Nucleic Acids Research, 2024, **52**, 462–473
https://doi.org/10.1093/nar/gkad1102
Advance access publication date: 30 November 2023
Structural Biology

Substrate selectivity and catalytic activation of the type III CRISPR ancillary nuclease Can2

Kenny Jungfer, Annina Sigg and Martin Jinek*,†

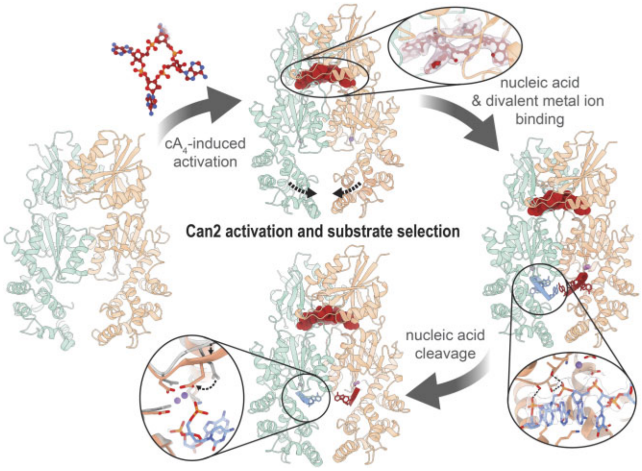
Department of Biochemistry, University of Zurich, Zurich, Switzerland

*To whom correspondence should be addressed. Tel: +41 44 635 5572; Fax: +41 44 63 56805; Email: jinek@bioc.uzh.ch
†Lead contact.

Abstract

Type III CRISPR-Cas systems provide adaptive immunity against foreign mobile genetic elements through RNA-guided interference. Sequence-specific recognition of RNA targets by the type III effector complex triggers the generation of cyclic oligoadenylate (cOA) second messengers that activate ancillary effector proteins, thus reinforcing the host immune response. The ancillary nuclease Can2 is activated by cyclic tetra-AMP (cA_4); however, the mechanisms underlying cA_4-mediated activation and substrate selectivity remain elusive. Here we report crystal structures of *Thermoanaerobacter brockii* Can2 (TbrCan2) in substrate- and product-bound complexes. We show that TbrCan2 is a single strand-selective DNase and RNase that binds substrates via a conserved SxTTS active site motif, and reveal molecular interactions underpinning its sequence preference for CA dinucleotides. Furthermore, we identify a molecular interaction relay linking the cA_4 binding site and the nuclease catalytic site to enable divalent metal cation coordination and catalytic activation. These findings provide key insights into the molecular mechanisms of Can2 nucleases in type III CRISPR-Cas immunity and may guide their technological development for nucleic acid detection applications.

Graphical abstract


cA4-induced activation
nucleic acid & divalent metal ion binding
Can2 activation and substrate selection
nucleic acid cleavage

Introduction

Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR) and CRISPR-associated (Cas) systems provide bacteria and archaea with adaptive immunity against invading mobile genetic elements (MGEs) through CRISPR RNA (crRNA)-guided interference (1,2). CRISPR-Cas systems are categorized into two main classes (class 1 and class 2) that can be further subdivided into six types: type I, III and IV in class 1, and type II, V and VI in class 2 systems (3). In contrast to class 2 systems that deploy single protein effectors during interference, class 1 systems comprise large, multisubunit complexes composed of several Cas proteins that perform different functions. The modularity of class 1 loci gives rise to unique interference mechanisms, which is particularly prevalent in type III CRISPR-Cas immune responses.

During type III CRISPR-Cas interference, the Csm (type III-A, -D and -E) or Cmr (type III-B and -C) effector complex recognizes and degrades foreign-derived RNAs complementary to their crRNA guides (4–6). Binding of cognate target RNAs further activates the Cas10 (Csm1/Cmr2) subunit of the Csm/Cmr complex, which uses its HD nuclease domain to non-specifically degrade single-stranded DNA (ssDNA) (7,8). In turn, the Palm (or cyclase) domain of activated Cas10 converts ATP to 3′-5′ cyclic oligoadenylate (cOA) nucleotides (9,10). The produced cOAs of system-specific length, commonly cyclic tri-(cA_3), tetra-(cA_4) or

Received: July 26, 2023. **Revised:** October 26, 2023. **Editorial Decision:** October 27, 2023. **Accepted:** November 9, 2023

© The Author(s) 2023. Published by Oxford University Press on behalf of Nucleic Acids Research.
This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial License (http://creativecommons.org/licenses/by-nc/4.0/), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited. For commercial re-use, please contact journals.permissions@oup.com

hexa-adenylates (cA_6), act as second messengers to stimulate downstream interference pathways by allosterically activating type III CRISPR-Cas ancillary effector proteins that contain CRISPR-associated Rossman-fold (CARF) or SMODS-associated (SAVED) regulatory domains fused to various effector domains (11–13). A broad spectrum of type III-associated CARF and SAVED domain-encoding proteins have been identified that additionally contain effector domains including nucleases, proteases, and transcription factors (14–16).

The Csm6/Csx1 family of nucleases constitutes the best studied class of ancillary type III effectors. Csm6-like nucleases form homodimeric complexes that harbor C-terminal higher eukaryotic and prokaryotic nuclease (HEPN) domains that non-specifically cleave RNAs in a metal-independent manner upon cOA binding (17–19). The stimulation of Csm6/Csx1 signaling cascades contributes to the inhibition of phage or plasmid proliferation through the induction of cell dormancy or abortive infection. Some Csm6/Csx1 proteins also have an intrinsic ring nuclease activity, thereby limiting their cOA-dependent activation and self-toxicity (20–22). The immune signal amplification afforded by the target-dependent generation of cOAs and induction of Csm6/Csx1 ribonucleases instigated the development of RNA detection tools for diagnostic applications (23–25). However, initial approaches failed to reach sufficient sensitivity in the absence of a prior amplification step, partly due to the autocatalytic inactivation of Csm6. An alternative RNA detection technology with substantially improved detection sensitivity has recently been developed by implementing a new RNA capture approach in combination with the replacement of Csm6 with the CRISPR-associated nuclease 2 (Can2), which lacks ring nuclease activity (26).

The Can2 family is a widely distributed class of type III CRISPR-Cas ancillary effector nucleases that contain PD-(D/E)xK nuclease domains with similarities to type II restriction endonucleases (REase) (27,28). Can2 form homodimeric complexes that bind a single cOA molecule in a central cavity formed at the dimer interface of the CARF and nuclease domains. Binding to cA_4, but not to cA_6, leads to conformational changes that activate the C-terminal nuclease domain, which results in the degradation of oligonucleotide substrates, conferring immunity against bacteriophages and invading plasmids through growth arrest (28). It remains unclear as to how the cA_4-induced conformational changes mediate nuclease activation. Interestingly, different Can2 variants exhibit varying nuclease substrate-specificities (26–28). The Can2-like nuclease from *Treponema succinifaciens* TresuCARD1 specifically degrades single-stranded DNA and RNA, whereas Can2 enzymes from *Thioalkalivibrio sulfidiphilus* (TsuCan2) and *Sulfobacillus thermosulfidooxidans* (SthCan2), as well as an archaeal Can2 from Archaeoglobi archaeon JdFR-42 (AaCan2), degrade single-stranded RNA and nick supercoiled plasmid DNA (26–28). Moreover, Can2 nuclease activity exhibits a certain degree of cleavage preference for specific oligonucleotide sequences. AaCan2 was shown to preferentially degrade substrates at poly-A and poly-T sites, whereas TresuCARD1 cleaved upstream of T-(G/A) dinucleotide sequences. However, the molecular basis for substrate preferences in Can2 nucleases remains unclear (26,28).

Here, we present multiple crystal structures of the cA_4-dependent type III-D ancillary nuclease Can2 from *Thermoanaerobacter brockii* (TbrCan2) in complexes with DNA substrates and products, revealing intricate molecular mechanisms that underpin substrate recognition. Complementing our structural observations with biochemical assays, we identify molecular determinants that account for the substrate selectivity for CA dinucleotide motifs. Furthermore, we reveal the catalytic activation mechanism of Can2 nucleases involving an allosteric relay from the cA_4 binding site to the nuclease catalytic site to enable divalent metal cation coordination and substrate binding.

Materials and methods

Protein expression and purification

The sequence of TbrCan2 was amplified from genomic DNA (*Thermoanaerobacter brockii* subsp. *finnii* Ako-1 DSM 1457) and cloned into a Macrolab 16B-series vector (N-terminal hexahistidine tag; TEV protease cleavage site) by ligation-independent cloning. QuikChange site-directed mutagenesis was performed to introduce point mutations. All plasmids generated for this work are listed in Supplementary Table S2. Can2 and its mutants were expressed in Rosetta2 (DE3) *Escherichia coli* cells grown at 37°C in LB medium. Protein expression was induced at OD_{600} of 0.6 by addition of 0.5 mM IPTG at 18°C overnight. Cells were harvested and resuspended in 20 mM HEPES–KOH pH 7.5, 500 mM KCl, 15 mM imidazole supplemented with protease inhibitors and lysed using a high-pressure cell lyser (Maximator HPL6). Cell lysates were clarified by centrifugation at 20 000 × g for 30 min and loaded on a 5 ml Ni-NTA Superflow column (QIAGEN) followed by extensive washing. Tagged Can2 was eluted with 20 mM HEPES–KOH pH 7.5, 250 mM KCl, 200 mM imidazole. The imidazole concentration was adjusted to a concentration of 30 mM and the affinity tag was removed in the presence of His-tagged TEV protease overnight at 4°C. His-tag-containing proteins were removed by reloading the sample on a Ni-NTA affinity column. The untagged protein was concentrated and further purified by size exclusion chromatography (SEC) on a HiLoad 16/600 Superdex 200 column (Cytiva) in 20 mM HEPES–KOH pH 7.5, 250 mM KCl, 1 mM DTT. Pure protein fractions were concentrated and flash-frozen in liquid nitrogen and stored at -80°C. Detailed purification conditions are listed in supplementary data (Supplementary Table S3).

Crystallization and structure determination

For crystallization, TbrCan2 was diluted to a final protein concentration of 150 μM diluted with buffer X (20 mM HEPES–KOH pH 7.5, 100 mM KCl, 1 mM DTT). The $TbrCan2^{WT}$-cA_4 complex was prepared by mixing 150 μM protein with 225 μM cA_4 in buffer X and assembled at RT for 10 min. The activated $TbrCan2^{WT/E341A}$ complexes in the presence of DNA substrates were assembled by mixing 150 μM of protein in a 1:1.5:3 molar ratio with cA_4 and substrate DNA oligonucleotide, followed by incubation at 37°C for 10 min in buffer X supplemented with 20 mM $MnCl_2$. Crystals were grown at 20°C by sitting-drop vapor diffusion by mixing protein samples in a 1:1 ratio (0.2 μl + 0.2 μl) with a reservoir solution (50 mM Tris pH 8, 17.1% PEG MME 550, 4.3% PEG 4000). Crystals were cryoprotected with 25% PEG 400 and flash-frozen in liquid nitrogen. Diffraction data was collected at the PXI and PXIII beamlines of the Swiss Light Source (Paul Scherrer Institute, Villigen, Switzerland)

and processed using the data-processing program package autoPROC (29). Structures were solved by molecular replacement (MR) with Phaser, as part of the PHENIX package (30). For apo Can2, the *in silico* generated AlphaFold2 model of TbrCan2^{WT} was used as a search model (31). The structures of the active Can2 complexes were solved by MR using the refined monomeric apo Can2 structure. Atomic model adjustment and refinement was conducted iteratively using Coot and PHENIX.refine (Supplementary Table S1) (32,33). Figures were generated with ChimeraX (34).

Plasmid nuclease assay

In vitro nuclease assays with supercoiled M13mp18 RF I dsDNA and M13mp18 ssDNA were performed with 250 nM Can2 in the presence of 2.5 μM cA_4 and 2 nM circular DNA in reaction buffer (20 mM HEPES pH7.5, 100 mM KCl,5 mM $MnCl_2$). The reaction was incubated at 45°C and was terminated by addition of 10 mM EDTA at indicated time points (control reactions were stopped after 45 min). 6× DNA loading buffer (10 mM Tris pH 7.6, 60 mM EDTA, 60% glycerol, 0.03% bromophenol blue, 0.03% xylene cyanole FF) was added and analyzed by 1% native agarose gel electrophoresis. Gels were scanned with a ChemiDoc Imager (Bio-Rad) and band intensities were quantified with Image Lab 6.1 (Bio-Rad).

RNase and DNase activity assays

RNA and DNA cleavage assays with linear substrates were carried out with 50 nM 5′-Atto532-labelled DNA and RNA substrates, 250 nM Can2 and 2.5 μM cA_4 in reaction buffer at 45°C. Reactions were stopped by addition of 2× denaturing loading dye (5% glycerol, 25 mM EDTA, 90% formamide, 0.15% OrangeG). The reactions were analyzed by 20% denaturing PAGE (20% acrylamide, 7 M urea, 0.5× TBE buffer) and bands were visualized on a Typhoon FLA 9500 laser scanner (GE Healthcare).

For determining *in vitro* nuclease cleavage kinetics under multiple-turnover conditions, 5 nM Can2 was mixed with 50 nM cA_4 and increasing concentrations of fluorophore-quencher oligonucleotides containing a 5′-6FAM fluorophore and a 3′-Black Hole Quencher 1 (50, 70, 100, 150, 250, 500, 1000, 2000 nM DNA or RNA) in a reaction buffer supplemented with 0.01% Tween-20. The triplicate experiments were monitored using a PHERAstar FSX microplate reader (BMG Labtech) measuring excitation and emission wavelength of 485 and 520 nm over 35 min at 45°C. RNA and DNA probes were purchased from Integrated DNA Technologies (IDT). Sequence information of DNA and RNA substrates used in this study are listed in Supplementary Table S4.

Electrophoretic mobility shift assays

For the detection of protein-RNA/DNA interactions, nucleic acid binding assays were performed with 25 nM 5′-Atto532 oligonucleotides and 50 μM cA_4 in the presence of different amounts of TbrCan2^{E341A}. Reactions were incubated in EMSA buffer (20 mM HEPES pH7.5, 25 mM KCl,1 mM DTT, 5 mM EDTA) for 10 min at 37°C. Samples were loaded on a native 10% polyacrylamide TBE gel and electrophoresis was carried out at 8°C. Gels were imaged on a Typhoon FLA 9500 and band intensities were quantified as described before. Assays to determine K_D were performed in triplicates.

Analytical SEC cOA binding assays

The binding of TbrCan2 to different cOAs (cA_3, cA_4, cA_6) was assayed by incubating 50 μM ligand with 100 μM (cA_4) or 300 μM (cA_3 and cA_6) TbrCan2 for 10 min at 45°C in binding buffer (20 mM HEPES pH7.5, 100 mM KCl,5 mM $MnCl_2$, 1 mM DTT). Analytical SEC was performed using an ÄKTA™ pure micro (Cytiva). The sample was separated on a Superdex 200 Increase 5/150 GL column that was pre-equilibrated with binding buffer and absorption was detected at a 260 nm and 280 nm wavelength.

Sequence analysis and construction of phylogenetic tree

A list of homologous sequences was generated by standard protein BLAST (35) using previously published Can2 sequences as query. Redundant sequences were removed using EMBOSS SkipRedundant (36) and CD-HIT (37). The resulting list of 828 Can2 sequences was used as input for multiple sequence alignment using MAFFT v7.429 (Supplementary Table S5) (38). From this alignment, a phylogenetic tree was generated via PHYLIP neighbor-joining (Supplementary Table S6) (39). CRISPR system assignment for the CRISPR loci containing *can2* genes was performed with CRISPRCasFinder (40).

Results

Phylogenetic analysis of Can2 family nucleases

The Can2 family of type III CRISPR-Cas ancillary effector nucleases is highly abundant in both archaeal and bacterial species. Recent studies have revealed that Can2-like nucleases exhibit distinctive targeting mechanisms despite their shared domain compositions (26–28). Can2 is typically encoded in canonical type III CRISPR-Cas loci, which comprise genes encoding the interference modules and additional accessory effectors, such as the non-specific RNase Csx1 (Figure 1A). Phylogenetic analysis of 828 orthologs identified seven distinctive Can2 clades with low sequence identity between them (Figure 1B). Analysis of the CRISPR loci that contain the respective *can2* genes shows that the clades often correspond to specific subtypes of type III CRISPR-Cas systems. Clade 5 primarily comprises Can2 proteins from type III-A systems, such as TresuCARD1, while Can2 proteins belonging to clades 2 and 6 originate from type III-B systems, including SthCan2 and TsuCan2 (Figure 1B). Clade 1 constitutes the largest cluster and encompasses orthologs that are encoded within type III-D loci (Figure 1A, B). Can2 proteins share a conserved domain architecture comprising an N-terminal CARF domain and a C-terminal PD-(D/E)xK nuclease (REase) domain, connected by an α-helical insertion (INS) domain (Figure 1C, D). In contrast, Can2-like nucleases of clade 4 lack the INS domain and only possess a short linker that connects the CARF and nuclease domains (Figure 1D). In all clades, Can2 harbors a conserved NxTGGTK motif within the CARF domain, which in SthCan2 and TresuCARD1 has been shown to mediate cyclic oligoadenylate (cOA) binding (Figure 1C-E) (28,41). A conserved E-ExD-(E/S)CK motif, present in most Can2 orthologs, has been shown to be essential for the nuclease activity of TresuCARD1 (28). However, Can2 nucleases from clade 1 instead contain a distinct E-ExD-SxTTS-K motif, suggesting that type III-D Can2 nucleases may differ mechanistically from previously characterized orthologs (Supplementary Figure S1e).

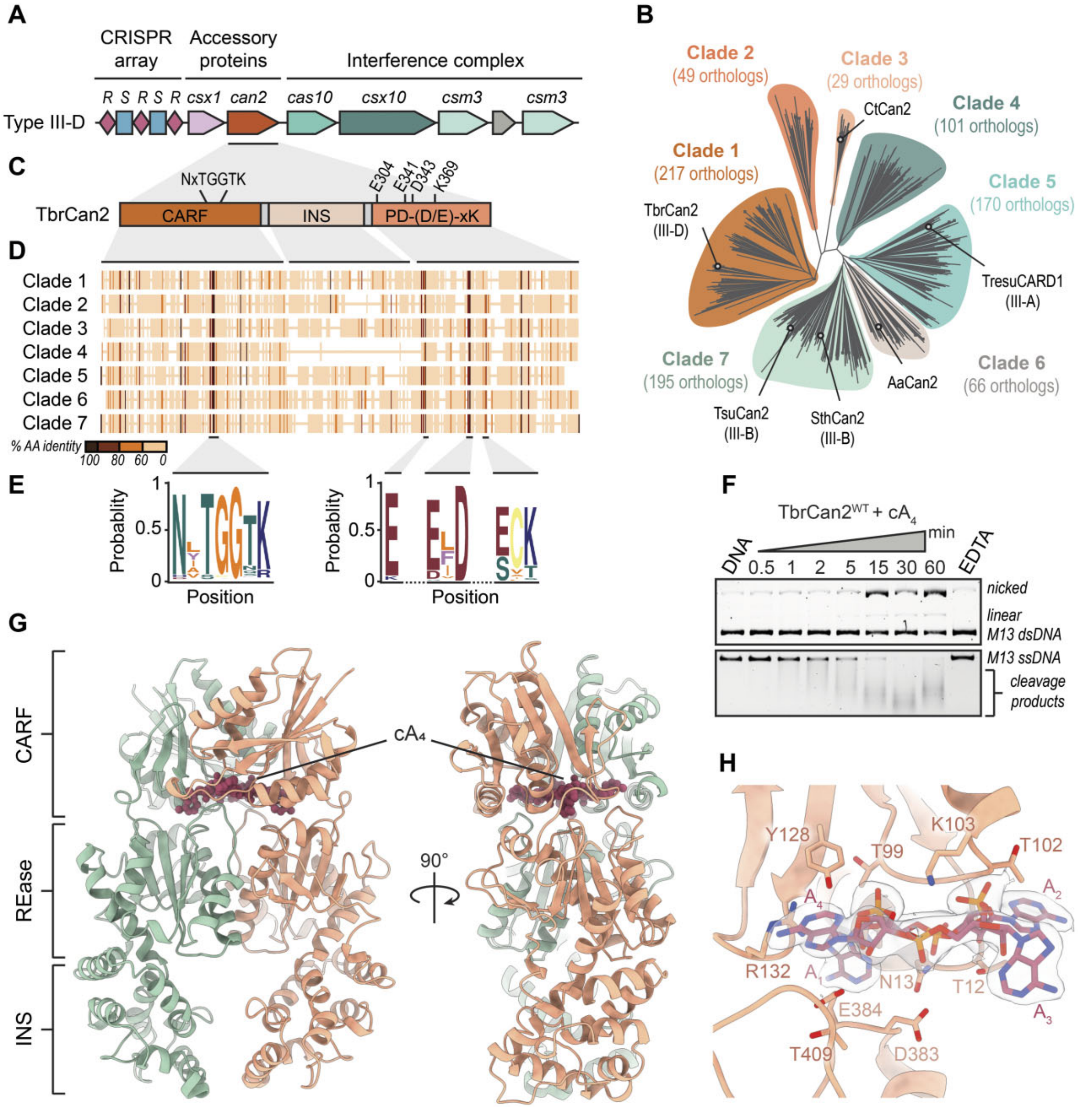

A
CRISPR array
Accessory proteins
Interference complex
R S R S R csx1 can2 cas10 csx10 csm3 csm3
Type III-D
B
Clade 2 (49 orthologs)
Clade 3 (29 orthologs)
CtCan2
Clade 4 (101 orthologs)
Clade 1 (217 orthologs)
Clade 5 (170 orthologs)
TbrCan2 (III-D)
TresuCARD1 (III-A)
Clade 7 (195 orthologs)
Clade 6 (66 orthologs)
AaCan2
TsuCan2 (III-B)
SthCan2 (III-B)
C
NxTGGTK
E304 E341 D343 K369
TbrCan2
CARF
INS
PD-(D/E)-xK
D
Clade 1
Clade 2
Clade 3
Clade 4
Clade 5
Clade 6
Clade 7
% AA identity
100 80 60 0
E
Probability
1
0.5
0
Position
F
TbrCan2^WT + cA4
min
DNA 0.5 1 2 5 15 30 60 EDTA
nicked
linear
M13 dsDNA
M13 ssDNA
cleavage products
G
CARF
REase
INS
cA4
90°
H
Y128
T99
K103
T102
A4
A2
R132
A1
N13
T12
E384
A3
T409
D383

Figure 1. Structure and activity of type III ancillary nuclease TbrCan2. **(A)** Schematic diagram of the *Thermoanaerobacter brockii finnii* Ako-1 type III-D CRISPR-Cas locus; *R*: direct repeat; *S*: spacer. **(B)** Phylogenetic clustering of Can2-like nucleases. **(C)** Schematic diagram of the domain composition of TbrCan2. **(D)** Amino acid sequence conservation among different Can2 clades. **(E)** Sequence logos of conserved motifs within CARF (left) and REase domains (right). **(F)** Nuclease activity assay of TbrCan2 in presence of single-stranded and double-stranded M13 DNA plasmids. **(G)** Overall views of the structure of TbrCan2 in complex with cA_4. **(H)** Close-up view of the cA_4 binding site at the CARF domain interface. The $2mF_O$-DF_C composite omit map is contoured at 1.0 σ and displayed within a radius of 2.0 Å around the cA_4 ligand.

Molecular architecture of the type III-D ancillary nuclease TbrCan2

Previous studies have reported distinct substrate selectivity of Can2 family nucleases (27,28). Given the distinctive catalytic site composition of clade 1 Can2 orthologs, we chose to characterize the activity and mechanism of the CRISPR type III-D ancillary nuclease Thebr_0943 (TbrCan2; Figure 1A–C). TbrCan2 exhibited ssDNase activity *in vitro*, completely degrading single-stranded M13 plasmid DNA in the presence of cA_4 and $MnCl_2$ (Figure 1F). Moreover, TbrCan2 nicked supercoiled double-stranded plasmid DNA, albeit with considerably lower efficiency as compared to ssDNA degradation. Nuclease activity was absent in the presence of EDTA, consistent with the divalent metal-dependency of PD-(D/E)xK family nucleases (42).

To gain molecular insights into TbrCan2 activity, we performed X-ray crystallographic studies and initially determined the atomic structure of wild-type TbrCan2 (TbrCan2^{WT}) in complex with cA_4 in the presence of Mn^{2+} ions, at a resolution of 3.2 Å (Figure 1G, Supplementary Figure S1a–d). TbrCan2 forms a symmetrical homodimeric complex, binding cA_4 in a central cavity located between the CARF and REase domains at the dimer interface, similar to other Can2 nucleases (Supplementary Figure S2) (27,28). The CARF domain of TbrCan2 (residues 1–145) exhibits a canonical Rossman fold conserved within the Can2 family, superimposing with TresuCARD1 and SthCan2 CARF domains with root-mean-square deviations (RMSD) of 2.2 and 2.3 Å over 139 and 136 residues, respectively (Supplementary Figure S2b, e). The cA_4 molecule is coordinated by an extensive network of hydrogen bonding and hydrophobic interactions involving residues from both the CARF and REase domains. The bases A1 and A3 of cA_4 are contacted by CARF domain residue Pro16 and REase domain residue Asp383 and Glu384, while nucleobases A2 and A4 are bound by CARF residues Gly11, Thr12 and Ser36 and the backbone of REase residues Gly410 and Thr409 (Figure 1H). The residues Thr99, Gly100, Gly101, Thr102 and Lys103 of the conserved NxTGGTK motif form a network of interactions with the ribose and phosphate moieties of the cA_4 ligand, further stabilizing its binding between the two CARF domains. The TbrCan2 REase domain (residues 290–437) exhibits considerable similarity with corresponding domains in previously described orthologs, superimposing with a RMSD of 2.2 Å (137 residues) and 2.4 Å (123 residues) with the nuclease domains of SthCan2 and TresuCARD1, respectively (Supplementary Figure S2c, f). The REase catalytic site in TbrCan2 features conserved residues Glu304, Glu341, Asp343, Lys369 and Glu372, of which Asp343 directly coordinates a single Mn^{2+} ion in the TbrCan2-cA_4 complex. In SthCan2 and TresuCARD1, divalent metal binding is additionally mediated by the ECK REase motif, which is absent in TbrCan2 and replaced with a SxTTS motif that is conserved among clade 1 Can2 orthologs (Supplementary Figure S2c, f). Ser356 within this motif, equivalent to Glu308 in CARD1 and Glu291 in SthCan2, is involved in water-mediated, outer-sphere coordination of the Mn^{2+} ion together with Glu304 and Lys369. Residue Glu372 is conserved among clade 1 Can2 variants and resides in a position equivalent to active site motif I in many PD-(D/E)xK nucleases (43). However, the sidechain of Glu372 is positioned too far (6.4 Å) from the metal ion to engage in inner- or outer-sphere coordination. The nuclease activity of PD-(D/E)xK enzymes occurs via a two-metal ion mechanism (42). Coordination of the second metal ion would be mediated by Asp343 and Glu341 in TbrCan2, based on structural superpositions with type II restriction endonucleases (Supplementary Figure S3c). The absence of the second metal ion in the structure implies that this binding site has intrinsically low affinity and that coordination of the second metal ion likely occurs cooperatively upon substrate nucleic acid binding.

TbrCan2 preferentially cleaves single-stranded nucleic acid substrates

To investigate the substrate selectivity of TbrCan2, we analyzed the cOA-dependent nuclease activity of TbrCan2 on linear single-stranded and double-stranded DNA and RNA substrates with similar sequences. A single-stranded DNA (ssDNA) substrate was completely degraded after 30 minutes, while TbrCan2 cleaved the RNA in less than 1 minute, demonstrating potent RNase activity (Figure 2A, B). Double-stranded DNA was cleaved at a low rate, comparable to the background nuclease activity observed in the absence of cA_4 (Figure 2A and Supplementary Figure S4b, e), indicating that TbrCan2 primarily targets single-stranded nucleic acids. The observed nuclease activity was dependent on the presence of manganese cations and was significantly reduced in the presence of magnesium (Figure 2B and Supplementary Figure S4). Moreover, catalytic activation required the presence of cA_4, and was not supported by cA_3 or cA_6 (Figure 2B and Supplementary Figure S4a, d). This was in agreement with analytical size-exclusion chromatography (SEC) binding assays, which revealed that TbrCan2 exhibits stable binding exclusively to cA_4 (Supplementary Figure S5), confirming cA_4 as the primary activator of TbrCan2. Mutations of conserved residues in the CARF cA_4 binding site, specifically T12A/N13A and Y128A, did not impede nuclease activation, indicating a certain degree of cooperativity among CARF residues (Figure 2C). Alanine substitutions of the catalytic residues Glu304, Asp343 and Lys369 abolished DNase activity. In contrast, RNase activity was partially perturbed by mutations of individual active site residues and was substantially reduced by simultaneous mutations of Glu341 and Lys369 (Figure 2C and Supplementary Figure S6). Next, we assessed the binding affinities of TbrCan2 for the oligonucleotide substrates by electrophoretic mobility shift assays (EMSAs), providing further evidence that TbrCan2 primarily functions as a single-strand selective nuclease (Figure 2D). Furthermore, we observed that $TbrCan2^{E341A}$ bound to DNA and RNA with apparent dissociation constants (K_D) of 2.9 μM and 0.24 μM, respectively, suggesting a higher overall binding affinity for RNA (Figure 2D). This was corroborated by complementary multiple-turnover kinetics experiments (Figure 2E), in which the observed rate constant for ssRNA cleavage ($k_{cat} = 7.6\ min^{-1}$) was approximately two orders of magnitude higher than that for ssDNA cleavage ($k_{cat} = 0.1\ min^{-1}$), indicating a strong selectivity for ssRNA substrates by TbrCan2 (Figure 2F).

Substrate binding mechanism of TbrCan2

We sought to elucidate the substrate binding mechanism of TbrCan2 by determining the structure of TbrCan2 in complex with a ssDNA substrate. To this end, we initially solved a 2.4 Å structure of the catalytically inactive E341A mutant ($TbrCan2^{E341A}$) in complex with cA_4 and a 5′-TTTAAA-3′ ssDNA (Figure 2G–I). The TbrCan2-cA_4 complex symmetrically binds two DNA molecules within the REase active sites which face a central substrate binding channel (Figure 2G). The bases of nucleotides dT0 to dA3 of both strands form a continuous π-π stack, with the 5′-terminal nucleotides facing the binding channel exit. The scissile phosphate group located between nucleotides dT0 and dA1 is positioned near the Mn^{2+} ion, hydrogen bonding to Glu304 and a metal-coordinating water molecule (Figure 2H, I). DNA binding is mediated by hydrogen bonding interactions between the phosphate groups of nucleotides dA2 and dA3, and side chain hydroxyl groups of Thr358, Thr359 and Ser360 from the conserved SxTTS motif (Figure 2H, I). Substrate binding is further stabilized by salt bridges between the scissile phosphate and Lys301, and between the dA3 phosphate and Arg213. In agreement with these observations, TbrCan2 nuclease activity was disrupted by mutations R213A, T358A and T359A and significantly reduced for S356A and S360A mutants, highlighting the importance of these residues for substrate nucleic acid binding (Figure 2C and Supplementary Figure S6).

The TbrCan2–ssDNA complex structure revealed that the two REase active sites in the TbrCan2 homodimer can engage two DNA strands simultaneously (Figure 2G), yet TbrCan2 preferentially cleaves single-stranded nucleic acids and exhibits only moderate nickase activity in the presence of

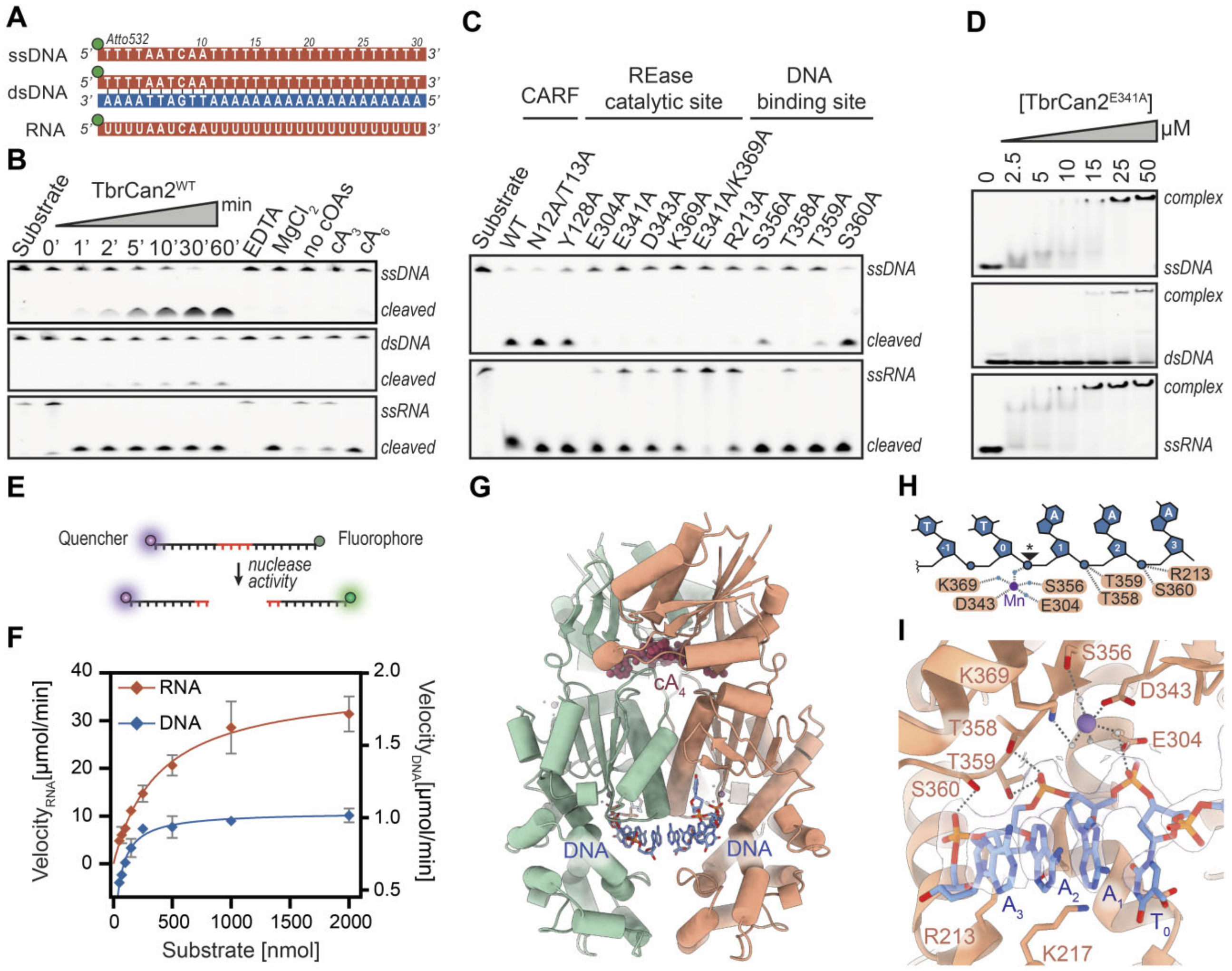


Figure 2. TbrCan2 has ssRNA- and ssDNA-endonuclease activities. **(A)** Schematic representation of nucleic acid substrates used for *in vitro* nuclease assays. **(B)** TbrCan2-catalyzed cleavage of 5′-fluorophore-labeled ssDNA, RNA and dsDNA substrates in the presence and absence of cOAs and $MnCl_2$ under single-turnover conditions. **(C)** DNase and RNase activities of structure-based mutant TbrCan2 proteins. **(D)** Electrophoretic mobility shift assay of TbrCan2^{E341A} binding to DNA and RNA substrates. **(E)** Schematic of the fluorophore-quencher reporter assay to determine enzyme kinetics of TbrCan2. **(F)** Multiple-turnover kinetics of TbrCan2 on ssDNA and ssDNA substrates. **(G)** Crystal structure of catalytically inactive E341A TbrCan2 mutant in complex with cA_4, Mn^{2+} and a ssDNA substrate (5′-TTTAAA-3′). **(H)** Schematic diagram of the DNA substrate and the protein-DNA interactions observed in the crystal structure. The scissile phosphate is denoted with an asterisk. Blue dots represent water molecules. **(I)** Close-up of the REase active site showing bound substrate. The $2mF_O$-DF_C composite omit map is contoured at 1.0 σ and displayed within a radius of 2.0 Å around the DNA.

dsDNA substrates (Figure 1F). To gain insight into the single-stranded target selectivity, we modeled a B-form DNA duplex into the TbrCan2 active site using the ssDNA-bound TbrCan2 structure as a reference. Positioning the substrate strand of the dsDNA duplex in the REase active site based on superposition with the 5′-TTTAAA-3′ ssDNA resulted in steric clashes of the non-substrate strand with the protein backbone of both TbrCan2 protomers (Supplementary Figure S7). This explains the lack of dsDNA endonuclease activity in TbrCan2 and implies that dsDNA binding and nicking may require conformational changes in TbrCan2 or structural distortions in the bound dsDNA. Attempts to elucidate the disparities in the double- and single-strand selective DNase and RNase activities by co-crystallizing TbrCan2 with dsDNA or RNA substrates remained unsuccessful, potentially due to the low binding affinities for dsDNA substrates and the residual RNase activity observed for TbrCan2^{E341A} (Figure 2D and Supplementary Figure S6).

Base selectivity for CA dinucleotides

The nuclease domain of Can2 nucleases shares similarities with restriction endonucleases, which exhibit strong sequence specificity. Previous studies have reported that Can2 nucleases show a certain bias for nucleotide sequences (26,28). However, the structure of the TTTAAA ssDNA-bound complex did not reveal specific contacts between the protein and the bases of the DNA substrate. To test for sequence selectivity of TbrCan2, we performed nuclease activity and substrate binding assays in the presence of 15-mer oligo-A, -C and -U/T nucleotides. TbrCan2 efficiently bound and degraded rA_{15} RNA, while it showed reduced turnover and binding of rC_{15} RNA (Figure 3A, B and Supplementary Figure S8a, b). In contrast, rU_{15} RNA, dT_{15} DNA or G-containing RNA and DNA (U/T_7-G_3-U/T_{20}) substrates were not cleaved in *in vitro* cleavage assays, nor were they stably bound by TbrCan2 in EMSA binding assays. Altogether, these results suggest a base preference of TbrCan2 for A and C nucleotides (Figure 3A, B).

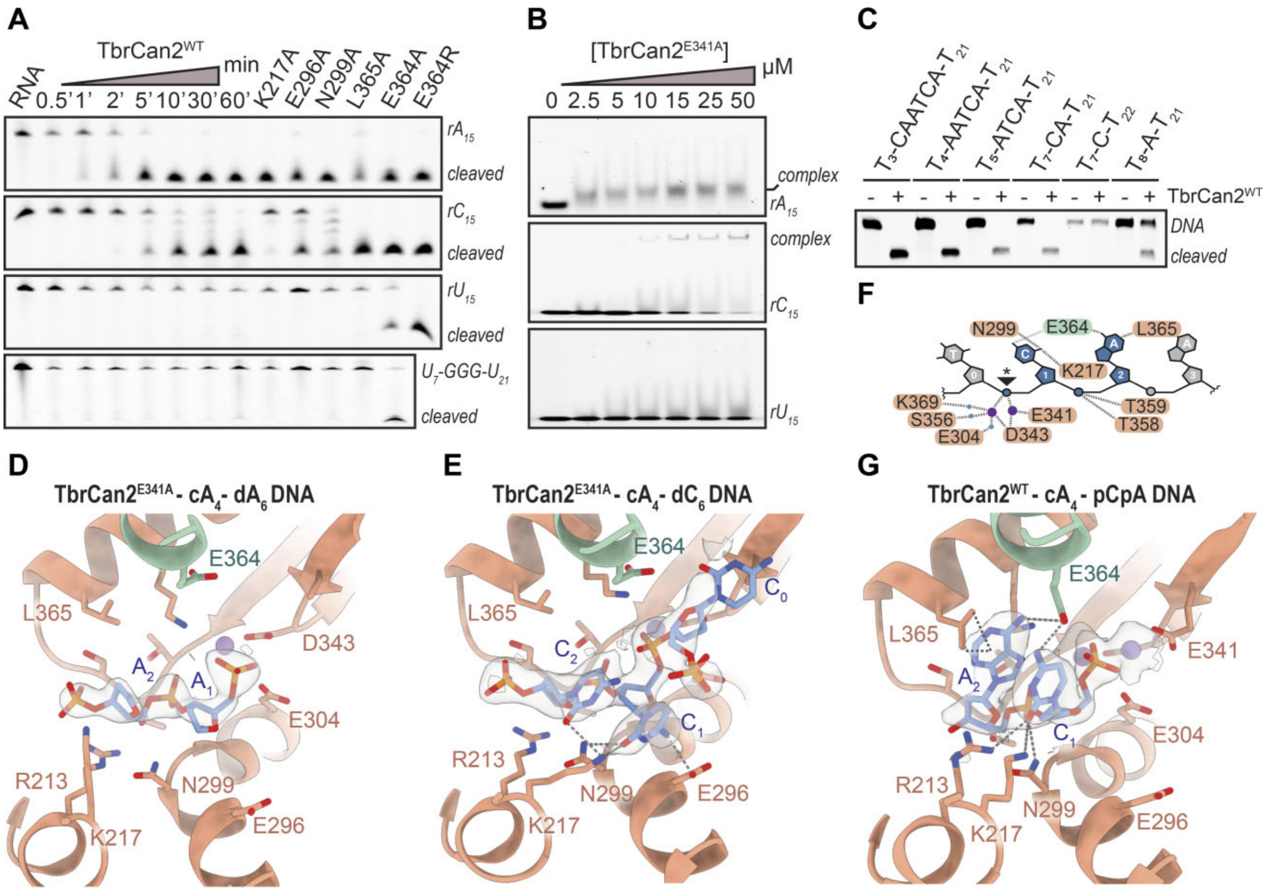

A
TbrCan2WT
min
RNA 0.5' 1' 2' 5' 10' 30' 60' K217A E296A N299A L365A E364A E364R
rA15
cleaved
rC15
cleaved
rU15
cleaved
U7-GGG-U21
cleaved
B
[TbrCan2E341A]
μM
0 2.5 5 10 15 25 50
complex
rA15
complex
rC15
rU15
C
T3-CAATCA-T21
T4-AATCA-T21
T5-ATCA-T21
T7-CA-T21
T7-C-T22
T8-A-T21
- + - + - + - + - + - +
TbrCan2WT
DNA
cleaved
F
N299
E364
L365
K217
K369
S356
E304
E341
D343
T359
T358
D
TbrCan2E341A - cA4 - dA6 DNA
E364
L365
A2
A1
D343
E304
R213
N299
K217
E296
E
TbrCan2E341A - cA4 - dC6 DNA
E364
C0
L365
C2
C1
R213
N299
E296
K217
G
TbrCan2WT - cA4 - pCpA DNA
E364
L365
E341
A2
C1
E304
R213
N299
K217
E296

Figure 3. TbrCan2 displays preference for CA dinucleotides. **(A)** Nuclease activity assay of WT and mutant TbrCan2 enzymes with rA_{15}, rC_{15}, rU_{15} and U_7-GGG-U_{21} RNA substrates. **(B)** EMSA binding assays of TbrCan2^{E341A} with RNA substrates. **(C)** Cleavage of ssDNA substrates by TbrCan2^{WT}. **(D)** Crystal structure of dA_6-bound TbrCan2^{E341A}-cA_4 complex. **(E)** Crystal structure of TbrCan2^{E341A}-cA_4 in complex with dC_6 DNA. **(F)** Schematic depicting interactions involved in substrate binding by TbrCan2. **(G)** Crystal structure of the TbrCan2-cA_4 complex bound to a 5′-pCpA-3′ dinucleotide cleavage product. The $2mF_O$-DF_C composite omit maps in **(D)**, **(E)** and **(G)** are shown contoured at 1.0 σ within a radius of 2.0 Å around the DNA.

Next, we tested the nuclease activity of TbrCan2 on ssDNA substrates designed to contain various sequence motifs (Figure 3C and Supplementary Figure S8d, e). We observed efficient degradation of sequences containing CA dinucleotide motifs, whereas cleavage of DNAs containing only A or C nucleotides was significantly slower (Figure 3C and Supplementary Figure S8d). Moreover, TbrCan2 cleaved substrates containing AC, GA, AG, GC or CG dinucleotides at considerably lower rates, suggesting selective recognition of the CA dinucleotide (Supplementary Figure S8e). To identify residues involved in the base selectivity, we solved the crystal structures of TbrCan2^{E341A}-cA_4 complex bound to dA_6 and dC_6 DNA oligonucleotides at a resolution of 2.0 Å and 2.3 Å, respectively (Figure 3D, E). In the dA_6-bound structure, only the DNA backbone is resolved, with no observable density for the adenine bases, suggesting that the nucleobases do not make specific contacts within the active site (Figure 3D). In contrast, four consecutive nucleotides are resolved in the dC_6-bound structure, revealing a distortion of the DNA backbone around dC0, where the bases are stabilized by hydrogen bonding interactions between the dC1 and dC2 bases and Lys217, Glu296, and Asn299 from the INS domain (Figure 3E). Alanine substitutions of these residues substantially reduced rC_{15} cleavage *in vitro*, supporting their involvement in cytosine base interactions, while rA_{15} degradation was only partly reduced (Figure 3A). Moreover, the dC_6-bound structure revealed that the acidic side chain of Glu364 from the opposite protomer protrudes into the active site in proximity to the 4-amino group of the dC nucleobases (Figure 3E and Supplementary Figure S8f). Intriguingly, the E364A mutant (TbrCan2^{E364A}) retained rA_{15} and rC_{15} cleavage activities and in addition degraded rU_{15} as well as U/T_7-G_3-U/T_{20} RNA and DNA substrates (Figure 3A and Supplementary Figure S8a). The E364R mutation further increased rU_{15} and T_7-G_3-T_{20} cleavage rates. Supporting these results, the crystal structure of TbrCan2^{E364R} bound to dT_6 DNA reveals that the sidechain of Arg364 is within hydrogen bonding distance of the carbonyl O4 atoms of the dT2 base (Supplementary Figure S8g). Together, these observations point to Glu364 as a major determinant of the nucleobase preference of TbrCan2.

The structures of the substrate-bound complexes of the catalytically inactive mutant TbrCan2^{E341A} reveal only one Mn^{2+} bound in the REase active site. In consequence, the observed mode of substrate binding might be perturbed due to the absence of the second divalent cation. To gain insights into nucleic acid binding in the context of the wild-type enzyme, we co-crystallized wild-type TbrCan2 (TbrCan2^{WT}) in the presence of cA_4, Mn^{2+} and a CA-containing oligonucleotide (5′-AATCAATCA-3′ ssDNA). The resulting structure, determined at a resolution of 2.2 Å, captures cA_4-activated TbrCan2^{WT} in a post-catalytic state, bound to the dCdA dinucleotide of the 3′-terminal cleavage product (Figure 3F, G).

The nucleobases of the cleavage product are recognized through hydrogen bonding interactions between the 6-amino group of dC0 and the 4-amino group of dA1 with Glu364 in the opposite TbrCan2 protomer. These observations confirm the role of this amino acid residue in the base selectivity of TbrCan2 (Figure 3F, G). Interestingly, the orientation of the dC_6 substrate DNA within the DNA binding channel is distinct from that of the dCdA dinucleotide cleavage product (Figure 3E, G and Supplementary Figure S8c). It remains uncertain whether these differences reflect distinct pre- and post-catalytic substrate/product binding modes, or differential binding of CA-containing nucleotides, as opposed to oligo-C DNA. Furthermore, the differences might be caused by the E341A mutation and the consequent loss of one catalytic Mn^{2+} ion in the dC_6-bound substrate complex. Taken together, the biochemical and structural data demonstrate that TbrCan2 has sequence preference for CA dinucleotides located immediately downstream of the scissile site, which can be attributed to direct interactions with the CA nucleobases.

Nuclease activation by cA_4-induced rearrangement of the catalytic site

Previous studies of CARD1 have revealed that cA_4 binding induces global conformational changes that bring the REase domains into closer proximity, which was postulated to result in catalytic activation (28). However, the precise mechanism of cA_4-induced activation of Can2/CARD1 enzymes remains elusive. To address this question, we additionally determined a crystal structure of $TbrCan2^{WT}$ in the apo state, at a resolution of 2.0 Å, to allow for comparisons with the structures of $TbrCan2^{WT}$-cA_4 and the product-bound $TbrCan2^{WT}$-cA_4 complex (Figure 4A and Supplementary Figure S9a–f). Structural overlay of the apo and cA_4-activated structures reveals pronounced conformational changes of the Can2 homodimer, whereby the two REase domains undergo a shearing motion (Figure 4A, Supplementary Figure S9a, d and Supplementary Video 1), similar to the structural changes observed in TresuCARD1 upon cA_4 binding (28). In addition to the global structural rearrangement, we observe local conformational changes in specific regions of the TbrCan2 nuclease domain. cA_4 binding leads to the repositioning of a loop comprising residues 406–415 (Figure 1H and Supplementary Video 2), which is displaced from the REase domain towards the cA_4 binding cavity of the CARF domains (Supplementary Figure S9b, e). However, mutations in the loop did not perturb substrate cleavage in the presence of cA_4, suggesting that this loop does not directly contribute to cA_4-mediated Can2 activation (Supplementary Figure S9g).

Notably, additional local restructuring occurs in a loop comprising residues 382–385 (Figure 1H) that connects α-helix 15 (α15) and β-strand 12 (β12) of the REase domain and protrudes into the cOA binding cavity formed by the two CARF domains (Figure 4B and Supplementary Video 3). Here, the binding of cA_4 is sensed by interactions between the A1/A3 base and Asp383 and Glu384, thereby pushing onto the loop, which repositions α15 within the REase domain. Through interactions between Ile380 (α15) and Phe340 (β10), this movement is transmitted onto the β10 strand, which contains catalytic residues Glu341 and Asp343 within the E-ExD-SxTTS-K REase motif. Comparison of the apo- and cA_4-bound structures reveals that β10 shifts by approximately 1.5 Å towards the active site upon cA_4 binding (Figure 4C and Supplementary Figure S9c, f). This restructuring enables the binding of a second catalytic Mn^{2+} ion ($Mn^{2+}{}_b$) by bringing Glu341 within inner-coordination distance, while Asp343 is positioned to simultaneously coordinate the $Mn^{2+}{}_a$ and $Mn^{2+}{}_b$ cations (Figure 4C, D and Supplementary Video 3). The active conformation of TbrCan2 is further stabilized by placing Arg376 (α15) within salt-bridging distance of Asp404 (α15) of the opposite protomer as the two REase domains move into proximity. Catalytic activation is thus allosterically coupled to cA_4 binding through a cascade of interactions, resulting in correct positioning of Glu341 and Asp343 for Mn^{2+} binding. Mutations of amino acid residues involved in the cascade abolished Can2 activation and impaired nuclease activity (Figure 4E), while the structural integrity and cA_4 binding of the respective mutants was maintained (Supplementary Figure S9h), thus supporting the allosteric activation mechanism.

Discussion

The cyclic oligoadenylate signaling mechanism of type III CRISPR-Cas systems is used to activate a broad spectrum of ancillary proteins, including CARF domain-containing effector nucleases (13,44,45). As the nuclease activity of these enzymes leads to growth arrest and cell death upon persistent activation of the type III CRISPR-Cas system (28,46), this necessitates stringent control of the cOA-dependent activation mechanism. Although previous structural studies have captured CARF-domain nucleases in the inactive and cOA-bound active conformational states, the details of the activation mechanism have remained elusive due to the lack of structural information on substrate binding and cleavage.

In this study, we provide critical insights into the nucleic acid binding and cA_4-dependent catalytic activation mechanism of the type III-D ancillary nuclease TbrCan2. We show that TbrCan2 is a cA_4-dependent, dual-specificity PD-(D/E)xK nuclease, capable of cleaving single-stranded DNA and RNA substrates through a two-metal-ion catalytic mechanism. Our structural analysis points to a mechanism in which cA_4 binding induces restructuring of the REase active site. This rearrangement is brought about by a cascade of interactions that link the cA_4 binding site with a key structural element (β-strand 10) in the REase domain that houses the metal-coordinating catalytic residues (Figure 5A, B), ultimately resulting in proper positioning of the divalent metal ions for substrate binding and catalysis. Although the specific amino acid residues involved in the allosteric relay are conserved among orthologs from clade 1, they are notably absent from other Can2 clades (Figure 5A, B), suggesting a distinct activation mechanism. Previously determined structures of TresuCARD1 in both apo and cA_4-bound states reveal that the REase loop connecting α-helix 15 and β-strand 14 is pulled towards the cOA binding cavity, resulting in a marked clockwise rotation of α15 during the transition from the inactive to the active state. This movement positions a conserved lysine residue, which is involved in coordinating a water molecule between the scissile phosphate and one of the two metal ions (Figure 5C). The active state conformation of α15 is stabilized through hydrophobic interactions with residues from the opposite protomer and enabled by the rotation and approximation of the two REase domains. These structural comparisons suggest that Can2-like auxiliary nucleases have evolved two distinct activation mechanisms: a 'push-and-flip' mechanism, exemplified by TbrCan2 from clade 1 (Figure 5D), and a

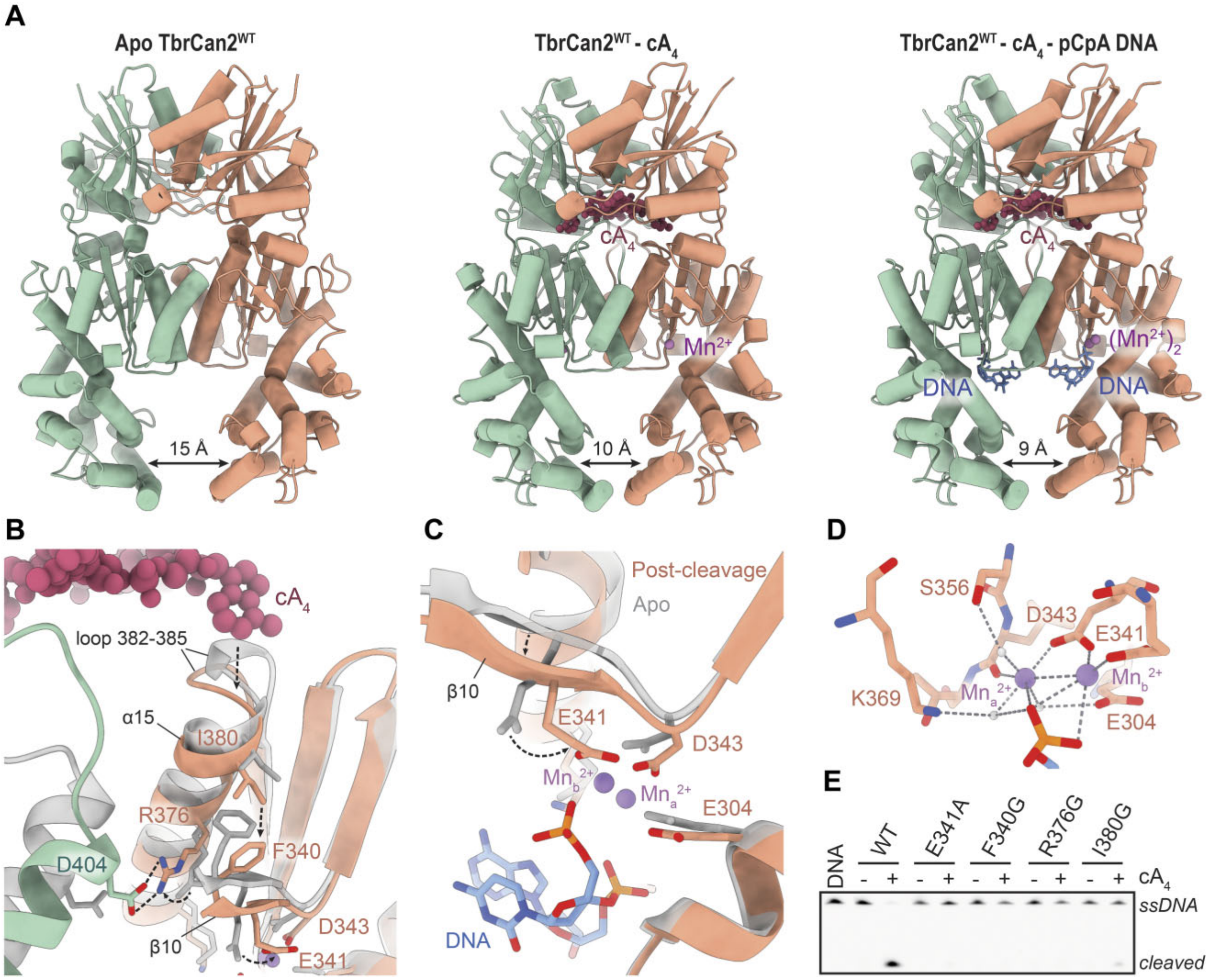


Figure 4. cA_4-induced rearrangement of divalent cation binding sites underpins catalytic activation of TbrCan2. **(A)** Crystal structures of apo $TbrCan2^{WT}$ (left), cA_4-bound $TbrCan2^{WT}$ (center) and $TbrCan2^{WT}$ (right) in the post-catalytic state with a 5′-pCpAp-3′ DNA fragment bound to the active site. The indicated distances refer to the distances between the Cα atoms of the Lys238 residues in the two TbrCan2 protomers. **(B)** Superposition of the structural elements involved in the allosteric relay between the cA_4 binding site and REase active site in the apo (gray) and post-catalytic (colored) states of TbrCan2. **(C)** Superposition of the TbrCan2 REase active sites in the apo (gray) and post-catalytic (colored) states, depicting local conformation transitions induced by cA_4 binding. **(D)** Coordination of two catalytic Mn^{2+} ions within the REase active site. **(E)** ssDNA cleavage activity of TbrCan2 enzymes with mutated residues involved in cA_4-induced allosteric activation.

'pull-and-rotate' mechanism, as observed in the clade 5 ortholog TresuCARD1 (Figure 5E).

A key feature of TbrCan2 is its selectivity for RNA and DNA substrates. The substrates are bound within a channel located between the two REase domains, predominantly lined with charged side chains. The spatial disposition of the two REase active sites enables the concurrent coordination of two DNA or RNA strands. The proper orientation of the substrate is ensured by a conserved SxTTS motif that binds to the 3′ end of the target nucleic acid, positioning the scissile phosphate within the catalytic site. The coordination of the substrate is further secured through interactions with active site residues and metal ions. Intriguingly, the SxTTS motif is absent in Can2 clades other than clade 1, indicating potential mechanistic differences in substrate coordination and likely reflecting differential substrate specificities observed in previously characterized Can2 orthologs (26–28).

The structures of TbrCan2 bound to DNA further rationalize the observed substrate selectivity for CA dinucleotide sequence motifs by hydrogen bonding interactions with Glu364, while T and U bases are discriminated by the absence of productive hydrogen bonding with Glu364 and its negative charge. In agreement with the structural observation, charge reversal at this position confers substrate preference for U- and G-containing substrates. Importantly, numerous clade 1 orthologs feature an arginine residue in the equivalent position, implying a distinct base preference. While the substrate selectivity of TbrCan2 is not as stringent as the sequence specificity displayed by PD-(D/E)xK type II restriction endonucleases, some similarities exist between the two. For example, the restriction enzyme NgoMIV targets a conserved 5′-GCCGGC-3′ hexanucleotide sequence, whose recognition is facilitated by an Arg-Ser-Asp-Arg (RSDR) amino acid motif located near the active site, at a position analogous to $\alpha15$ in TbrCan2 (Supplementary Figure S3d, f) (47). In NgoMIV, the RSDR residue Asp193 connects the amino groups of the C bases of the DNA target strand within the opposing REase active site, a function comparable to Glu364 in TbrCan2

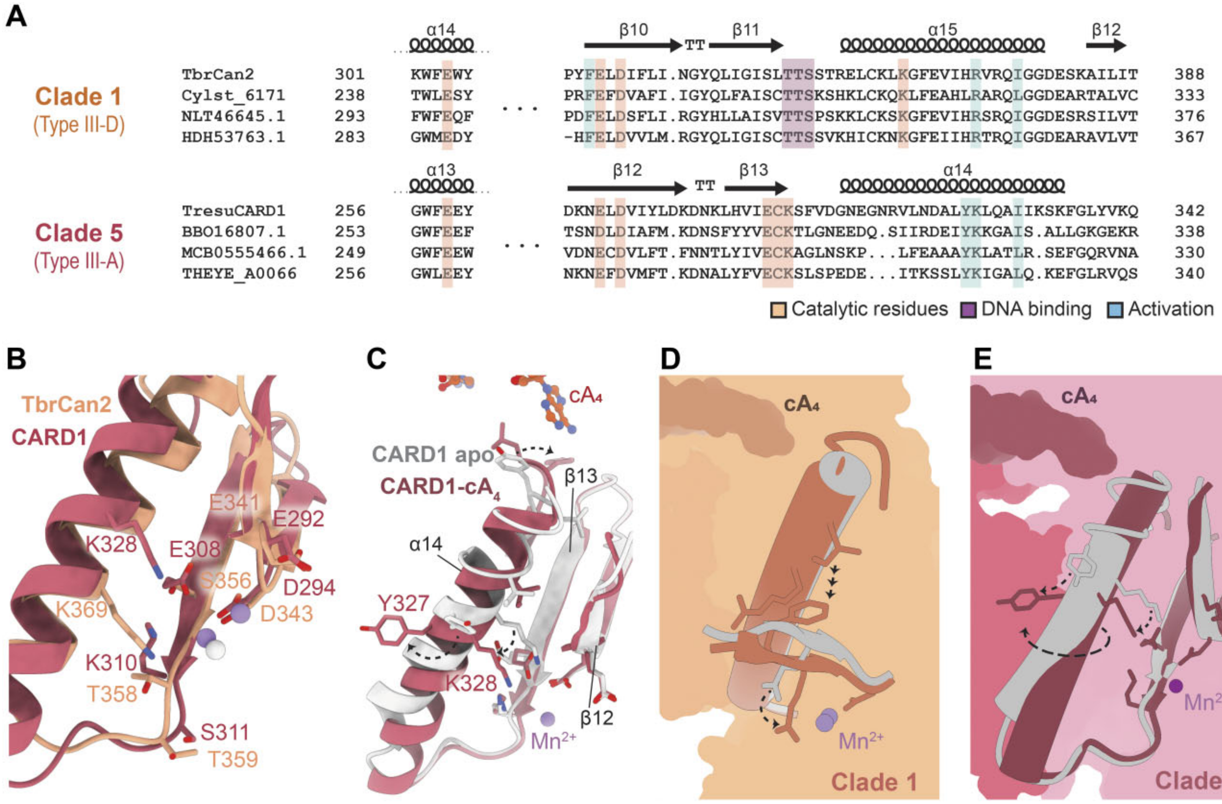

A
α14
β10
TT
β11
α15
β12
Clade 1
(Type III-D)
TbrCan2 301 KWFEWY ... PYFELDIFLI.NGYQLIGISLTTSSTRELCKLKGFEVIHRVRQIGGDESKAILIT 388
Cylst_6171 238 TWLESY ... PRFEFDVAFI.IGYQLFAISCTTSKSHKLCKQKLFEAHLRARQLGGDEARTALVC 333
NLT46645.1 293 FWFEQF ... PDFELDSFLI.RGYHLLAISVTTSPSKKLCKSKGFEVIHRSRQIGGDESRSILVT 376
HDH53763.1 283 GWMEDY ... -HFELDVVLM.RGYQLIGISCTTSSVKHICKNKGFEIIHRTRQIGGDEARAVLVT 367
α13
β12
TT
β13
α14
Clade 5
(Type III-A)
TresuCARD1 256 GWFEEY ... DKNELDVIYLDKDNKLHVIECKSFVDGNEGNRVLNDALYKLQAIIKSKFGLYVKQ 342
BBO16807.1 253 GWFEEF ... TSNDLDIAFM.KDNSFYYVECKTLGNEEDQ.SIIRDEIYKKGAIS.ALLGKGEKR 338
MCB0555466.1 249 GWFEEW ... VDNECDVLFT.FNNTLYIVECKAGLNSKP...LFEAAAYKLATLR.SEFGQRVNA 330
THEYE_A0066 256 GWLEEY ... NKNEFDVMFT.KDNALYFVECKSLSPEDE...ITKSSLYKIGALQ.KEFGLRVQS 340
Catalytic residues
DNA binding
Activation
B
TbrCan2
CARD1
E341
E292
K328
E308
S356
D294
K369
D343
K310
T358
S311
T359
C
cA₄
CARD1 apo
CARD1-cA₄
β13
α14
Y327
K328
β12
Mn²⁺
D
cA₄
Mn²⁺
Clade 1
E
cA₄
Mn²⁺
Clade 5

Figure 5. Allosteric activation mechanisms of Can2-family nucleases. **(A)** Sequence alignments of clade 1 and clade 5 Can2 enzymes. Strictly conserved active site residues involved in catalysis (orange), DNA binding (purple) and cA$_4$-mediated activation (green) are highlighted. **(B)** Superposition of active site residues in TbrCan2 (orange) and TresuCARD1 (magenta). **(C)** Superposition of α-helix 14 in apo- and cA$_4$-bound TresuCARD1 structures. (D, E) Schematic representation of the conformational activation mechanisms of **(D)** clade 1 (TbrCan2) and **(E)** clade 5 Can2 nucleases (TresuCARD1).

(Supplementary Figure S3d, f), highlighting the silimilarities between substrate recognition mechanisms of restriction enzymes and Can2. However, TbrCan2 lacks the extensive network of interactions that enable specific sequence recognition in NgoMIV, thus resulting in less stringent substrate selectivity (Supplementary Figure S3f). Another property that sets TbrCan2 apart from PD-(D/E)xK domain restriction endonucleases is its selectivity for single-stranded RNA and DNA substrates. While TbrCan2 can bind to double-stranded DNA with low affinity and can catalyze nicking at a low rate, it is incapable of generating double-strand breaks. In type II restriction endonucleases such as EndoMS and NgoMIV, both active sites bind and cleave respective strands simultaneously (Supplementary Figure S3) (47,48). In contrast, structural modeling of dsDNA binding to TbrCan2 results in steric clashes between dsDNA duplex and the protein backbone (Supplementary Figure S7), indicating that TbrCan2 is unable to catalyze staggered nicks in a dsDNA substrate due to the specific arrangement of its two catalytic sites and the orientation at which the substrate DNA is coordinated by the SxTTS motif in each REase active site.

We have been unable to determine the structure of TbrCan2 in complex with an RNA substrate, necessitating further investigations into the mechanisms underlying the ribonuclease and deoxyribonuclease activities of CARD1/Can2 enzymes. This is particularly important in light of contradictory reports of the *in vitro* and *in vivo* nuclease activities of TresuCARD1 (28). Although CARD1, like TbrCan2, cleaves RNA targets more rapidly than DNA *in vitro*, its expression in a heterologous host did not affect RNA transcript levels. Although this would suggest that CARD1-mediated growth inhibition might primarily be due to the ssDNase activity, the relative contribution of the DNase and RNase activities to the functional mechanism of CARD1/Can2 enzymes is presently unclear. Future work will thus be needed to define the functional landscape of CARD1/Can2 accessory nucleases so as to comprehend their biological functions in their native hosts.

Data availability

The data underlying this article are available in the article and in its online supplementary material. Atomic coordinates and structure factors are available in the Protein Data Bank under accession codes 8Q3Y, 8Q3Z, 8Q40, 8Q41, 8Q42, 8Q43 and 8Q44.

Supplementary data

Supplementary Data are available at NAR Online.

Acknowledgements

We would like to thank Beat Blattmann (Protein Crystallization Center, UZH) for the crystallographic support. We also thank Vincent Olieric and Takashi Tomizaki (Swiss Light Source, Paul Scherrer Institute, Villigen, Switzerland) for the assistance with crystallographic data collection. We thank members of the Jinek laboratory for helpful discussions and critical reading of the paper. M.J. is International Research Scholar of the Howard Hughes Medical Institute and Vallee

Scholar of the Bert L & N Kuggie Vallee Foundation. K.J. is a member of the Biomolecular Structure and Mechanism PhD Program of the Life Science Zurich Graduate School.

Funding

European Research Council (ERC) Consolidator Grant CRISPR2.0 [ERC-CoG-820152 to M.J.]; M.J. is a member of the NCCR RNA & Disease. Funding for open access charge: University of Zurich.

Conflict of interest statement

M.J. is a co-founder and Scientific Advisory Board member of Caribou Biosciences. All other authors declare no competing interests.

References

1. Koonin,E.V., Makarova,K.S. and Zhang,F. (2017) Diversity, classification and evolution of CRISPR-Cas systems. *Curr. Opin. Microbiol.*, **37**, 67–78.
2. Barrangou,R. and Marraffini,L.A. (2014) CRISPR-Cas systems: prokaryotes upgrade to adaptive immunity. *Mol. Cell*, **54**, 234–244.
3. Makarova,K.S., Wolf,Y.I., Iranzo,J., Shmakov,S.A., Alkhnbashi,O.S., Brouns,S.J.J., Charpentier,E., Cheng,D., Haft,D.H., Horvath,P., *et al.* (2019) Evolutionary classification of CRISPR–Cas systems: a burst of class 2 and derived variants. *Nat. Rev. Microbiol.*, **18**, 67–83.
4. Molina,R., Sofos,N. and Montoya,G. (2020) Structural basis of CRISPR-Cas Type III prokaryotic defence systems. *Curr. Opin. Struct. Biol.*, **65**, 119–129.
5. Hale,C.R., Zhao,P., Olson,S., Duff,M.O., Graveley,B.R., Wells,L., Terns,R.M. and Terns,M.P. (2009) RNA-guided RNA cleavage by a CRISPR RNA-Cas protein complex. *Cell*, **139**, 945–956.
6. Kazlauskiene,M., Tamulaitis,G., Kostiuk,G., Venclovas,Č. and Siksnys,V. (2016) Spatiotemporal control of type III-A CRISPR-Cas immunity: coupling DNA degradation with the target RNA recognition. *Mol. Cell*, **62**, 295–306.
7. Elmore,J.R., Sheppard,N.F., Ramia,N., Deighan,T., Li,H., Terns,R.M. and Terns,M.P. (2016) Bipartite recognition of target RNAs activates DNA cleavage by the type III-B CRISPR–Cas system. *Genes Dev.*, **30**, 447–459.
8. Estrella,M.A., Kuo,F.T. and Bailey,S. (2016) RNA-activated DNA cleavage by the type III-B CRISPR–Cas effector complex. *Genes Dev.*, **30**, 460–470.
9. Niewoehner,O., Garcia-Doval,C., Rostøl,J.T., Berk,C., Schwede,F., Bigler,L., Hall,J., Marraffini,L.A. and Jinek,M. (2017) Type III CRISPR–Cas systems produce cyclic oligoadenylate second messengers. *Nature*, **548**, 543–548.
10. Kazlauskiene,M., Kostiuk,G., Venclovas,Č., Tamulaitis,G. and Siksnys,V. (2017) A cyclic oligonucleotide signaling pathway in type III CRISPR-Cas systems. *Science*, **357**, 605–609.
11. Jia,N., Jones,R., Yang,G., Ouerfelli,O. and Patel,D.J. (2019) 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.
12. Hogrel,G., Guild,A., Graham,S., Rickman,H., Grüschow,S., Bertrand,Q., Spagnolo,L. and White,M.F. (2022) Cyclic nucleotide-induced helical structure activates a TIR immune effector. *Nature*, **608**, 808–812.
13. Makarova,K.S., Timinskas,A., Wolf,Y.I., Gussow,A.B., Siksnys,V., Venclovas,Č. and Koonin,E.v. (2020) Evolutionary and functional classification of the CARF domain superfamily, key sensors in prokaryotic antivirus defense. *Nucleic Acids Res.*, **48**, 8828–8847.
14. Shah,S.A., Alkhnbashi,O.S., Behler,J., Han,W., She,Q., Hess,W.R., Garrett,R.A. and Backofen,R. (2019) Comprehensive search for accessory proteins encoded with archaeal and bacterial type III CRISPR-cas gene cassettes reveals 39 new cas gene families. *RNA Biol*, **16**, 530–542.
15. Shmakov,S.A., Makarova,K.S., Wolf,Y.I., Severinov,K.V. and Koonin,E.V. (2018) Systematic prediction of genes functionally linked to CRISPR-Cas systems by gene neighborhood analysis. *Proc. Natl. Acad. Sci. U.S.A.*, **115**, E5307–E5316.
16. Steens,J.A., Bravo,J.P.K., Raymund,C., Salazar,P., Yildiz,C., Amieiro,A.M., Köstlbacher,S., Prinsen,S.H.P., Patinios,C., Bardis,A., *et al.* (2023) Type III-B CRISPR-Cas signaling-based cascade of proteolytic cleavages. bioRXiv doi: https://doi.org/10.1101/2023.06.23.546230, 23 June 2023, preprint: not peer reviewed.
17. Molina,R., Stella,S., Feng,M., Sofos,N., Jauniskis,V., Pozdnyakova,I., López-Méndez,B., She,Q. and Montoya,G. (2019) Structure of Csx1-cOA4 complex reveals the basis of RNA decay in type III-B CRISPR-Cas. *Nat. Commun.*, **10**, 4302.
18. Niewoehner,O. and Jinek,M. (2016) Structural basis for the endoribonuclease activity of the type III-A CRISPR-associated protein Csm6. *RNA*, **22**, 318–329.
19. Foster,K., Kalter,J., Woodside,W., Terns,R.M. and Terns,M.P. (2019) The ribonuclease activity of Csm6 is required for anti-plasmid immunity by Type III-A CRISPR-Cas systems. *RNA Biol*, **16**, 449–460.
20. Garcia-Doval,C., Schwede,F., Berk,C., Rostøl,J.T., Niewoehner,O., Tejero,O., Hall,J., Marraffini,L.A. and Jinek,M. (2020) Activation and self-inactivation mechanisms of the cyclic oligoadenylate-dependent CRISPR ribonuclease Csm6. *Nat. Commun.*, **11**, 1596.
21. Athukoralage,J.S., Graham,S., Grüschow,S., Rouillon,C. and White,M.F. (2019) A type III CRISPR ancillary ribonuclease degrades its cyclic oligoadenylate activator. *J. Mol. Biol.*, **431**, 2894–2899.
22. Smalakyte,D., Kazlauskiene,M., Havelund,J.F., Rukšenaite,A., Rimaite,A., Tamulaitiene,G., Færgeman,N.J., Tamulaitis,G. and Siksnys,V. (2020) Type III-A CRISPR-associated protein Csm6 degrades cyclic hexa-adenylate activator using both CARF and HEPN domains. *Nucleic Acids Res.*, **48**, 9204–9217.
23. Santiago-Frangos,A., Hall,L.N., Nemudraia,A., Nemudryi,A., Krishna,P., Wiegand,T., Wilkinson,R.A., Snyder,D.T., Hedges,J.F., Cicha,C., *et al.* (2021) Intrinsic signal amplification by type III CRISPR-Cas systems provides a sequence-specific SARS-CoV-2 diagnostic. *Cell Rep Med*, **2**, 100319.
24. Steens,J.A., Zhu,Y., Taylor,D.W., Bravo,J.P.K., Prinsen,S.H.P., Schoen,C.D., Keijser,B.J.F., Ossendrijver,M., Hofstra,L.M., Brouns,S.J.J., *et al.* (2021) SCOPE enables type III CRISPR-Cas diagnostics using flexible targeting and stringent CARF ribonuclease activation. *Nat. Commun.*, **12**, 5033.
25. Sridhara,S., Goswami,H.N., Whyms,C., Dennis,J.H. and Li,H. (2021) Virus detection via programmable Type III-A CRISPR-Cas systems. *Nat. Commun.*, **12**, 5653.
26. Nemudraia,A., Nemudryi,A., Buyukyoruk,M., Scherffius,A.M., Zahl,T., Wiegand,T., Pandey,S., Nichols,J.E., Hall,L.N., McVey,A., *et al.* (2022) Sequence-specific capture and concentration of viral RNA by type III CRISPR system enhances diagnostic. *Nat. Commun.*, **13**, 7762.
27. Zhu,W., McQuarrie,S., Grüschow,S., McMahon,S.A., Graham,S., Gloster,T.M. and White,M.F. (2021) The CRISPR ancillary effector Can2 is a dual-specificity nuclease potentiating type III CRISPR defence. *Nucleic Acids Res.*, **49**, 2777–2789.
28. Rostøl,J.T., Xie,W., Kuryavyi,V., Maguin,P., Kao,K., Froom,R., Patel,D.J. and Marraffini,L.A. (2021) The Card1 nuclease provides defence during type III CRISPR immunity. *Nature*, **590**, 624–629.
29. Vonrhein,C., Flensburg,C., Keller,P., Sharff,A., Smart,O., Paciorek,W., Womack,T. and Bricogne,G. (2011) Data processing and analysis with the autoPROC toolbox. *Acta. Crystallogr. D Biol. Crystallogr.*, **67**, 293–302.

30. Liebschner,D., Afonine,P.v., Baker,M.L., Bunkoczi,G., Chen,V.B., Croll,T.I., Hintze,B., Hung,L.W., Jain,S., McCoy,A.J., *et al.* (2019) Macromolecular structure determination using X-rays, neutrons and electrons: recent developments in Phenix. *Acta Crystallogr D Struct Biol*, **75**, 861–877.
31. Mirdita,M., Schütze,K., Moriwaki,Y., Heo,L., Ovchinnikov,S. and Steinegger,M. (2022) ColabFold: making protein folding accessible to all. *Nat. Methods*, **19**, 679–682.
32. Emsley,P. and Cowtan,K. (2004) Coot: model-building tools for molecular graphics. *Acta. Crystallogr. D Biol. Crystallogr.*, **60**, 2126–2132.
33. Afonine,P.v., Grosse-Kunstleve,R.W., Echols,N., Headd,J.J., Moriarty,N.W., Mustyakimov,M., Terwilliger,T.C., Urzhumtsev,A., Zwart,P.H. and Adams,P.D. (2012) Towards automated crystallographic structure refinement with phenix.Refine. *urn:issn:0907-4449*, **68**, 352–367.
34. Pettersen,E.F., Goddard,T.D., Huang,C.C., Meng,E.C., Couch,G.S., Croll,T.I., Morris,J.H. and Ferrin,T.E. (2021) UCSF ChimeraX: structure visualization for researchers, educators, and developers. *Protein Sci.*, **30**, 70–82.
35. Altschul,S.F., Madden,T.L., Schäffer,A.A., Zhang,J., Zhang,Z., Miller,W. and Lipman,D.J. (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.*, **25**, 3389–3402.
36. Rice,P., Longden,L. and Bleasby,A. (2000) EMBOSS: the European Molecular Biology Open Software Suite. *Trends Genet.*, **16**, 276–277.
37. Li,W. and Godzik,A. (2006) Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinformatics*, **22**, 1658–1659.
38. Katoh,K. and Standley,D.M. (2013) MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol. Biol. Evol.*, **30**, 772.
39. Felsenstein,J. (1989) PHYLIP - phylogeny inference package (Version 3.5c). *Cladistics*, **5**, 164–166.
40. Couvin,D., Bernheim,A., Toffano-Nioche,C., Touchon,M., Michalik,J., Néron,B., Rocha,E.P.C., Vergnaud,G., Gautheret,D. and Pourcel,C. (2018) CRISPRCasFinder, an update of CRISRFinder, includes a portable version, enhanced performance and integrates search for Cas proteins. *Nucleic Acids Res.*, **46**, W246–W251.
41. McMahon,S.A., Zhu,W., Graham,S., Rambo,R., White,M.F. and Gloster,T.M. (2020) Structure and mechanism of a type III CRISPR defence DNA nuclease activated by cyclic oligoadenylate. *Nat. Commun.*, **11**, 500.
42. Yang,W. (2011) Nucleases: diversity of structure, function and mechanism. *Q. Rev. Biophys.*, **44**, 1–93.
43. Knizewski,L., Kinch,L.N., Grishin,N.V., Rychlewski,L. and Ginalski,K. (2007) Realm of PD-(D/E)XK nuclease superfamily revisited: detection of novel families with modified transitive meta profile searches. *BMC Struct. Biol.*, **7**, 40.
44. Athukoralage,J.S. and White,M.F. (2021) Cyclic oligoadenylate signaling and regulation by ring nucleases during type III CRISPR defense. *RNA*, **27**, 855–867.
45. Steens,J.A., Salazar,C.R.P. and Staals,R.H.J. (2022) The diverse arsenal of type III CRISPR–Cas-associated CARF and SAVED effectors. *Biochem. Soc. Trans.*, **50**, 1353.
46. Rostøl,J.T. and Marraffini,L.A. (2019) Non-specific degradation of transcripts promotes plasmid clearance during type III-A CRISPR–Cas immunity. *Nat. Microbiol.*, **4**, 656–662.
47. Deibert,M., Grazulis,S., Sasnauskas,G., Siksnys,V. and Huber,R. (2000) Structure of the tetrameric restriction endonuclease NgoMIV in complex with cleaved DNA. *Nat. Struct. Biol.*, **7**, 792–799.
48. Nakae,S., Hijikata,A., Tsuji,T., Yonezawa,K., Kouyama,K.-I., Mayanagi,K., Ishino,S., Ishino,Y. and Shirai,T. (2016) Structure of the EndoMS-DNA complex as mismatch restriction endonuclease. *Structure*, **24**, 1960–1971.

Received: July 26, 2023. **Revised:** October 26, 2023. **Editorial Decision:** October 27, 2023. **Accepted:** November 9, 2023

© The Author(s) 2023. Published by Oxford University Press on behalf of Nucleic Acids Research.
This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial License (http://creativecommons.org/licenses/by-nc/4.0/), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited. For commercial re-use, please contact journals.permissions@oup.com