TITLE

DNA interference states of the hypercompact CRISPR-CasΦ effector

AUTHORS

Patrick Pausch[1,2,3*], Katarzyna M. Soczek[1,2,3,*], Dominik A. Herbst[3,4], Connor A. Tsuchida[1,5], Basem Al-Shayeb[1,6], Jillian F. Banfield[1,7], Eva Nogales[2,3,4,8], Jennifer A. Doudna[1,2,3,4,8,9,10,†]

AFFILIATIONS

[1] Innovative Genomics Institute, University of California, Berkeley, CA, USA

[2] Department of Molecular and Cell Biology, University of California, Berkeley, CA, USA

[3] California Institute for Quantitative Biosciences (QB3), University of California, Berkeley, CA 94720

[4] Molecular Biophysics and Integrated Bioimaging Division, Lawrence Berkeley National Laboratory, Berkeley, CA, USA

[5] University of California, Berkeley-University of California, San Francisco Graduate Program in Bioengineering, Berkeley, CA, USA

[6] Department of Plant and Microbial Biology, University of California, Berkeley, CA, USA

[7] Department of Earth and Planetary Sciences, University of California, Berkeley, CA, USA

[8] Howard Hughes Medical Institute, USA

[9] Department of Chemistry, University of California, Berkeley, CA, USA

[10] Gladstone Institute of Data Science and Biotechnology, Gladstone Institutes, San Francisco, CA, USA

* Contributed equally to this manuscript

† Correspondence: doudna@berkeley.edu

ABSTRACT

CRISPR-CasΦ, a small RNA-guided enzyme found uniquely in bacteriophages, achieves programmable DNA cutting as well as genome editing. To investigate how the hypercompact enzyme recognizes and cleaves double-stranded DNA, we determined cryo-EM structures of CasΦ (Cas12j) in pre- and post-DNA binding states. The structures reveal a streamlined protein architecture that tightly encircles the CRISPR RNA and DNA target to capture, unwind and cleave DNA. Comparison of the pre- and post-DNA binding states reveals how the protein rearranges for DNA cleavage upon target recognition. Based on these structures, we created and tested mutant forms of CasΦ that cut DNA up to 20-fold faster relative to wildtype, showing how this system may be naturally attenuated to improve the fidelity of DNA interference. The structural and mechanistic insights into how CasΦ binds and cleaves DNA should allow for protein engineering for both *in vitro* diagnostics and genome editing.

MAIN TEXT

INTRODUCTION

CRISPR-Cas systems (clustered regularly interspaced short palindromic repeats, CRISPR-associated proteins) enable microbes to acquire immunity against bacteriophages and plasmids [1,2], and they also function as transformative tools for genome editing in a wide range of cell types [3]. Fundamental to these systems are RNA-guided nucleases, such as the Cas9 (type II) and Cas12 (type V) families that use CRISPR RNA (crRNA) to recognize foreign double-stranded DNA (dsDNA) by forming an R-loop structure in which 20 nucleotides (nts) of the crRNA (the crRNA "spacer") base pair with one strand of the target DNA [2,4]. In addition, both protein families must bind to a protospacer-adjacent motif (PAM), a short DNA sequence next to the crRNA-complementary sequence, to prevent autoimmunity and initiate R-loop formation for DNA interference [5].

Although CRISPRs are typically encoded in the genomes of bacteria and archaea [6], the hypercompact CRISPR-CasΦ system (type V-J, Cas12j) occurs exclusively in bacteriophages [7]. CRISPR-CasΦ genomic loci comprise solely the *casΦ* effector gene and a short CRISPR sequence array containing 36-base pair sequence repeats flanking an average of five unique spacer sequences (**Fig. 1a**) [8]. The CasΦ protein, which is distantly related to other type V (Cas12) enzymes based on homology of its Mg^{2+}-dependent RuvC endonuclease domain [7], initiates dsDNA unwinding by binding to a 5′-TBN-3′ PAM sequence (where B is G, T, or C), next to a crRNA-complementary DNA target sequence (**Fig. 1a**) [8]. Similar to other Cas12 enzymes, CasΦ catalyzes guide RNA hybridization to ~20 nts of the DNA target strand and consecutive cleavage of the DNA non-target stand (NTS) and target strand (TS) to introduce staggered cuts (**Fig. 1a**) [8]. CasΦ also exhibits target-activated collateral single-stranded DNA cleavage [8], an activity associated with the Cas12 enzyme family [9]. Uniquely, however, CasΦ generates mature crRNAs from CRISPR-array transcribed precursor-cRNA (pre-crRNA) using the same RuvC active site that it uses for DNA cutting (**Fig. 1a**) [8]. The ~100 kDa CasΦ-crRNA effector complex is substantially smaller than other dsDNA cutting CRISPR effectors [8], such as crRNA-guide bound SpyCas9 (~190 kDa) [10–13], FnCas12a (~165 kDa) [14,15], DpbCasX (~155 kDa) [16,17], and Cas14 dimers (~180 kDa) [18,19], but maintains the ability to unwind and cut dsDNA to protect its host from superinfection by other mobile genetic elements [8]. How CasΦ achieves RNA-guided DNA unwinding and cleavage despite its compact size and evolutionary remoteness from much larger type V CRISPR-Cas proteins is not known.

We show here that CasΦ forms a compact architecture in which protein and crRNA intertwine to enable RNA-guided dsDNA unwinding and cleavage. Cryo-EM-based structural studies of CasΦ in three states, one pre- and two post-DNA binding, reveal conformational changes the enzyme undergoes during DNA recognition for nuclease activation. Using two distinct strategies to trap CasΦ in DNA-bound states prior to and upon DNA recruitment to the active site, we identified key structures necessary for DNA recognition and cutting. Our results

support a model in which CasΦ shares surprising structural and mechanistic similarities with much larger type V CRISPR-Cas enzymes. This analysis enabled the design and testing of CasΦ amino-acid substitution and deletion mutants that have enhanced DNA cutting kinetics relative to the wildtype enzyme and show how CasΦ is naturally inhibited by an α-helix to improve the fidelity of DNA cutting. The results suggest that Cas proteins enhance DNA-interference fidelity at the expense of rapid DNA cleavage kinetics.

RESULTS

CasΦ-crRNA forms a compact surveillance complex poised for DNA binding

To determine the structural basis for CasΦ activity, we focused on the CasΦ-2 ribonucleoprotein (RNP), which possesses a robust DNA cutting and genome editing activity relative to other CasΦ homologs [8]. The purified protein was complexed with a crRNA comprising a 24 nt repeat sequence and a 20 nt spacer sequence [8] (**Fig. 1b**) to form an RNA-guided enzyme competent for DNA recognition. Cryo-EM structure determination (**Fig. 1c, Table 1, Extended Data Fig. 1**) at a resolution of 3.6 Å revealed the streamlined architecture of CasΦ, which wraps tightly around the crRNA guide in the binary state (**Fig. 1d**). Missing density for residues 1-52 and 717-757 indicated structural flexibility in the N- and C-terminal regions (**Fig. 1c**).

Despite its small size and sequence divergence, the CasΦ-crRNA structure revealed overall similarities to other type V proteins that were not anticipated due to the absence of sequence conservation outside of the RuvC domain. The architecture resembles a distinct miniature version of the recognition (Rec) and nuclease (Nuc) bilobed structure of large type V CRISPR-Cas enzymes, including Cas12a (**Extended Data Fig. 2**), that structurally separates target DNA recognition and cleavage functions. The Rec lobe of CasΦ includes 10 α-helices, as compared to the ~20 α-helices of Cas12a and other type V enzymes (**Extended Data Fig. 2**). Similarly reduced in size, the Nuc lobe is solely formed by the RuvC domain and a minimal zinc

ribbon (ZR) Cys4-type zinc finger [20], and lacks the structurally distinct Nuc domain of large Cas12 proteins (**Extended Data Fig. 2**). Zinc finger and Nuc domains (*i.e.*, Nuc of Cas12a, Cas12b and Cas12i; target strand loading domain (TSL) of CasX; target nucleic-acid binding domain (TNB) of Cas14) have been previously suggested to assist in recruitment of the DNA substrate to the RuvC active site for DNA cleavage [17,18,21–24]. In CasΦ, a Cas12-typical oligonucleotide binding domain (OBD), characterized by its central 7-stranded β-barrel, accommodates the crRNA hairpin (**Fig. 1d and Extended Data Fig. 3**), and an inserted four α-helical bundle of the Rec lobe channels nucleotides 1-12 of the crRNA spacer via an elongated α-helix (α7) in a channel above the RuvC domain (**Fig. 1d**). The crRNA spacer occurs in a pre-ordered A-form geometry in positions 1-5 (**Fig. 1d**), which could offset the entropic penalty of crRNA:TS hybridization during target recognition, providing the seed for R-loop formation [25]. We hypothesize that the α-helical bundle assists in target recognition, and we refer to it as RecI domain, consistent with the nomenclature of large type V enzymes. A second recognition domain (RecII) is inserted between the RuvC-I and RuvC-II subdomains of the discontinuous RuvC domain and encloses the crRNA spacer up to nucleotide 12 (**Fig. 1b,c**). We did not observe defined density for nucleotides 14-20 of the crRNA (**Fig. 1c**), indicating flexibility in the binary state.

Notably, spacer lengths ≥ 14 nt, or ≥ 16 nt, are required to bind and cleave supercoiled plasmid DNA [8], or linear dsDNA (**Supplementary Figs. 1 and 2a**), respectively. It is conceivable that encasing of the crRNA spacer might protect the RNA from RNase-mediated degradation *in vivo*, to maintain the crRNA integrity for genome surveillance.

A prominent feature of the CasΦ-crRNA surveillance complex is an elongated protein loop structure (amino acids 607-634) that associates with the top of the RuvC active site, between the mixed beta-sheet of RuvC and the crRNA spacer, and extends to the RecII domain (**Fig. 1e**). The low resolution of the EM density map around the loop and the RecII domain indicates structural flexibility of this region (**Extended Data Fig. 1b**). The loop is reminiscent of the lid structures that occlude the RuvC active site for nuclease repression in other Cas12 enzymes [26,27], and might

therefore prohibit association of the magnesium cofactors and DNA substrate to the catalytic RuvC residues D394, E606 and D695 within the binary state. This suggests that enzymatic activity is inhibited in the absence of a DNA target. This observation helps explain previous results showing that CasΦ does not indiscriminately cleave DNA in the absence of a crRNA-bound DNA target [8].

Structure of CasΦ-crRNA bound to DNA

To understand how the minimal domains in CasΦ enable sequence-specific DNA unwinding, we reconstituted the DNA-bound ternary complex for cryo-EM structure determination by assembling CasΦ-crRNA with DNA containing a target strand complementary to the 20-nucleotide guide sequence (**Fig. 2a**). To prevent DNA cutting for entrapment of the ternary complex in the pre-cleavage state, Mg^{2+} was omitted from the complex; to favor R-loop formation, the DNA lacked complementarity between the TS and NTS (**Fig. 2a**). The resulting ~3 Å resolution structure revealed CasΦ bound to crRNA and DNA (**Fig. 2b,c, Table 1, Extended Data Fig. 4**).

Similar to other Cas12 enzymes, the bound DNA bifurcates downstream of the PAM to enable base pairing between the DNA target strand and the crRNA spacer to form an A-form RNA-DNA heteroduplex for sequence-specific target recognition (**Fig. 2a,c**). The non-target DNA strand is channeled towards the RuvC nuclease active site where, in the presence of Mg^{2+}, it would undergo first-strand cleavage as part of a two-step reaction (**Fig. 2c**). Notably, we did not observe defined EM density corresponding to the PAM-distal dsDNA, nucleotides 11-20 of the single-stranded NTS, the ZR domain and the ZR-adjacent RuvC active site residue D695 (**Fig. 2b**), suggesting that these elements are highly flexible in this state. The RecII domain caps the crRNA:TS duplex at base pair (bp) 18 (**Fig. 2a,c**) via a loop structure, in contrast to the capping at bp ≥ 20 observed for large type V enzymes [17,18,23,26,28].

PAM recognition by three domains initiates DNA interference

CRISPR-Cas proteins initiate local DNA unwinding upon PAM recognition [5]. CasΦ recognizes its 5'-TTA-3' PAM in a shape- and sequence-specific manner using the OBD and RecI domains and an N-terminal three α-helical bundle (referred to as the PAM-Interacting domain; PI) (**Fig. 2c,d**); the PI domain was not visible in the binary state EM map, indicating flexibility (**Fig. 1c**). Conserved residues K29 and K33 of the PI, residue Q202 of the OBD, and residues V126, Q127 and N130 of RecI, directly probe the identity or lie in proximity to the PAM base pairs (**Fig. 2d**), while the surrounding protein scaffold accommodates the negatively charged sugar phosphate backbone (**Supplementary Fig. 3**). Individual and domain-centric combinatorial alanine substitution of the PAM-interacting residues results in a decreased ability of CasΦ to bind DNA (**Fig. 2d and Supplementary Fig. 2b**). Analytical size exclusion chromatography and DNA cleavage assays confirmed that these CasΦ mutants are properly folded and cut DNA (**Supplementary Figs. 1 and 4**), supporting the conclusion that the mutated residues are critical for PAM recognition. However, the V126A, Q127A and N130A RecI triple mutant does not cut DNA (**Supplementary Figs. 1)**, likely due to a strong DNA binding defect, as observed in the filter binding assay (**Supplementary Fig. 2b).** Although 5'-TGN-3' and 5'-TCN-3' PAMs promote DNA interference *in vivo* [8], we observed that neither of these two alternative PAMs are bound *in vitro* using small oligonucleotide duplexes as targets (**Supplementary Fig. 2c**). It is possible that other factors, such as the superhelicity of the target DNA, contribute to PAM-dependent DNA unwinding *in vivo* after recognition of suboptimal PAMs. This idea is supported by the observation that superhelical DNA targets containing alternative PAMs are efficiently cut, in contrast to linear DNA (**Extended Data Fig. 5**). Protein engineering, enabled by the structural knowledge provided here, might broaden the PAM promiscuity of CasΦ and increase the number of targetable sequences for genome editing, as demonstrated for Cas9 and Cas12a [29–31].

Coupled DNA recognition and positioning in the RuvC active site

We next aligned an ideal B-form dsDNA target to our structure, based on the PAM-proximal dsDNA conformation observed in the ternary state, in order to determine which elements might be involved in DNA unwinding (**Fig. 3a**). Following the theoretical path of the ideal B-form DNA duplex downstream of the PAM in the ternary state, helix α2 of the PI domain clashes with the DNA NTS (**Fig. 3a,b**). Deletion of the PI domain (Δ1-54) upstream of a di-proline hinge that connects it to the OBD (prolines 55, 56, 60 and 61, (**Supplementary Fig. 5**)) reduced the DNA binding affinity of CasΦ more than was observed for the PI domain-PAM interaction alanine variant K29A/K33A (**Fig. 3b**). It is conceivable that α2 acts as a helicase wedge structure that disrupts the PAM-downstream DNA duplex to unwind the dsDNA, functioning analogously to the loop-lysine helix–loop (LKL) region of Cas12a's DNA unwinding PI domain [27,32]. Helix α7 of the RecI domain also clashes with the dsDNA and might thus be involved in dsDNA interactions (**Fig. 3a,b**). However, helix α7 is associated with the crRNA spacer and RecII domain in the CasΦ binary state (**Fig. 1d**) and might therefore not engage in DNA interactions prior to R-loop formation.

The CasΦ ternary structure additionally revealed that three lysines (K367, K371, K373) and tryptophan W368, emanating from the loop adjacent to helix α13 of the RuvC domain, lie proximal to or contact the NTS downstream of the PAM (**Supplementary Fig. 6**). Interestingly, those residues may contact the DNA prior to unwinding, as suggested by the alignment of the B-form DNA to the ternary complex (**Fig. 3b**). Alanine substitution of the lysines drastically impaired the DNA binding ability of CasΦ (**Fig. 3b**), suggesting that those residues are critical for DNA recognition and that the RuvC domain itself is involved in R-loop formation.

We next superimposed the binary and ternary states of CasΦ to assess how DNA target identification and cleavage are coupled. The superposition revealed that RecI and RecII reposition ~50 Å away from each other to accommodate the crRNA:TS duplex (**Fig. 3c)**. This conformational change rotates helix α7 of the RecI domain in close proximity to the NTS near helix α13 of the RuvC domain, possibly to guide the strand to the active site (**Extended Data Fig. 6)**. The

comparison further revealed that the RuvC lid loop is displaced from the RuvC active site and associates with the crRNA:TS duplex (**Extended Data Fig. 7a**), likely involving rearrangement of the RecII domain to enable RuvC activation (**Fig. 3d)**. Concomitantly, parallel strands β13 and β14 of the RuvC domain mixed β-sheet, which directly connect to the lid loop, come closer to the active center in the ternary state (**Fig. 3d**). This rearrangement repositions the β13-emanating RuvC active side residue E606, relative to the catalytic center residue D394A (**Fig. 3d**), possibly for recruitment of the magnesium cofactor. Alanine substitution of the lid residues 610-614, or replacement of the lid-loop structure (aa 610-638) by a short glycine-serine linker (GSSG), abolished CasΦ's ability to cut dsDNA (**Extended Data Fig. 7b**). Analytical size exclusion chromatography confirmed that both CasΦ mutants are properly folded (**Extended Data Fig. 7c**), supporting the conclusion that the lid-loop structure might be critical for RuvC activation.

Upon R-loop formation, type V enzymes position the DNA NTS into the Mg^{2+}-dependent RuvC active site to catalyze first-strand endonucleolytic cleavage [22,24]. We found that the DNA cleavage- and Mg^{2+} binding-defective RuvC-I motif variant D394A [8] was also impaired in DNA binding (**Fig. 3e**). This observation implies that Mg^{2+}-mediated association of the NTS to the RuvC active site might contribute to R-loop formation. We tested a single amino acid substitution of the RuvC-II (E606Q) that is anticipated to bind Mg^{2+} and DNA while inhibiting DNA cutting, as shown for Cas12a [22]. Indeed, the E606Q variant had higher binding affinity for DNA relative to the D394A variant (**Fig. 3e,f**), while being strongly impaired for DNA cleavage relative to WT CasΦ (**Supplementary Fig. 1**). Moreover, in the absence of Mg^{2+}, the DNA binding affinities of the WT, D394A and E606Q variants were reduced to a level similar to that of the Mg^{2+}-binding defective D394A variant in the presence of Mg^{2+} (**Fig. 3e**). These data show that Mg^{2+}-mediated binding of DNA within the RuvC active site stabilizes the R-loop and highlights the contribution of CasΦ's RuvC domain to DNA binding in addition to its catalytic roles in pre-crRNA processing and DNA cleavage [8].

Structure of CasΦ-crRNA in complex with an active-site-bound DNA analog

To understand how the RuvC active site facilitates DNA binding and cleavage, we determined the CasΦ ternary structure in the presence of Mg^{2+} and a hydrolysis-inhibiting phosphorothioate-modified NTS to trap CasΦ in the catalytic state (**Fig. 4a,b, Table 1 and Extended Data Fig. 8**). The 2.9 Å resolution structure revealed the architecture of the RuvC active site with the ZR domain stabilized in close proximity (**Fig. 4c**). Based on the conserved RuvC fold, three negatively charged residues are likely to position the two scissile phosphate-coordinating cofactors Mg^{2+}_A and Mg^{2+}_B and emanate from the RuvCI (D394), RuvCII (E606) and RuvCIII (D695) motifs (**Fig. 4d,e**). Additionally, polar side chains of T663, S664 and R678 may contribute to the active site to stabilize Mg^{2+}_A and the scissile phosphate (**Fig. 4e**). The resolution of the map was not sufficient to model water molecules that contribute to the expected octahedral Mg^{2+} coordination sphere [24]. Based on the mechanism described for Cas12i [24], we propose that Mg^{2+}_A activates the water nucleophile for scissile phosphate attack, while Mg^{2+}_B stabilizes the 5' phosphate leaving group, to achieve DNA hydrolysis (**Fig. 4f**).

Surprisingly, the cryo-EM map revealed a density above the active site connecting to the crRNA bound TS DNA (**Extended Data Fig. 9**). Modeling of the TS into this density suggested that the PAM-distal TS DNA is single-stranded (**Supplementary note and Extended Data Fig. 9**), as previously shown for Cas12a [33]. PAM-distal dsDNA-unwinding and orientation of the single-stranded TS in a parallel orientation, relative to the NTS, may facilitate TS recruitment to the RuvC active site in a geometry permitting the DNA-cleavage catalysis.

Helix α7 enhances accuracy by attenuating cleavage kinetics

Inspection of the overall architecture of CasΦ revealed that helix α7 of the RecI domain blocks the path of the TS towards a PAM-proximal position (**Fig. 5a**). To assess the possible dynamics of the DNA and conformational heterogeneity around the active site, we performed a 3D variability analysis (3DVA) [34] on our cryo-EM datasets (**Supplementary note and Extended**

Data Fig. 10). The analysis revealed several distinct DNA-conformations within and above the RuvC active site, and a corresponding structural heterogeneity of the ZR (**Supplementary note and Extended Data Fig. 10**). Notably, the 3DVA suggested that helix α7 of the RecI domain and the loop connecting helices α17 and α18 of the RecII domain confine the DNA-sampling above the active site (**Supplementary note and Extended Data Fig. 10**).

The structural variability analysis implied that helix α7 might regulate the accessibility of the RuvC domain for association of single-stranded (ss)DNA. We wondered if mutations in helix α7 could increase active site accessibility and, correspondingly, CasΦ-catalyzed DNA cleavage kinetics. To test this idea, we introduced multiple alanine substitutions into the positively charged center (K146A, R150A, K153A, R157A) and negatively charged tip (E159A, S160A, S164A, D167A, E168A) of α7 in an effort to modulate nucleic acid binding interactions (**Fig. 5b**). Filter binding and DNA cleavage assays revealed that the CasΦ variants maintained the ability to bind and cut dsDNA (**Supplementary Figs. 1 and 2g**). While mutation of the positively charged center did not affect TS and NTS cleavage (**Supplementary Fig. 7**), variation of the negatively charged tip residues resulted in a variant that cleaved the NTS at a rate faster than observed for wildtype (WT) CasΦ (**Fig. 5c**). Replacement of the weakly conserved negatively charged tip (**Supplementary Fig. 8**) by a short glycine-serine linker (Δ155-176 (GSSG)) resulted in a variant that cleaved both the NTS and TS faster, leading to the introduction of dsDNA breaks with an ~18-fold higher rate relative to WT CasΦ (**Fig. 5c**). We hereafter refer to the engineered variants as nCasΦ (NTS accelerating; E159A, S160A, S164A, D167A, E168A) and vCasΦ (velocity; Δ155-176 (GSSG)).

It is conceivable that α7 sterically blocks access of the DNA to the RuvC active site, and helix truncation might thus facilitate faster association of the DNA strands for DNA cleavage. However, α7 interacts with the RecII domain in the binary state (**Fig. 1d**) and the truncation variant may be biased towards a RuvC-activated state, resulting in faster cleavage kinetics. To test this idea, we incubated the variant CasΦ RNPs with a ssDNA oligonucleotide in the absence of a

spacer-complementary DNA activator. The assay revealed that the variant CasΦ RNPs do not cleave ssDNA non-specifically and may therefore not reside in a pre-activated state (**Supplementary Fig. 9**). We hypothesized that the slower strand cleavage kinetics of WT CasΦ enable a higher on-target specificity by allowing CasΦ to dissociate from off-target DNA. To test this idea, we set up a crRNA:TS duplex base pair mismatch assay to assess the effects of single A-G, or C-T/C-U mismatches between the crRNA spacer and TS in positions 1-22 for WT CasΦ and the dsDNA cleavage accelerating vCasΦ variant. The assay revealed that vCasΦ is more tolerant of single mismatches in positions 1-15, as evidenced by the high DNA cleavage activity for mismatched substrates compared to WT CasΦ (**Fig. 6a**). This result suggests that helix α7 attenuates CasΦ-catalyzed DNA cutting kinetics to allow time for dissociation of mismatched targets. Helix α7 might thus improve the fidelity of DNA interference, presumably to reduce off-target toxicity in its host.

CasΦ variants display improved ssDNA detection activities

We wondered whether the two faster variants, vCasΦ and nCasΦ, might improve sensitivity in an *in vitro* nucleic acid detection assay. After target DNA recognition and cleavage, CasΦ remains in a nuclease-activated state to non-specifically degrade ssDNA in *trans* [8]. We employed a fluorophore quencher (FQ)-assay [35–37] to assess the ability of wildtype and variant CasΦ to non-specifically degrade a fluorophore conjugated reporter DNA after activation (**Fig. 6b**), which can be used to differentiate between active and inactive variants (**Extended Data Fig. 7d**). Incubation of the wildtype and engineered CasΦ RNPs in presence of a crRNA spacer complementary ssDNA oligonucleotide activator revealed that nCasΦ catalyzes fluorophore reporter degradation faster than vCasΦ and WT CasΦ (**Fig. 6b**), allowing detection of the activator in the low picomolar-range (**Fig. 6c**). These data show that CasΦ can be engineered for more efficient nucleic acid detection *in vitro* and suggests that other type V enzymes could be similarly engineered for more sensitive and ‘time-saving’ *in vitro* diagnostics.

To test whether the vCasΦ and nCasΦ variant proteins display an altered specificity for the activator, we set up a mismatch-FQ-assay by incubating CasΦ in the presence of activator ssDNA oligonucleotides with single and tiled double mismatches in positions 1-20 (**Fig. 6d**). The assay revealed that the mismatch-tolerance profiles of vCasΦ and nCasΦ are comparable to wild type CasΦ (**Fig. 6d**). These data show that the mutations within helix α7 do not alter the substrate specificity for matched and mismatched ssDNA activators and suggests that the observed differences in reporter turnover might be solely caused by faster cleavage kinetics, in contrast to an increased substrate affinity.

DISCUSSION

Our study revealed an unexpected convergence of the CasΦ structure with structures of much larger Cas12-family enzymes, despite protein size differences and sequence homology limited to the RuvC domain [7,8]. Accordingly, our data support a model for RNA-dependent DNA binding and cleavage by CasΦ that is analogous to other Cas12-family proteins. DNA interference begins with PAM-dependent dsDNA association to CasΦ, triggering structural rearrangements that locally melt the DNA for initial crRNA-DNA hybridization (**Fig. 6e**). Directional DNA unwinding, upon successive base pairing of the DNA to the crRNA spacer sequence, creates an R-loop structure in which the CasΦ Rec domains rearrange and the lid loop blocking the RuvC nuclease active site repositions to allow DNA access. Local DNA strand displacement by crRNA-DNA target strand hybridization enables the DNA non-target strand to enter the RuvC active site for first-strand DNA cutting. After cleavage and dissociation of the DNA NTS, the PAM-distal DNA strands must rearrange to position the DNA TS in the RuvC active site for second-strand cutting, producing the observed staggered DNA cleavage product [8].

Despite its compact size and unique consolidation of crRNA maturation and DNA cutting within a single RuvC active site [8], we found that the CasΦ protein architecture creates a natural

kinetic attenuation that enhances accuracy of RNA-guided dsDNA cleavage. Comparisons of CasΦ structures before and after dsDNA binding suggested that helix α7 of the CasΦ RecI domain constrains the active site, a prediction borne out by the finding that CasΦ variants containing helix α7 mutations (vCasΦ and nCasΦ) increase DNA cutting rates by up to ~18-fold relative to the wildtype enzyme. The increased kinetics, and a corresponding increase in the enzyme's tolerance of mismatches between the crRNA and the dsDNA substrate (**Fig. 6a**), support the conclusion that CasΦ evolved slow DNA cleavage kinetics for enhanced cutting accuracy. Notably, the amino acid sequence and primary structure of helix α7 is only weakly conserved among CasΦ homologs (**Supplementary Fig. 8**), suggesting ongoing evolution and potentially different cleavage kinetics within the CasΦ enzyme family.

Relationships between kinetics and accuracy have been established for other enzymes and processes, such as the kinetic proofreading model first proposed for protein synthesis [38–40]. Moreover, kinetic measurements showed that the fidelity of DNA cleavage by engineered high fidelity Cas9 (Cas9-HF1 [41]) and hyper-accurate Cas9 (HypaCas9 [42]) proteins is enhanced due to slower DNA cleavage rates that favor off-target DNA release [43]. In the case of CasΦ, however, the relationship between DNA binding kinetics and DNA cutting is not yet clear. We have shown that the rates of single-stranded DNA *trans*-cleavage catalyzed by wildtype CasΦ and the faster variants in the presence of matched versus mismatched ssDNA activators are comparable (**Fig. 6d**), while the rates of double stranded DNA cleavage appear to be different for mismatched substrates (**Fig. 6a**). This suggests that R-loop formation, which is not required for ssDNA binding, contributes to the DNA interference fidelity. A detailed analysis of cleavage kinetics in the presence of mismatched dsDNA substrates will be needed to determine the effects of reaction speed on both on-target dsDNA cutting and *trans*-ssDNA cutting accuracy.

The compact size of CasΦ makes it attractive for genome editing applications in human cells and plants based on the potential for tissue delivery [8]. However, variability of guide RNA efficiency currently limits the utility of CasΦ relative to larger enzymes such as Cas9 [8]. The

structural and mechanistic insights reported here should enable engineering of improved CasΦ variants for use in diagnostics and genome editing applications, as demonstrated for other Cas proteins [5,29–31,44]. In addition, the streamlined architecture of CasΦ offers exciting potential for creating fusion proteins with activities including base editing and transcriptional modulation that will provide hypercompact CRISPR-Cas systems for therapeutic and agricultural use in the future.

REFERENCES

1. Barrangou, R. *et al.* CRISPR provides acquired resistance against viruses in prokaryotes. *Science* **315**, 1709–1712 (2007).
2. 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).
3. Knott, G. J. & Doudna, J. A. CRISPR-Cas guides the future of genetic engineering. *Science* **361**, 866–869 (2018).
4. Swarts, D. C. & Jinek, M. Cas9 versus Cas12a/Cpf1: Structure-function comparisons and implications for genome editing. *Wiley Interdiscip. Rev. RNA* e1481 (2018).
5. Gleditzsch, D. *et al.* PAM identification by CRISPR-Cas effector complexes: diversified mechanisms and structures. *RNA Biol.* **16**, 504–517 (2019).
6. Westra, E. R., Dowling, A. J., Broniewski, J. M. & van Houte, S. Evolution and Ecology of CRISPR. *Annu. Rev. Ecol. Evol. Syst.* **47**, 307–331 (2016).
7. Al-Shayeb, B. *et al.* Clades of huge phages from across Earth's ecosystems. *Nature* **578**, 425–431 (2020).
8. Pausch, P. *et al.* CRISPR-CasΦ from huge phages is a hypercompact genome editor. *Science* **369**, 333–337 (2020).
9. Swarts, D. C. Stirring Up the Type V Alphabet Soup. *The CRISPR journal* vol. 2 14–16

(2019).

10. Jinek, M. *et al.* A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science* **337**, 816–821 (2012).
11. Jinek, M. *et al.* Structures of Cas9 endonucleases reveal RNA-mediated conformational activation. *Science* **343**, 1247997 (2014).
12. Nishimasu, H. *et al.* Crystal structure of Cas9 in complex with guide RNA and target DNA. *Cell* **156**, 935–949 (2014).
13. Anders, C., Niewoehner, O., Duerst, A. & Jinek, M. Structural basis of PAM-dependent target DNA recognition by the Cas9 endonuclease. *Nature* **513**, 569–573 (2014).
14. Zetsche, B. *et al.* Cpf1 is a single RNA-guided endonuclease of a class 2 CRISPR-Cas system. *Cell* **163**, 759–771 (2015).
15. Yamano, T. *et al.* Crystal Structure of Cpf1 in Complex with Guide RNA and Target DNA. *Cell* **165**, 949–962 (2016).
16. Burstein, D. *et al.* New CRISPR-Cas systems from uncultivated microbes. *Nature* **542**, 237–241 (2017).
17. Liu, J.-J. *et al.* CasX enzymes comprise a distinct family of RNA-guided genome editors. *Nature* **566**, 218–223 (2019).
18. Takeda, S. N. *et al.* Structure of the miniature type V-F CRISPR-Cas effector enzyme. *Mol. Cell* (2020) doi:10.1016/j.molcel.2020.11.035.
19. Karvelis, T. *et al.* PAM recognition by miniature CRISPR–Cas12f nucleases triggers programmable double-stranded DNA target cleavage. *Nucleic Acids Res.* **48**, 5016–5023 (2020).
20. Kluska, K., Adamczyk, J. & Krężel, A. Metal binding properties, stability and reactivity of zinc fingers. *Coord. Chem. Rev.* **367**, 18–64 (2018).
21. Yang, H., Gao, P., Rajashankar, K. R. & Patel, D. J. PAM-Dependent Target DNA Recognition and Cleavage by C2c1 CRISPR-Cas Endonuclease. *Cell* **167**, 1814–1828.e12

(2016).

22. Swarts, D. C., van der Oost, J. & Jinek, M. Structural Basis for Guide RNA Processing and Seed-Dependent DNA Targeting by CRISPR-Cas12a. *Mol. Cell* **66**, 221–233.e4 (2017).
23. Swarts, D. C. & Jinek, M. Mechanistic Insights into the cis- and trans-Acting DNase Activities of Cas12a. *Molecular Cell* vol. 73 589–600.e4 (2019).
24. Huang, X. *et al.* Structural basis for two metal-ion catalysis of DNA cleavage by Cas12i2. *Nat. Commun.* **11**, 5241 (2020).
25. Gorski, S. A., Vogel, J. & Doudna, J. A. RNA-based recognition and targeting: sowing the seeds of specificity. *Nat. Rev. Mol. Cell Biol.* **18**, 215–228 (2017).
26. Zhang, H., Li, Z., Xiao, R. & Chang, L. Mechanisms for target recognition and cleavage by the Cas12i RNA-guided endonuclease. *Nat. Struct. Mol. Biol.* (2020) doi:10.1038/s41594-020-0499-0.
27. Stella, S. *et al.* Conformational Activation Promotes CRISPR-Cas12a Catalysis and Resetting of the Endonuclease Activity. *Cell* **175**, 1856–1871.e21 (2018).
28. Wu, D., Guan, X., Zhu, Y., Ren, K. & Huang, Z. Structural basis of stringent PAM recognition by CRISPR-C2c1 in complex with sgRNA. *Cell Res.* **27**, 705–708 (2017).
29. Kleinstiver, B. P. *et al.* Engineered CRISPR-Cas9 nucleases with altered PAM specificities. *Nature* **523**, 481–485 (2015).
30. Kleinstiver, B. P. *et al.* Engineered CRISPR–Cas12a variants with increased activities and improved targeting ranges for gene, epigenetic and base editing. *Nat. Biotechnol.* **37**, 276–282 (2019).
31. Walton, R. T., Christie, K. A., Whittaker, M. N. & Kleinstiver, B. P. Unconstrained genome targeting with near-PAMless engineered CRISPR-Cas9 variants. *Science* **368**, 290–296 (2020).
32. Stella, S., Alcón, P. & Montoya, G. Structure of the Cpf1 endonuclease R-loop complex after target DNA cleavage. *Nature* **546**, 559–563 (2017).

33. Cofsky, J. C. *et al.* CRISPR-Cas12a exploits R-loop asymmetry to form double-strand breaks. *Elife* **9**, (2020).
34. Punjani, A. & Fleet, D. J. 3D variability analysis: Resolving continuous flexibility and discrete heterogeneity from single particle cryo-EM. *J. Struct. Biol.* **213**, 107702 (2021).
35. Chen, J. S. *et al.* CRISPR-Cas12a target binding unleashes indiscriminate single-stranded DNase activity. *Science* **360**, 436–439 (2018).
36. Harrington, L. B. *et al.* Programmed DNA destruction by miniature CRISPR-Cas14 enzymes. *Science* **362**, 839–842 (2018).
37. East-Seletsky, A. *et al.* Two distinct RNase activities of CRISPR-C2c2 enable guide-RNA processing and RNA detection. *Nature* vol. 538 270–273 (2016).
38. Hopfield, J. J. Kinetic proofreading: a new mechanism for reducing errors in biosynthetic processes requiring high specificity. *Proc. Natl. Acad. Sci. U. S. A.* **71**, 4135–4139 (1974).
39. Ninio, J. Kinetic amplification of enzyme discrimination. *Biochimie* **57**, 587–595 (1975).
40. Mallory, J. D., Igoshin, O. A. & Kolomeisky, A. B. Do We Understand the Mechanisms Used by Biological Systems to Correct Their Errors? *J. Phys. Chem. B* **124**, 9289–9296 (2020).
41. Kleinstiver, B. P. *et al.* High-fidelity CRISPR–Cas9 nucleases with no detectable genome-wide off-target effects. *Nature* **529**, 490–495 (2016).
42. Chen, J. S. *et al.* Enhanced proofreading governs CRISPR–Cas9 targeting accuracy. *Nature* **550**, 407–410 (2017).
43. Liu, M.-S. *et al.* Engineered CRISPR/Cas9 enzymes improve discrimination by slowing DNA cleavage to allow release of off-target DNA. *Nat. Commun.* **11**, 3576 (2020).
44. Strecker, J. *et al.* Engineering of CRISPR-Cas12b for human genome editing. *Nat. Commun.* **10**, 212 (2019).

ACKNOWLEDGEMENTS

We thank Emeric Charles, Carolyn Huang, Gavin Knott, Dylan Smock and members of the Doudna and Nogales laboratories for discussion. We thank Josh Cofsky and Gavin Knott for critical reading and comments on the manuscript. We thank Jonathan Remis, Daniel Toso, and Gavin Knott for electron microscopy assistance and Abhiram Chintangal for computational support. EM data were collected at the Cal-Cryo facility located at UC Berkeley. PP is supported by the NIH Somatic Cell Genome Editing consortium (NIH U01AI142817-02). CAT is supported by Campus Executive Grants 2101705 and 1655264 through Sandia National Laboratories. BAS was supported by an NSF Graduate Research Fellowship (No. DGE 1752814). JAD receives funding from the William M. Keck Foundation, the National Multiple Sclerosis Society, the NIH (RM1HG009490, U01AI142817-02), NSF (1817593). D.A.H. was supported by the EMBO (ALTF 1002-2018) and SNSF (P2BSP3_181878). JAD and EN are Howard Hughes Medical Institute Investigators.

AUTHOR CONTRIBUTIONS STATEMENT

P.P. conceived the study with input from J.A.D.. P.P. designed experiments and analyzed data. P.P. cloned constructs, purified proteins and performed biochemical experiments. K.M.S. and P.P. prepared cryo-EM grids. K.M.S. collected and processed cryo-EM data for 3D image reconstruction and 3DVA with input from E.N. and D.A.H.. P.P. built and refined structure models. C.A.T. provided materials. B.A.S. and J.F.B. provided the sequence information and bioinformatics analysis for CRISPR-CasΦ homologs, prior to publication of [7,8]. P.P. wrote the manuscript and prepared figures with input from K.M.S. and J.A.D.. The manuscript was reviewed and approved by all co-authors.

COMPETING INTERESTS STATEMENT

The Regents of the University of California, Berkeley have patents pending for CRISPR technologies on which the authors are inventors. J.A.D. is a co-founder of Caribou Biosciences, Editas Medicine, Intellia Therapeutics, Scribe Therapeutics, and Mammoth Biosciences. J.A.D. is a scientific advisory board member of Caribou Biosciences, Intellia Therapeutics, eFFECTOR Therapeutics, Scribe Therapeutics, Synthego, Inari Agriculture and Mammoth Biosciences. J.A.D. is a director at Johnson & Johnson and has sponsored research projects by Pfizer, Roche Biopharma, and Biogen. J.F.B. is a founder of Metagenomi. The remaining authors declare no competing interests.

FIGURE LEGENDS/CAPTIONS

Fig. 1: Structure of the crRNA-bound CasΦ poised for DNA recognition. a, Scheme illustrating the genomic locus and function of CRISPR-CasΦ. **b**, Above: Domain organization of CasΦ. Domain coloring is used throughout the manuscript. Purple hexagons highlight the position of the active site. Below: crRNA sequence and secondary structure. **c**, Cryo-EM maps of CasΦ-crRNA. The high resolution LocSpiral (colored surface), contoured at 11 σ (map level/root-mean-square deviation from zero (RMS)) and unfiltered cryoSPARC (translucent surface), contoured at 3.4 σ are shown, to highlight defined areas and flexible regions, respectively. **d**, Structure model of CasΦ-crRNA. **e**, Close-up view centered on the RuvC lid loop structure (dashed circle).

Fig. 2: Minimal domains mediate DNA recognition by CasΦ. a, R-loop organization scheme. **b**, Cryo-EM maps of CasΦ-crRNA:DNA. The LocSpiral (colored surface), contoured at 10 σ, and unfiltered cryoSPARC (translucent surface), contoured at 3.3 σ, are shown. **c**, Structure model of CasΦ-crRNA:DNA in two 90°-rotated orientations. **d**, Left: Close up view of the PAM base pairs in positions -2 (left panel), -1 (middle panel) and 0 (right panel). Amino acid side chains in proximity to the base pairs are shown as sticks. The LocSpiral map (colored surface) is shown as a translucent surface. Right: Filter binding assay testing for the ability of variant CasΦ proteins

to bind DNA (n = 3 independent reaction replicates; means ± SD). Raw data are shown in **Supplementary Fig. 2b**. Numerical source data for panel d are available online.

Fig. 3: DNA unwinding and target recognition activate CasΦ for DNA cutting. a, Structural alignment of an ideal B-form DNA duplex to the PAM-proximal DNA segment of the CasΦ ternary complex. Base pairs are not shown for clarity. **b**, Left: Close-up views onto the PI domain and RuvC helix α13, close to the ideal B-form DNA. Right: Filter binding assay testing for the ability of variant CasΦ proteins to bind DNA. (n = 3 independent reaction replicates; means ± SD). Raw data are shown in **Supplementary Fig. 2d**. **c**, Superimposition of the binary state (orange) and ternary state (blue) highlighting the rearrangement (arrows) of RecI and RecII. Structures were aligned via the RuvC β-strands 10 and 11 to visualize the rearrangement relative to the RuvC center. **d**, Left: Close up view of the RecII and RuvC domains superposition highlighting the lid loop rearrangement (bold arrow). Right: Close up view of the RuvC mixed β-sheet. The peptide backbones of parallel β-strands 10 and 13, including the side chains of E606 and D394 (inset), are shown as sticks. **e**, Filter binding assay testing for the ability of variant CasΦ proteins to bind DNA. (n = 3 independent reaction replicates; means ± SD). Active site mutants were assayed in the presence (left panel) and absence (right panel) of the RuvC Mg^{2+} cofactor. Raw data are shown in **Supplementary Fig. 2e,f**. Numerical source data for panels b and e are available online.

Fig. 4: Structure of CasΦ with a trapped substrate in the active site. a, R-loop organization scheme. Asterisks indicate the positions of phosphorothioate(PS)-DNA modifications. **b**, Cryo-EM maps of CasΦ-crRNA:PS-DNA in presence of Mg^{2+}. The LocSpiral (colored surface), contoured at 10 σ, and unfiltered cryoSPARC (translucent surface), contoured at 4.2 σ, are shown. **c**, Structure model of CasΦ-crRNA:PS-DNA + Mg^{2+} (cartoon). The PS-DNA nucleoside moieties are shown as sticks within the active site. Mg^{2+} cofactors are colored in magenta. **d**,

Close-up view on the PS-DNA substrate and Mg^{2+} cofactors within the RuvC active site. The LocSpiral map (blue mesh) is shown contoured at 7 σ. **e**, Overview of the RuvC active site. Only amino acid side chains close to the two Mg^{2+} cofactors are shown for clarity. **f**, DNA-cleavage mechanism. Waters that contribute to the expected octahedral coordination sphere of the two magnesium cofactors are not shown.

Fig. 5: Helix α7 of the RecI domain regulates substrate accessibility of the RuvC. a, Close up on the nucleic acids and α7 above the RuvC active site. The bold line highlights the steric α7 barrier, blocking the TS path (arrow). Nucleic acid backbones are shown as bands. **b**, Detailed view of helix α7 as seen from the RuvC. Biochemically analyzed residues are shown as sticks. Arrows indicate directions of the nucleic acids. **c**, TS (dark blue curve) and NTS (light blue curve) cleavage efficiency of variant CasΦ. Derived reaction rate constants are shown above the X-axis. (n = 3 independent reaction replicates; means ± SD). Raw data are shown in **Supplementary Fig. 7**. Numerical source data for panel c are available online.

Fig. 6: Helix α7 adjusts fidelity and can be engineered for sensitive nucleic acid detection. a, crRNA:TS duplex base pair mismatch assay for WT (orange bars) and vCasΦ (blue bars). (n = 3 independent reaction replicates; means ± SD). Raw data are shown in **Supplementary Fig. 10**. **b**, Left: Scheme illustrating the *in vitro* nucleic acid detection FQ assay. Right: FQ assay for detection of 2 nM ssDNA-activator by WT and engineered CasΦ. (n = 3 independent reaction replicates; means ± SD). **c**, FQ-assay for detection of pico-molar ssDNA-activator concentrations by WT and nCasΦ. (n = 3 independent reaction replicates; means ± SD). **d**, Mismatch-FQ-assay probing the FQ-reporter cleavage in presence of activators with single (left) and double (right) mismatches. Only end-point data at time = 2 h are shown. Graphs of the full time course for the various mismatches are shown in **Supplementary Figs. 11 and 12**. **e**, Model for CasΦ mediated DNA interference. Numerical source data for panels a-d are available online.

TABLES

Table 1: Cryo-EM data collection, refinement and validation statistics

	CasΦ ternary (EMD- 23600, PDB 7LYS)	**CasΦ ternary + Mg^{2+}** (EMD- 23601, PDB 7LYT)	**CasΦ binary** (EMD- 23678, PDB 7M5O)
Data collection and processing			
Magnification	36,000	36,000	36,000
Voltage (kV)	200	200	200
Electron exposure (e^{-}/Å^2)	50	50	50
Defocus range (μm)	0.8 - 2	0.5 - 1.8	1 - 2.2
Pixel size (Å)	1.115	1.115	1.115
Symmetry imposed	*C1*	*C1*	*C1*
Initial particle images (no.)	2,666,893	1,059,466	2,349,114
Final particle images (no.)	396,531	410,553	298,646
Map resolution (Å)	3.05	2.84	3.54
FSC threshold	0.143	0.143	0.143
Refinement			
Initial model used	*Ab initio*	7LYS	7LYS
Model resolution (Å)	3.3	3.2	4.1
FSC threshold	0.5	0.5	0.5
Model composition			
Non-hydrogen atoms	13575	14355	11649
Protein residues	669	712	655
Nucleotide	89	92	37
Ligands	0	SC 3, ZN 1, MG 2	ZN 1
B factors (Å^2)			
Protein	85.54	69.97	126.55
Nucleotide	56.28	45.66	91.86
Ligand	n/a	76.81	168.51
R.m.s. deviations			
Bond lengths (Å)	0.003	0.003	0.003
Bond angles (°)	0.535	0.539	0.537
Validation			
MolProbity score	1.32	1.38	1.6
Clashscore	5.82	6.95	10.22
Poor rotamers (%)	0.17	0.33	1.25
Ramachandran plot			
Favored (%)	100	100	99.85
Allowed (%)	0	0	0.15
Disallowed (%)	0	0	0
Rama-Z score, whole (r.m.s. Rama-Z)	-1.46 (0.3)	-1.59 (0.29)	-2.03 (0.31)
Map CC (box)	0.83	0.82	0.79
Map CC (mask)	0.81	0.79	0.75

METHODS

Generation of *casΦ-2* mutant expression vectors

Plasmids were cloned and mutagenized via Golden Gate assembly as previously described [8]. In brief, the pRSFDuet-1 derived *casΦ-2* overexpression vector pPP085 [8] (Addgene #158795) was amplified around the horn using primers containing the desired mutation and AarI Golden Gate cloning sites. The resulting fragment was circularized using the restriction enzyme AarI (Thermo Fisher Scientific) and T4 ligase (NEB). Plasmids were propagated in *Escherichia coli* MachI (QB3-Macrolab, UC Berkeley). Generated plasmids (Supplementary Tab. 1) were sequenced across the coding sequence of *casΦ-2.*

CasΦ-2 protein production and purification

C-terminally hexa-histidine-tagged CasΦ-2 was produced by heterologous expression in *E. coli* and purified as previously described [8]. In brief, overexpression plasmids were transformed into *E. coli* BL21(DE3)-Star (QB3-Macrolab, UC Berkeley). Expression cultures were grown shaking vigorously at 37 °C in 1.5 L TB-Kan (50 µg/mL Kanamycin) media to an OD_{600} of 0.6. Subsequently, cultures were cooled down on ice for 15 min and gene expression was induced with 0.5 mM IPTG before incubation overnight at 16 °C. Cells were harvested by centrifugation and resuspended in wash buffer (50 mM HEPES-Na pH 7.5 RT, 1 M NaCl, 20 mM imidazole, 5 % glycerol and 0.5 mM TCEP), subsequently lysed by sonication, followed by lysate clarification by centrifugation. The soluble fraction was loaded on a 5 mL Ni-NTA Superflow Cartridge (Qiagen) pre-equilibrated in wash buffer. Bound proteins were washed with 20 column volumes (CV) wash buffer and subsequently eluted in 4 CV elution buffer (50 mM HEPES-Na pH 7.5 RT, 500 mM NaCl, 500 mM imidazole, 5% glycerol and 0.5 mM TCEP). The eluted proteins were concentrated to 1-2 mL before injection into a HiLoad 16/600 Superdex 200pg column (GE

Healthcare) pre-equilibrated in size-exclusion chromatography (SEC) buffer (20 mM HEPES-Na pH 7.5 RT, 500 mM NaCl, 5 % glycerol and 0.5 mM TCEP). Peak fractions were concentrated to 1 mL and concentrations were determined based on the absorbance at 280 nm using a NanoDrop 8000 Spectrophotometer (Thermo Scientific). Proteins were purified at a constant temperature of 4 °C and concentrated proteins were kept on ice to prevent aggregation, snap frozen in liquid nitrogen and stored at -80 °C.

Binary and ternary complex reconstitution for cryo-EM

CasΦ-2 was produced as described above. crRNA (rPP012; (Supplementary Tab. 2)) was ordered as a synthetic RNA oligonucleotide from IDT (Integrated DNA Technologies) and dissolved in DEPC treated ddH_20 to a concentration of 0.5 mM. Subsequently, the crRNA was heated to 65 °C for 3 min and cooled down to RT to allow for hairpin formation. DNA oligonucleotides (Supplementary Tab. 3) were designed to contain a non-complementary protospacer segment to produce “bubbled” substrates and facilitate rapid R-loop formation during ternary complex reconstitution. Oligonucleotides were ordered from and synthesized by IDT. DNA oligonucleotides were combined in a 1:1.2 molar ratio (Target strand:non-target strand) and annealed to a final DNA-duplex concentration of 0.5 mM in hybridization buffer (10 mM Hepes-Na pH 7.5 RT, 150 mM NaCl) by heating for 5 min at 95 °C and a subsequent slow cool down in a thermocycler.

The CasΦ-2 binary complex was reconstituted by incubation of 20 µM CasΦ-2 and 24 µM crRNA for 10 min at RT in a total volume of 250 µL SEC buffer. For formation of the ternary complex, 40 µM DNA-duplex was added to the assembly reaction after formation of the binary complex and the sample was incubated for an additional 10 min at RT. Subsequently, assembly reactions were injected into a Superdex 200 10/300 GL column (GE Healthcare) pre-equilibrated in low salt (LS) buffer (10 mM Hepes-Na pH 7.5, 150 mM NaCl) at 4 °C to separate complexes from excess nucleic acids. Peak fractions were concentrated to 250 µL at 4 °C and concentrations

were estimated using the absorbance at 280 nm as measured on a Nanodrop 8000 Spectrophotometer (Thermo Scientific), based on the absorbance at 280 nm of a sample containing the individual components mixed in the expected molecular ratio at a known concentration. Assembled complexes were kept on ice to prevent aggregation.

Electron microscopy grid preparation and data collection

CasΦ-2 complexes were frozen using an FEI Vitrobot Mark IV cooled to 8 °C at 100% humidity. The ternary complex was frozen at a concentration of 9 μM on carbon 1.2/1.3 400 mesh C-flat grids (Electron Microscopy Sciences #CF413-50), which were glow discharged at 15 mA for 25 sec prior to sample application using a PELCO easyGLOW. 4 μL of sample was applied to the grid and immediately blotted for 5 sec with blot force 8. The ternary complex in the presence of magnesium was prepared by addition of $MgCl_2$ to a final magnesium concentration of 5 mM and subsequently incubated for 15 min at RT before application onto 1.2/1.3 300 mesh UltrAuFoil gold grids (Electron Microscopy Sciences #Q350AR13A). Grids were glow discharged with 15 mA for 40 sec, 4 μL of sample was applied and blotted for 5 sec with blot force 8. The binary complex was frozen at a concentration of 3.7 μM on 1.2/1.3 300 mesh UltrAuFoil gold grids (Electron Microscopy Sciences #Q350AR13A), which were glow discharged with 15 mA for 25 sec. 4 μL of sample were applied to grids and grids were immediately blotted for 4 sec with blot force 8. Micrographs for all data sets were collected on a Talos Arctica operated at 200 kV and 36,000x magnification (1.115 Å pixel size), using the super resolution camera setting (0.5575 Å pixel size) on a K3 Direct Electron Detector. Cryo-EM data was collected using SerialEM v. 3.8.7 software. Images were collected using beam shift.

Single particle cryo-EM data processing and 3D volume reconstruction

For the binary complex, 2,934 movies in super resolution were collected with defocus ranging from -0.7 to -1.8 μm. Movies were processed in cryoSPARC (Structura Biotechnology

Inc.) and resulted in a map with a preferred orientation. In order to overcome this problem a tilted data set with a 20° tilt was collected. The tilted data set was processed in cryoSPARC v3.1, independently from the untitled dataset and resulted in the final map and structure presented in this study (**Extended Data Fig. 1**). For the tilted data set, 1,641 movies were collected in super resolution with defocus ranging from -1.0 to -2.2 µm. Movies were corrected for beam induced motion with patch motion and manually curated. After curation, 1,233 micrographs were used for CTF parameter calculation with patch CTF. Then, particles were picked with a blob picker from all 1,233 micrographs and 2,349,114 particles were extracted with bin factor 2. After one round of 2D classification, 435,648 particles were selected and re-extracted with re-centering and then used for *ab initio* reconstruction with 2 classes. Particles (298,679) from the best class were again re-extracted with re-centering. After local motion correction in cryoSPARC, the corrected particles were downsampled using bin factor 2 and the data sign was flipped. These particles were used to refine the best *ab initio* class using non-uniform refinement [45] with per group CTF optimization including beam tilt and trefoil fitting. The resulting map reached 3.54 Å resolution. Half maps from this refinement were used for LocSpiral [46] enhancement of weaker density regions. The LocSpiral map was used for model building.

For the ternary complex (**Extended Data Fig. 4**), 4,374 movies were collected with defocus ranging from -0.8 to -2 µm. Data processing was performed with cryoSPARC v3.0.1. Movies were corrected for beam induced motion using patch motion, and CTF parameters were calculated using patch CTF. After micrograph curation, 4,090 micrographs were included for further data analysis. 20 micrographs were used for particle picking with blob picker for preliminary 2D classification and template selection for Topaz particle training [47]. The resulting Topaz model was used to pick particles from all micrographs. 2,666,893 particles were extracted from micrographs with binning factor 2 and used for further analysis. Two rounds of 2D classification were performed, leading to a set of 1,733,254 selected particles. These particles were re-extracted with re-centering and used for *ab initio* reconstruction with 5 classes. The classes were

evaluated visually and the class with the best defined densities for all domains and the largest number of particles (819,071 particles) was chosen and used as a seed for 3 classes in heterogeneous refinement performed on a set of particles from this class. Upon visual inspection, a class from heterogeneous refinement with the highest number of particles (419,818) and 3.64 Å resolution was chosen for map refinement. Particles from this class were again re-extracted with re-centering and duplicate particles were removed (23,287 particles removed). We performed a homogeneous refinement and non-uniform refinement with per group CTF correction without spherical aberration fitting and without tetrafoil fit and obtained volumes with resolutions of 3.13 Å, and 3.05 Å, respectively. We chose the 3.05 Å resolution map as our final map. Half maps from the final volume were used in LocSpiral in order to enhance weaker map regions. The LocSpiral volume was used for model building. For 3D variability analysis with cryoSPARC v3.1, the best class from *ab initio* reconstruction (819,071 particles) was used. First, the map from this class was homogeneously refined with particles from this class. The output from this refinement was used for 3DVA [48] with 3 modes of motion. The 3 modes of motion were visualized by 3DVA simple display with 20 frames each (two frames from two modes of motion are included in **Extended Data Fig. 10)**.

For the ternary complex in the presence of magnesium (**Extended Data Fig. 8**), 4,827 micrographs were collected with -0.5 to -1.8 µm defocus range. Micrographs in super resolution were corrected for beam induced motion with patch motion, and CFT was estimated using patch CTF in cryoSPARC v2.15. After curation, 1,962 micrographs were retained for further analysis. An initial set of particles was picked using Blob picker from a small subset of micrographs and used for initial 2D classification and template selection for Topaz particle picking training. The optimized Topaz picking model was used to pick particles from all micrographs. A total of 1,059,466 particles were extracted from micrographs with pixel size binning factor 2 and used for 2D classification. Promising classes with 679,252 particles were selected for *ab initio* reconstruction using 3 classes. One class volume (410,553 particles) appeared to correspond to

the full-length complex and was used for homogeneous refinement, resulting in a map at 2.87 Å resolution. The homogeneous refinement output was used for non-uniform refinement, which improved the resolution to 2.84 Å. Half maps from non-uniform refinement reconstruction were used for LocSpiral map enhancement in order to improve regions with weaker density. The LocSpiral map was used for model building. Outputs from homogenous refinement were used for 3D variability analysis (3DVA) with 5 motion modes. The 5 modes of motion were visualized by 3DVA simple display with 20 frames each. Two frames from each mode of motion are shown in the supplement (**Extended Data Fig. 10)**.

Model building and refinement

The model of the CasΦ-2 ternary complex was built *de novo* in Coot [49] using LocSpiral improved and unfiltered cryoSPARC maps, and refined using the Real-space refinement tool [50] as implemented in Phenix (Version 1.18, [51]). The CasΦ-2 binary complex structure was modelled into the respective LocSpiral map, based on the ternary structure of CasΦ using rigid body fit and Real-space refinement as implemented in Coot and refined using the Real-space refinement tool as implemented in Phenix. The model of CasΦ-2 bound to phosphorothioate-modified DNA in the presence of magnesium was built based on the CasΦ-2 ternary model into the respective LocSpiral map, and refined using the Real-space refinement tool as implemented in Phenix. 2-deoxy-cytidine-5'-thiophosphorate (SC ligand) restraints were generated using Elbow [52] as implemented in Phenix. 3'-oxygen to 5'-phosphate bond angles involving phosphorothioate-modified DNA, as well as distances between the two magnesium cofactors and D394/D695, were restrained using custom geometry restraints in the Real-space refinement tool as implemented in Phenix. Custom angle geometry restraints were estimated by measuring the angles of 3'-oxygen to 5'-phosphate bonds observed for unmodified DNA, as refined in the corresponding cryoEM map using the Phenix implemented Real-space refinement tool and PyMol (pymol.org). For the binary model refinement, distances between the zinc cofactor and the surrounding cysteine

residues were restrained based on distances measured for the ternary model. Models were refined using atomic displacement parameters, global minimization and local grid search refinement strategies as implemented in the Phenix Real-space refinement tool. The models were additionally refined using secondary structure, rotamer and ramachandran restraints as implemented in the Phenix Real-space refinement tool. Prior to the final refinement, hydrogens were added to the models using ReadySet as implemented in Phenix and local_grid_search refinement was disabled.

Data deposition and figure preparation

Cryo-EM maps and model coordinates were deposited to the EMDB (codes EMD-23600, EMD-23601 and EMD-23678) and PDB (codes 7LYS, 7LYT and 7M5O). Figures were prepared in UCSF ChimeraX [53], UCSF Chimera [54] and Coot [49] . Cryo-EM map σ-levels were calculated as: map level/root-mean-square deviation from zero (RMS). Orientation distribution plot was generated using pyem csparc2star.py and star2bild.py programs [55]. Map vs. model FSC graphs were calculated in Mtriage, as implemented in Phenix [56]. Gold standard FSC and map vs model FSC were replotted using Prism 8 (GraphPad).

RNP complex reconstitution for DNA cleavage and binding assays

CasΦ-2 was produced as described above. crRNA guides (Supplementary Tab. 2) were ordered as synthetic RNA oligonucleotides from IDT (Integrated DNA Technologies) and dissolved in DEPC treated dd$H_2$0 to a concentration of 0.5 mM. Subsequently, the crRNA was heated to 65 °C for 3 min and cooled down to RT to allow for hairpin formation. CasΦ-2 RNP complexes were reconstituted at a concentration of 10 µM by incubation of 10 µM CasΦ-2 and 12 µM crRNA for 10 min at RT in 2x cleavage buffer (2xCB) (20 mM Hepes-Na pH 7.5, 300 mM KCl, 10 mM $MgCl_2$, 20 % glycerol, 1 mM TCEP). Formed RNPs were aliquoted to a volume of 10

µL, flash frozen in liquid nitrogen and stored at -80 °C. Before usage, RNP aliquots were thawed on ice.

DNA cleavage and mismatch tolerance assay

DNA targets (Supplementary Tab. 4) were cloned as previously described [8]. Mutations (A->C, T->G, C->A, G->T) were introduced in the target region by Golden Gate mutagenesis as described above. DNA targets were produced by PCR as previously described [8] and diluted to a final concentration of 20 nM in CB buffer (10 mM Hepes-Na pH 7.5, 150 mM KCl, 5 mM $MgCl_2$, 10 % glycerol, 0.5 mM TCEP). CasΦ-2 RNPs were prepared as described above with either oligonucleotide rPP012 (20 nt spacer) for variant protein cleavage assays, or rPP013 (22 nt spacer) for mismatch tolerance assays. RNPs were diluted on ice in CB buffer to a final concentration of 2 µM. Reactions were initiated by addition of the RNP (1 µM final) to the target DNA (10 nM final) and incubation at 37 °C in a thermocycler for 1 h. Subsequently 0.08 units/µL Proteinase K (NEB) were added and the reaction was incubated for 30 min at 37 °C in a thermocycler. Loading dye (Gel Loading Dye Purple 6X, NEB) was added and samples (10 µL) were analyzed by electrophoresis on an 1% agarose gel, stained with SYBR Safe (Thermo Fisher Scientific). For the cleavage assay involving alternative PAMs and supercoiled DNA, 2 µL samples were separated on an 1% agarose gel, stained with SYBR Gold (Thermo Fisher Scientific). Gels were imaged using a ChemiDoc MP and Image Lab v5.2.1 software (Biorad). Substrate and product intensities were quantified using ImageJ (Version 1.51, https://imagej.nih.gov/ij/) [57] and the cleaved fraction was calculated as the product intensity sum divided by the combined substrate and product intensity sum. Data was plotted using Prism 8 (Graphpad).

Filter binding assay

Non-target DNA strands (Supplementary Tab. 3) were 5′-end-labelled using T4-PNK (NEB) in the presence of ^{32}P-γ-ATP to track binding of the PAM-proximal DNA segment. Oligo duplex targets were generated by combining ^{32}P-labelled and unlabelled complementary oligonucleotides in a 1:1.5 molar ratio and annealed to a final duplex concentration of 50 nM in hybridization buffer (10 mM Hepes-Na pH 7.5 RT, 150 mM NaCl) by heating for 5 min at 95 °C and a subsequent slow cool down in a heating block. Binding reactions were initiated by combining CasΦ-2 with 0.1 nM DNA duplex in CB buffer. Reactions were subsequently incubated for 1 h at 37 °C in a thermocycler. For the binding experiment in the absence of magnesium, 12.5 mM EDTA was supplemented in CB buffer (w/o magnesium) throughout the experiment. Binding reactions were subsequently blotted at RT using a 96-well dot blot apparatus (Minifold I, S&S) assembled from below to top with two layers Whatman 1 filter paper (Whatman), one layer Amersham Hibond-N (GE Healthcare), one layer Amersham Protran 0.1 µm NC (GE Healthcare) and one layer HT-450 Tuffryn (PALL Life Sciences). Membranes and filter paper were pre-equilibrated for 1 h in CB buffer prior to assembly of the dot blot apparatus. Membranes were washed with 50 µL CB buffer per well prior to application of 30 µL CasΦ-2 binding reaction at an approximate flow rate of 15 µL/min . Bound proteins and nucleic acids were subsequently washed with 100 µL CB buffer. Membranes were air dried before phosphor-imaging visualization using an Amersham Typhoon scanner, v2.0.0.6 firmware version 208 (GE Healthcare). Assessment of the HT-450 Tuffryn membrane showed no aggregated fraction, containing the radiolabeled substrate. Spot intensities were quantified using ImageQuant TL 8.1 (Cytivia) and the bound DNA fraction for each spot was calculated relative to the mean (n = 3) of the intensities observed at the maximum concentration of CasΦ-2 WT RNP. Curves were fitted using a sigmoidal four parameter logistic curve model in Prism 8 (Graphpad). Of note, “apparent K_d’s” were not derived from the data, since the assay was performed under non-equilibrium conditions using a final 100 µL wash step with CB buffer w/o DNA.

In vitro cleavage assays - radiolabeled nucleic acids

CasΦ RNP were assembled as described above. Substrates were 5′-end-labelled using T4-PNK (NEB) in the presence of ^{32}P-γ-ATP (Substrate sequences are given in Supplementary Tab. 3). Oligonucleotide-duplex targets were generated by combining ^{32}P-labelled and unlabelled complementary oligonucleotides in a 1:1.5 molar ratio. Oligos were hybridized to a DNA-duplex concentration of 50 nM in hybridization buffer (10 mM Hepes-Na pH 7.5 RT, 150 mM NaCl), by heating for 5 min to 95 °C and a slow cool down to RT in a heating block. Cleavage reactions were initiated by combining 200 nM RNP with 2 nM substrate in CB buffer and subsequently incubated at 37 °C. Reactions were stopped by addition of two volumes formamide loading buffer (96 % formamide, 100 µg/mL bromophenol blue, 50 µg/mL xylene cyanol, 10 mM EDTA, 50 µg/mL heparin), heated to 95 °C for 5 min, and cooled down on ice before separation on a 12.5 % denaturing urea-PAGE. Gels were dried for 4 h at 80 °C before phosphor-imaging visualization using an Amersham Typhoon scanner, v2.0.0.6 firmware version 208 (GE Healthcare). Bands were quantified using ImageQuant TL 8.1 (Cytivia) and the cleaved fraction was calculated as the product intensity sum divided by the combined substrate and product intensity sum. Curves were fitted to a One-Phase-Decay model in Prism 8 (graphpad) to derive the rate of cleavage.

Fluorophore quencher assay

CasΦ RNP were assembled as described above. Reactions were initiated by combining 100 nM RNP (100 nM CasΦ, 120 nM rPP012 crRNA), 100 nM DNase Alert (IDT) FQ probe, with and without activator ssDNA (Supplementary Tab. 3) in cleavage buffer (10 mM Hepes-Na pH 7.5, 150 mM KCl, 5 mM $MgCl_2$, 10 % glycerol, 0.5 mM TCEP) in a 384 well flat bottom black polystyrene assay plate (#3820, Corning). Three replicates for each reaction were monitored (λ_{ex}: 530 nm; λ_{ex}: 590 nm) in a Cytation 5 plate reader (BioTek, software Gen v3.04) at 37 °C every 1.5 min for the activator titration experiment. For the FQ-mismatch-assay, 2 nM activator oligonucleotides were used in singlicates. The data were background-subtracted using the mean

values of the measurements taken for three no-activator controls at the respective time point. To derive the % relative fluorescence, the values were calculated for each variant as the ratio of the intensity measured for the respective mismatched activator divided by the intensity measured for the no-mismatch activator. Data were plotted in Prism 8 (graphpad).

Data Availability

The cryo-EM maps and model coordinates have been deposited to the EMDB (codes EMD-23600, EMD-23601 and EMD-23678) and wwPDB (codes PDB 7LYS, PDB 7LYT and PDB 7M5O), respectively. Source data are available online.

METHODS REFERENCES

45. Punjani, A., Zhang, H. & Fleet, D. J. Non-uniform refinement: adaptive regularization improves single-particle cryo-EM reconstruction. *Nat. Methods* **17**, 1214–1221 (2020).
46. Kaur, S. *et al.* Local computational methods to improve the interpretability and analysis of cryo-EM maps. *Cold Spring Harbor Laboratory* 2020.05.11.088013 (2020) doi:10.1101/2020.05.11.088013.
47. Bepler, T. *et al.* Positive-unlabeled convolutional neural networks for particle picking in cryo-electron micrographs. *Nat. Methods* **16**, 1153–1160 (2019).
48. Punjani, A. & Fleet, D. J. 3D Variability Analysis: Resolving continuous flexibility and discrete heterogeneity from single particle cryo-EM. *Cold Spring Harbor Laboratory* 2020.04.08.032466 (2021) doi:10.1101/2020.04.08.032466.
49. Emsley, P. & Cowtan, K. Coot: model-building tools for molecular graphics. *Acta Crystallogr. D Biol. Crystallogr.* **60**, 2126–2132 (2004).
50. Afonine, P. V. *et al.* Real-space refinement in PHENIX for cryo-EM and crystallography. *Acta Crystallogr D Struct Biol* **74**, 531–544 (2018).
51. Liebschner, D. *et al.* Macromolecular structure determination using X-rays, neutrons and electrons: recent developments in Phenix. *Acta Crystallogr D Struct Biol* **75**, 861–877 (2019).
52. Moriarty, N. W., Grosse-Kunstleve, R. W. & Adams, P. D. electronic Ligand Builder and Optimization Workbench (eLBOW): a tool for ligand coordinate and restraint generation. *Acta Crystallogr. D Biol. Crystallogr.* **65**, 1074–1080 (2009).
53. Pettersen, E. F. *et al.* UCSF ChimeraX: Structure visualization for researchers, educators, and developers. *Protein Sci.* **30**, 70–82 (2021).
54. Pettersen, E. F. *et al.* UCSF Chimera?A visualization system for exploratory research and analysis. *Journal of Computational Chemistry* vol. 25 1605–1612 (2004).

55. Asarnow, D., Palovcak, E. & Cheng, Y. *asarnow/pyem: UCSF pyem v0.5*. (2019). doi:10.5281/zenodo.3576630.

56. Afonine, P. V. *et al.* New tools for the analysis and validation of cryo-EM maps and atomic models. *Acta Crystallogr D Struct Biol* **74**, 814–840 (2018).

57. Schneider, C. A., Rasband, W. S. & Eliceiri, K. W. NIH Image to ImageJ: 25 years of image analysis. *Nat. Methods* **9**, 671–675 (2012).

58. Madeira, F. *et al.* The EMBL-EBI search and sequence analysis tools APIs in 2019. *Nucleic Acids Res.* **47**, W636–W641 (2019).

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

60. Laskowski, R. A., Jabłońska, J., Pravda, L., Vařeková, R. S. & Thornton, J. M. PDBsum: Structural summaries of PDB entries. *Protein Sci.* **27**, 129–134 (2018).