG2pT3 in a single-stranded form with idealized B-DNA geometry and was positioned in one (labeled directionality 1) of two alternating strand directionalities. The only close contact observed in the complex of ssDNA bound to monomeric cA_4-Card1 obtained by docking was that of Ile190 and the base C upstream of T1. It has been resolved by selecting an alternative rotamer for Ile190. The conformations of N193 and R251 have been changed by the selection of the allowed conformers from the library of Schrodinger program. As a result of such a concerted conformer search, the conformation of R251 interacting with dG2 was identified. The orientations of NH_2 and C=O groups of N194 and N199 interacting with T1 and dT3 respectively were selected by the hydrogen-bond optimizer module of the protein preparation wizard. The docked structure of ApCpT1pG2pT3 with bound directionality 1 positioned onto the monomer of cA_4-Card1 is shown in Extended Data Fig. 10a and its strand directionality is similar to that observed in the EcoR5-dsDNA complex.

The dimeric structure of cA_4-Card1 complex was obtained by copying the ssDNA (directionality 1)-bound monomeric form into the coordinates of the free monomer's position via structural alignment of two Card1 monomers by all their C-alpha atoms.

We also modeled ApCpT1pG2pT3 bound with the opposite directionality (labeled directionality 2) onto the monomer of cA_4-Card1 with the results shown in Extended Data Fig. 10b. Such a strand directionality has been observed in the crystal structure of most type II restriction enzyme-dsDNA complexes[35]. The close contact with the side chain of Ile190 has been resolved by selecting an alternative conformer. The close contacts between the C3´ and O3´ atoms of T3 and Cβ and Cγ carbons of K187 were resolved by energy minimization.

Bacterial growth

S. aureus strain RN4220[44] was grown in brain heart infusion (BHI) medium at 37 °C, supplemented with chloramphenicol at 10 μg/ml for maintaining pCRISPRs, and erythromycin at 10 μg/ml for maintaining pTarget. 5 μM $CaCl_2$ was supplemented in phage experiments.

Molecular cloning

The plasmids used in this study are listed in Supplementary Methods Table 2. The oligonucleotides used for this cloning are listed in Supplementary Methods Table 3. The cloning strategies for generating these plasmids are listed in Supplementary Methods Table 4. For obtaining the coding sequence of Card1, the amino acid sequence of Tresu_2185 (NCBI Reference Sequence WP_013702306.1) from *Treponema succinifaciens* DSM 2489 (Genbank accession number CP002631) was codon optimised for expression in *S. aureus* and synthesised by Genewiz (NJ, USA).

Growth curves

For *in vivo* Card1 toxicity induction, triplicate RN4220 overnight cultures harbouring pTarget and a pCRISPR are diluted 1:100, outgrown for about an hour, and normalised for OD. Cells are then seeded in a 96 well plate. To induce targeting, 6.25-12.5 ng/ml of anhydrotetracycline (aTc) is added to the appropriate wells. Absorbance at 600 nm is then measured every 10 minutes by a microplate reader (TECAN Infinite 200 PRO). To analyse targeting escapers, cells from the end of the experiment (either cells from wells without aTc, i.e. naïve cells, or cells from wells that recovered later in the time course due to Card1 toxicity) are re-streaked on BHI agar plates, and individual colonies were launched in liquid culture, diluted the next day, and used for a new time course experiment. From these overnight cultures, plasmid DNA was isolated (QIAgen Spin Miniprep Kit), digested by BamHI-HF (single-cutter for both pTarget and pCRISPR) (New England Biolabs), and visualised by gel electrophoresis. The deletion of important features in pTarget (making it unable to be targeted by pCRISPR) or pCRISPR was confirmed by Sanger sequencing.

For *in vivo* anti-phage immunity, cells harbouring various pCRISPRs were launched in triplicate overnight, diluted 1:100, outgrown for about

ACCEPTED ARTICLE PREVIEW

an hour, and normalised for OD. Cells were seeded into a 96 well plate. Phage Φ12γ3[25] ΦNM1γ6[26] or was added at the specified multiplicity of infection, and OD measurements were taken every 10 minutes.

Card1 toxicity assay

To measure the effect of Card1 activity on *S. aureus* viability over time, colonies of *S. aureus* harbouring pTarget and the specified pCRISPR were launched in liquid culture overnight in triplicate. The next day, cells were diluted 1:100 and grown out for about an hour, and normalised for OD. One aliquot was taken from each culture, and then aTc was added to induce CRISPR targeting and Card1 activity (to a concentration of 3 ng/ml in Fig. 3b or 125 ng/ml in Extended Data Fig. 7h). At each timepoint, cell aliquots were removed, centrifuged, resuspended in media lacking aTc, and serial dilutions were plated on solid BHI agar plates with or without aTc. All viable cells should grow on the solid agar plates, but only targeting escapers (cells that recover due to mutations in pTarget or pCRISPR) should form CFUs on plates with aTc.

Liquid anti-phage infection

To obtain CFU and PFU counts over time from cultures infected with phage, RN4220 cultures harbouring various pCRISPRs were launched overnight, diluted 1:100, and outgrown for about one hour. Cells were then infected with phage Φ12γ3 at an MOI of 10, and an aliquot was taken shortly after to obtain PFUs at time 0. The cultures were then incubated further, with aliquots taken at 1 and 4 hours.

Plasmid curing

To assess Card1's ability to promote plasmid curing under low transcription conditions, a plasmid curing assay was performed, similarly to previously described[5]. Briefly, overnight cultures of *S. aureus* cells harbouring pTarget and a pCRISPR containing either Card1, dCard1, or no spacer (*Δspc*) were diluted to exactly OD 0.15 in tryptic soy broth with 10 ug/ml chloramphenicol. After removing a cell aliquot for the 0 timepoint, aTc was added to a concentration of 9.3 ng/ml (Fig. 3c) or 125 ng/ml (Fig. 3d) and the cells were incubated at 37 °C, with further aliquots taken at the indicated times. The cells were then lysed, and the plasmid DNA isolated using a QIAprep Miniprep kit (Qiagen) according to the manufacturer's protocol (Qiagen). 400 ng of plasmid was then linearised using BamHI-HF (NEB), which cuts both pTarget and pCRISPR once, followed by visualisation by gel electrophoresis.

Next-generation sequencing of ssDNA degradation products

To assess the ssDNA cleavage patterns of Card1, 2 ug of ΦX174 virion DNA (NEB) or M13 ssDNA (M13mp18) (NEB) was first digested by 250 nM Card1 with 2.5 uM of cA4. At the specified time points, the reaction was quenched by adding 25 mM of EDTA. Half the reaction was visualised by agarose gel electrophoresis. The remaining digestion products from the 2 hour timepoint were purified by phenol chloroform extraction.

Without further fragmentation, the purified digested DNA was subjected to the Accel-NGS 1S Plus DNA Library Kit (Swift Biosciences), proceeding according to the manufacturer's protocol, using a 1.5 x ratio of magnetic beads (AMPure XP beads by Beckman Coulter) to also include small DNA fragments. One of the library preparation steps involves the addition of on average 8 nucleotides to the 3′ end of the DNA. The 5′ end of the input DNA molecules remains unchanged. Paired-end sequencing was performed on an Illumina MiSeq. The 5′ end of each read R1 represents the start of a DNA molecule, and thus a Card1 cut site. Using a custom python script, the location of 7,020,067 ΦX174 reads (mapping to Genbank reference NC_001422) and 7,670,616 M13 reads (mapping to Genbank reference X02513) was determined. To account for reads mapping at the circular junctions, 65 nucleotides of the first 5′ end of the maps were copied and added at the 3′ end of the maps. The DNA sequence 20 nucleotides upstream and downstream of the cut sites was extracted using a custom Python script, and the Card1 cleavage motifs for ΦX174 and the M13 were determined separately using Weblogo3[45], with basal nucleotide compositions determined by the base compositions in each map (ΦX174 with A:24.0, C:21.4, G:23.3, T:31.3, and M13 with A:24.4, C:21.1, G:21.1, T:33.4). For the fragment size analysis, 8 nucleotides were removed from all the reads from the 3′ end pair mate by the "Trim Ends" option in the Geneious Bioinformatics Software platform[46]. Using the STAR aligner (version 2.7.3)[47], 7,505,136 reads were successfully mapped to ΦX174, and 8,179,356 reads were successfully mapped to M13, using default arguments with the addition of --alignIntronMax 1 --alignMatesGapMax 6000 --peOverlapNbasesMin 5 --alignEndsProtrude 10 ConcordantPair.

Northern blotting

Northern blotting was done as previously described[5]. Briefly, overnight RN4220 cultures harbouring pTarget and pCRISPR (carrying Cas10^{HD} and either Card1 or dCard1) were diluted to exactly OD 0.15 before targeting was induced by the addition of 125 ng/ml of aTc. After three minutes, cells were spun down, lysed in PBS with 1 mg/ml of lysostaphin and 2mg/ml of lysozyme for 10 minutes. TRIzol (Thermo Fisher Scientific) was then added, and RNA was purified according to the manufacturer's protocol. The RNA was separated by gel electrophoresis on a 6% acrylamide gel, and blotted onto a nylon filters (Invitrogen BrightStar Plus) with a semi-dry blotting unit (Fisher Scientific). The filters were incubated with ^{32}P-labelled oligonucleotides (IDT) overnight at 42 °C, and visualised on an Amersham Typhoon 5 Biomolecular Imager.

RNA-seq of +Card1 and dCard1 cells

To investigate the RNase activity of Card1 *in vivo*, RNA-seq using spike-in normalisation was performed similarly to described before[5]. In duplicate, cells containing pTarget and pCRISPR with Cas10^{HD} and either +Card1 or dCard1 were grown overnight and normalised to an OD of exactly 0.15. Cells were harvested (quenching growth with the addition of ice-cold media) for the 0 minute timepoint, and aTc was added to 125 ng/ml to initiate pTarget transcription and Card1 activation. At 3 minutes, the cell growth was quenched. The cells were lysed in PBS with 1 mg/ml of lysostaphin and 2 mg/ml of lysozyme at 37 °C for 5 minutes before adding TRI reagent. At this step, 3 µg of Listeria seeligeri was added to each sample to serve as an internal normalisation control. RNA was then purified with the Quick-RNA Miniprep Plus kit according to the manufacturer's protocol (Zymo Research). The RNA was DNase treated (Invitrogen TURBO DNA-free kit) and rRNA depleted (Ribo-Zero Plus rRNA Depletion Kit (Illumina)), before undergoing library preparation with a TruSeq Stranded mRNA kit (Illumina). The samples were sequenced using an Illumina NextSeq at the Rockefeller University Genomics Core. Mapping was performed by STAR v2.7.5 allowing one mismatch. To determine the scaling factor to apply to the reads (to normalise the read depth for spike reads), the reads were first mapped to the genome of *S. aureus* NCTC 8325 (GenBank reference CP000253) to remove non-spike reads, and the remaining reads were mapped to the *L. seeligeri* genome (NC_013891.1). This process was then reversed, first mapping all the original reads to *L. seeligeri* (to remove spike reads), and the remaining reads were mapped to *S. aureus*. Reads were assigned to each gene with featureCounts[48] with largestOVERLAP set to TRUE. The number of reads per gene for each sample was then scaled according to the previously calculated scaling factor, allowing direct comparison of read depth per gene between all samples.

Reporting summary

Further information on research design is available in the Nature Research Reporting Summary linked to this paper.

Data availability

The atomic coordinates have been deposited in the Protein Data Bank with the codes 6WXW (apo Card1), 6WXX (cA_4-Card1 complex), 6WXY (cA_6-Card1 complex), and 6XL1 (cA_4-Card1(D294N) complex). The raw

ACCEPTED ARTICLE PREVIEW

data for the DNAseq and RNAseq experiments performed in this study is found at the Sequence Read Archive (NIH) through accession code PRJNA672128.

Code availability

Custom Python code is available upon request.

37. Madeira, F. *et al*. The EMBL-EBI search and sequence analysis tools APIs in 2019. *Nucleic Acids Res*. **47**, W636-W641, https://doi.org/10.1093/nar/gkz268 (2019).
38. Kabsch, W. Xds. *Acta Crystallogr D Biol Crystallogr* **66**, 125-132, https://doi.org/10.1107/S0907444909047337 (2010).
39. McCoy, A. J. *et al*. Phaser crystallographic software. *J Appl Crystallogr* **40**, 658-674, https://doi.org/10.1107/S0021889807021206 (2007).
40. Adams, P. D. *et al*. PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr D Biol Crystallogr* **66**, 213-221, https://doi.org/10.1107/S0907444909052925 (2010).
41. Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. Features and development of Coot. *Acta Crystallogr D Biol Crystallogr* **66**, 486-501, https://doi.org/10.1107/S0907444910007493 (2010).
42. Afonine, P. V. *et al*. Towards automated crystallographic structure refinement with phenix.refine. *Acta Crystallogr D Biol Crystallogr* **68**, 352-367, https://doi.org/10.1107/S0907444912001308 (2012).
43. Goddard, T. D. *et al*. UCSF ChimeraX: Meeting modern challenges in visualization and analysis. *Protein Sci*. **27**, 14-25, https://doi.org/10.1002/pro.3235 (2018).
44. Kreiswirth, B. N. *et al*. The toxic shock syndrome exotoxin structural gene is not detectably transmitted by a prophage. *Nature* **305**, 709-712 (1983).
45. Crooks, G. E., Hon, G., Chandonia, J. M. & Brenner, S. E. WebLogo: a sequence logo generator. *Genome Res*. **14**, 1188-1190, https://doi.org/10.1101/gr.849004 (2004).
46. Kearse, M. *et al*. Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics* **28**, 1647-1649, https://doi.org/10.1093/bioinformatics/bts199 (2012).
47. Dobin, A. *et al*. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15-21, https://doi.org/10.1093/bioinformatics/bts635 (2013).
48. Liao, Y., Smyth, G. K. & Shi, W. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* **30**, 923-930, https://doi.org/10.1093/bioinformatics/btt656 (2014).

Acknowledgements We thank the HTS core facility of the Rockefeller University for assistance with ITC experiments, and the Rockefeller University Genomics Resource Center for performing the Card1 *in vivo* RNase NextSeq NGS experiment. JTR was supported by a Boehringer Ingelheim Fonds PhD fellowship. LAM is supported by a Burroughs Wellcome Fund PATH Award and an NIH Director's Pioneer Award (DP1GM128184). LAM is an investigator of the Howard Hughes Medical Institute. DJP is supported by funds from the Geoffrey Beene Cancer Research Center, by NIH GM129430 and by Memorial Sloan-Kettering Cancer Center Core Grant (P30CA008748). This work is based upon research conducted at the Northeastern Collaborative Access Team beamlines, which are funded by the US National Institutes of Health (NIGMS P30 GM124165). The Pilatus 6M detector on 24-ID-C beamline is funded by a NIH-ORIP HEI grant (S10 RR029205). This research used resources of the Advanced Photon Source, a US Department of Energy (DOE) Office of Science User Facility operated for the DOE Office of Science by Argonne National Laboratory under Contract No. DE-AC02-06CH11357, and those of the Minnesota Supercomputing Institute.

Author contributions JTR and LAM conceived the study. JTR performed in vitro Card1 cleavage assays and NGS of its products, as well as all the in vivo experiments with the help of KK and RF. WX performed Card1 biochemistry assays and all structural experiments. VK performed structural simulations and modeling. PM provided reagents and wrote a custom Python code. JTR, WX, DP and LAM wrote the manuscript. All authors read and approved the manuscript.

Competing interests LAM is a cofounder and Scientific Advisory Board member of Intellia Therapeutics and a cofounder of Eligo Biosciences. A patent related to this work has been filed.

Additional information

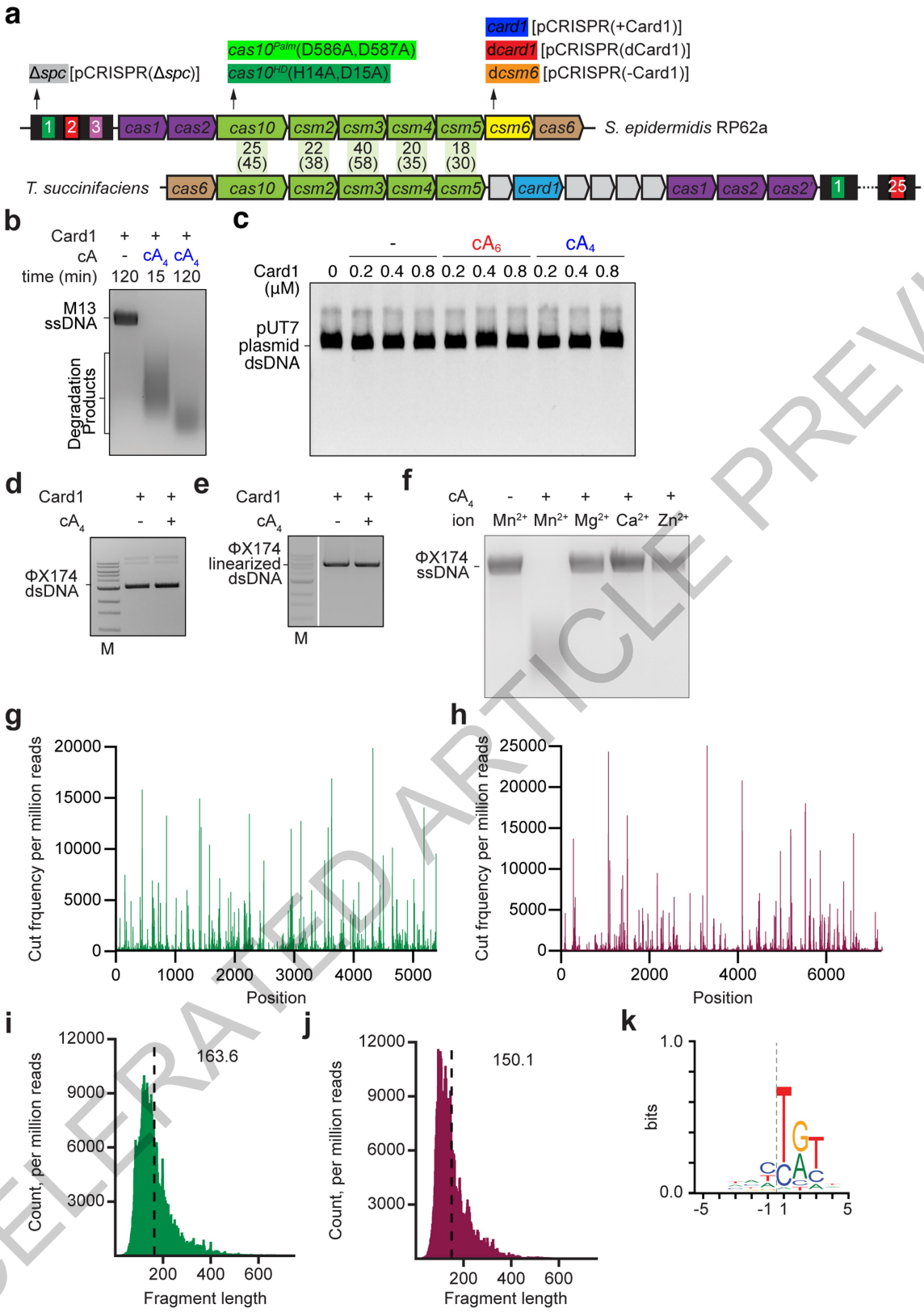
Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41586-021-03206-x.

Correspondence and requests for materials should be addressed to D.J.P. or L.A.M.

Peer review information *Nature* thanks Philip Kranzusch, John van der Oost and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

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

ACCELERATED ARTICLE PREVIEW

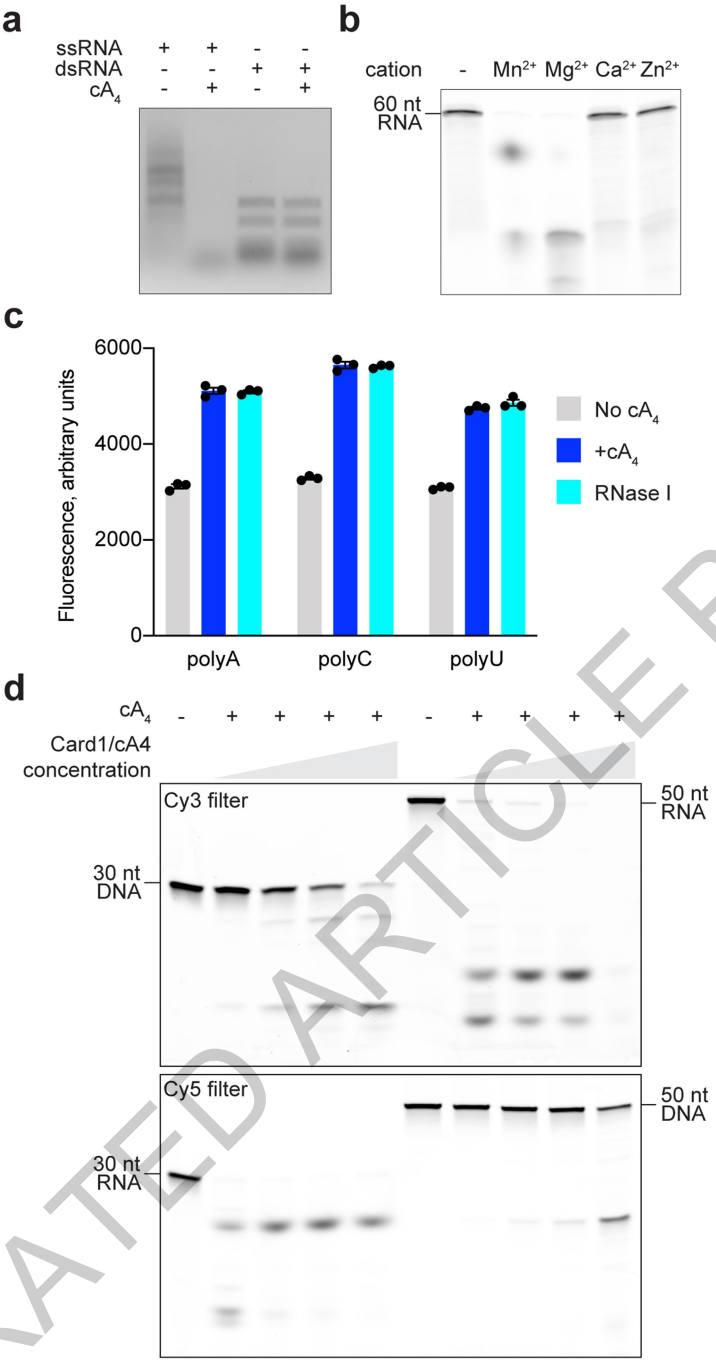


Extended Data Fig. 1 | See next page for caption.

ACCELERATED ARTICLE PREVIEW

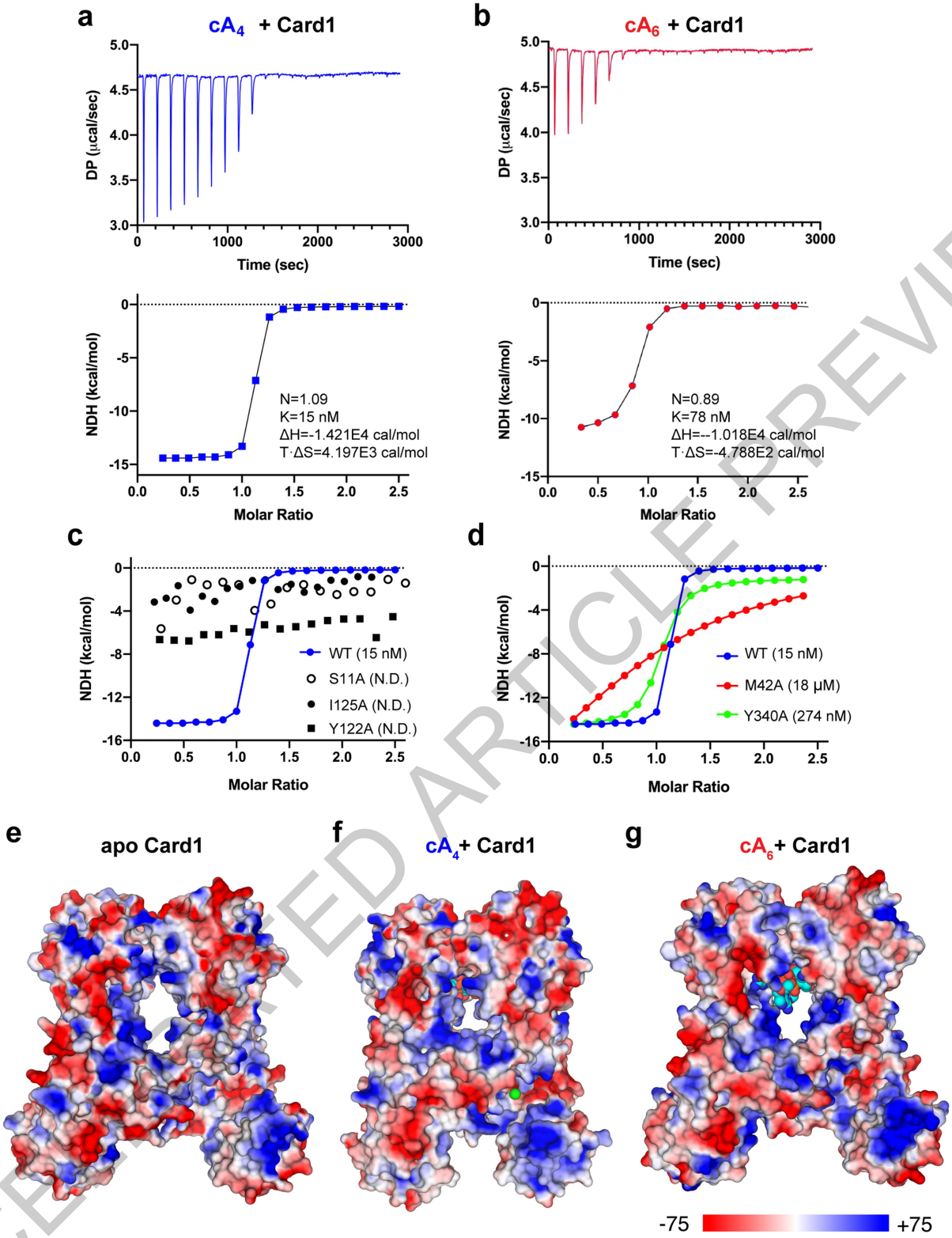
Extended Data Fig. 1 | cA_n-mediated cleavage of ssDNA by Card1 at 37 °C. **(a)** Schematic of the *S. epidermidis* type III-A locus showing the replacement of *csm6* by *card1* and the different mutations investigated in this study. A comparison with the T. succinifaciens type III-A locus is provided. Numbers indicate the % identity between the genes that encode the Cas10 complex (with % similarity in parenthesis. **(b-e)** Card1 digestion of **(b)** M13 ssDNA, performed two times, **(c)** pUT7 dsDNA plasmid (30 minutes), performed two times, **(d)** ΦX174 supercoiled dsDNA (30 minutes), performed two times, or **(e)** ΦX174 linearized dsDNA (30 minutes), performed two times, in the presence or absence of cA_4, visualized by agarose gel electrophoresis. **(f)** Card1 digestion of ΦX174 ssDNA (60 minutes) in the presence of cA_4 and different divalent cations, visualized by agarose gel electrophoresis; performed two times. **(g)** Overview of Card1 cleavage sites across the ΦX174 genome based on the 5′ end mapping of DNA degradation products obtained after 2 hours of digestion, per 1 million reads. There appears to be preferential cleavage sites that may reflect lack of Card1 access to secondary structures formed within the ssDNA molecule. 26.7% of cuts occur at the 25 most frequent positions. **(h)** Same as **(g)** but after the analysis of M13 ssDNA degradation products obtained after 2 hours of digestion. 31.1% of cuts occur at the 25 most frequent positions. **(i)** Fragment size distribution of the ΦX174 degradation products after 2 hours of Card1 digestion. The average fragment length (163.6 nucleotides) is marked by the dotted line. **(j)** Same as **(i)** but analyzing M13 degradation products after 2 hours of digestion. The average fragment length (150.1 nucleotides) is marked by the dotted line. **(k)** Cleavage preference of Card1, represented as a WebLogo, determined after NGS of M13 degradation products. Five nucleotide positions upstream (-5 to -1) and downstream (1 to 5) of the detected cleavage sites are shown.

ACCELERATED ARTICLE PREVIEW



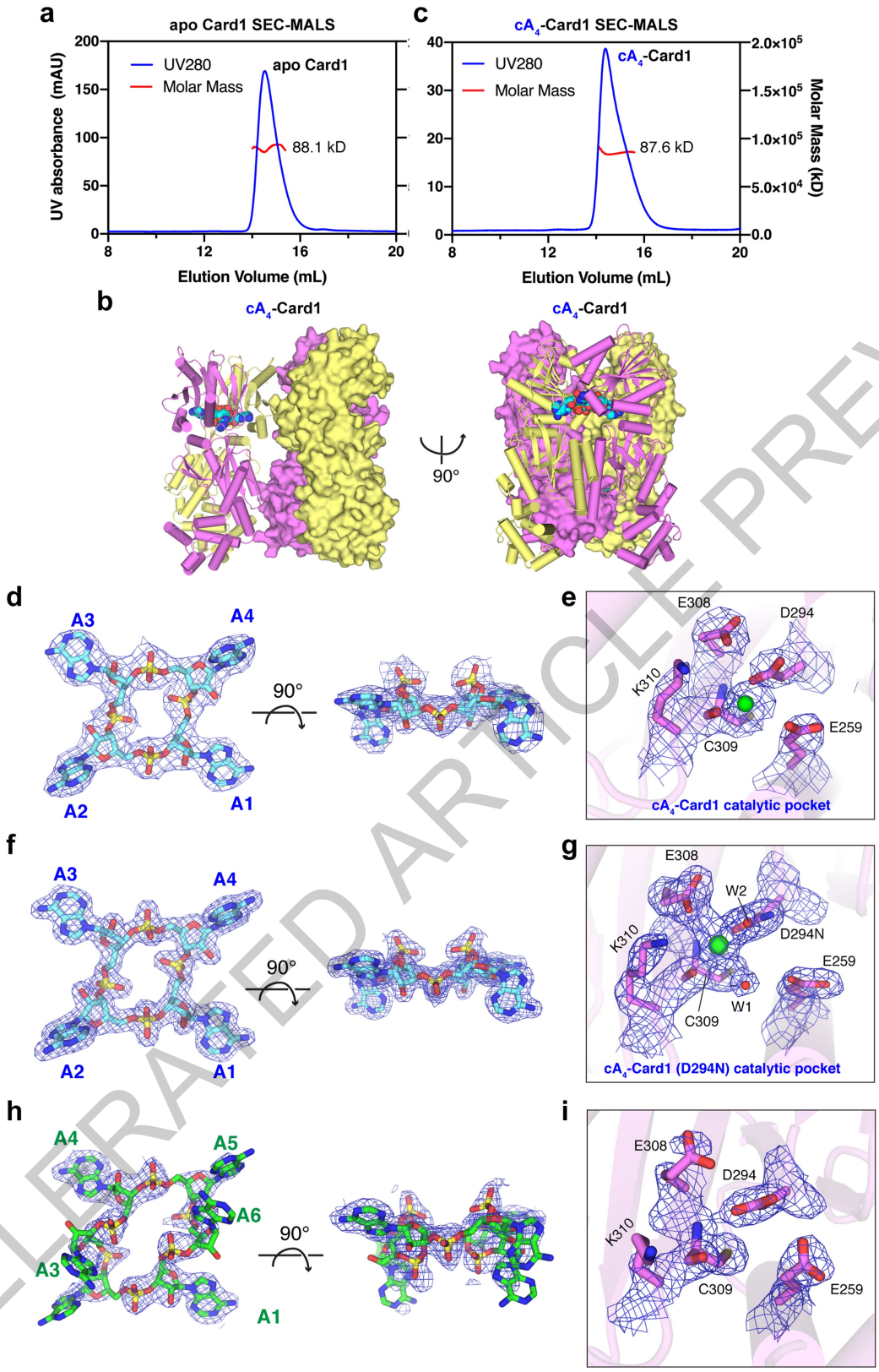
Extended Data Fig. 2 | cA_n-mediated cleavage of ssRNA by Card1 at 37 °C. **(a)** Card1 digestion of a ssRNA or a dsRNA molecular weight ladder. Card1 rapidly degrades the ssRNA, but not the dsRNA, ladder; performed three times. **(b)** Digestion of a 60-nucleotide (nt) RNA species for 15 minutes in buffers containing either no divalent cation, or either Mn^{2+}, Mg^{2+}, Ca^{2+}, or Zn^{2+}; performed two times. **(c)** Cleavage of RNA oligonucleotides containing a fluorophore-quencher pair, measured as the increase in fluorescence, by Card1 with or without cA_4, or with the non-specific RNase I as a positive control. The RNA oligonucleotides are either poly-A_{15}, poly-C_{15}, or poly-U_{15}. Poly-G could not be synthesized, and cleavage of G_5-A-G_5 nor G_5-C-G_5 could not be tested due to its resistance to cleavage by RNases. Each bar represents the mean of three replicates ± s.e.m., given as relative fluorescent units **(d)** Simultaneous Card1 digestion of a pair of 30-nt DNA and RNA oligonucleotides, or of a pair of 50-nt DNA or RNA oligonucleotides, with increasing concentration of Card1 and cA_4 in the presence of Mn. This results in direct competition between the DNase and RNase activities of Card1 in each reaction. For each pair, one oligonucleotide is labelled with Cy3 and the other with Cy5 fluorescent groups, and the two panels display the same gel imaged through different filters. All reactions were quenched after 15 minutes; performed one time.

ACCELERATED ARTICLE PREVIEW



Extended Data Fig. 3 | Energetics of binding of cAs to dimeric Card1 and Electrostatic surface representation of Card1 and its cA_n-bound complexes. (a, b) ITC curves for binding of cA_4 **(a)** and cA_6 **(b)** to dimeric Card1. **(c)** No binding observed by ITC for cA_4 binding to selected Card1 mutants S11A, Y122A and I125A. **(d)** K_D values determined from ITC binding studies of selected Card1 mutants Y340A and M42A. **(e-g)** Electrostatic surface views of apo (panel e), cA_4-bound (panel f) and cA_6-bound (panel g) Card1. Electrostatic surface potentials were calculated in PyMol and contoured at ± 75.

ACCEPTED ARTICLE PREVIEW

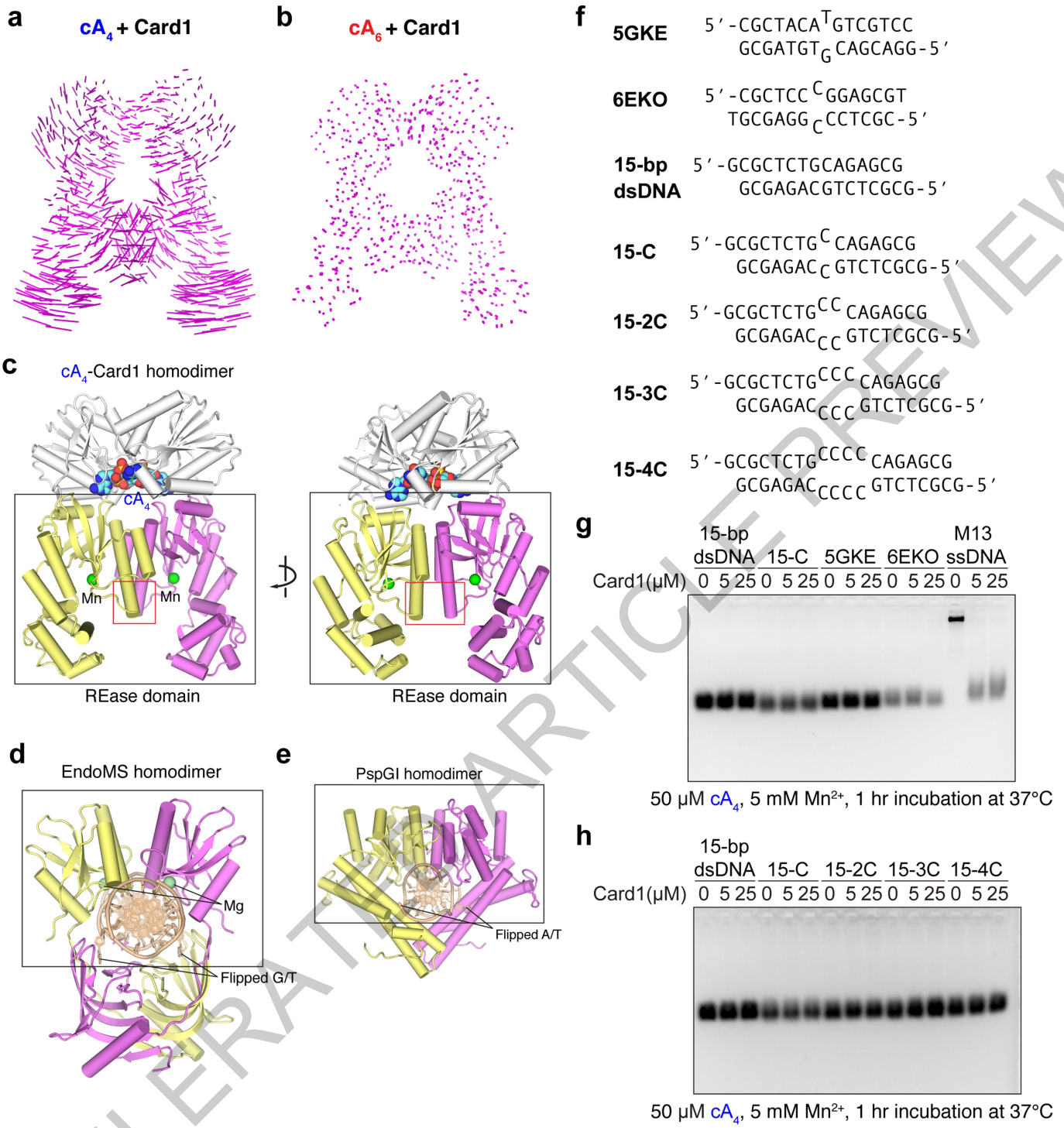


Extended Data Fig. 4 | See next page for caption.

ACCELERATED ARTICLE PREVIEW

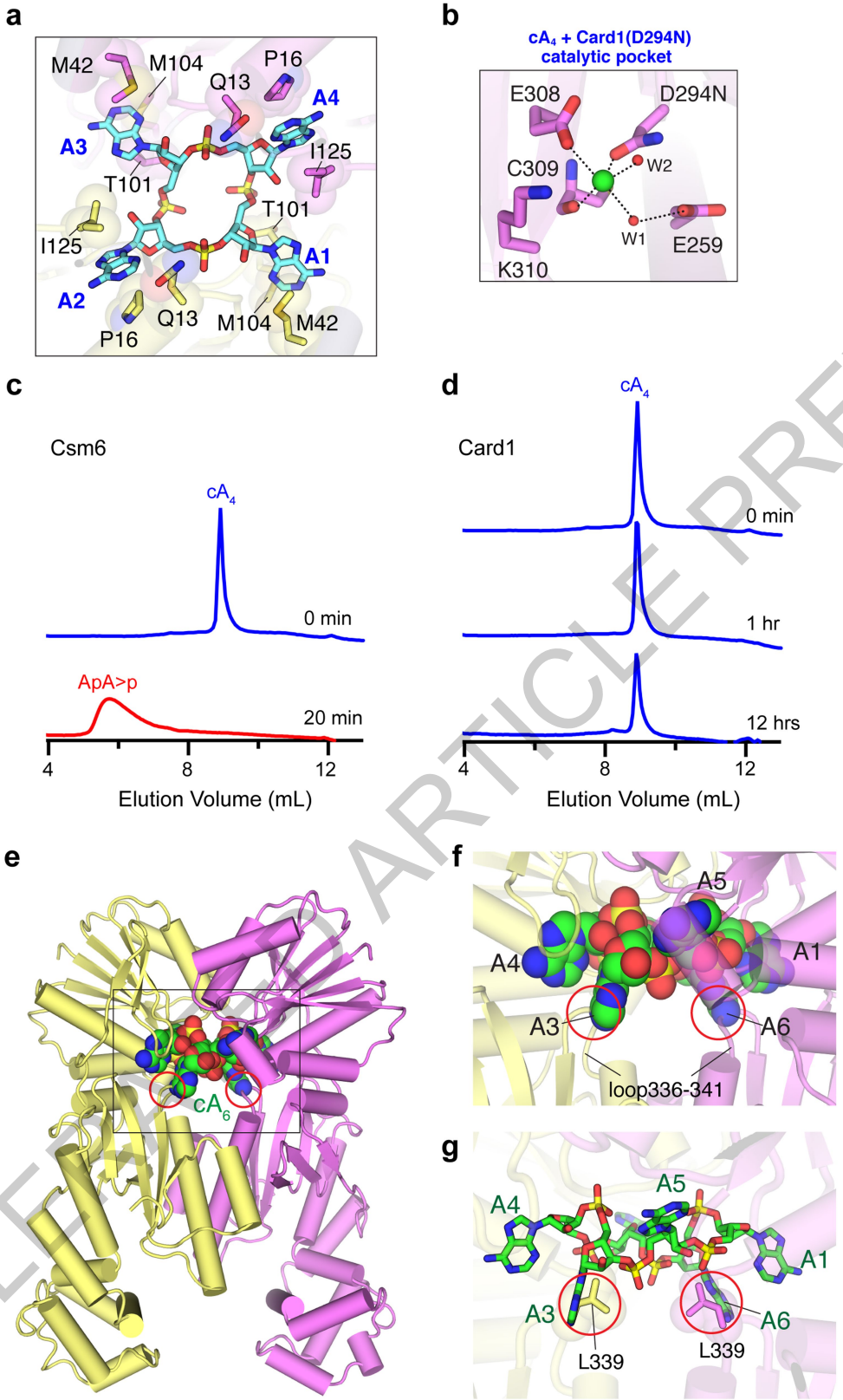
Extended Data Fig. 4 | Oligomeric composition of the apo-Card1 and cA_4-Card1 complex in the crystal and in solution and quality of 2Fo-Fc density maps in cA_n- and cA_6-bound structures of Card1. (a) SEC-MALS measurement of the molecular weight of apo-Card1 in solution. The measured solution molecular weight of 88.1 kDa is close to the calculated molecular weight of the dimeric Card1 (90.0 kDa). **(b)** Two alternate views of the head-to-tail dimer of dimers alignment of cA_4-Card1 complex in the crystal. One dimer is shown in a ribbon while the other dimer is shown in a surface representation. **(c)** SEC-MALS measurement of the molecular weight of cA_4-Card1 complex in solution. The measured solution molecular weight of 87.6 kDa is close to the calculated molecular weight of the dimeric cA_4-Card1 complex (91.2 kDa) rather than the tetrameric complex. **(d, e)** Fitting of electron density contoured at 1.2 σ for (panel d) two orthogonal views of bound cA_4 and (panel e) the Mn in the catalytic pocket in the cA_4-dimeric Card1 complex. (**f, g**) Fitting of electron density contoured at 1.2 σ for the CARF pocket (panel f) and the catalytic pocket with bound Mn cation (panel g, shown as a green ball) in the structure of the cA_4-dimeric Card1 D294N mutant complex. Bound waters (shown as red balls) can be observed in this 1.95 Å high-resolution structure of this mutant complex. **(h, i)** Fitting of electron density contoured at 1.2 σ for two orthogonal views of bound cA_6 (panel h) and key residues in the catalytic pocket (panel i) in the cA_6-dimeric Card1 complex.

ACCELERATED ARTICLE PREVIEW



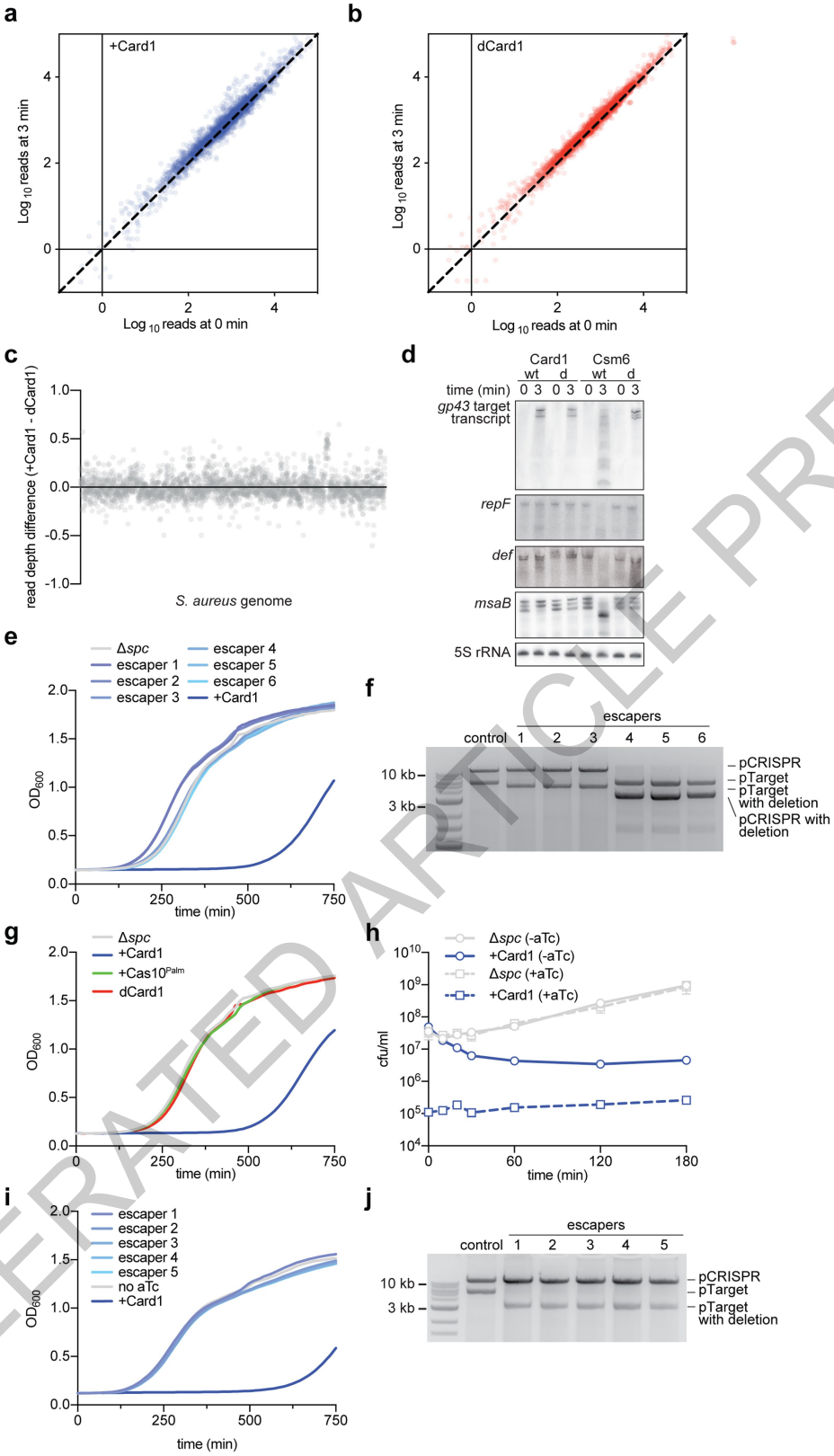
Extended Data Fig. 5 | Conformational changes between apo- and cA_n-bound states of dimeric Card1, comparison of the dimeric REase pockets in cA_4-bound Card1 and type II restriction enzyme complexes with bound mismatch-containing dsDNA and attempts at cA_4-mediated cleavage of dsDNA containing central mismatches by Card1. (a, b) Vector lengths identify degree of conformational changes between dimeric apo-Card1 and cA_4-Card1 complex (panel a) and cA_6-Card1 complex (panel b). **(c)** Two views of the structure of cA_4-bound Card1. The dimeric alignment of the REase domains of Card1 are shown in a black box. Note the position of one end of helical segments shown in a red box. **(d, e)** Structures of type II restriction enzymes EndoMS (PDB: 5GKE) (panel d) and PspGI (PDB: 3BM3) (panel e) bound to a single central mismatch containing dsDNA. The black boxed regions highlight their pair of REase domains. (**f**) dsDNA substrates containing mismatches used for cleavage assays in (panel g) and (panel h). (**g**) cA_4-activated Card1 digestion of dsDNA containing either no mismatch or a central C•C mismatch, as well as central mismatch-containing DNAs used in EndoMS (PDB 5GKE) and PfoI (PDB 6EKO) complexes. Cleavage of M13 ssDNA is shown as a control; performed two times. (**h**) cA_4-activated Card1 digestion of dsDNA containing either single, double, triple or quadruple central C•C mismatches; performed two times.

ACCELERATED ARTICLE PREVIEW



Extended Data Fig. 6 | Intermolecular contacts and Mn coordination in the cA_4-Card1 complex, comparative cleavage propensity of conversion of cA_4 to ApA>p by Csm6 and Card1 in a time-dependent manner and structure of the cA6-Card1 complex and clashes between cA6 and dimeric Card1 loop residues in a model of cA6 in the bound state.. **(a)** Intermolecular hydrophobic interactions between bound cA_4 and amino acids of dimeric Card1. **(b)** Amino acids lining the catalytic pocket of the structure of the cA_4-Card1(D294) complex. The bound Mn is shown as a green ball. The water molecules are shown as red balls. **(c)** Mono-Q column analysis of cA_4 cleavage by Csm6. **(d)** Mono-Q column analysis of the time-dependent stability of cA_4 incubating with Card1. **(e)** Structure of the cA_6-dimeric Card1 complex. **(f, g)** Clash between cA_6 and L339 of loop residues of Card1 in space-filling (panel f) and ribbon (panel g) representations in a model of the cA_6-dimeric Card1 complex.

ACCELERATED ARTICLE PREVIEW

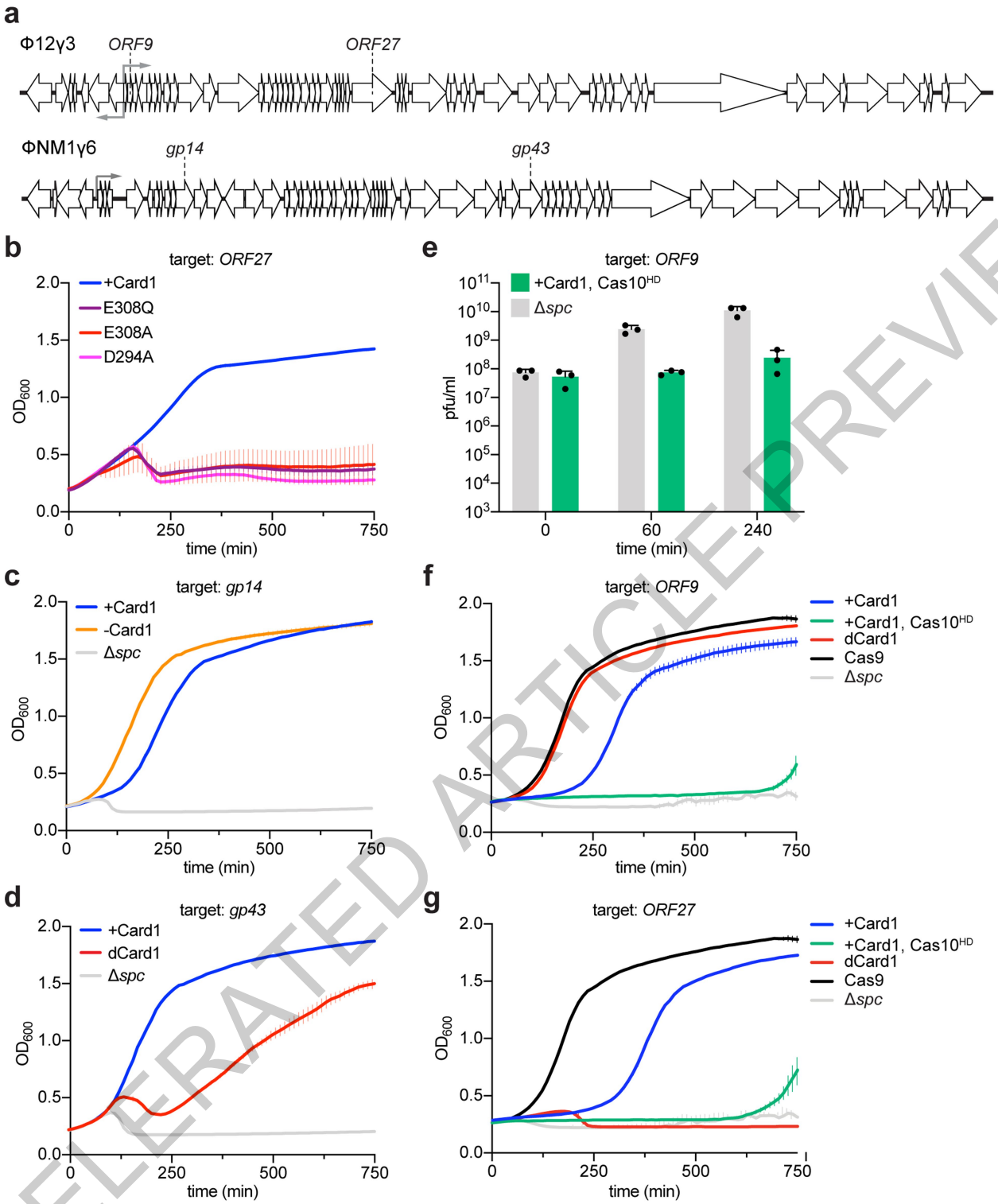


Extended Data Fig 7 | See next page for caption.

ACCELERATED ARTICLE PREVIEW

Extended Data Fig 7 | The RNase activity of Card1 is not detected *in vivo*, however its activation leads to a growth arrest. (a) RNA-seq of staphylococci harbouring pTarget and pCRISPR-Cas10^{HD}, and +Card1. At 0 minutes, targeting is induced by the addition of aTc, and cells are harvested after 3 minutes. An equal amount of RNA from *Listeria seeligeri* was added to all samples prior to RNA purification to allow absolute comparison between timepoints. Each dot represents a gene, and is the average of two biological replicates. Genes that fall on or near the identity line are unchanged by 3 minutes of Card1 activity. **(b)** Like (**a**), but in cells carrying a catalytically dead Card1 (dCard1). **(c)** A comparison between the $\log_{10}$ read depth for all individual chromosomal genes between +Card1 cells and dCard1 cells, at 3 minutes. A value of 0 means that a gene showed no difference between +Card1 and dCard1 cells. Overall, there is no clear trend for depletion (or enrichment) in +Card1 cells relative to dCard1 cells. **(d)** Northern blot analysis of cells carrying pTarget and pCRISPR-Cas10^{HD}, with either +Card1, dCard1, +Csm6, or dCsm6; performed three times. Targeting was induced at time 0 with the addition of aTc, and RNA was analysed with probes specific to the protospacer target transcript (in pTarget), the plasmid replication gene *repF* (in pTarget), the *def* gene (peptide deformylase, in the *S. aureus* chromosome), or the *msaB* gene (in the *msaABCR* operon, in the *S. aureus* chromosome). 5S rRNA is used as a loading control. Card1 activation showed no detectable RNA degradation, in contrast to robust RNA depletion following Csm6 activation. OD_{600} measurements confirmed that the +Card1 and +Csm6 cells both experienced growth arrest. **(d)** Growth of staphylococci carrying different pCRISPR(+Card1) taken from six escaper colonies obtained at the end of the experiment in Figure 3c, measured as OD_{600} after the addition of aTc to induce the production of cA_4 by the Cas10 complex. Mean of three biological triplicates ± s.e.m. are reported. **(e)** Agarose gel electrophoresis of plasmid DNA was extracted from escaper cells grown in **(d)**, showing deletions in pTarget or pCRISPR. Sanger sequencing determined the same promoter deletion in pTarget escapers 1-3, and similar pCRISPR deletions in escapers 4-6, all comprising the whole CRISPR-*cas* locus. **(f)** Growth of staphylococci carrying different pCRISPR variants expressing Cas10^{HD}, measured as OD_{600} after the addition of aTc to induce the production of cA_4 by the Cas10^{HD} complex. Mean of three biological triplicates ± s.e.m. are reported. **(g)** Enumeration of colony-forming units (cfu) within staphylococcal cultures carrying different pCRISPR variants expressing Cas10^{HD} where cA_4 production was activated by the addition of aTc. At the indicated times after induction aliquots were removed and plated on solid media with or without aTc to count the remaining viable cells. Mean of three biological replicates ± s.e.m are reported. **(h)** Growth of staphylococci carrying different pCRISPR(+Card1, Cas10^{HD}) taken from five escaper colonies obtained in **(g)**, measured as OD_{600} after the addition of aTc to induce the production of cA_4 by the Cas10^{HD} complex. Mean of three biological triplicates ± s.e.m. are reported. **(j)** Agarose gel electrophoresis of plasmid DNA was extracted from escaper cells grown in **(i)**, showing deletions in pTarget. Sanger sequencing determined the same deletion in pTarget escapers 1-5, comprising both the promoter and target sequences.

ACCELERATED ARTICLE PREVIEW

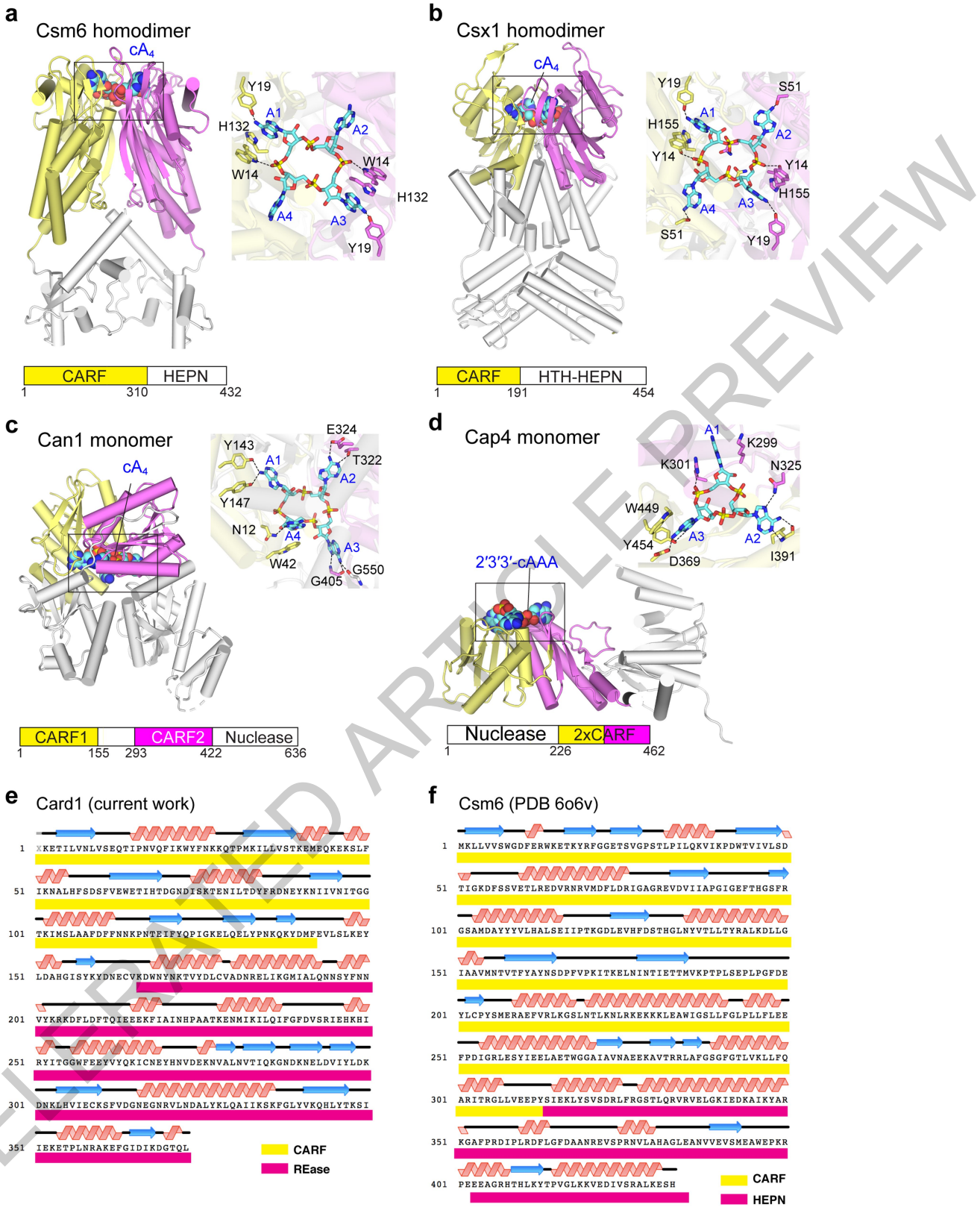


Extended Data Fig 8 | See next page for caption.

ACCELERATED ARTICLE PREVIEW

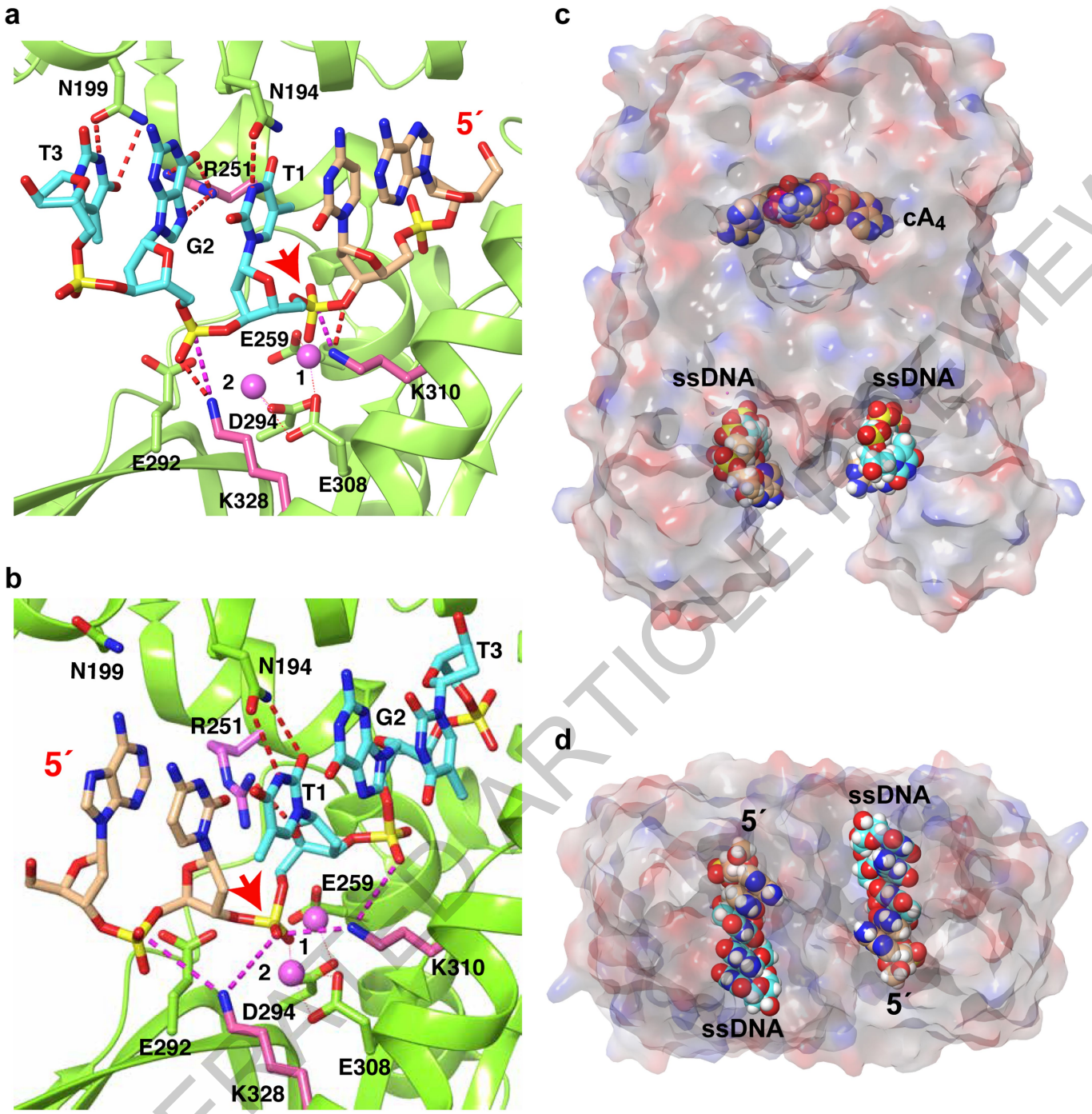
Extended Data Fig 8 | Card1-mediated anti-phage immunity. (a) Schematic of the genomes of the staphylococcal phages used in this study, Φ12γ3 and ΦNM1γ6, showing the location of the transcripts targeted by the type III-A CRISPR-Cas system. Grey arrows indicate promoters. **(b)** Growth of staphylococci carrying different pCRISPR variants with mutations in the catalytic pocket of Card1, programmed to target the *ORF27* transcript of Φ12γ3, measured as OD_{600} at different times after infection, at an MOI ~15. Mean of three biological triplicates ± s.e.m. are reported. **(c)** Growth of staphylococci carrying different pCRISPR variants programmed to target the *gp14* transcript of ΦNM1γ6, measured as OD_{600} at different times after infection, at an MOI ~15. Mean of three biological triplicates ± s.e.m. are reported. **(d)** Same as in **(c)** but targeting the *gp43* transcript, at an MOI ~2. Mean of three biological triplicates ± s.e.m. are reported. (**e**) Enumeration of plaque-forming units (pfu) within staphylococcal cultures carrying different pCRISPR variants after infection with Φ12γ3 at an MOI ~10. At the indicated times after infection aliquots were removed and plated on top agar media seeded with a susceptible strain. Mean of three biological replicates ± s.e.m are reported. **(f)** Growth of staphylococci carrying different pCRISPR variants programmed to target the *ORF9* transcript of Φ12γ3, measured as OD_{600} at different times after infection at an MOI ~25. The immunity provided by the Cas9 nuclease, which directly recognizes and cleaves the phage genome shortly after its injection and therefore allows the survival of the infected cells, is used as a control to show that the observed growth delays are not due to an excessive amount of phage added in the experiment. Mean of three biological triplicates ± s.e.m. are reported. **(g)** Same as in **(f)** but targeting the *ORF27* transcript. In both **(f)** and **(g)**, $Cas10^{HD}$ cells with +Card1 do not lyse from infection (as it is the case for *Δspc* cells), indicating an incomplete phage life cycle. Mean of three biological triplicates ± s.e.m. are reported. In **(f)** and **(g)**, the data representing Cas9 and *Δspc* is from the same experiment.

ACCELERATED ARTICLE PREVIEW



Extended Data Fig 9 | Comparison of the cA bound structures of dimeric Csm6, dimeric Csx1, monomeric Cam1 and monomeric Cap4 with the emphasis on domain alignment, overall structure, and interactions between bound cA and catalytic residues in the dimeric CARF pockets and a comparative study of sequence and topology between cA_4 complexes with dimeric Card1 and dimeric Csm6.. (**a**) cA_4-dimeric Csm6 complex (PDB: 6O6V). (**b**) cA_4-dimeric Csx1 complex (PDB: 6R9R). (**c**) cA_4-monomeric Can1 complex (PDB: 6SCE). (**d**) cA_3-Cap4 monomeric complex (PDB: 6VM6). There is a trimeric alignment of monomers in this complex. **(e, f)** Sequence and topology of cA_4-dimeric Card1 complex (panel e) and cA_4-dimeric Csm6 complex (PDB: 6O6V) (panel f). The secondary structure (helices and sheets) is shown above the sequences. The CARF domains are highlighted by yellow. The images were generated by PHENIX.refine and modified.

ACCEPTED ARTICLE PREVIEW



Extended Data Fig 10 | Docking model of ssDNA two possible strand directionalities positioned in the catalytic pocket of one monomer of the structure of the cA_4-bound Card1 complex and model of alignment of a pair of ssDNAs (strand directionality 1) within the catalytic pockets of the opposing REase domains of cA_4-Card1 complex. (a) We used the HDOCK program[47] to position a B-form ssDNA with the sequence ApC**p**T1pG2pA3 with one (labeled strand directionality 1) of two possible strand directionalities in the catalytic pocket whereby the cleavable phosphate (in bold) was positioned relative to the pair of divalent cations coordinated to the catalytic acidic residues (see red arrow). One divalent cation (labeled 1) was observed in the x-ray structure of cA_4-bound Card1, while the other (labeled 2) was modeled based on its position in the structure of the *Ngo*MIV restriction enzyme-DNA complex (PDB: 1FIU) [49], which exhibits a similar catalytic residue alignment. The model outlines hydrogen bonding alignments (dashed lines) in the model of the T1-G2-T3 segment interacting with side chains of the REase domain. It also shows positioning of the sugar-phosphate backbone flanking the cleavable phosphate (red arrow) of the modeled bound ssDNA within the REase catalytic pocket, thereby outlining the alignment of the cleavable phosphate relative to the pair of divalent cations and acidic catalytic residues (E259 coordinated to cation E292 coordinated to cation 2, while D294 and E308 coordinated to both cations). A pair of lysine side chains (K310 and K328) form salt bridges (dashed lines) to the cleavage site and flanking phosphates. **(b)** The ApCpT1pG2pT3 is bound to the REase domain with an opposite directionality (labeled strand directionality 2). See the Methods section for modeling computations using strand directionality 2. **(c)** Side view of the modeled complex emphasizing the space available to readily accommodate ssDNAs (strand directionality 1) positioned in the pair of opposing REase pockets. **(d)** A top-down view of the modeled complex emphasizing the strand directionalities of the bound ssDNAs (strand directionality 1) positioned in the pair of opposing REase pockets.

ACCEPTED ARTICLE PREVIEW

nature research

Corresponding author(s): Luciano Marraffini

Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see Authors & Referees and the Editorial Policy Checklist.

Statistical parameters

When statistical analyses are reported, confirm that the following items are present in the relevant location (e.g. figure legend, table legend, main text, or Methods section).

n/a	Confirmed	
☐	☒	The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☐	☒	An indication of whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒	The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☒	☐	A description of all covariates tested
☒	☐	A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☐	☒	A full description of the statistics including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☐	☒	For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
☒	☐	For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☒	☐	For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☒	☐	Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated
☐	☒	Clearly defined error bars *State explicitly what error bars represent (e.g. SD, SE, CI)*

Our web collection on statistics for biologists may be useful.

Software and code

Policy information about availability of computer code

Data collection	custom Python code mapping ssDNA NGS reads to reference genome available upon request
Data analysis	Geneious version 2020.1.1 STAR mapper 2.7.3 WebLogo3 version 3.7 COOT Version 0.8.9.2 PyMol Version 2.2.3 Phenix version 1.18 GraphPad Prism Version 8 UCSF Chimera version 1.13.1

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors/reviewers upon request. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

The atomic coordinates have been deposited in the Protein Data Bank with the codes 6WXW (apo Card1), 6WXX (cA4-Card1 complex), 6WXY (cA6-Card1 complex), and 6XL1 (cA4-Card1(D294N) complex). The raw data for the DNAseq and RNAseq experiments performed in this study is found at the Sequence Read Archive (NIH) through accession code PRJNA672128.

Field-specific reporting

Please select the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/authors/policies/ReportingSummary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	Sample sizes were not statistically predetermined , standard 3 biological replicates were used
Data exclusions	No data was excluded from the analysis
Replication	All experimental findings were reliably replicated as indicated in the figure legends
Randomization	Not relevant to this study as there are no animal nor human experiments, and the experimental outcome does not depend on the order in which samples were analysed in the experiments
Blinding	Not relevant to this study as there are no animal nor human experiments, and the knowledge the order or identity of a sample does not change the experimental outcome

Reporting for specific materials, systems and methods

Materials & experimental systems

n/a	Involved in the study
☒	☐ Unique biological materials
☒	☐ Antibodies
☒	☐ Eukaryotic cell lines
☒	☐ Palaeontology
☒	☐ Animals and other organisms
☒	☐ Human research participants

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☒	☐ Flow cytometry
☒	☐ MRI-based neuroimaging