RESEARCH ARTICLE

CRISPR

RNA-activated protein cleavage with a CRISPR-associated endopeptidase

Jonathan Strecker[1,2,3,4,5]*†, F. Esra Demircioglu[1,2,3,4,5]†, David Li[1,2,3,4,5,6], Guilhem Faure[1,2,3,4,5], Max E. Wilkinson[1,2,3,4,5], Jonathan S. Gootenberg[3], Omar O. Abudayyeh[3], Hiroshi Nishimasu[7,8,9,10], Rhiannon K. Macrae[1,2,3,4,5], Feng Zhang[1,2,3,4,5]*

In prokaryotes, CRISPR-Cas systems provide adaptive immune responses against foreign genetic elements through RNA-guided nuclease activity. Recently, additional genes with non-nuclease functions have been found in genetic association with CRISPR systems, suggesting that there may be other RNA-guided non-nucleolytic enzymes. One such gene from *Desulfonema ishimotonii* encodes the TPR-CHAT protease Csx29, which is associated with the CRISPR effector Cas7-11. Here, we demonstrate that this CRISPR-associated protease (CASP) exhibits programmable RNA-activated endopeptidase activity against a sigma factor inhibitor to regulate a transcriptional response. Cryo–electron microscopy of an active and substrate-bound CASP complex reveals an allosteric activation mechanism that reorganizes Csx29 catalytic residues upon target RNA binding. This work reveals an RNA-guided function in nature that can be leveraged for RNA-sensing applications in vitro and in human cells.

Prokaryotes possess a multitude of defense systems against foreign genetic elements, including CRISPR and CRISPR-associated proteins (Cas) systems (*1*–*3*). Although the predominant function of CRISPR-Cas systems is to provide adaptive immunity through RNA-guided DNA or RNA nuclease activity, additional proteins have been identified in genetic association with CRISPR loci (*3*–*5*). One example is the CRISPR-associated transposase (CAST) systems (*6*, *7*), which perform RNA-guided DNA insertion whereby nuclease-inactive CRISPR effectors guide Tn7-like mobile genetic elements to specific DNA sequences (*8*, *9*). CAST systems have evolved on at least three separate occasions (*10*), highlighting the ability of diverse CRISPR effectors to acquire, or be acquired by, other bacterial enzymes. Beyond CAST systems, additional functions genetically linked to CRISPR-Cas systems are beginning to emerge, and it is likely that more remain to be discovered and characterized.

Previous work has uncovered several RNA-targeting type III CRISPR-associated protease (CASP) systems (*3*, *4*), including a Lon protease that responds to cyclic oligoadenylate second messengers (cA_4) to cleave the CRISPR-T protein (*11*). A recently characterized subtype III-E effector Cas7-11 (*12*, *13*) (also referred to as gRAMP) is likewise associated with a protease, a CHAT family member containing tetratricopeptide repeats (TPR-CHAT, or Csx29). In contrast to prototypical type III CRISPR systems that consist of multisubunit Csm or Cmr complexes (*14*), Cas7-11 effectors contain naturally fused Cas7 and Cas11 domains (*3*). Members of the CHAT family of proteases harbor catalytic cysteine residues and include eukaryotic caspases that are involved in programmed cell death (*15*), and Cas7-11-Csx29 was previously hypothesized to act as a bacterial caspase and support viral immunity (*12*, *13*). Notably, Cas7-11 and Csx29 from *Candidatus Scalindua brodae* were shown to form a stable protein complex (*13*), but the substrate and function of the associated protease is unknown.

Here, we determined the protein substrate, structure, and mechanism of a type III-E CASP system from the marine anaerobe *Desulfonema ishimotonii*, which revealed insights into its natural function in coordinating a transcriptional response to foreign genetic material, and engineered it for new RNA sensing applications in vitro and in human cells.

A Cas7-11–Csx29 complex cleaves the Csx30 protein

The reported cleavage of CRISPR-T by the neighboring Lon protease (*11*) inspired us to look more closely at type III-E loci for potential substrates. In addition to the associated Csx29 protease, these loci frequently contain three additional genes [*csx30*, *csx31*, and a predicted sigma factor (*3*), hereafter CASP-σ] that we hypothesized were prime candidates (Fig. 1A and fig. S1). Starting from a system found in *D. ishimotonii* (DiCASP) (*12*), we purified a stable Cas7-11–Csx29–CRISPR RNA (crRNA) complex [as previously reported for *Candidatus S. brodae* (*13*)] (fig. S2A) and performed in vitro reactions by adding the proteins expressed from the three upstream genes in the presence or absence of a target RNA complementary to the crRNA. We identified that the largest protein, Csx30, is specifically cleaved in response to a target RNA (Fig. 1, B and C). Moreover, in vitro reactions yielded two precise protein products, indicating a single cleavage event within Csx30 as opposed to processive protein degradation.

We determined the requirements of Csx30 cleavage and found that whereas mutating the catalytic residues of the Csx29 protease [H615A/C658A (His^{615}→Ala/Cys^{658}→Ala)] abolished activity, disrupting the catalytic sites of the Cas7-11 endonuclease (dCas7-11:D429A/D654A; D, Asp) (*12*) did not (Fig. 1D and fig. S2B). This result indicates that target RNA binding alone is sufficient for Csx29 activation and that RNA cleavage is dispensable. In vitro characterization revealed that DiCASP is a highly active adenosine triphosphate (ATP)–independent protease that can cleave a 100-fold molar excess of Csx30 substrate in minutes, with an optimal activity between 37° and 45°C (fig. S2, C to F). Full Csx30 cleavage activity required 22 nucleotides of complementarity between the crRNA and target RNA, and we detected low tolerance to base-pair mismatches, particularly at the 5′ end of the target RNA (fig. S3).

Characterization of Csx30 proteolytic processing

Structural prediction of the Csx30 protein revealed two domains separated by a flexible linker (Fig. 1, E and F), which we hypothesized to be the site of cleavage. However, mass spectrometry analysis (and the estimated 48- and 16-kDa gel products) indicates that Csx30 is cleaved further downstream between residues 427 and 429 (fig. S4), placing the cleavage site within a small flexible loop (residues 423 to 437) in the C-terminal domain (CTD) of the structural model. By generating truncation mutations of Csx30, we determined that the N-terminal domain (NTD) is dispensable for processing by Cas7-11–Csx29 because Csx30 fragments containing residues 396 to 565 were efficiently cleaved in vitro (Fig. 1G and fig. S5). By contrast, we observed that Csx30 C-terminal residues are strictly required and that even a

[1]Howard Hughes Medical Institute, Massachusetts Institute of Technology, Cambridge, MA 02139, USA. [2]Broad Institute of MIT and Harvard, Cambridge, MA 02142, USA. [3]McGovern Institute for Brain Research, Massachusetts Institute of Technology, Cambridge, MA 02139, USA. [4]Department of Brain and Cognitive Sciences, Massachusetts Institute of Technology, Cambridge, MA 02139, USA. [5]Department of Biological Engineering, Massachusetts Institute of Technology, Cambridge, MA 02139, USA. [6]Department of Electrical Engineering and Computer Science, Massachusetts Institute of Technology, Cambridge, MA 02139, USA. [7]Structural Biology Division, Research Center for Advanced Science and Technology, The University of Tokyo, Tokyo 153-8904, Japan. [8]Department of Chemistry and Biotechnology, Graduate School of Engineering, The University of Tokyo, Tokyo 113-8656, Japan. [9]Department of Biological Sciences, Graduate School of Science, The University of Tokyo, Tokyo 113-0033, Japan. [10]Inamori Research Institute for Science, Kyoto 600-8411, Japan.
*Corresponding author. Email: zhang@broadinstitute.org (F.Z.); strecker@broadinstitute.org (J.S.)
†These authors contributed equally to this work.

Downloaded from https://www.science.org at Shanghai Normal University on November 25, 2022

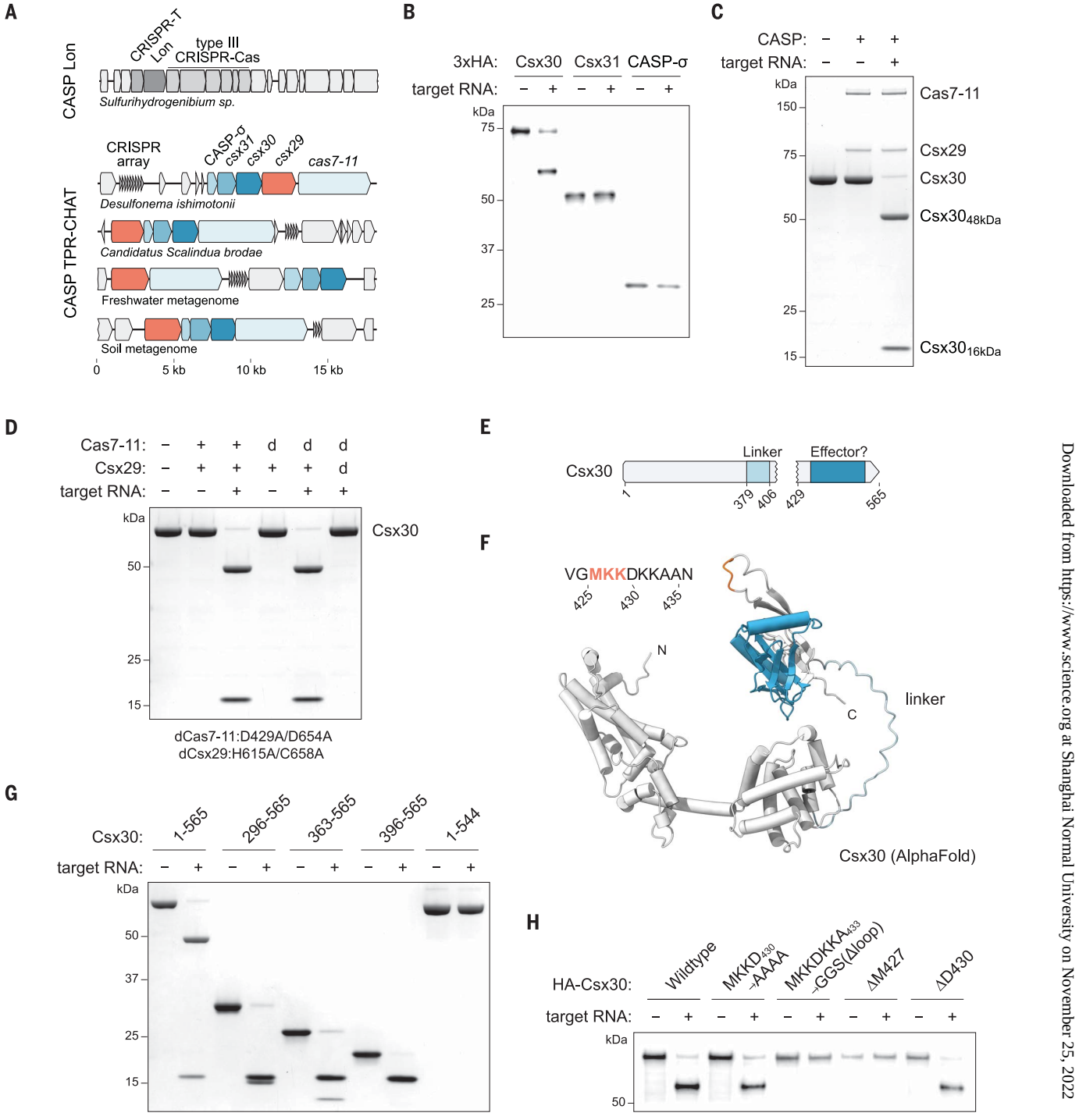


Fig. 1. The type III-E CASP Csx29 cleaves Csx30. (**A**) Schematic of selected CASP loci and three additional conserved genes in type III-E loci. (**B**) Immunoblot analysis of in vitro reactions with Cas7-11–Csx29 and HA-tagged Csx30, Csx31, and CASP-σ produced by cell-free transcription-translation. (**C**) A Cas7-11–Csx29–crRNA complex cleaves Csx30 protein in response to target RNA. (**D**) Csx30 cleavage requires target RNA and the Csx29 protease catalytic residues but not the catalytic residues of Cas7-11. (**E**) Schematic of Csx30 highlighting the cleavage site (amino acids 427 to 429), linker (amino acids 377 to 406), and a potential effector domain annotated from HHpred (amino acids 452 to 545). (**F**) AlphaFold2 prediction of Csx30. N, NTD; C, CTD. (**G**) Analysis of dCas7-11–Csx29 (Cas7-11: D429A/D654A) proteolytic activity on truncated Csx30 proteins. (**H**) Immunoblot analysis of HA-tagged Csx30 mutants produced by cell-free transcription-translation. G, Gly. (C), (D), and (G) are SDS–polyacrylamide gel electrophoresis (SDS-PAGE) gels stained with Coomassie brilliant blue.

20–amino acid truncation ($Csx30_{1\text{-}544}$) abolished cleavage activity (Fig. 1G).

Mutational analysis by alanine substitutions revealed no Csx30 residues that are essential for cleavage, although some reduced the efficiency (Fig. 1H and fig. S6). Instead, the size of the cleaved loop appears important for processing. We observed that truncating the loop by four residues, or deleting M427 (M, Met) alone, prevented Csx30 cleavage, whereas the deletion of D430 had no effect (Fig. 1H). Using an uncleavable $Csx30_{\Delta loop}$

Downloaded from https://www.science.org at Shanghai Normal University on November 25, 2022

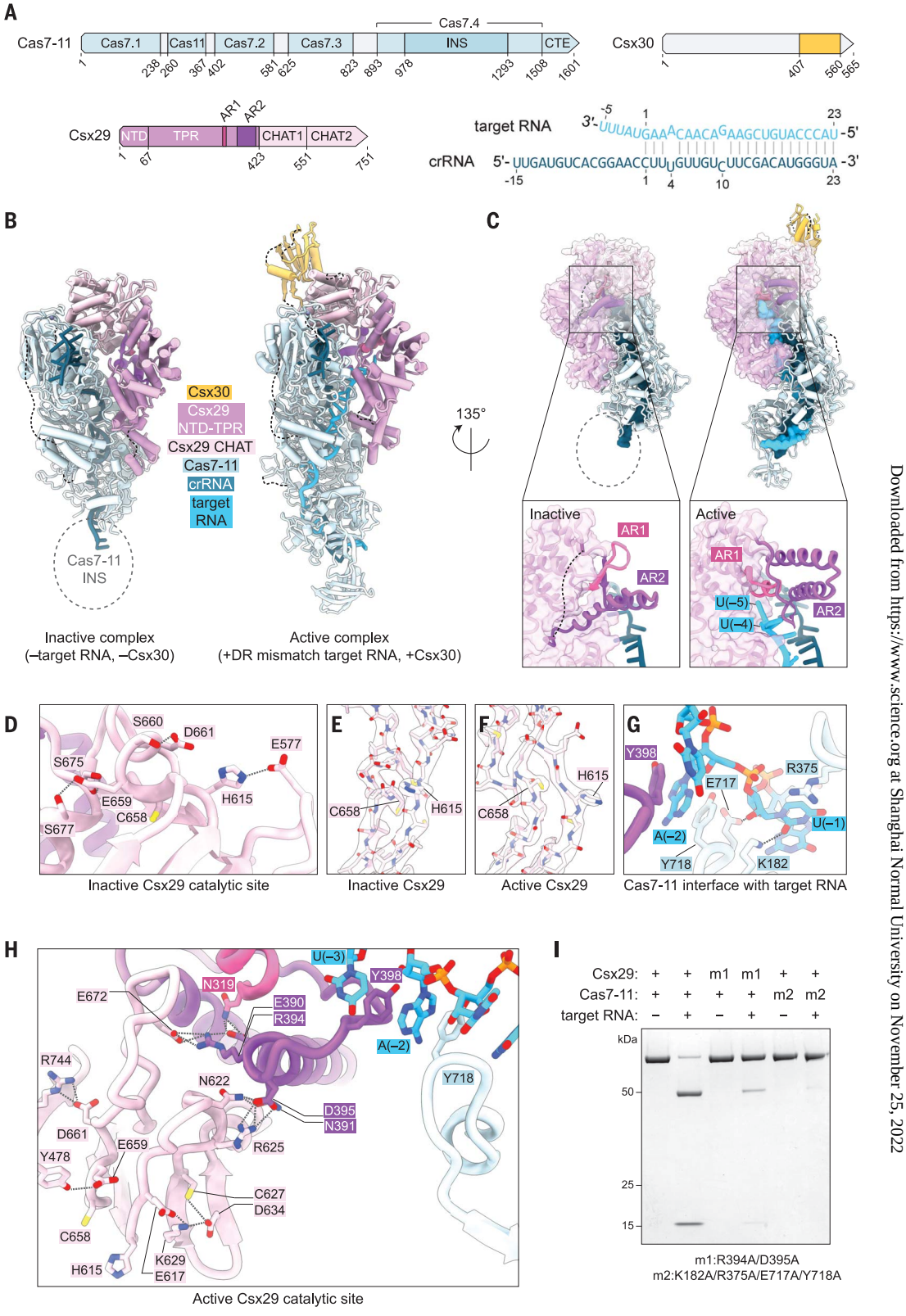


Fig. 2. Allosteric activation of Csx29 upon RNA binding. (**A**) Schematic of Cas7-11, Csx29, and Csx30 proteins domains and the crRNA and target RNA used in the structural studies. (**B**) Structures of the inactive (Cas7-11–Csx29–crRNA) and active (Cas7-11–Csx29–crRNA–target RNA–Csx30) CASP complexes. (**C**) Structural organization of the Csx29 AR in the inactive and active CASP complexes. The dashed circle indicates the Cas7-11 INS domain, which is also indicated in (B). (**D**) Electrostatic and hydrogen-bonded network within the Csx29 catalytic site in the inactive state. (**E** and **F**) Catalytic H615 and C658 residues in inactive and active Csx29 shown with EM density. (**G**) Contacts between Cas7-11 and the DR-mismatched portion of the target RNA in the active state. (**H**) Electrostatic and hydrogen-bonded network extending from the AR to the Csx29 catalytic site in the active state. (**I**) SDS-PAGE gel stained with Coomassie brilliant blue showing that mutations that disrupt allosteric activation residues impair Csx30 cleavage by Cas7-11–Csx29.

mutant as bait, we pulled down the Cas7-11–Csx29 complex both in the presence and absence of target RNA, suggesting that Csx30 binding to Cas7-11–Csx29 is not regulated by target RNA recognition or activation of the protease (fig. S7). By contrast, we did not detect Cas7-11–Csx29 binding using a truncated $Csx30_{1\text{-}544}$ mutant, revealing that an intact CTD is required for substrate binding (fig. S7).

Allosteric activation of Csx29 upon target RNA binding

To gain insight into the activation mechanism of Cas7-11–Csx29 and substrate recognition of Csx30, we solved single-particle cryo–electron

Downloaded from https://www.science.org at Shanghai Normal University on November 25, 2022

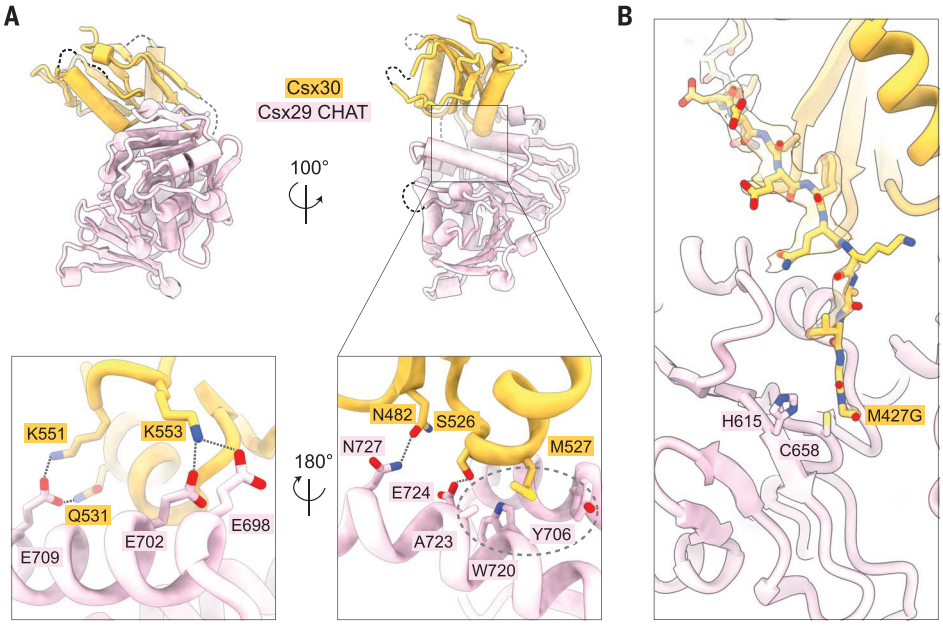


Fig. 3. Csx30 substrate recognition by Csx29. (**A**) Csx29-Csx30 interface in the active CASP structure. Electrostatic interactions and hydrogen bonds are drawn as dashed lines, and the hydrophobic pocket is drawn as a dashed oval. (**B**) Close-up of the Csx29-Csx30 interface near the catalytic H615 and C658 residues.

microscopy (cryo-EM) structures of $Csx30_{\Delta loop}$ bound to Cas7-11–Csx29 with target RNA and an inactive complex of Cas7-11–Csx29 alone at 2.5- and 3.0-Å resolution, respectively (Fig. 2, A to C; figs. S8 to S10; and table S1). The overall architecture of Cas7-11 in both complexes resembles the reported DiCas7-11 structure (*16*), in which the Cas7.1 to Cas7.4 domains organize into a filament around the crRNA core with Cas11 at the midpoint. The insertion (INS) domain within Cas7.4 was visible only in the active state (Fig. 2, B and C). Csx29 consists of a three-helix bundle NTD, a TPR domain with eight repeats, and a protease region containing a pseudo-caspase (CHAT1) and active-caspase (CHAT2) domain that resembles that of separases (*17*, *18*). In both complexes, Cas7.2 to Cas7.4 interface with the NTD, TPR, and CHAT1 domains of Csx29. Although the overall organization of Cas7-11 remains the same upon Csx29 binding, linker L2 and the Cas7.4 zinc-finger loop undergo structural changes that look similar in both the active and inactive states (fig. S11).

In the inactive state, the catalytic residues of CHAT2 are improperly positioned; C658 is turned downward away from the catalytic H615, and the catalytic histidine is instead positioned toward D661 (fig. S12). However, they are repositioned upon target RNA binding to resemble the geometry of active caspases (Fig. 2, D to F, and figs. S12 and S13). Because CHAT2 makes no direct contact with Cas7-11 or target RNA, we hypothesized that conformational changes likely occur in other regions of Csx29 and transduce an allosteric signal to the catalytic core. By comparing the inactive and active complexes we observed a major structural change within the eighth repeat of the TPR domain, which we term the activation region (AR). The AR is bipartite, composed of AR1 (amino acids 313 to 325) and AR2 (amino acids 356 to 411), which stack with each other in the inactive state (Fig. 2C). In the active complex, AR1 senses the 3′ end of target RNA (positions −4 and −5) through base-stacking interactions and pushes the AR2 helices away, preventing a steric clash (Fig. 2C).

The target RNA in our active complex is noncomplementary to the direct repeat (DR), and the structure reveals that this is an important feature. In this state, the 3′ portion of the target RNA is separated from the crRNA, and it makes a sharp kink at position −2, enabling it to traverse the TPR domain of Csx29 and reach AR1 (fig. S14A). This observation suggests that a DR-matched RNA might not activate Csx29 because it could stay hybridized with the crRNA at position −2 and beyond. Supporting this model, a target RNA that fully matched the DR strongly reduced Csx30 cleavage (fig. S14, B and C). Mismatches at positions −1 and −2 alone were only able to partially activate Csx29, and mismatches at positions −1 to −4 were required to restore full Csx30 cleavage (fig. S14C). Eliminating base pairing between the DR and the target RNA is therefore crucial for CASP activation and highlights the importance of the AR1–target RNA interaction. Of note, noncomplementarity between the DR and target RNA also plays an important role in type III-A and III-B CRISPR systems to suppress the response against host-derived transcripts (*19*, *20*) and thus is a generalized component of signal transduction in type III systems.

In addition to target RNA sensing by Csx29 AR1, we identified contacts between Cas7-11 and target RNA at the DR-mismatched site. In addition to Y718, which base-stacks with the nucleotide at position −2, we identified K182, R375, and E717 as contacting the nucleotide at position −1 (Y, Tyr; K, Lys; R, Arg; E, Glu) (Fig. 2G and fig. S13). To better understand CASP activation and the AR-induced signal transduction, we examined downstream allosteric events in Csx29. In the active complex, the kinked target RNA site at position −2 is stabilized by base-stacking interactions, provided by both Cas7-11–Y718 and Csx29-Y398 within AR2. Adjacent residues at the tip of the AR2 helix—E390, N391, R394, and D395—initiate a network of electrostatic and hydrogen-bonded contacts that extend all the way to the CHAT2 active site (N, Asn) (Fig. 2H and fig. S13). Prominent salt bridges formed between R394-E672 and D395-R625 help position the loop that contains the catalytic C658 and the strand that contains the catalytic H615, respectively. Further down, the active-site H615 is positioned by E617 contacts, whereas the active-site C658 is kept in place by E659-Y478 and D661-R744. In the inactive state, these same residues that position C658 in the active complex make entirely different contacts: E659 forms hydrogen bonds with S675 and S677, and D661 instead bonds with S660 (S, Ser) (Fig. 2, D and H, and fig. S13). We note the similarity of this mechanism to that in eukaryotic caspases, which are also thought to be regulated by the conformation of the L4 loop that contains their catalytic cysteine (*21*). Together, these structures reveal an allosteric cascade initiated by the 3′ end of DR-mismatched target RNA, which triggers the AR within the Csx29 TPR domain and transduces structural changes to the Csx29 CHAT2 domain to coordinate active-site residues.

Downloaded from https://www.science.org at Shanghai Normal University on November 25, 2022

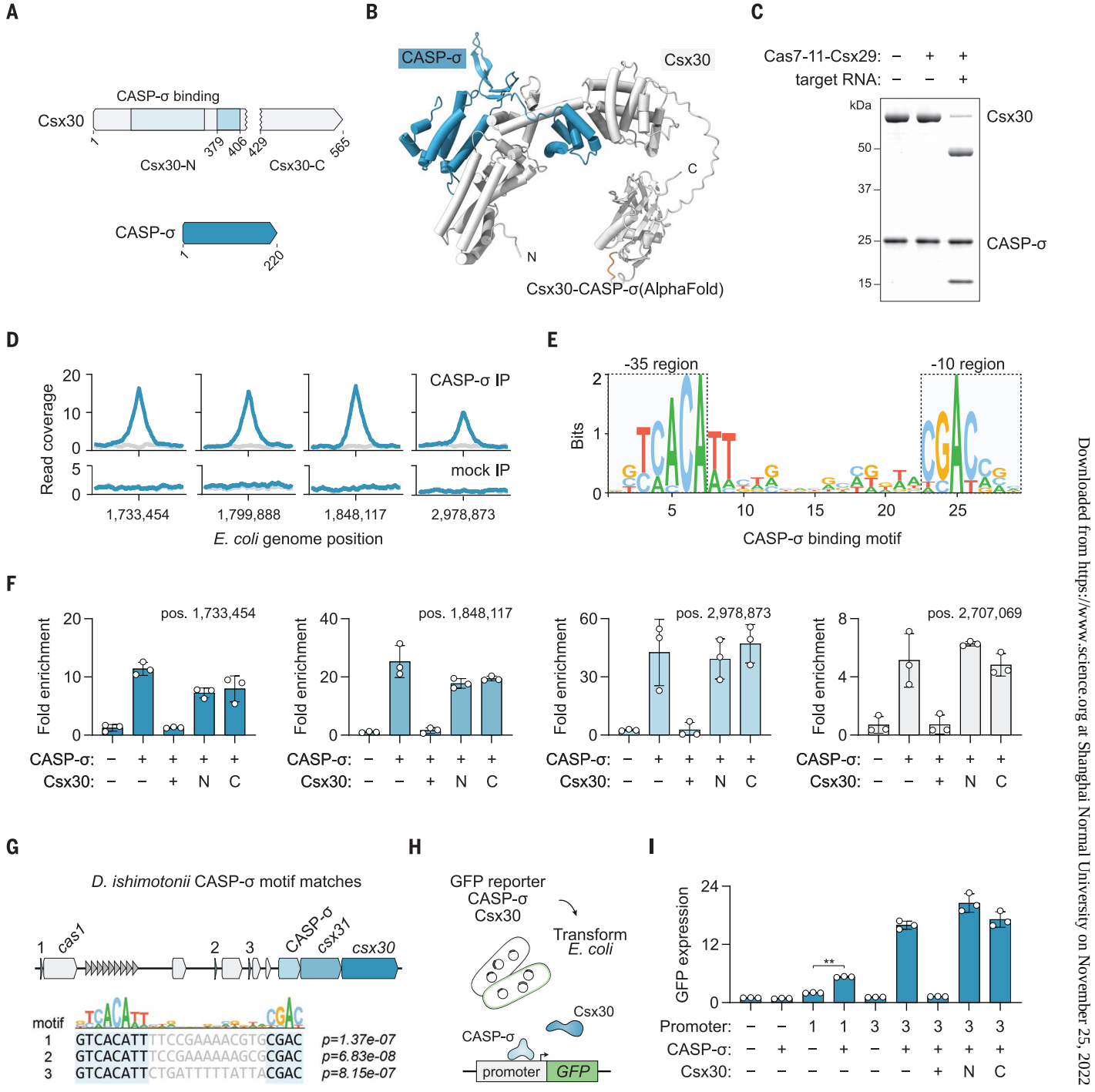


Fig. 4. Csx30 binds and inhibits the transcription factor CASP-σ. (**A**) Schematic of Csx30 and CASP-σ proteins. (**B**) AlphaFold2 prediction of a Csx30-CASP-σ interaction. (**C**) SDS-PAGE gel stained with Coomassie brilliant blue showing purification of a Csx30–CASP-σ complex that is cleaved by dCas7-11–Csx29. (**D**) Representative CASP-σ ChIP-seq peaks in *E. coli* with a 1-kb window; input coverage is shown in gray. (**E**) Identification of a CASP-σ binding motif from ChIP-seq peaks. (**F**) Enrichment of CASP-σ at four *E. coli* peaks by ChIP–quantitative polymerase chain reaction (qPCR). $n = 3$ replicates. pos., *E. coli* genome position. (**G**) Predicted CASP-σ targets in the *D. ishimotonii* CASP locus. (**H**) Schematic of the CASP-σ transcriptional reporter assay. (**I**) CASP-σ–mediated transcriptional activity in *E. coli*. GFP expression was normalized to cells with a scrambled promoter sequence. $n = 3$ replicates. $^{**}p < 0.01$ using Student's t test. In (F) and (I), error bars represent standard deviation from the mean.

To test this model, we made mutations in the allosteric network. A Csx29-R394A/D395A double mutant within AR2 formed a stable Cas7-11–Csx29 complex, but Csx30 cleavage was greatly impaired (Fig. 2I and fig. S14D). Further down the allosteric cascade, mutating Csx29 residues E659 and D661 in the vicinity of the catalytic C658 likely disrupted Csx29 folding, and we were unable to purify a Cas7-11–Csx29 complex. Finally, we tested the importance of contacts between Cas7-11 and target RNA at the DR-mismatched site. Mutating

Downloaded from https://www.science.org at Shanghai Normal University on November 25, 2022

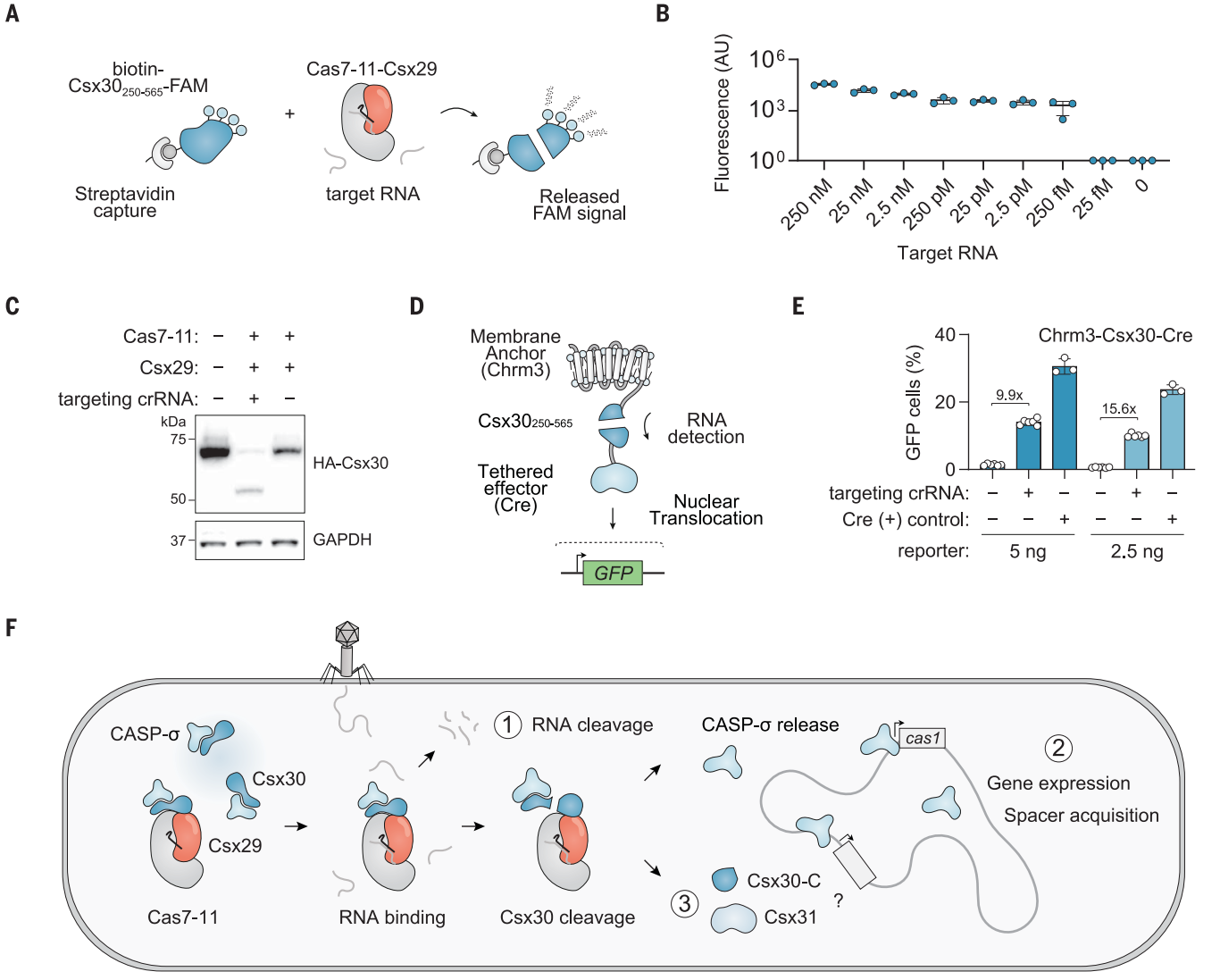


Fig. 5. RNA-sensing applications and a proposed model for CASP systems. (**A**) Schematic of in vitro RNA detection using CASP systems and fluorescent Csx30 reporters. FAM, 5′ 6-FAM (fluorescein). (**B**) In vitro detection of RNA as measured by released fluorescence. n = 3 replicates. AU, arbitrary units. (**C**) Immunoblot analysis of HA-tagged Csx30 in HEK293T human cells transfected with DiCASP components. GAPDH, glyceraldehyde phosphate dehydrogenase. (**D**) Schematic of engineered proteins containing a cell membrane tether, a Csx30 linker, and an effector domain. (**E**) Flow cytometry analysis of DiCASP activity in mouse Neuro2A loxP:GFP cells using a Chrm3-$Csx30_{250\text{-}565}$-Cre reporter. n = 3 to 6 replicates. (**F**) Model for a three-pronged strategy of CASP systems in the defense against foreign genetic elements, including Cas7-11–mediated RNA endonuclease activity, a Csx30-regulated CASP-σ transcriptional response, and a possible third arm involving Csx31. In (B) and (E), error bars represent standard deviation from the mean.

Cas7-11 residues K182, E717, R375, and Y718 into alanines did not impair Cas7-11–Csx29 complex assembly but strongly reduced CASP activation upon target RNA binding (Fig. 2I and fig. S14D). Thus, target RNA stabilization by Cas7-11 on the DR-mismatched end is also critical for protease activation.

Csx30 recognition by Cas7-11–Csx29

In addition to revealing insight into CASP activation, our active complex also provides structural details regarding the interaction with Csx30. Despite using a full-length $Csx30_{\Delta loop}$ mutant for complex assembly, only a small portion (amino acids 407 to 560) is visible in our structure (Fig. 3A and fig. S15A), and the remaining residues must therefore be flexible with respect to Cas7-11–Csx29. This region mirrors the minimal substrate we identified through truncation experiments and confirms that recognition of Csx30 is mediated through its CTD. In our structure, Csx30 is bound only to the Csx29 CHAT2 domain and does not interact with Cas7-11.

There is notable charge complementarity at the Csx29-Csx30 interface, and substrate recognition is likely electrostatically driven through the negatively charged surface of Csx29 and positively charged surface of Csx30 (fig. S15B). Detailed analysis of the interface reveals that Csx30 polar and positively charged residues (N482, S526, Q531, K551, and K553; Q, Gln) make contact with the Csx29 CHAT2 domain (Fig. 3A and fig. S16). In addition, Csx30-M527 is enclosed in a tight hydrophobic pocket lined with Csx29's Y706, W720, and A723 (W, Trp). The major determinant of Csx30 engagement is likely a cumulative effect of these interactions, because mutating individual regions of the Csx29-Csx30 interface did not affect Csx30 cleavage (fig. S15C). Consistent with our ability to pull

Downloaded from https://www.science.org at Shanghai Normal University on November 25, 2022

down a Cas7-11–Csx29–Csx30$_{\Delta loop}$ complex in the presence and absence of target RNA (fig. S7), the interfacing residues of Csx29 adopt a similar organization in both the active and inactive complexes, and therefore we conclude that Csx30 binding is not allosterically regulated.

We also examined the position of the Csx30 cleavage site within the active complex. One limitation of our structure is that the cleavage loop is mutated (and slightly shortened), and thus, we cannot observe substrate engagement in the active site in detail. Because the loop is also flexible, it is not well resolved, but its density places it near the active site of Csx29, positioning it for cleavage (Fig. 3B).

Csx30 binds and inhibits the transcription factor CASP-σ

We next sought to explore the biological function of Csx30 and understand how cleavage might regulate its activity. Because the Cas7-11 effector alone provides defense against phage (*12*), we reasoned that additional functions of DiCASP would similarly be involved in the immune response. One possibility is that processed Csx30 fragments, Csx30-N (residues 1 to 428) or Csx30-C (residues 429 to 565), promote cell death or an abortive infection response to prevent phage propagation. However, we did not observe defense against three tested phage (fig. S17A). Homology searches revealed a match of Csx30-C to a peptidoglycan *N*-acetylglucosamine deacetylase (HHpred probability: 92.85%; e-value: 0.56), but we did not detect modification of peptidoglycan or its components with cleaved Csx30 in vitro (fig. S17B). Overexpression of Csx30 fragments was not toxic in *Escherichia coli*, and we only observed a slight growth defect in cells expressing full-length Csx30, which was temperature dependent and suppressed by the addition of Csx31 and CASP-σ (fig. S18).

We next turned to the other proteins encoded in the locus to gain insight into Csx30 function. We predicted a strong binding interaction between the NTD of Csx30 and CASP-σ, which strongly resembles an extracytoplasmic function (ECF) sigma factor (*3*) (HHpred probability: 100%; e-value: 3.4×10^{-31}) (Fig. 4, A and B, and fig. S19). Sigma factors are transcription-initiation proteins that bind DNA and recruit the RNA polymerase catalytic core to specific promoters (*22*), hinting that Csx30 might be involved in regulating a transcriptional response. Consistent with our computational prediction, purification of CASP-σ in the presence of Csx30 yielded a Csx30–CASP-σ complex, in which Csx30 could still be cleaved by Cas7-11–Csx29 (Fig. 4C). Csx30-N was sufficient for the interaction with CASP-σ, although at considerably lower yield (fig. S20).

Although *D. ishimotonii* CASP-σ is unlikely to regulate its target genes heterologously in *E. coli*, we reasoned that the identification of putative CASP-σ binding sites might yield insight into its preferred sequence motif and function in the natural host. We performed chromatin immunoprecipitation sequencing (ChIP-seq) in *E. coli* with hemagglutinin (HA)–tagged CASP-σ and identified 13 high-confidence peaks compared with input and mock immunoprecipitation (IP) controls (Fig. 4D and fig. S21A). Motif analysis of ChIP-seq peaks yielded a clear hit (Fig. 4E and fig. S21B), which was similar to a de novo predicted motif (fig. S21C) (*23*).

Sigma factors are frequently regulated by inhibitors (anti-sigma factors), and there are examples in bacteria in which a protease cleaves an anti-sigma factor to activate a transcriptional stress response, including the anti-sigma factors RseA in *E. coli* (*24*) and RsiW in *Bacillus subtilis* (*25*). In *E. coli*, the DegS protease senses cell envelope stress and cleaves a transmembrane segment of RseA (*26*), resulting in the eventual release of the sequestered sigma factor RpoE. Based on our structural model, we predict that the Csx30–CASP-σ interaction would block CASP-σ DNA binding based on steric clashes to sigma factor–bound DNA in experimental structures (*27*) (fig. S22). To test whether Csx30 inhibits CASP-σ, we repeated ChIP experiments in *E. coli* coexpressing Csx30 and found that CASP-σ DNA binding was blocked at all four tested loci (Fig. 4F). This inhibition was dependent on full-length Csx30 because both Csx30-N and Csx30-C fragments were unable to antagonize CASP-σ binding (Fig. 4F). Together, our results suggest that Csx30 is an inhibitor of CASP-σ and that processing by Cas7-11–Csx29 alleviates this inhibition.

Csx30 processing regulates CASP-σ transcriptional activity

We next sought to identify potential CASP-σ targets in the natural host *D. ishimotonii*. Because many ECF sigma factors autoregulate their own expression (*28*), we first searched the DiCASP locus. We identified three strong matches in the promoters of *cas1* and two genes of unknown function (Fig. 4G and table S2), indicating that CASP-σ likely coordinates additional defense functions, including CRISPR spacer acquisition. Genome-wide searches for motifs in *D. ishimotonii* promoter regions yielded several candidates, although only one site, upstream of the *nhaA* gene, was below a *q* value of 0.6 (tables S3 and S4). To test these predictions, we constructed transcriptional reporters by placing putative CASP-σ promoters upstream of green fluorescent protein (GFP) and measured the resulting fluorescence in *E. coli* (Fig. 4H and fig. S23, A and B). We observed GFP expression with both tested promoters as compared with a random DNA control and found that fluorescence was fully dependent on CASP-σ expression (Fig. 4I). Consistent with our previous results, coexpression of full-length Csx30 was able to completely inhibit CASP-σ–mediated GFP expression, whereas processed Csx30 fragments had no effect (Fig. 4I). In support of a role for CASP-σ–mediated transcription in the immune response, we could computationally identify one of the two unknown open reading frames—a predicted membrane protein—in other CRISPR and defense loci (fig. S23C).

RNA-sensing applications with DiCASP

The high proteolytic activity of Cas7-11–Csx29 in response to a target RNA enables numerous biological applications. In addition, the ability to uncouple RNA cleavage from activation of the Csx29 protease allows for nondestructive sensing of RNA. Although the collateral nuclease activity of CRISPR effectors has been used to cleave nucleic acid–based reporters for diagnostic applications (*29*), CASP systems allow for a new modality of substrates using engineered Csx30 proteins. As a proof of concept, we generated a fluorescently labeled engineered variant of Csx30 and demonstrated its ability to detect RNA in vitro down to 250 fM without nucleic acid amplification (Fig. 5, A and B, and fig. S24).

We also sought to apply DiCASP for RNA-transcript sensing in live cells. To determine whether DiCASP can mediate RNA-activated proteolytic cleavage in human cells, we transfected plasmids expressing Cas7-11, Csx29, crRNA, a synthetic target RNA, and Csx30 fused to an HA epitope tag into human embryonic kidney (HEK) 293T cells. Immunoblots of cell lysate revealed processing of Csx30 that was dependent on a targeting crRNA and the catalytic residues of the Csx29 protease (Fig. 5C and fig. S25, A and B). Testing DiCASP activity across a panel of endogenous transcripts revealed Csx30 cleavage efficiencies ranging from 2 to 20% (fig. S25, C and D).

To convert RNA sensing with DiCASP into a discrete and readily detectable signal, we sought to design reporters containing effector domains that could be activated by Csx30 cleavage. We transfected plasmids encoding a fusion protein in which Cre recombinase is tethered to membrane anchors [e.g., the cholinergic receptor, muscarinic 3 (Chrm3) G protein–coupled receptor] by a Csx30-derived linker, sequestering Cre from the nucleus (Fig. 5D). Mouse Neuro-2A cells harboring an inactive loxP-GFP reporter cassette were transfected with DiCASP components and synthetic target RNA. Flow cytometry analysis revealed crRNA-dependent GFP expression in 10% of cells and a 15-fold increase over nontargeting crRNA controls under optimal conditions (Fig. 5E and fig. S25, E and F).

Downloaded from https://www.science.org at Shanghai Normal University on November 25, 2022

Discussion

Here, we demonstrate that the Csx29 protease associated with the type III-E effector Cas7-11 mediates RNA-activated endopeptidase activity and elucidate its substrate, structure, and mechanism. Although the full biological consequence of Csx30 processing in *D. ishimotonii* is unknown, our work supports a model in which Csx30 inhibits the sigma factor CASP-σ and proteolytic cleavage by the Csx29 protease acts to relieve this inhibition. The parallels between DiCASP and other protease-regulated anti-sigma factors, such as DegS and RseA (*26*), reveal convergent mechanisms for modulating gene expression in response to cellular threats. The NTD of Csx30 is sufficient for binding to CASP-σ, and it is therefore unclear how proteolytic cleavage within the Csx30 CTD would release CASP-σ or why expression of Csx30-N is unable to inhibit CASP-σ. One possibility is that the processed Csx30 fragments are unstable and that the exposed termini are subject to further degradation by host proteins. Consistent with this hypothesis, immunoblots of *E. coli* cell lysates that harbor HA-tagged isoforms of Csx30 revealed expression of full-length Csx30 and Csx30-C, but not Csx30-N, and blocking the "cleaved" termini with an epitope tag increased expression (fig. S26). We note potential similarities to other protease-regulated anti-sigma factor systems; DegS cleavage of RseA is insufficient to release the sigma factor RpoE, and the remaining RseA fragment is further processed by the RseP (*30*, *31*) and ClpXP proteases (*32*) to liberate RpoE.

Our identification of three CASP-σ binding motifs within the CASP locus points to the positive autoregulation of defense genes, including *cas1*, which may be a mechanism to acquire new spacers during active infection and to safeguard against the acquisition of self-targeting spacers during normal growth. This result is consistent with the reported up-regulation of *cas1* in *Pseudomonas aeruginosa* by the ECF sigma factor PvdS (*33*). The functions of the two other predicted up-regulated genes in the locus are unknown, although one has strong homology to a membrane transporter component EcsC (HHpred probability: 99.9; e-value: 3.1×10^{-22}). Interestingly, the top motif match outside of the CASP locus is upstream of *nhaA* (table S3), a Na^+-H^+ antiporter known to be up-regulated during phage infection (*34*), indicating that CASP-σ may also regulate targets elsewhere in the genome.

Together, our results suggest that the subtype III-E CASP systems use a three-pronged strategy to defend against foreign genetic material: (i) targeted RNA cleavage through the RNA endonuclease Cas7-11; (ii) a Csx30–CASP-σ regulated transcriptional response that leads to, among other possibilities, spacer acquisition; and (iii) a potential third arm mediated by Csx31 and possibly Csx30-C (Fig. 5F). The clear conservation of Csx31 (fig. S1) is a strong indication of its biological importance, and future work will be required to determine its role in the immune response.

We predict similar interactions between Csx30 and CASP-σ in other type III-E systems as well as putative CASP-σ binding motifs at *cas1* within the *Candidatus S. brodae* locus (fig. S27). There may also be parallels between DiCASP and the type III CRISPR-associated Lon protease (*11*). We note that CRISPR-T is also associated with a neighboring sigma factor and is predicted to physically interact (fig. S28). We hypothesize that cleavage of CRISPR-T could similarly trigger transcriptional changes and may reflect a common functional theme across diverse CASP families. This work reveals an example of CRISPR systems coordinating a wider cellular response beyond nuclease activity, and we expect that the continued investigation of CRISPR-associated enzymes will uncover many interesting, and potentially useful, RNA-activated biological processes.

REFERENCES AND NOTES

1. A. Bernheim, R. Sorek, *Nat. Rev. Microbiol.* **18**, 113–119 (2020).
2. L. Gao *et al.*, *Science* **369**, 1077–1084 (2020).
3. K. S. Makarova *et al.*, *Nat. Rev. Microbiol.* **18**, 67–83 (2020).
4. S. A. Shmakov, K. S. Makarova, Y. I. Wolf, K. V. Severinov, E. V. Koonin, *Proc. Natl. Acad. Sci. U.S.A.* **115**, E5307–E5316 (2018).
5. S. A. Shah *et al.*, *RNA Biol.* **16**, 530–542 (2019).
6. J. E. Peters, K. S. Makarova, S. Shmakov, E. V. Koonin, *Proc. Natl. Acad. Sci. U.S.A.* **114**, E7358–E7366 (2017).
7. G. Faure *et al.*, *Nat. Rev. Microbiol.* **17**, 513–525 (2019).
8. J. Strecker *et al.*, *Science* **365**, 48–53 (2019).
9. S. E. Klompe, P. L. H. Vo, T. S. Halpin-Healy, S. H. Sternberg, *Nature* **571**, 219–225 (2019).
10. E. V. Koonin, K. S. Makarova, *PLOS Biol.* **20**, e3001481 (2022).
11. C. Rouillon *et al.*, bioRxiv 2021.12.06.471393 [Preprint] (2021). https://doi.org/10.1101/2021.12.06.471393.
12. A. Özcan *et al.*, *Nature* **597**, 720–725 (2021).
13. S. P. B. van Beljouw *et al.*, *Science* **373**, 1349–1353 (2021).
14. J. van der Oost, E. R. Westra, R. N. Jackson, B. Wiedenheft, *Nat. Rev. Microbiol.* **12**, 479–492 (2014).
15. L. Aravind, E. V. Koonin, *Proteins* **46**, 355–367 (2002).
16. K. Kato *et al.*, *Cell* **185**, 2324–2337.e16 (2022).
17. A. Boland *et al.*, *Nat. Struct. Mol. Biol.* **24**, 414–418 (2017).
18. Z. Lin, X. Luo, H. Yu, *Nature* **532**, 131–134 (2016).
19. L. You *et al.*, *Cell* **176**, 239–253.e16 (2019).
20. N. Sofos *et al.*, *Mol. Cell* **79**, 741–757.e7 (2020).
21. K. McLuskey, J. C. Mottram, *Biochem. J.* **466**, 219–232 (2015).
22. A. Feklístov, B. D. Sharon, S. A. Darst, C. A. Gross, *Annu. Rev. Microbiol.* **68**, 357–376 (2014).
23. H. Todor *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **117**, 33496–33506 (2020).
24. N. P. Walsh, B. M. Alba, B. Bose, C. A. Gross, R. T. Sauer, *Cell* **113**, 61–71 (2003).
25. S. Schöbel, S. Zellmeier, W. Schumann, T. Wiegert, *Mol. Microbiol.* **52**, 1091–1105 (2004).
26. S. E. Ades, L. E. Connolly, B. M. Alba, C. A. Gross, *Genes Dev.* **13**, 2449–2461 (1999).
27. W. J. Lane, S. A. Darst, *PLOS Biol.* **4**, e269 (2006).
28. D. Casas-Pastor *et al.*, *Nucleic Acids Res.* **49**, 986–1005 (2021).
29. J. S. Gootenberg *et al.*, *Science* **356**, 438–442 (2017).
30. B. M. Alba, J. A. Leeds, C. Onufryk, C. Z. Lu, C. A. Gross, *Genes Dev.* **16**, 2156–2168 (2002).
31. K. Kanehara, K. Ito, Y. Akiyama, *Genes Dev.* **16**, 2147–2155 (2002).
32. J. M. Flynn, I. Levchenko, R. T. Sauer, T. A. Baker, *Genes Dev.* **18**, 2292–2301 (2004).
33. S. D. Ahator, W. Jianhe, L.-H. Zhang, bioRxiv 2020.01.31.929752 [Preprint] (2020). https://doi.org/10.1101/2020.01.31.929752.
34. L. M. Malone, H. G. Hampton, X. C. Morgan, P. C. Fineran, *Nucleic Acids Res.* **50**, 160–174 (2022).
35. J. Strecker, D. Li, F. Zhang, Code and processed data for: RNA-activated protein cleavage with a CRISPR-associated endopeptidase, Version 1.0, Zenodo (2022). https://doi.org/10.5281/zenodo.7221526.

ACKNOWLEDGMENTS

We thank L. Gao, J. Schmid-Burgk, and D. Strebinger for valuable discussions; B. Lash for providing Neuro2A cells; E. Spooner and the Whitehead Institute Proteomics Core Facility for mass spectrometry analysis; E. Brignole and C. Borsa at the MIT.nano CryoEM Center; staff at the Arnold and Mabel Beckman Foundation for funding the MIT.nano CryoEM Center; and the entire Zhang lab for support and advice. **Funding:** F.Z. is supported by NIH grants 1DP1-HL141201 and 2R01HG009761-05; the Howard Hughes Medical Institute; the Yang-Tan Molecular Therapeutics Center at McGovern; and the Phillips family and J. and P. Poitras. J.S. is a Charles A. King Trust Postdoctoral Research Fellow. **Author contributions:** J.S. designed and performed experiments; F.E.D. solved cryo-EM structures with input from M.E.W.; J.S. and D.L. analyzed data; G.F. performed bioinformatic analysis and structural predictions; F.Z. supervised the research and experimental design with support from R.K.M.; J.S.G., O.O.A., and H.N. provided input and discussed unpublished work; and J.S., F.E.D., and D.L. wrote the manuscript with input from all authors. **Competing interests:** F.Z. is a scientific advisor for Editas Medicine, Beam Therapeutics, Pairwise Plants, Arbor Biotechnologies, Proof Diagnostics, and Octant. J.S., F.E.D., and F.Z. are inventors on US patent application 63/337,386 submitted by The Broad Institute, which covers programmable protease compositions. **Data and materials availability:** ChIP-seq data are available at the NCBI Sequence Read Archive (BioProject IDs PRJNA888197 and SAMN31210314-21). Code and processed data are available at Zenodo (*35*). Plasmids are available from Addgene. The cryo-EM maps have been deposited in the Electron Microscopy Data Bank with the following codes: EMD-28064 (inactive CASP, focused refinement of Cas7-11, crRNA, and Csx29 NTD), EMD-28070 (inactive CASP, focused refinement of Csx29 TPR-CHAT), EMD-28065 (active CASP, focused refinement of Cas7-11, except INS, crRNA, and target RNA), EMD-28071 (active CASP, focused refinement of Cas7-11 INS), EMD-28072 (active CASP, focused refinement of Csx29 NTD and TPR), and EMD-28073 (active CASP, focused refinement of Csx29 CHAT and Csx30). The coordinates for the composite atomic models have been deposited in the Protein Data Bank under accession codes 8EEX (inactive CASP) and 8EEY (active CASP). **License information:** Copyright © 2022 the authors, some rights reserved; exclusive licensee American Association for the Advancement of Science. No claim to original US government works. https://www.science.org/about/science-licenses-journal-article-reuse

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.add7450
Materials and Methods
Figs. S1 to S28
Tables S1 to S7
References (*36*–*56*)
MDAR Reproducibility Checklist

View/request a protocol for this paper from *Bio-protocol*.

Submitted 30 June 2022; accepted 24 October 2022
10.1126/science.add7450

Downloaded from https://www.science.org at Shanghai Normal University on November 25, 2022

Science

RNA-activated protein cleavage with a CRISPR-associated endopeptidase

Jonathan StreckerF. Esra DemircioğluDavid LiGuilhem FaureMax E. WilkinsonJonathan S. GootenbergOmar O. AbudayyehHiroshi NishimasuRhiannon K. MacraeFeng Zhang

Science, 378 (6622), • DOI: 10.1126/science.add7450

RNA-activated protein cleavage

CRISPR-Cas systems typically function as RNA-guided nucleases to protect bacteria against foreign genetic material; however, diverse associated genes have been identified within CRISPR-Cas systems, raising the possibility that new RNA-guided enzymatic functions exist in nature. One example is a CHAT family protease found next to the RNA-targeting CRISPR effector gRAMP. Strecker *et al*. found that this CRISPR-associated protease performs RNA-activated protein cleavage of a sigma factor inhibitor to activate a transcriptional response after infection. Characterization and engineering of this natural system allows for new possibilities in biology, including RNA-sensing applications in human cells. —DJ

View the article online
https://www.science.org/doi/10.1126/science.add7450
Permissions
https://www.science.org/help/reprints-and-permissions

Use of this article is subject to the Terms of service

Science (ISSN) is published by the American Association for the Advancement of Science. 1200 New York Avenue NW, Washington, DC 20005. The title *Science* is a registered trademark of AAAS.
Copyright © 2022 The Authors, some rights reserved; exclusive licensee American Association for the Advancement of Science. No claim to original U.S. Government Works

Downloaded from https://www.science.org at Shanghai Normal University on November 25, 2022