

RNA-triggered protein cleavage and cell death by the RNA-guided type III-E CRISPR-Cas nuclease-protease complex

Kazuki Kato^{1,9}, Sae Okazaki^{1,9}, Cian Schmitt-Ulms^{2,9}, Kaiyi Jiang^{2,9}, Wenyan Zhou², Junichiro Ishikawa¹, Yukari Isayama¹, Shungo Adachi³, Tomohiro Nishizawa⁴, Kira S. Makarova⁵, Eugene V. Koonin⁵, Omar O. Abudayyeh^{2,*}, Jonathan S. Gootenberg^{2,*}, Hiroshi Nishimasu^{1,6,7,8,*}

¹ Structural Biology Division, Research Center for Advanced Science and Technology, The University of Tokyo, 4-6-1 Komaba, Meguro-ku, Tokyo 153-8904, Japan.

² McGovern Institute for Brain Research at MIT, Massachusetts Institute of Technology, Cambridge, MA 02139, USA.

³ Cellular and Molecular Biotechnology Research Institute, National Institute of Advanced Industrial Science and Technology, 2-4-7 Aomi, Koto-ku, Tokyo 135-0064, Japan.

⁴ Graduate School of Medical Life Science, Yokohama City University, 22-2 Seto, Kanazawa-ku, Yokohama, Kanagawa 236-0027, Japan.

⁵ National Center for Biotechnology Information, National Library of Medicine, National Institutes of Health, Bethesda, MD 20894, USA.

⁶ Department of Chemistry and Biotechnology, Graduate School of Engineering, The University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo 113-8656, Japan.

⁷ Department of Biological Sciences, Graduate School of Science, The University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo 113-0033, Japan.

⁸ Inamori Research Institute for Science, 620 Suiginya-cho, Shimogyo-ku, Kyoto 600-8411, Japan.

⁹ Co-first authors

* Correspondence:

omar@abudayyeh.science (O.O.A.)

jgoot@mit.edu (J.S.G.)

nishimasu@g.ecc.u-tokyo.ac.jp (H.N.)

Abstract

The type III-E Cas7-11 effector nuclease forms a complex with a CRISPR RNA (crRNA) and the putative caspase-like protease Csx29, catalyzes crRNA-guided target RNA cleavage, and has been used for RNA targeting in eukaryotic cells. Here, we report cryo-electron microscopy structures of the Cas7-11–crRNA–Csx29 complex with and without target RNA, and demonstrate that target RNA binding induces a conformational change in Csx29 and results in the protease activation. Biochemical analysis confirmed that Cas7-11-bound Csx29 cleaves Csx30 in a target RNA-dependent manner. Reconstitution of the system in bacteria uncovered Csx30-dependent cellular toxicity regulated by Csx31, and that Csx29-mediated cleavage produces toxic Csx30 fragments, promoting cell death. We find that Csx30 can bind both Csx31 and the associated sigma factor RpoE, suggesting Csx30 can inhibit RpoE and modulate cellular stress response towards infection. Overall, the RNA-guided nuclease-protease activities of the Cas7-11–Csx29 effector complex facilitate protease-based programmed cell death.

Introduction

Prokaryotic CRISPR-Cas systems provide adaptive immunity against foreign nucleic acids, including phages and mobile genetic elements, via diverse mechanisms of programmed nucleic-acid cleavage. CRISPR-Cas systems are divided into two classes based on the number of components in the effector complexes responsible for defense via cleavage of invading nucleic acids programmed by a CRISPR RNA (crRNA) guide(1, 2). In Class 1 systems, which encompass types I, III, and IV, target nucleic acids are degraded by multi-protein effector complexes, whereas, in Class 2 systems, including types II, V, and VI, the effector complexes are formed by a single multidomain Cas protein (Cas9, Cas12, and Cas13, respectively). Beyond primary effector nuclease function, both Class 1 and Class 2 CRISPR-Cas systems deploy a wide-array of accessory proteins to enhance the antiviral activity of the primary effector nuclease, including secondary nuclease activation via cyclic oligoadenylate generation in type III-A/B/D systems(3–6) and target RNA-dependent pore formation by Csx28 in type VI-B systems(7).

Unlike typical Class 1 effectors, the type III-E effector Cas7-11 (also known as gRAMP) is a single-protein, multidomain effector that consists of four Cas7 domains (Cas7.1–Cas7.4) and a Cas11 domain and likely evolved from the more complex type III-D multi-subunit effectors via domain fusions(8, 9). Cas7-11 associates with a crRNA and cleaves complementary single-stranded RNA (ssRNA) targets at two defined positions, using the Cas7.2 and Cas7.3 domains, respectively. Whereas the type VI effector Cas13 displays promiscuous RNase activity(10, 11), Cas7-11 exhibits specific, guide RNA-dependent RNA cleavage activity in human cells, and has been used as a novel RNA-targeting tool with high specificity and low cell toxicity(8, 9). The type III-E locus contains multiple conserved accessory proteins, including Csx29 (a caspase-like putative protease with fused

TPR and CHAT domains), Csx30 and Csx31 (proteins with unknown functions), and RpoE (an alternative sigma factor). Cas7-11 associates with Csx29(8, 9), suggesting a potential mechanism of RNA-guided protease activity for antiviral immunity. Recently, we reported the cryo-electron microscopy (cryo-EM) structure of *Desulfonema ishimotonii* Cas7-11 in complex with its cognate crRNA and target RNA (tgRNA)(12), providing mechanistic insights into the pre-crRNA processing and tgRNA cleavage. However, how Cas7-11 cooperates with the other proteins encoded in the type III-E locus (Csx29, Csx30, Csx31, and RpoE), and how Cas7-11 binds to Csx29 and potentially activates its protease activity remain unknown.

Structures of Cas7-11 in complex with Csx29

We reasoned that structural insights would allow for mechanistic understanding of the Cas7-11–Csx29 effector complex. To prepare the Cas7-11–crRNA–Csx29 complex for structural analysis, we co-expressed the catalytically inactive *D. ishimotonii* Cas7-11 mutant (referred to as Cas7-11 for simplicity), with D429A (Cas7.2) and D654A (Cas7.3) mutations introduced to prevent tgRNA cleavage by Cas7-11, together with Csx29 and a crRNA transcribed from a CRISPR array containing two repeat–spacer units. We determined the cryo-EM structures of the Cas7-11–crRNA–Csx29 complex with and without a 27-nucleotide tgRNA at 2.8- and 2.5-Å resolutions, respectively (Fig. 1A–1D, fig. S1A–S1D and S2A–S2D, and Table S1). In the Cas7-11–crRNA–Csx29 complex, Cas7-11 adopts a modular architecture consisting of four Cas7 domains (Cas7.1–Cas7.4) with a zinc finger (ZF) motif, a Cas11 domain, an insertion (INS) domain inserted within the Cas7.4 domain, a C-terminal extension (CTE) domain, and four interdomain linkers (L1–L4) (Fig. 1C and 1D), as in the Csx29-unbound Cas7-11–crRNA–tgRNA structure(12) (fig. S3A–S3C).

The 15-nt 5' tag region (U(–15)–C(–1)) in the 38-nt crRNA (U(–15)–A23) is anchored by the Cas7.1 and Cas7.2 domains (Fig. 1C and 1D). Nucleotide U(–16) was not resolved in the density map (fig. S4A), suggesting that the co-expressed pre-crRNA was processed by Cas7-11 into the mature crRNA. U(–15) is surrounded by H43, R53, Y55, N152, and S158 in the Cas7.1 domain (fig. S4A), consistent with the proposed pre-crRNA processing mechanism, in which H43 functions as a general base to deprotonate the 2'-hydroxy group of U(–16) for pre-crRNA maturation(12). In the tgRNA-free structure, the 23-nt crRNA spacer region (C1–A23) is recognized by the Cas7.2–Cas7.4 domains (Fig. 1C and fig. S3B), while in the tgRNA-bound structure, the crRNA spacer region (C1–A23, except for U4 and C10) hybridizes with the tgRNA (G1*–U23*, except for A4* and G10*) to form a guide-target duplex (Fig. 1D and fig. S3C), as in the Cas7-11–crRNA–tgRNA structure(12) (fig. S3A). The thumb-like β -hairpins in the Cas7.2 and Cas7.3 domains penetrate the guide-target duplex, inducing base flipping at fourth and tenth positions (U4–A4* and C10–G10*), respectively (fig. S4B). The catalytic residues D429 (Cas7.2) and D654 (Cas7.3) (D429A and D654A in the present structures) are located in the vicinity of the scissile phosphodiester bonds before A4* and G10* of the tgRNA,

respectively (fig. S4B), as previously observed(12). These observations explain why Csx29 binding does not substantially affect tgRNA cleavage by Cas7-11(8, 9). In the tgRNA-free structure, the peripheral region (residues 1044–1124) of the INS domain was less resolved in the density map (fig. S2A), probably due to its flexibility resulting from the lack of the interaction with the guide-target duplex. Thus, the peripheral region of the INS domain was not included in the final model of the tgRNA-free structure.

Csx29 structure

Csx29 consists of an N-terminal domain (NTD) (residues 1–67) and a C-terminal domain (CTD) (residues 75–751), connected with a flexible interdomain linker (residues 68–74) that is disordered in the Cas7-11–crRNA–Csx29 structure (Fig. 2A). The NTD adopts a three-helix bundle and interacts with the Cas7.4 domain of Cas7-11 (Fig. 2A). The CTD can be divided into a TRP (tetratricopeptide repeat) domain consisting of eight TPR units (TPR1–TPR8) (residues 75–423) and a CHAT (Caspase HetF Associated with TPRs) protease domain (residues 424–751) (Fig. 2A). In Csx29, each TPR unit contains two α helices, similar to canonical TPR-containing proteins where TPRs interact with their protein targets(13). In the Cas7-11–crRNA–Csx29 structure, the TPR domain of Csx29 interacts with the L2 linker of Cas7-11 (Fig. 1C). The CHAT domain of Csx29 consists of a central 11-stranded mixed β -sheet and flanking α -helices, and can be divided into a pseudo-protease domain (residues 424–551) and an active-protease domain (residues 552–751) with the conserved putative catalytic residues H615 and C658 (Fig. 2A). A Dali search(14) confirmed that the CHAT domain of Csx29 structurally resembles caspase-like cysteine proteases, such as human separase(15) (fig. S5).

Interactions between Cas7-11 and Csx29

In the tgRNA-free structure, Csx29 extensively interacts with Cas7-11 at multiple regions (Fig. 2B and fig. S6A). The L2 linker (residues 367–401) and an α -helical insertion (residues 1313–1341) in the Cas7.4 ZF motif, which are disordered in the Csx29-unbound Cas7-11 structure(12) (fig. S6B), are ordered and form interactions with Csx29 in the Cas7-11–crRNA–Csx29 structure (Fig. 2B and fig. S6A and S6C). This α -helical insertion in the ZF motif is unique to Cas7.4 and absent in Cas7.1–Cas7.3 (fig. S6C). The NTD of Csx29 mainly interacts with the Cas7.4 domain of Cas7-11 (Fig. 2B and fig. S6A). I5, I8, L30, Y33, L50, R53, F57, L60, S61, and R64 of Csx29 hydrophobically interact with W1316, L1322, L1325, Y1328, L1333, and L1334 of the α -helical insertion region of Cas7-11, while R53 and R64 form hydrogen bonds with R1336 and E1330 of Cas7-11, respectively (fig. S7A). In addition, the Cas7.2 thumb-like β -hairpin and the L2/L4 linkers contribute to the binding to the Csx29 NTD (Fig. 2B). N505 and F507 of Cas7-11 (Cas7.2) interact with T44 and E42/L45 of Csx29, respectively, and K879 and E878 of Cas7-11 (L4) hydrogen bond with E42 and N3/Q47 of Csx29, respectively (fig. S7B). Furthermore, L370 of Cas7-11 (L2) is accommodated within a hydrophobic pocket at the NTD–TPR1 interface of Csx29 (fig. S7C).

TPR1 and TPR2 interact with Cas7.3 (ZF) and L2 of Cas7-11, respectively (Fig. 2B and fig. S6A). R97/E101 of TPR1 and R136 of TPR2 hydrogen bond with D705/Y718 and D705 of Cas7.3 (fig. S7D). TPR2 also interacts with Cas7.1 (thumb-like β -hairpin) and Cas7.2 (ZF) (Fig. 2B). TPR1 and TPR2 are the only TPR domains that mediate the Cas7-11–Csx29 interaction, and TPR3–TPR8 do not contact Cas7-11. The CHAT protease domain of Csx29 interacts with Cas7.1 (thumb-like β -hairpin), Cas7.2 (ZF), Cas7.3 (ZF), and L2 of Cas7-11 (Fig. 2B and fig. S6A). Notably, the protease active site of Csx29 is located in the vicinity of the Cas7.2 domain of Cas7-11 (Fig. 2C), suggesting limited accessibility for the peptide substrate in this conformation. Furthermore, unlike in the separase–securin structure(15), the side chain of the catalytic residue C658 is buried inside the CHAT domain in the present structure (fig. S5), indicating that a structural rearrangement of C658 would be required for the substrate cleavage. These observations suggest that the Cas7-11–crRNA–Csx29 structure represents the inactive state of the Csx29 putative protease.

Target RNA binding-induced structural change in the Cas7-11–Csx29 complex

A comparison of the Cas7-11–crRNA–Csx29 structures with and without the tgRNA revealed a notable conformational difference in Csx29 (Fig. 1C and 1D). While the Csx29 NTD similarly binds Cas7-11 in the two structures (Fig. 1C and 1D), the Csx29 CTD is not resolved in the density map in the tgRNA-bound structure (fig. S2B), indicating that upon the tgRNA binding, the Csx29 CTD dissociates from Cas7-11 and adopts a flexible conformation. This conformational change releases the steric block on the Csx29 active site, allowing access to the substrate protein. A comparison between the two structures indicated that the tgRNA (G1* and A2*) sterically clashes with Cas7-11 (L2) and Csx29 (TPR1) in the Cas7-11–crRNA–Csx29 structure (Fig. 2D, 2E, and fig. S6E), consistent with the tgRNA binding-induced structural change in Csx29. These structural observations suggest that the Cas7-11–crRNA–Csx29–tgRNA structure represents the active state of the Csx29 protease and that Csx29 is a target RNA-triggered protease.

Target RNA-triggered Csx30 cleavage by Csx29

Given our structural findings, we explored whether proteins encoded in the endogenous *D. ishimotonii* type III-E CRISPR locus are substrates for Csx29 cleavage. Neither Cas7-11 nor Csx29 were degraded in the presence of the tgRNA in biochemical assays (Fig. 3A and 3B), suggesting that the Cas7-11–Csx29 complex itself is not a Csx29 substrate. Given that Csx30 and Csx31 are encoded together with Cas7-11 and Csx29 in the CRISPR operon and are highly conserved among the type III-E systems(8), we hypothesized that Csx29 could target either Csx30 or Csx31. To test this hypothesis, we attempted to prepare the recombinant Csx30 and Csx31 proteins and examine whether they are cleaved by Csx29 in a tgRNA-dependent manner. We found that while Csx30 could be purified as a soluble protein, Csx31 was expressed in an insoluble fraction. We examined the *in vitro* cleavage of

Csx30 by Cas7-11–crRNA–Csx29 in the absence and presence of the tgRNA, and found that Cas7-11–crRNA–Csx29 cleaves Csx30 into two fragments, Csx30-1 (~50 kDa) and Csx30-2 (~15 kDa), only in the presence of the tgRNA (Fig. 3B). The C615A/H658A mutations in Csx29 abolished the Csx30 cleavage (Fig. 3B), but did not affect the tgRNA cleavage by Cas7-11 (fig. S8), indicating the separable nuclease and protease activities. Furthermore, the D429A/D654A catalytic mutations in Cas7-11 abolished tgRNA cleavage (fig. S8), as previously observed(12), and, unexpectedly, improved the Csx30 cleavage by Csx29 to almost complete proteolysis (Fig. 3B). This improvement in the proteolytic activity suggests that the tgRNA dissociates from the effector complex after the Cas7-11-mediated cleavage and that the Csx29 protease is only active as long as a target RNA is bound to the Cas7-11–Csx29 complex. These results demonstrated that Csx30 is cleaved by the CHAT protease domain of Csx29 in a target RNA-dependent manner.

N-terminal analysis of the Csx30-2 fragment showed that it begins with K428 (fig. S9), indicating that Csx30 is cleaved by Csx29 between M427 and K428 (Fig. 3C). A structural prediction using AlphaFold2(16) indicated that Csx30 consists of an N-terminal domain (NTD) and a C-terminal domain (CTD), which are connected by a linker region. The NTD (residues 1–377) contains two α -helical subdomains, whereas the CTD (residues 418–565) comprises a core β -barrel with flanking α helices (fig. S10). The cleavage site between M427 and K428 is located at a β -hairpin in the Csx30 CTD (fig. S10). We examined the *in vitro* Csx29-mediated cleavage of eight Csx30 mutants, in which V425–K431 were individually replaced with an alanine. The M427A mutation substantially reduced the Csx30 cleavage, whereas the other mutations had no effect (Fig. 3D). Thus, Csx29 seems to primarily recognize M427 at the P1 site within the AVGM|KKDK sequence in Csx30 and cleaves Csx30 between M427 (P1) and K428 (P1'). Together, these results demonstrated that the Cas7-11–crRNA–Csx29 complex catalyzes target RNA-triggered Csx30 proteolytic cleavage.

Effects of Csx30 and Csx31 on bacterial cell growth

To explore the physiological relevance of the Csx29-mediated Csx30 cleavage, we overexpressed in *Escherichia coli* the full-length Csx30 (referred to as Csx30 for simplicity), the N-terminal fragment of Csx30 (residues 1–427, Csx30-1), or the C-terminal fragment of Csx30 (residues 428–565, Csx30-2), and monitored the cell growth (Fig. 4A). Overexpression of Csx30 substantially inhibited the cell growth compared to uninduced controls (Fig. 4B, 4C, and fig. S11A), demonstrating the cell toxicity of Csx30. Overexpression of Csx30-1 similarly caused significant growth inhibition, whereas Csx30-2 displayed only non-significant, mild inhibition (Fig. 4B, 4C, and fig. S11A), indicating that Csx30-1 harbors the toxicity of this protein. As a structural prediction using AlphaFold2(16) suggested that Csx30 and Csx31 have oppositely charged surfaces and electrostatically interact with each other (fig. S11B), we also explored the role of Csx31 in our bacterial growth assays. Overexpression of Csx31 rescued the Csx30-mediated growth defect (Fig. 4D, 4E, and fig. S11C), suggesting that Csx31

interacts with Csx30 and suppresses Csx30-induced toxicity.

Interactions between Csx30, Csx31, and RpoE

Csx30, Csx31, and RpoE (a sigma factor involved in the regulation of extracytoplasmic stress response pathways(17)) frequently co-occur the type III-E CRISPR loci(2, 8). Based on the results of our bacterial growth assays (Fig. 4), including involvement of Csx31 in modulating Csx30 toxicity, we hypothesized further interplay between the three proteins in the locus. To test this hypothesis, we leveraged the fact that RpoE is involved in the heat shock response in *E. coli* and tested for the effect of Csx30 and Csx31 in *E. coli* at different temperatures from 30°C to 42°C, hypothesizing that the growth defects might be more pronounced at higher temperatures due to inhibition of RpoE by the Csx30 and Csx31 proteins (Fig. 4F). Corroborating our hypothesis, we found that at higher temperatures, the toxicity of Csx30 was more dramatic compared to lower temperatures across all the combinations of proteins tested, suggesting the involvement of RpoE in relation to Csx30 and Csx31 (Fig. 4G). To examine direct interactions between Csx30, Csx31, and RpoE, we co-expressed the three proteins in *E. coli* and analyzed complex formation using gel-filtration. Csx30, Csx31, and RpoE eluted as a single peak from the column (fig. S12A), indicating that they form a stable complex. Like isolated Csx30, Csx30 in the Csx30–Csx31–RpoE complex was cleaved by the Cas7-11–Csx29 complex, and Csx30 (Csx30-1), Csx31, and RpoE co-eluted from the column (fig. S12B), indicating that Csx30-1, Csx31, and RpoE maintain a complex formation after Csx29 cleavage, with separation from Csx30-2. Consistently, structural prediction using AlphaFold2(16) suggested that Csx30, Csx31, and RpoE form a ternary complex, in which the Csx30 NTD extensively interacts with RpoE (fig. S12C). *D. ishimotonii* RpoE (DiRpoE) shares structural similarity with *E. coli* RpoE (EcRpoE) (fig. S12D), suggesting that, like the anti-sigma factor RseA(18), Csx30 interacts with EcRpoE in our growth assays, causing the cell growth inhibition. While EcRpoE is involved in extracytoplasmic stress response in *E. coli*, the associated regulatory proteins like RseA are not present in *Desulfohalobium* strains(19) and there are different paralogs of RpoE in *E. coli* with varied functions, such as FecI(20), suggesting that DiRpoE could mediate an unknown transcriptional response in its host.

A Dali search revealed that the Csx30 CTD is structurally similar to pore-forming proteins in type IV secretion systems, such as CagX(21) (fig. S12E). Given its less toxicity in *E. coli* cells, the Csx30 CTD might function as a membrane anchor, rather than a pore-forming protein, consistent with the role of RseA, which is also membrane-localized(17, 19). The CTD and NTD of Csx30 are connected via a flexible linker, suggesting that the Csx29-mediated cleavage releases the N-terminal fragment of Csx30 bound to Csx31 and RpoE into the cytoplasm, thereby modulating gene expression.

Sequence analysis revealed that Csx30 NTDs are highly conserved, whereas Csx30 CTDs are divergent and can be divided into seven distinct groups (fig. S13 and S14), two of which belong to

unrelated protein domains found in other contexts. One is an uncharacterized DUF4384 family, which is often fused to different protease domains (see domain architectures for DUF4384 in the CDD database(22)). Another group is similar to pilus assembly protein PilP, which forms a periplasmic ring of bacterial type IV pili(23). These observations suggest the mechanistic diversity of Csx30-mediated RpoE interaction and programmed gene expression modulation.

Discussion

In this study, we determined the cryo-EM structures of the Cas7-11–crRNA–Csx29 complex with and without a target RNA. Comparison of the two structures showed that target RNA-binding induces a structural change in the Csx29 protease domain, likely activating the Csx29 protease activity. Consistent with this structural finding, our biochemical analysis demonstrated that Csx29 is a target RNA-triggered protease that cleaves another type III-E associated protein, Csx30, at a unique site. Analysis of the effects of Csx30 and Csx31 on bacterial growth suggested that the Csx29-mediated Csx30 cleavage releases the toxic N-terminal fragment of Csx30 in complex with Csx31 and causes growth inhibition in bacterial cells. Furthermore, our biochemical and structural analyses indicated that Csx30, Csx31, and RpoE can form a ternary complex, in which Csx30 extensively interacts with RpoE, suggesting that Csx30 inhibits the RpoE activity and causes cell growth arrest. Taken together, these findings revealed that the type III-E Cas7-11–Csx29 effector complex is an RNA-triggered programmable nuclease-protease capable of cleaving ssRNA targets and the Csx30 protein, unleashing a downstream signaling cascade that affects cell growth via potential transcriptional regulation.

Thus, in the type III-E CRISPR-Cas systems, the Cas7-11–Csx29 effector complex likely degrades ssRNA transcripts of phage genes and stimulates potentially toxic host cell stress responses through the Csx29-mediated Csx30 cleavage (Fig. 5). Although the exact mechanism of this effect on cell growth remains to be investigated, our finding that Csx30, Csx31, and RpoE form a ternary complex suggests a potential pathway. Conceivably, in the absence of Csx31, the toxic N-terminal fragment of Csx30 might act as an anti-sigma factor, thereby inhibiting the RpoE activity, modulating gene expression, and leading to cell death or growth arrest. Moreover, given unexpected structural similarity between the Csx30 CTD and pore-forming proteins in type IV secretion systems, it is possible that Csx29 proteolysis might liberate the Csx30 NTD from being near the cell membrane into the cytoplasm to exert its cellular effects. This type of programmed cell death might be analogous to that caused by the bacterial membrane pore-forming toxins gasdermins that are activated via proteolytic cleavage and release of auto-inhibitory peptides by associated proteases activated during phage infection(24, 25). Moreover, given the high diversity of Csx30 CTDs (fig. S14), additional exploration of other type III-E systems might reveal other functions associated with Cas7-11-mediated target RNA recognition.

Among the CRISPR-Cas systems, a biological if not mechanistic analogy can be found in the type VI system, where the Cas13–crRNA effector complex recognizes complementary phage mRNAs and cleaves both phage (specifically and in *cis*) and host (indiscriminately and in *trans*) transcripts, stalling the cell growth and with it, the infectious cycle(26). Similarly, within type III systems, the CRISPR-Lon protease can be activated via cyclic oligoadenylates upon RNA recognition by type III effector complexes, activating nearby protein toxins via proteolytic cleavage(27). Our characterization of the type III-E system highlighted the remarkable diversity of CRISPR-associated functions activated by programmable nucleic-acid recognition, thereby motivating continued exploration of CRISPR-associated proteins and their potential programmable functions that may have useful roles for biology applications. Our findings that the type III-E Cas7-11–Csx29 effector complex is a so far unique RNA-triggered nuclease-protease establish a new paradigm of prokaryotic signal transduction, and could pave the way for the development of new RNA/protein-targeting technologies, including *in vitro* diagnostics and cellular RNA sensing.

Acknowledgements

We thank the staff scientists at The University of Tokyo’s cryo-EM facility, especially Y. Sakamaki, for help with cryo-EM data collection, and Dr. Nobuaki Okumura for help with N-terminal analysis. H.N. is supported by the Platform Project for Supporting Drug Discovery and Life Science Research (Basis for Supporting Innovative Drug Discovery and Life Science Research (BINDS)) from AMED under grant number JP21am0101115 (support number 2792); AMED under grant number JP21wm0325048h0001; JSPS KAKENHI Grant Numbers 20K20579 and 21H05281; Takeda Medical Research Foundation; and the Inamori Research Institute for Science. J.S.G. and O.O.A. are supported by NIH grants 1R21-AI149694, R01-EB031957, and R56-HG011857; The McGovern Institute Neurotechnology (MINT) program; the K. Lisa Yang and Hock E. Tan Center for Molecular Therapeutics in Neuroscience; the G. Harold & Leila Y. Mathers Charitable Foundation; MIT John W. Jarve (1978) Seed Fund for Science Innovation; FastGrants; the Cystic Fibrosis Foundation; Google Ventures; Longevity Impetus Grant from Norn Group; NHGRI/TDCC Opportunity Fund; and the McGovern Institute.

Author contributions

K.K. and S.O. performed sample preparation and biochemical analysis with assistance from Y.I. and J.I.; K.K., S.O., and H.N. performed structural analysis with assistance from J.I. and T.N.; C.S., K.J., W.Z., J.S.G., and O.O.A. performed bacterial experiments; S.A. performed N-terminal analysis; K.S.M. and E.V.K. performed computational analysis; J.S.G., O.O.A., and H.N. wrote the manuscript with help from K.K. and E.V.K.; J.S.G., O.O.A., and H.N. conceived the project and supervised the research.

Declaration of interests

The authors have filed a patent application related to this work. J.S.G. and O.O.A. are co-founders of Sherlock Biosciences, Proof Diagnostics, Moment Biosciences, and Tome Biosciences.

Data and code availability

The EM densities have been deposited in the Electron Microscopy Public Image Archive under the accession codes 33695 and 33696. The model coordinates have been deposited in the Protein Data Bank under the accession codes 7Y7X and 7Y7Y.

References

1. F. Hille, H. Richter, S. P. Wong, M. Bratovič, S. Ressel, E. Charpentier, The Biology of CRISPR-Cas: Backward and Forward. *Cell*. **172**, 1239–1259 (2018).
2. K. S. Makarova, Y. I. Wolf, J. Iranzo, S. A. Shmakov, O. S. Alkhnbashi, S. J. J. Brouns, E. Charpentier, D. Cheng, D. H. Haft, P. Horvath, S. Moineau, F. J. M. Mojica, D. Scott, S. A. Shah, V. Siksnys, M. P. Terns, Č. Venclovas, M. F. White, A. F. Yakunin, W. Yan, F. Zhang, R. A. Garrett, R. Backofen, J. van der Oost, R. Barrangou, E. V. Koonin, Evolutionary classification of CRISPR-Cas systems: a burst of class 2 and derived variants. *Nat. Rev. Microbiol.* **18**, 67–83 (2020).
3. M. Kazlauskienė, G. Kostiuk, Č. Venclovas, G. Tamulaitis, V. Siksnys, A cyclic oligonucleotide signaling pathway in type III CRISPR-Cas systems. *Science*. **357**, 605–609 (2017).
4. O. Niewoehner, C. Garcia-Doval, J. T. Rostøl, C. Berk, F. Schwede, L. Bigler, J. Hall, L. A. Marraffini, M. Jinek, Type III CRISPR-Cas systems produce cyclic oligoadenylate second messengers. *Nature*. **548**, 543–548 (2017).
5. C. Rouillon, J. S. Athukoralage, S. Graham, S. Grüşchow, M. F. White, Control of cyclic oligoadenylate synthesis in a type III CRISPR system. *Elife*. **7** (2018), doi:10.7554/eLife.36734.
6. W. Han, S. Pan, B. López-Méndez, G. Montoya, Q. She, Allosteric regulation of Csx1, a type IIIB-associated CARF domain ribonuclease by RNAs carrying a tetraadenylate tail. *Nucleic Acids Res.* **45**, 10740–10750 (2017).
7. A. R. VanderWal, J.-U. Park, B. Polevoda, E. H. Kellogg, M. R. O’Connell, CRISPR-Csx28 forms a Cas13b-activated membrane pore required for robust CRISPR-Cas adaptive immunity. *bioRxiv* (2021), p. 2021.11.02.466367.
8. A. Özcan, R. Krajcski, E. Ioannidi, B. Lee, A. Gardner, K. S. Makarova, E. V. Koonin, O. O. Abudayyeh, J. S. Gootenberg, Programmable RNA targeting with the single-protein CRISPR effector Cas7-11. *Nature*. **597**, 720–725 (2021).
9. S. P. B. van Beljouw, A. C. Haagsma, A. Rodríguez-Molina, D. F. van den Berg, J. N. A. Vink, S. J. J. Brouns, The gRAMP CRISPR-Cas effector is an RNA endonuclease complexed with a caspase-like peptidase. *Science*. **373**, 1349–1353 (2021).
10. O. O. Abudayyeh, J. S. Gootenberg, P. Essletzbichler, RNA targeting with CRISPR–Cas13. *Nature* (2017) (available at <https://www.nature.com/articles/nature24049>).
11. J. S. Gootenberg, O. O. Abudayyeh, J. W. Lee, P. Essletzbichler, A. J. Dy, J. Joung, V. Verdine, N. Donghia, N. M. Daringer, C. A. Freije, C. Myhrvold, R. P. Bhattacharyya, J. Livny, A. Regev, E. V. Koonin, D. T. Hung, P. C. Sabeti, J. J. Collins, F. Zhang, Nucleic acid detection with CRISPR-Cas13a/C2c2. *Science*. **356**, 438–442 (2017).
12. K. Kato, W. Zhou, S. Okazaki, Y. Isayama, T. Nishizawa, J. S. Gootenberg, O. O. Abudayyeh, H. Nishimasu, Structure and engineering of the type III-E CRISPR-Cas7-11 effector complex. *Cell* (2022).
13. N. Zeytuni, R. Zarivach, Structural and functional discussion of the tetra-trico-peptide repeat, a protein interaction module. *Structure*. **20**, 397–405 (2012).
14. L. Holm, Using Dali for Protein Structure Comparison. *Methods Mol. Biol.* **2112**, 29–42 (2020).
15. J. Yu, P. Raia, C. M. Ghent, T. Raisch, Y. Sadian, S. Cavadini, P. M. Sabale, D. Barford, S.

- Raunser, D. O. Morgan, A. Boland, Structural basis of human separase regulation by securin and CDK1-cyclin B1. *Nature*. **596**, 138–142 (2021).
16. J. Jumper, R. Evans, A. Pritzel, T. Green, M. Figurnov, O. Ronneberger, K. Tunyasuvunakool, R. Bates, A. Židek, A. Potapenko, A. Bridgland, C. Meyer, S. A. A. Kohli, A. J. Ballard, A. Cowie, B. Romera-Paredes, S. Nikolov, R. Jain, J. Adler, T. Back, S. Petersen, D. Reiman, E. Clancy, M. Zielinski, M. Steinegger, M. Pacholska, T. Berghammer, S. Bodenstein, D. Silver, O. Vinyals, A. W. Senior, K. Kavukcuoglu, P. Kohli, D. Hassabis, Highly accurate protein structure prediction with AlphaFold. *Nature*. **596**, 583–589 (2021).
17. L. Connolly, A. De Las Penas, B. M. Alba, C. A. Gross, The response to extracytoplasmic stress in *Escherichia coli* is controlled by partially overlapping pathways. *Genes Dev.* **11**, 2012–2021 (1997).
18. E. A. Campbell, J. L. Tupy, T. M. Gruber, S. Wang, M. M. Sharp, C. A. Gross, S. A. Darst, Crystal structure of *Escherichia coli* sigmaE with the cytoplasmic domain of its anti-sigma RseA. *Mol. Cell.* **11**, 1067–1078 (2003).
19. V. Schnaars, L. Wöhlbrand, S. Scheve, C. Hinrichs, R. Reinhardt, R. Rabus, Proteogenomic Insights into the Physiology of Marine, Sulfate-Reducing, Filamentous *Desulfonema limicola* and *Desulfonema magnum*. *Microb Physiol*, 1–20 (2021).
20. S. Mahren, V. Braun, The FecI extracytoplasmic-function sigma factor of *Escherichia coli* interacts with the beta' subunit of RNA polymerase. *J. Bacteriol.* **185**, 1796–1802 (2003).
21. J. M. Chung, M. J. Sheedlo, A. M. Campbell, N. Sawhney, A. E. Frick-Cheng, D. B. Lacy, T. L. Cover, M. D. Ohi, Structure of the *Helicobacter pylori* Cag type IV secretion system. *Elife*. **8** (2019), doi:10.7554/eLife.47644.
22. S. Lu, J. Wang, F. Chitsaz, M. K. Derbyshire, R. C. Geer, N. R. Gonzales, M. Gwadz, D. I. Hurwitz, G. H. Marchler, J. S. Song, N. Thanki, R. A. Yamashita, M. Yang, D. Zhang, C. Zheng, C. J. Lanczycki, A. Marchler-Bauer, CDD/SPARCLE: the conserved domain database in 2020. *Nucleic Acids Res.* **48**, D265–D268 (2020).
23. Y.-W. Chang, L. A. Rettberg, A. Treuner-Lange, J. Iwasa, L. Søgaard-Andersen, G. J. Jensen, Architecture of the type IVa pilus machine. *Science*. **351**, aad2001 (2016).
24. A. G. Johnson, T. Wein, M. L. Mayer, B. Duncan-Lowey, E. Yirmiya, Y. Oppenheimer-Shaanan, G. Amitai, R. Sorek, P. J. Kranzusch, Bacterial gasdermins reveal an ancient mechanism of cell death. *Science*. **375**, 221–225 (2022).
25. C. Clavé, W. Dyrka, E. A. Turcotte, A. Granger-Farbos, L. Ibarlosa, B. Pinson, R. E. Vance, S. J. Saupe, A. Daskalov, Fungal gasdermin-like proteins are controlled by proteolytic cleavage. *Proc. Natl. Acad. Sci. U. S. A.* **119** (2022), doi:10.1073/pnas.2109418119.
26. A. J. Meeske, S. Nakandakari-Higa, L. A. Marraffini, Cas13-induced cellular dormancy prevents the rise of CRISPR-resistant bacteriophage. *Nature*. **570**, 241–245 (2019).
27. C. Rouillon, N. Schneberger, H. Chi, M. F. Peter, M. Geyer, W. Boenigk, R. Seifert, M. F. White, G. Hagelueken, SAVED by a toxin: Structure and function of the CRISPR Lon protease. *bioRxiv* (2021), p. 2021.12.06.471393.
28. R. C. Edgar, MUSCLE v5 enables improved estimates of phylogenetic tree confidence by ensemble bootstrapping. *bioRxiv* (2021), p. 2021.06.20.449169.
29. A. Punjani, J. L. Rubinstein, D. J. Fleet, M. A. Brubaker, cryoSPARC: algorithms for rapid

- unsupervised cryo-EM structure determination. *Nat. Methods*. **14**, 290–296 (2017).
30. A. Punjani, H. Zhang, D. J. Fleet, Non-uniform refinement: adaptive regularization improves single-particle cryo-EM reconstruction. *Nat. Methods*. **17**, 1214–1221 (2020).
31. P. B. Rosenthal, R. Henderson, Optimal determination of particle orientation, absolute hand, and contrast loss in single-particle electron cryomicroscopy. *J. Mol. Biol.* **333**, 721–745 (2003).
32. E. F. Pettersen, T. D. Goddard, C. C. Huang, E. C. Meng, G. S. Couch, T. I. Croll, J. H. Morris, T. E. Ferrin, UCSF ChimeraX: Structure visualization for researchers, educators, and developers. *Protein Sci.* **30**, 70–82 (2021).
33. D. Liebschner, P. V. Afonine, M. L. Baker, G. Bunkóczi, V. B. Chen, T. I. Croll, B. Hintze, L. W. Hung, S. Jain, A. J. McCoy, N. W. Moriarty, R. D. Oeffner, B. K. Poon, M. G. Prisant, R. J. Read, J. S. Richardson, D. C. Richardson, M. D. Sammito, O. V. Sobolev, D. H. Stockwell, T. C. Terwilliger, A. G. Urzhumtsev, L. L. Videau, C. J. Williams, P. D. Adams, Macromolecular structure determination using X-rays, neutrons and electrons: recent developments in Phenix. *Acta Crystallogr D Struct Biol.* **75**, 861–877 (2019).
34. P. Emsley, B. Lohkamp, W. G. Scott, K. Cowtan, Features and development of Coot. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 486–501 (2010).
35. R. Sanchez-Garcia, J. Gomez-Blanco, A. Cuervo, J. M. Carazo, C. O. S. Sorzano, J. Vargas, DeepEMhancer: a deep learning solution for cryo-EM volume post-processing. *Commun Biol.* **4**, 874 (2021).
36. P. V. Afonine, B. K. Poon, R. J. Read, O. V. Sobolev, T. C. Terwilliger, A. Urzhumtsev, P. D. Adams, Real-space refinement in PHENIX for cryo-EM and crystallography. *Acta Crystallogr D Struct Biol.* **74**, 531–544 (2018).
37. C. J. Williams, J. J. Headd, N. W. Moriarty, M. G. Prisant, L. L. Videau, L. N. Deis, V. Verma, D. A. Keedy, B. J. Hintze, V. B. Chen, S. Jain, S. M. Lewis, W. B. Arendall III, J. Snoeyink, P. D. Adams, S. C. Lovell, J. S. Richardson, D. C. Richardson, MolProbity: More and better reference data for improved all-atom structure validation. *Protein Sci.* **27**, 293–315 (2018).
38. P. V. Afonine, B. P. Klaholz, N. W. Moriarty, B. K. Poon, O. V. Sobolev, T. C. Terwilliger, P. D. Adams, A. Urzhumtsev, New tools for the analysis and validation of cryo-EM maps and atomic models. *Acta Crystallogr D Struct Biol.* **74**, 814–840 (2018).
39. P. Matsudaira, Sequence from picomole quantities of proteins electroblotted onto polyvinylidene difluoride membranes. *J. Biol. Chem.* **262**, 10035–10038 (1987).

Figure legends

Fig. 1 | Cryo-EM structures of the Cas7-11–crRNA–Csx29 complexes with and without the target RNA.

- (A) Domain structures of Cas7-11 and Csx29.
- (B) Nucleotide sequences of the crRNA and its target RNA. The pre-crRNA processing site and the target RNA cleavage sites are indicated by cyan and yellow/green triangles, respectively.
- (C and D) Overall structures of Cas7-11–crRNA–Csx29 (C) and Cas7-11–crRNA–Csx29–tgRNA (D). The bound zinc ions are shown as orange spheres. The disordered L1 and L2 linkers are not shown for clarity. The disordered CTD of Csx29 is indicated by a dashed circle in (D).

Fig. 2 | Interaction between Cas7-11 and Csx29.

- (A) Structure of Csx29 in the Cas7-11–crRNA–Csx29 complex.
- (B) Interface between Cas7-11 and Csx29 in the Cas7-11–crRNA–Csx29 complex. The Cas11 and INS domains are omitted for clarity.
- (C) Location of the Csx29 active site. The catalytic residues (H615 and C658) of the Csx29 protease are colored orange.
- (D and E) Close-up views of the interface between Cas7-11 (Cas7.4 ZF) and Csx29 (NTD). The Cas7-11 L2 linker, crRNA, and tgRNA are shown as space-filling models. The Cas7-11 L2 and Csx29 CTD are disordered in the Cas7-11–crRNA–Csx29–tgRNA structure in (E).

Fig. 3 | Target RNA-triggered Csx30 cleavage by Csx29.

- (A) Schematic of the RNA-triggered Csx30 cleavage by the Cas7-11–crRNA–Csx29 complex.
- (B) RNA-triggered Csx30 cleavage by the Cas7-11–crRNA–Csx29 complex. The Cas7-11–crRNA–Csx29 complex was incubated with Csx30 at 37°C for 2 h in the presence or absence of the target RNA, and then the proteins were analyzed by SDS-PAGE. The wild-type (W) and catalytically inactivated (D) versions of Cas7-11 and Csx29 were used. The gel was stained with CBB.
- (C) Proteolytic cleavage site in Csx30. The Csx30 site cleaved by Csx29 is indicated by a triangle.
- (D) Csx29-mediated cleavage of the Csx30 mutants. The dCas7-11–crRNA–Csx29 complex was incubated with the Csx30 mutants at 37°C for 2 h in the presence or absence of the target RNA, and then the proteins were analyzed by SDS-PAGE. The gel was stained with CBB.

Fig. 4 | Effects of Csx30 and Csx31 on bacterial cell growth.

- (A) Schematic of bacterial growth assays for studying the Csx30 and Csx31 functions.
- (B and C) Growth curves (B) and end-point analysis (C) of *E. coli* expressing either full-length Csx30, the N-terminal fragment (residues 1–427) of Csx30 (Csx30-1), or the C-terminal fragment (residues 428–565) of Csx30 (Csx30-2).
- (D and E) Growth curves (D) and end-point analysis (E) of *E. coli* expressing either Csx30-1, full-

length Csx30 and Csx31, or Csx30-1 and Csx31. In (B)–(E), growth was compared between induced and uninduced expression conditions. In (C) and (E), significance was calculated via two tailed Student's t test (****, $p < 0.0001$; n.s., not significant). Data are shown as mean \pm s.e.m. ($n = 3$). (F) Schematic of bacterial growth assays at different temperatures ranging from 30°C to 42°C. (G) Heatmap comparing the percent survival of bacteria expressing either Csx30-1, CSX30-2, full length CSX30 and CSX31, full length CSX30, CSX30-1 and CSX31, or CSX30-2 and CSX31 at three different temperatures. (percent survival is calculated by the ratio of OD600 of the bacteria in the induced condition over the non-induced condition). Color scale shows percent survival from 0 to 100 percent.

Fig. 5 | Potential mechanism of cell growth inhibition by the Cas7-11–Csx29 effector complex.

The schematic presents a proposed mechanism of the RNA-triggered proteolytic activation of Csx30 by the Cas7-11–Csx29 complex, which induces cell death as part of anti-viral immunity. Csx31 likely functions as an antitoxin, thereby protecting the cell from the toxic effect of the Csx30 NTD. The Csx30 NTD bound to RpoE could affect cell growth and viability through unknown mechanisms.

Supplemental figure legends

fig. S1 | Cryo-EM analysis.

(A and B) Single-particle cryo-EM image processing workflows for Cas7-11–crRNA–Csx29 (A) and Cas7-11–crRNA–Csx29–tgRNA (B).

(C and D) Fourier shell correlation curves calculated between the half-maps and between the refined model and the density map for Cas7-11–crRNA–Csx29 (C) and Cas7-11–crRNA–Csx29–tgRNA (D).

fig. S2 | Cryo-EM density maps.

(A and B) Cryo-EM density maps, colored according to the protein domains, for Cas7-11–crRNA–Csx29 (A) and Cas7-11–crRNA–Csx29–tgRNA (B).

(C and D) Cryo-EM density maps, colored according to the local resolution, for Cas7-11–crRNA–Csx29 (C) and Cas7-11–crRNA–Csx29–tgRNA (D).

fig. S3 | Structural comparison of the Cas7-11 complexes in different states.

(A–C) Structures of Cas7-11–crRNA–tgRNA (PDB ID: 7WAH) (A), Cas7-11–crRNA–Csx29 (B), and Cas7-11–crRNA–Csx29–tgRNA (C). The bound zinc ions are shown as orange spheres. The disordered L1 and L2 linkers are not shown for clarity. The bound RNA molecules are shown on the right of the complexes.

fig. S4 | RNA recognition by Cas7-11.

(A) Recognition of the crRNA 5' end by the Cas7.1 domain. The density map is shown as a gray

mesh. The possible location of U(-16) and the pre-crRNA processing site are indicated by a dashed circle and a cyan triangle, respectively.

(B) Recognition of the guide-target duplex by the Cas7.2 and Cas7.3 domains. The catalytic residues (D429A and D654A) are depicted as space-filling models. The target RNA cleavage sites are indicated by yellow and green triangles.

fig. S5 | Structural comparison between Csx29 and human separase.

Overall structures of Csx29 and human separase (PDB ID: 7NJ1). The catalytic residues are depicted as space-filling models. Securin (separase inhibitor) is colored gray. The close-up views of the protease active sites are shown in insets.

fig. S6 | Interaction between Cas7-11 and Csx29.

(A) Interface between Cas7-11 and Csx29 in the Cas7-11-crRNA-Csx29 complex. Cas7-11 and Csx29 are shown as ribbon and surface representations, respectively. The INS and CTE domains of Cas7-11 are omitted for clarity.

(B–D) Structures of the Cas7.1–Cas7.4 domains in Cas7-11-crRNA-tgRNA (PDB ID: 7WAH) (B), Cas7-11-crRNA-Csx29 (C), and Cas7-11-crRNA-Csx29-tgRNA (D). The bound zinc ions are shown as orange spheres. The α -helical insertion in the Cas7.4 ZF motif is highlighted by magenta.

(E) Potential steric clash between Cas7-11/Csx29 and a bound target RNA. The Cas7-11-crRNA-Csx29 structure was superimposed onto the Cas7-11-crRNA-Csx29-tgRNA structure, and then only Csx29 and Cas7-11 L2 are shown in semi-transparent gray. A potential steric clash between Cas7-11 L2/Csx29 TPR1 and the bound target RNA is indicated by a dashed orange circle.

fig. S7 | Interface between Cas7-11 and Csx29.

(A) Interface between Cas7-11 Cas7.4 and Csx29 NTD.

(B) Interface between Cas7-11 Cas7.3/L2 and Csx29 NTD.

(C) Interface between Cas7-11 L2 and the Csx29 NTD/TPR.

(D) Interface between Cas7-11 Cas7.3 and Csx29 TPR1/2.

fig. S8 | Target RNA cleavage by the Cas7-11–Csx29 complex.

The Cas7-11-crRNA-Csx29 complex was incubated with a 5'-Cy5-labeled ssRNA target at 37°C for 10 min, and then analyzed by 15% TBE-urea PAGE. The gels were visualized, using either Cy5 or SYBR Gold fluorescence. The wild-type (W) and catalytically inactivated (D) versions of Cas7-11 and Csx29 were used.

fig. S9 | N-terminal analysis of Csx30.

Elution profiles for N-terminal seven residues in the ~15 kDa Csx30 fragment (Csx30-2) were shown.

fig. S10 | Predicted structures of Csx30.

The Csx30 structure was predicted using AlphaFold2(16). The C α atoms of M427 and K428 at the cleavage site are indicated by spheres.

fig. S11 | Effects of Csx30 and Csx31 on bacterial cell growth.

(A) Growth curves of *E. coli* expressing the non-induced full-length Csx30, the N-terminal fragment (residues 1–427) of Csx30 (Csx30-1), or the C-terminal fragment (residues 428–565) of Csx30 (Csx30-2). These curves serve as non-induced controls for the curves in Fig. 4B.

(B) Electrostatic surface potential of the Csx30 and Csx31 structures predicted using AlphaFold2(16). The predicted structures suggested that Csx30 and Csx31 have negatively and positively charged surfaces, respectively.

(C) Growth curves of *E. coli* expressing non-induced Csx30-1, full-length Csx30 and Csx31, or Csx30-1 and Csx31. These curves serve as non-induced controls for the curves in Fig. 4D.

fig. S12 | Interaction between Csx30, Csx31, and RpoE.

(A and B) Elution profiles of the Csx30–Csx31–RpoE complex from a gel-filtration column. Csx30, Hig₆-tagged Csx31, and Hig₆-tagged RpoE were co-expressed in *E. coli*, and purified by Ni-NTA and HiLoad 16/600 Superdex 200. In (A), the Csx30–Csx31–RpoE complex was loaded onto Superdex 200 Increase. In (B), the Csx30–Csx31–RpoE complex was incubated with the Cas7-11–crRNA–Csx29–tgRNA complex, and then loaded onto Superdex 200 Increase. The fractions indicated by orange lines were analyzed by SDS-PAGE. The gels were stained with CBB.

(C) Predicted structure of the Csx30–Csx31–RpoE complex. The structures of Csx30–Csx31 and Csx30–RpoE were predicted using AlphaFold2(16), and then they are superimposed based on the Csx30 NTDs. The Csx30 CTD in Csx30–RpoE is omitted for clarity.

(D) Structural comparison of *D. ishimotonii* RpoE (model) and *E. coli* RpoE (PDB ID: 6JBQ).

(E) Structural comparison of the Csx30 CTD (model) and CagX (PDB ID: 6OEG).

fig. S13 | Multiple sequence alignment of the N-terminal domain of the Csx30 orthologs.

Multiple alignment of the N-terminal domain of Csx30 family proteins was built using the Muscle5 program(28), and colored using Sequence Manipulation Suite (http://www.bioinformatics.org/sms2/color_align_cons.html). The Csx30 from *D. ishimotonii* is marked by an asterisk. Lysine residues, equivalent to K427 of *D. ishimotonii* Csx30 at the cleavage site by Csx29, is highlighted by red.

fig. S14 | Multiple sequence alignment of the C-terminal domain of the Csx30 orthologs.

Multiple alignment of the C-terminal domains of Csx30 family proteins were built using the Muscle5

program and colored using http://www.bioinformatics.org/sms2/color_align_cons.html server. Three families are represented by a single sequence and are not therefore aligned.

Method details

Plasmid construction

For the bacterial expression of the *D. ishimotonii* Cas7-11–crRNA–Csx29 complex, the gene encoding Cas7-11 was amplified by PCR and cloned into the modified pACYCDuet-1 plasmid vector (Novagen), expressing Cas7-11 with an N-terminal maltose-binding protein (MBP) and a C-terminal His₆-tag (MBP–Cas7-11–His₆). The gene encoding Csx29 was cloned into the His₆-Twin-Strep–SUMO-pET28a vector, expressing Csx29 with an N-terminal His₆-Twin-Strep–SUMO tag (His₆–Twin-Strep–SUMO–Csx29). The CRISPR array containing two direct repeats interspaced by a spacer with the 5' LacI-repressed T7 promoter and 3' T7 terminator sequences was synthesized by Eurofins Genomics. For the bacterial expression of Csx30 and Csx30–Csx31–RpoE, the gene encoding Csx30 or Csx30 and Csx31 was amplified from the type III-E *D. ishimotonii* CRISPR locus and cloned into the modified pE-SUMO vector (LifeSensors), in which the SUMO-coding region is replaced with the HRV3C protease recognition site. The gene encoding RpoE was cloned into the pACYCDuet-1 vector, expressing RpoE with an N-terminal His₆-tag. The mutants of Cas7-11, Csx29, and Csx30 were generated by a PCR-based method, and the sequences were confirmed by DNA sequencing.

Sample preparation

Cas7-11, Csx29, and the CRISPR array were co-expressed in *E. coli* BL21 (DE3) (Novagen) by induction with 0.25 mM isopropyl β-D-thiogalactopyranoside (Nacalai Tesque) at 18°C overnight. The *E. coli* cells were lysed by sonication in buffer A (20 mM Tris-HCl, pH 7.5, 20 mM imidazole, 150 mM NaCl, 10% glycerol, and 3 mM 2-mercaptoethanol), and the lysate was clarified by centrifugation at 40,000 g. The supernatant was applied to Ni-NTA Superflow resin (QIAGEN), and the bound protein was eluted with buffer A containing 300 mM imidazole. The eluted fraction was diluted 2-fold with buffer B (20 mM Tris-HCl, pH 7.5, 150 mM NaCl, 10% glycerol, and 3 mM 2-mercaptoethanol), and applied to Strep-Tactin XT high capacity (IBA), and the bound protein was eluted with buffer C (100 mM Tris-HCl, pH 8.0, 150 mM NaCl, 10% glycerol, 1 mM EDTA, 50 mM biotin, and 3 mM 2-mercaptoethanol). The eluted protein was applied to Amylose resin (NEB) equilibrated with buffer D (20 mM Tris-HCl, pH 7.5, 150 mM NaCl, 10% glycerol, and 1 mM DTT). The resin was washed with buffer D (20 mM Tris-HCl, pH 7.5, 150 mM NaCl, 10% glycerol, and 1 mM DTT), and the bound protein was eluted with buffer D containing 10 mM D-maltose. For biochemical experiments, the eluted protein was dialyzed against buffer D to remove D-maltose. The concentration of the Cas7-11–Csx29–crRNA complex was measured using Pierce 660-nm Protein Assay Reagent (Thermo Fisher Scientific). Csx30 was expressed in *E. coli* Rosetta2 (DE3), and purified by Ni-NTA Superflow resin and a Superdex 200 Increase column (GE Healthcare). Csx30, Csx31, and RpoE were co-expressed in *E. coli* BL21 (DE3), and the Csx30–Csx31–RpoE complex was purified by Ni-NTA Superflow resin and a HiLoad 16/600 Superdex 200 column (GE Healthcare). The target RNA (25 nucleotides plus 5' GG) was transcribed *in vitro* with T7 RNA

polymerase, and purified by 10% denaturing (7 M urea) polyacrylamide gel electrophoresis. The purified materials were stored at -80°C until use.

Cryo-EM grid preparation and data collection

To prevent tgRNA cleavage, the catalytically inactive Cas7-11 (D429A/D654A) was used for cryo-EM studies. The Cas7-11–crRNA–Csx29–tgRNA complex was reconstituted by mixing the purified Cas7-11–crRNA–Csx29 complex and the target RNA, at a molar ratio of 1:4. The complex was purified by size-exclusion chromatography on a Superose 6 Increase 10/300 column (GE Healthcare), equilibrated with buffer E (20 mM HEPES-NaOH, pH 7.0, 150 mM NaCl, 1 mM MgCl_2 , and 1 mM DTT). The peak fraction containing Cas7-11–crRNA–Csx29–tgRNA was analyzed by TBE-urea gel, and concentrated to an A_{260} of 3.0, using an Amicon Ultra-4 Centrifugal Filter Unit (MWCO 50 kDa). The sample (3 μL) was applied to freshly glow-discharged Au 300 mesh R1.2/1.3 grids (Quantifoil) in a Vitrobot Mark IV (FEI) at 4°C , with a waiting time of 10 s and a blotting time of 4 s under 100% humidity conditions. The grids were plunge-frozen into liquid ethane cooled at liquid nitrogen temperature. For the cryo-EM analysis of the Cas7-11–crRNA–Csx29 complex, the purified complex was further polished by size-exclusion chromatography on a Superose 6 Increase 10/300 column, equilibrated with buffer E. The peak fraction containing Cas7-11–crRNA–Csx29 was concentrated to an A_{260} of 2.5. The sample was applied onto freshly glow-discharged Au 300 mesh R0.6/1 grids (Quantifoil) in a Vitrobot Mark IV at 4°C , with a waiting time of 10 s and a blotting time of 4 s under 100% humidity conditions. The cryo-EM data were collected using a Titan Krios G3i microscope (Thermo Fisher Scientific), running at 300 kV and equipped with a Gatan Quantum-LS Energy Filter (GIF) and a Gatan K3 Summit direct electron detector. Micrographs were recorded at a nominal magnification of $\times 105,000$ with a pixel size of 0.83 Å in a total exposure of $48\text{ e}^{-}/\text{\AA}^2$ per 48 frames. The data were automatically acquired by the image shift method using the EPU software (Thermo Fisher Scientific), with a defocus range of -0.8 to $-2.0\text{ }\mu\text{m}$, and 2,924 and 3,179 movies were acquired for Cas7-11–crRNA–Csx29 and Cas7-11–crRNA–Csx29–tgRNA, respectively.

Image processing

The data processing was performed using cryoSPARC v3.3.1 software packages(29). The dose-fractionated movies were aligned using the Patch motion correction and the contrast transfer function (CTF) parameters were estimated using Patch-Based CTF estimation. Particles were automatically picked using Blob picker and template picker followed by reference free 2D classification to curate particle sets. The particles were further curated by Heterogeneous Refinement using the map derived from cryoSPARC *Ab initio* Reconstruction as a template. The selected particles after Heterogeneous Refinement were refined using Non-uniform refinement(30). Local motion correction followed by Non-uniform refinement with optimization of CTF value yielded a map at 2.48 Å and 2.77 Å resolution for Cas7-11–crRNA–Csx29 and Cas7-11–crRNA–Csx29–tgRNA, respectively, according

to the Fourier shell correlation (FSC) = 0.143 criterion(31). The local resolution was estimated by BlocRes in cryoSPARC.

Model building and validation

For model building for Cas7-11, the previously published Cas7-11 structure (PDB ID: 7WAH) was rigid-body fitted into the reconstructed density maps in UCSF ChimeraX(32). The initial model of Csx29 was predicted using AlphaFold2(16), and fitted into the density map using the Dock Predicted Model in PHENIX package(33). These models were manually modified using COOT(34) against the density map sharpened using DeepEMhancer(35). The models were refined using Real-space refinement in PHENIX(36) with the secondary structure and the Ramachandran restraints. Since the MBP, His₆ and SUMO tags were not resolved in the density map, they were not included in the final models. The structures were validated using MolProbity(37) from the PHENIX package. The curve representing model vs. full map was calculated using phenix.mtriage(38), based on the final model and the full, filtered and sharpened map. The statistics of the 3D reconstruction and model refinement are summarized in Table S1. The cryo-EM density maps were calculated with UCSF ChimeraX, and molecular graphics figures were prepared with CueMol (<http://www.cuemol.org>).

***In vitro* Csx30 cleavage experiment**

The purified Cas7-11–crRNA–Csx29 complex (750 nM) was incubated at 37°C for 2 h with the purified Csx30 protein (15 μM) in the presence or absence of the tgRNA (3 μM) in reaction buffer (20 mM HEPES-NaOH, pH 7.5, 150 mM NaCl, 5 mM MgCl₂, and 2 mM DTT). The reaction was quenched by the addition of an SDS-PAGE sample buffer, and the mixture was then analyzed by SDS-PAGE. The gels were stained with Bullet CBB Stain One (Nacalai Tesque), and then imaged using a FUSION Solo S system (Vilber Bio Imaging).

***In vitro* target RNA cleavage experiment**

The purified Cas7-11–crRNA–Csx29 complex (200 nM) was incubated at 37°C for 10 min with a 5'-Cy5-labeled ssRNA target (600 nM) in reaction buffer (20 mM HEPES-NaOH, pH 7.5, 50 mM NaCl, 5 mM MgCl₂, and 2 mM DTT). The reaction was quenched by the addition of quenching solution (0.45 mg/mL proteinase K (Nacalai Tesque), 6 mM EDTA, and 200 μM urea), and then incubated at 50°C for 15 min. The mixture was incubated at 100°C for 2 min with 4.5 M urea denaturing buffer, and then analyzed using a 15% Novex PAGE Tris–borate–EDTA (TBE)–urea gel (Invitrogen). The gels were imaged using a FUSION Solo S system, using either Cy5 fluorescence or SYBR Gold fluorescence (Thermo Fisher Scientific).

N-terminal analysis

The purified nuclease-inactivated Cas7-11–crRNA–Csx29 complex (3 μM) was incubated at 37°C for

1.5 h with the purified Csx30 protein (15 μ M) in the presence of the tgRNA (12 μ M) in reaction buffer (20 mM HEPES-NaOH, pH 7.5, 150 mM NaCl, and 2 mM DTT). The proteins were separated by SDS-PAGE, blotted onto a PVDF membrane, and stained with Bullet CBB Stain One. The protein band corresponding to a ~15 kDa Csx30 fragment (Csx30-2) was cut out from the membrane, and its N-terminal amino-acid sequence was analyzed by Edman microsequencing(39) on a Procise 494 cLC protein sequencer (Applied Biosystems), using the standard pulsed-liquid program for PVDF-blotted proteins.

Bacterial cell growth experiment

E. coli DH5 α Competent Cells (Thermo Fisher Scientific) were transformed with a Csx30-expressing plasmid alone or plasmids expressing both Csx30 and Csx31 (Table S2). Single colonies were picked into Terrific broth (Fisher Scientific) containing relevant antibiotics and 1% v/v glucose, and then cultured at 37°C overnight. Following overnight culture, bacterial OD₆₀₀ values were measured and normalized to an initial OD₆₀₀ of 0.00046 in 200 μ L final volume by dilution in Terrific broth containing relevant antibiotics as well as 1% glucose for non-induced conditions and 1% arabinose for induced conditions. Using BioCoat Cellware 96-well tissue culture plates (Corning), growth assays were performed using a BioTek Synergy Neo2 at 37°C with continuous shaking, reading OD₆₀₀ values at 10 min intervals for up to 22 hours. For growth curves with different temperatures, the same culturing conditions were used and the BioTek Synergy Neo2 is set to either 30°C, 37°C, or 42°C with continuous shaking for up to 22 hours during the measurement.

***In vitro* binding experiment**

The purified Csx30–Csx31–RpoE complex (15 μ M) was incubated at 37°C for 2 h with the purified nuclease-inactivated Cas7-11–crRNA–Csx29 complex (750 nM) in the presence of the tgRNA (3 μ M) in reaction buffer (20 mM HEPES-NaOH, pH 7.5, 150 mM NaCl, 5 mM MgCl₂, and 2 mM DTT). The mixture was separated on a Superdex 200 Increase 10/300 column equilibrated with buffer (20 mM HEPES-NaOH, pH 7.5, 150 mM NaCl, and 2 mM DTT). The peak fractions were analyzed by SDS-PAGE. The gels were stained with Bullet CBB Stain One, and then imaged using a FUSION Solo S system.

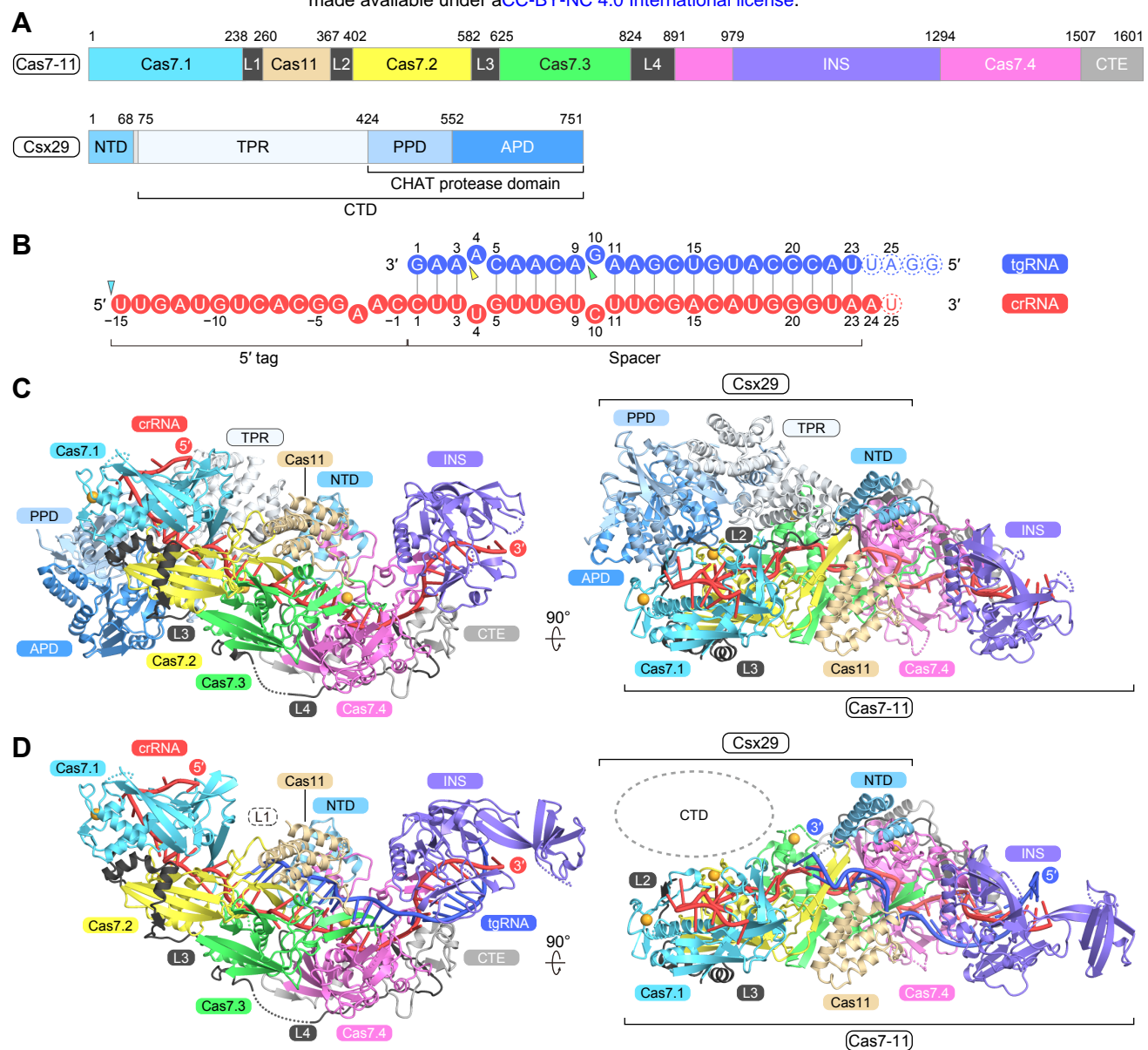


Figure 1

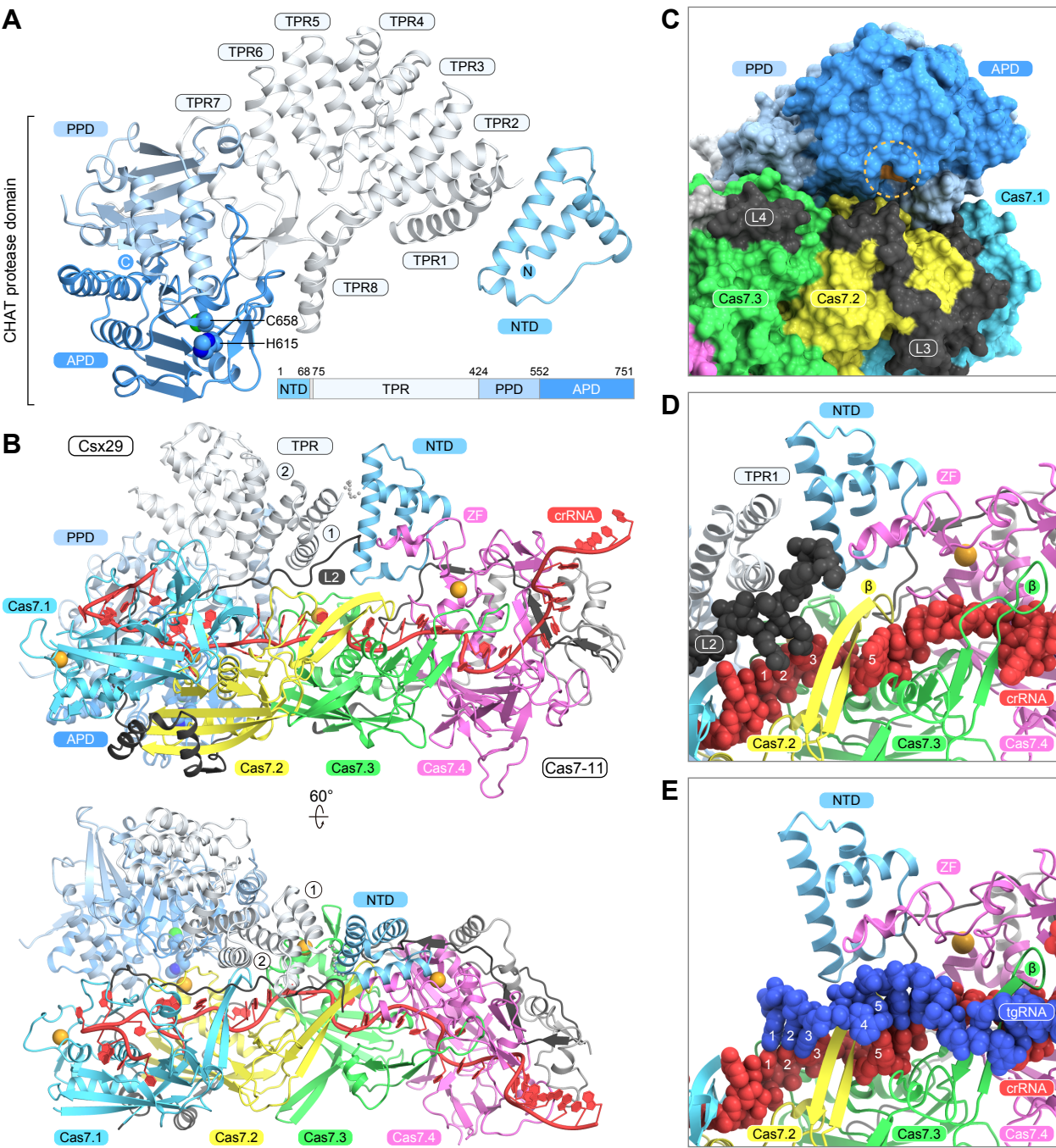
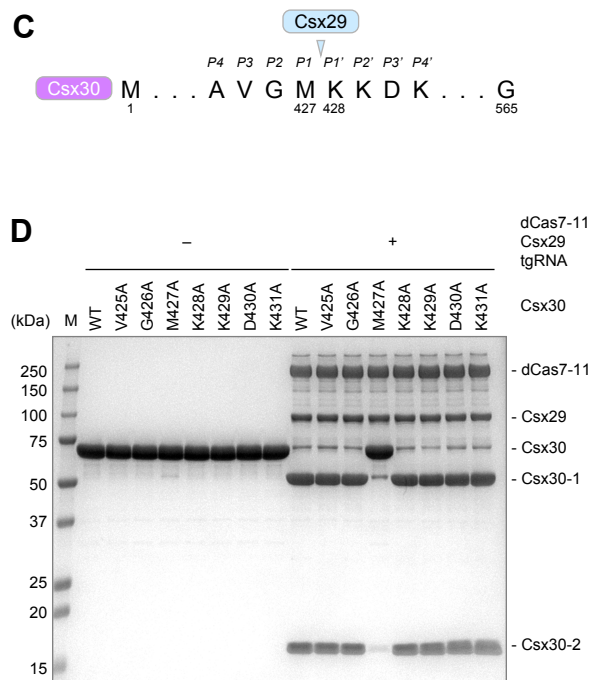


Figure 2



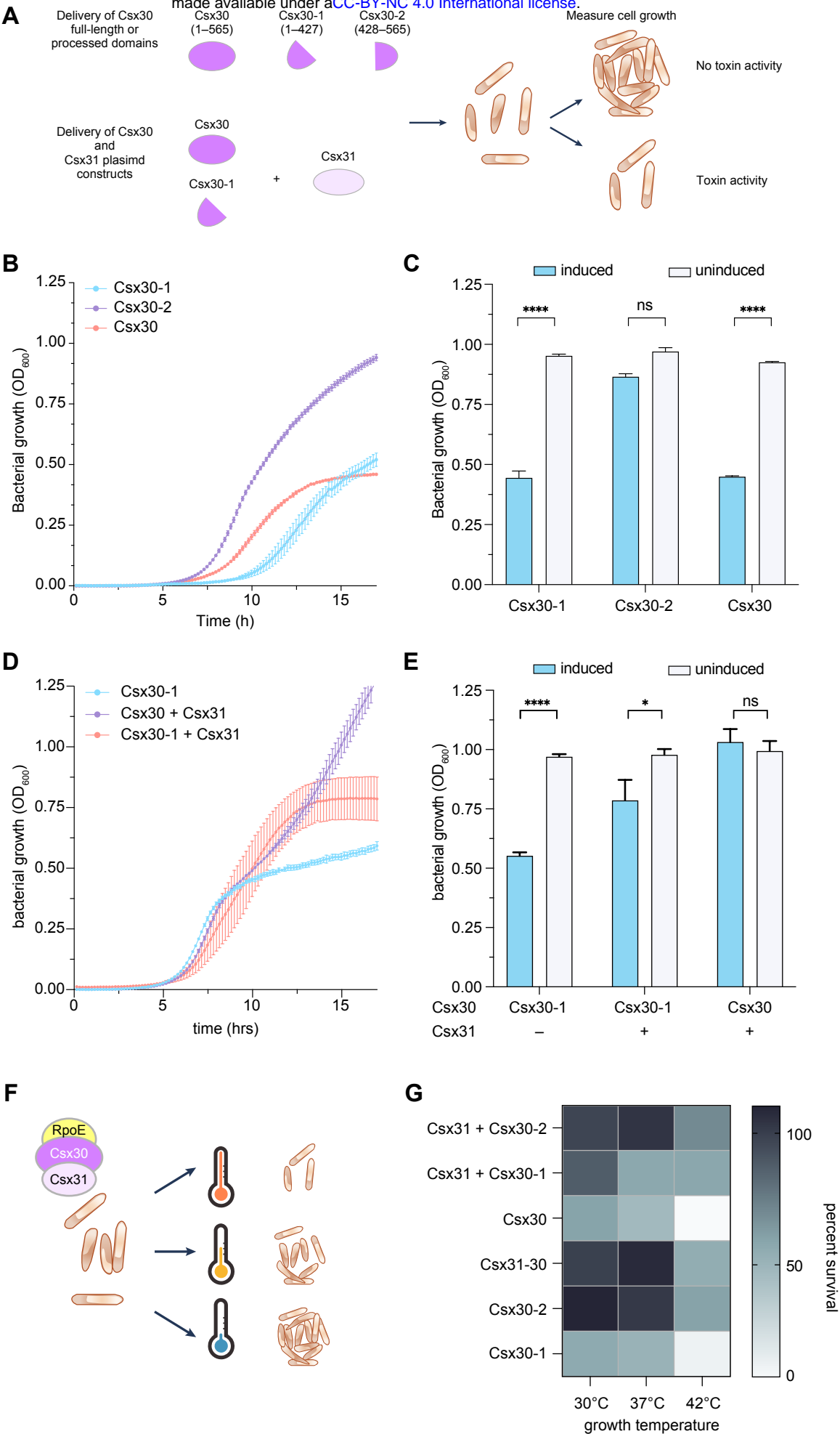


Figure 4

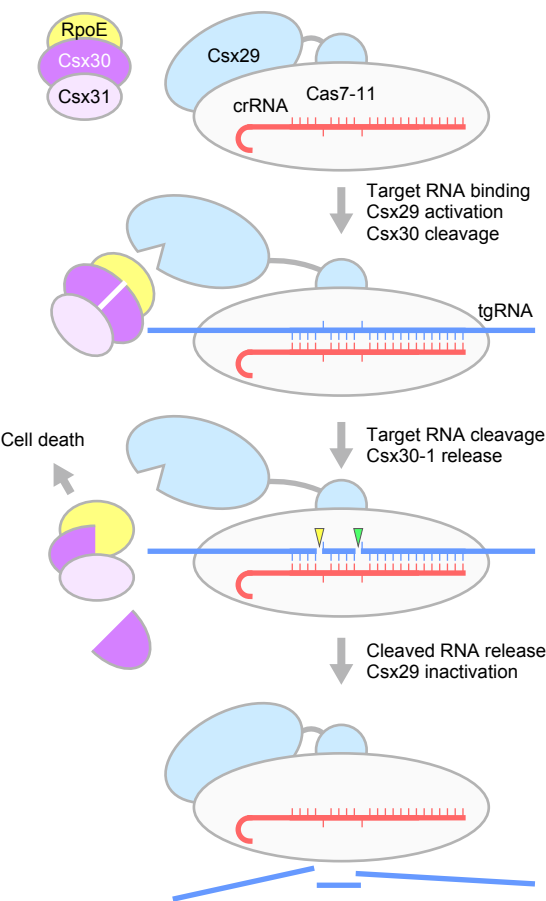


Figure 5

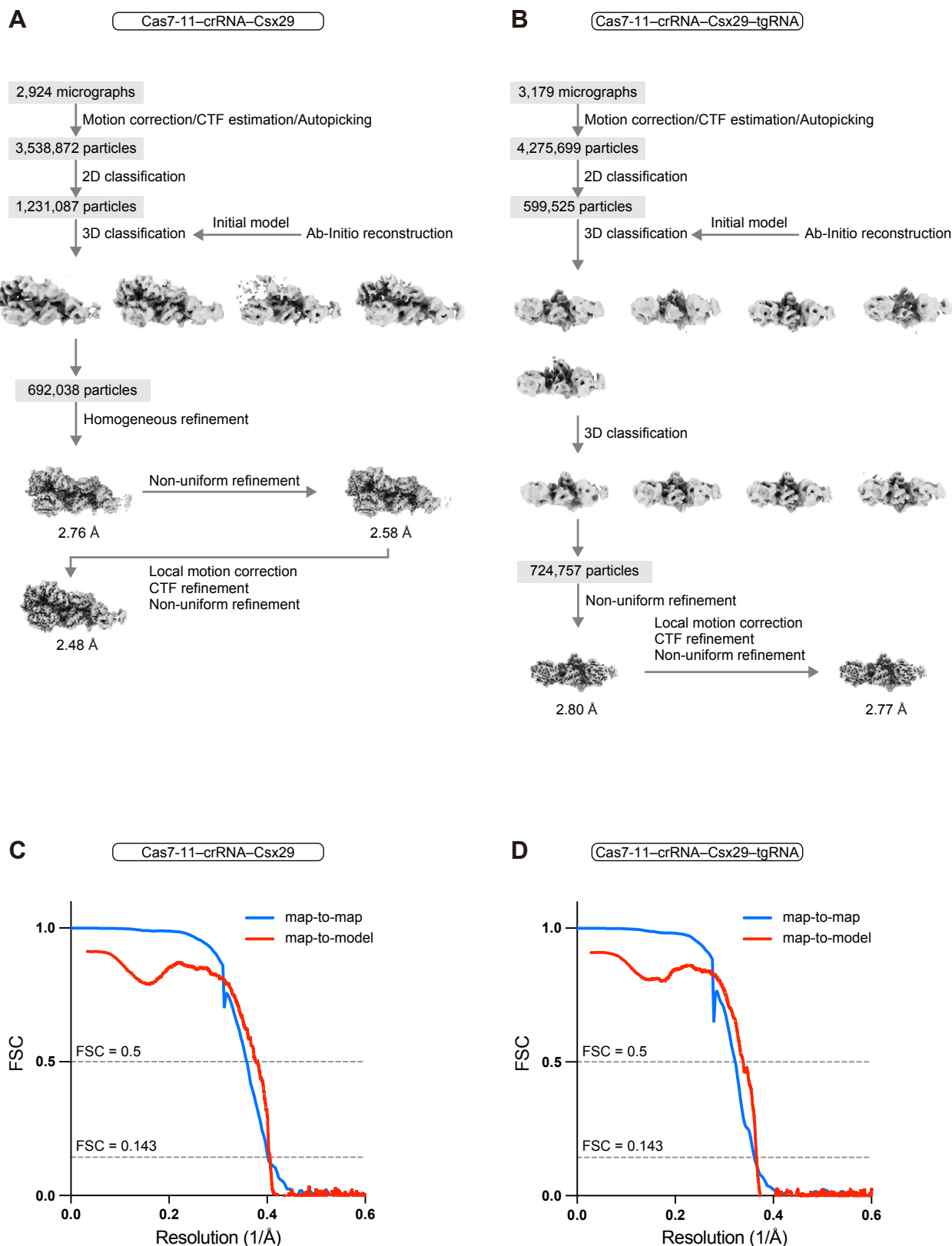


Figure S1

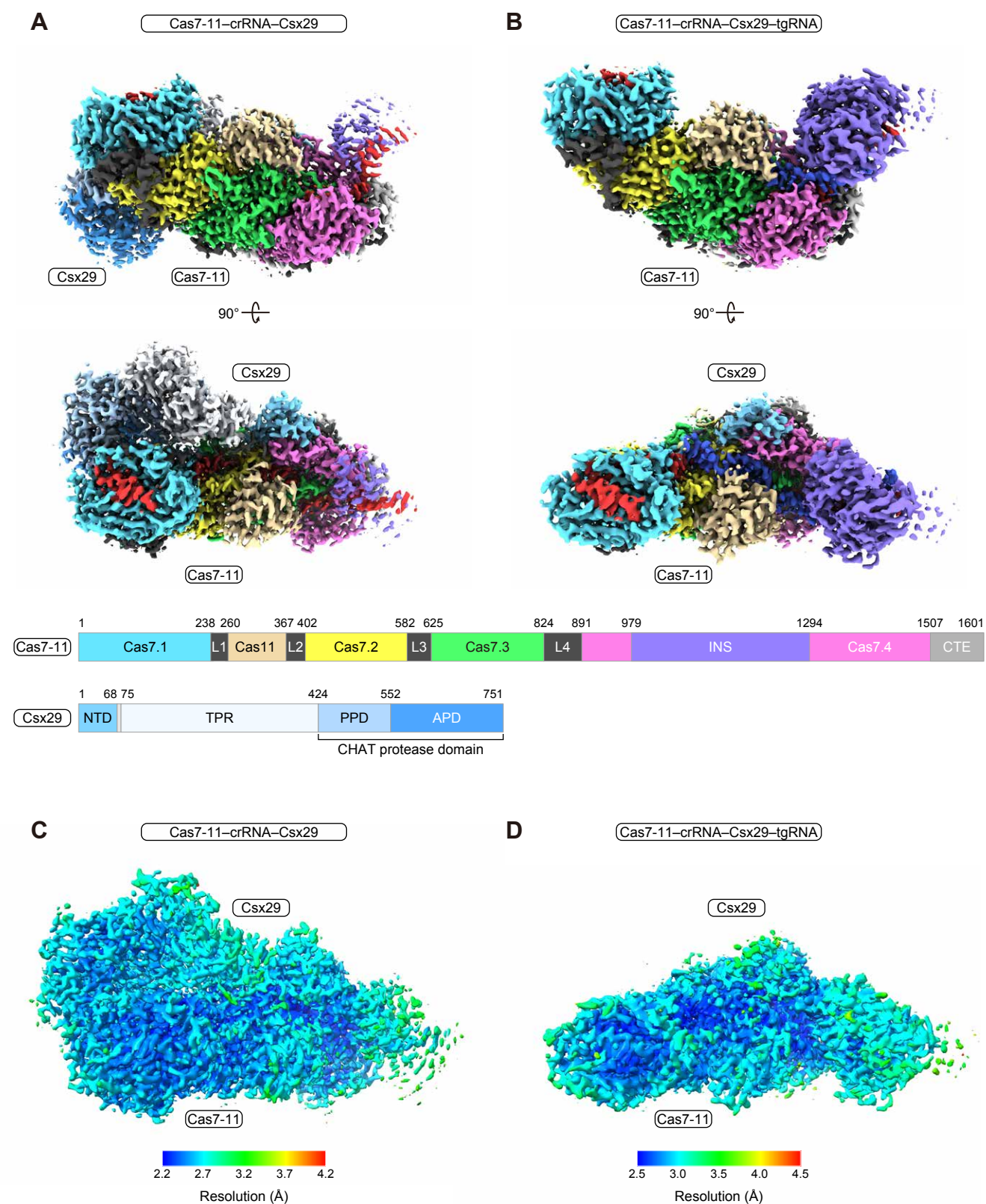


Figure S2

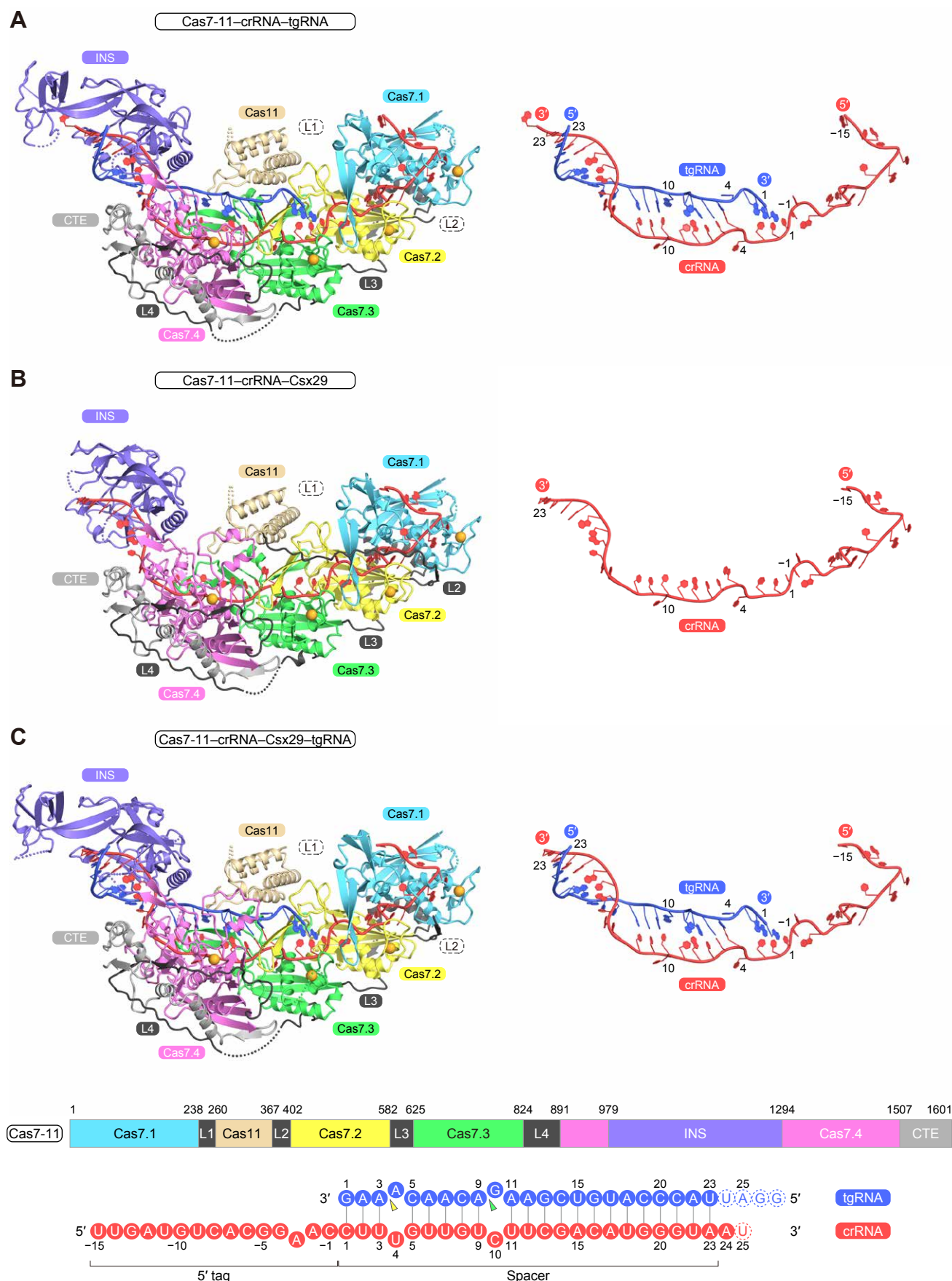


Figure S3

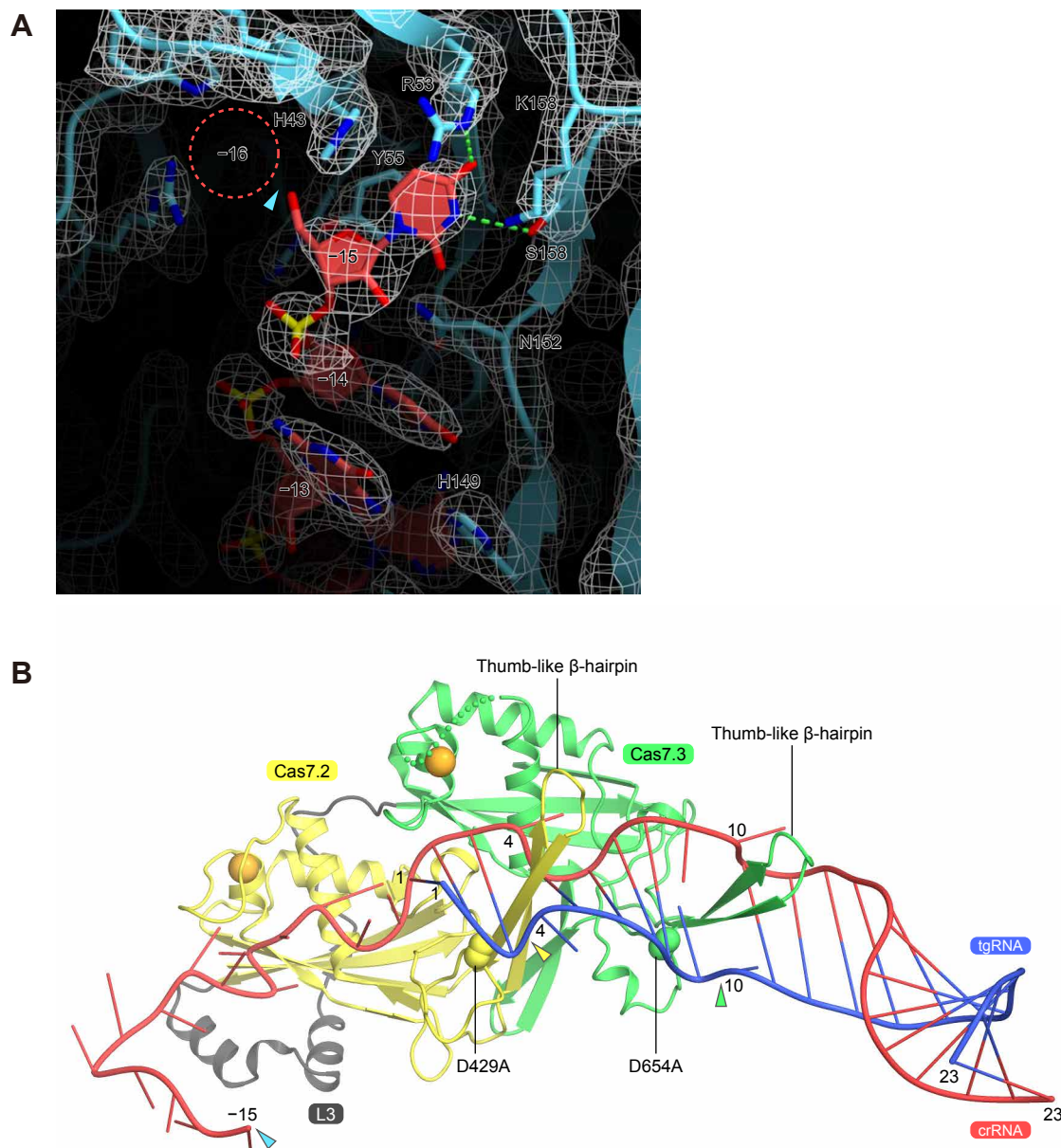


Figure S4

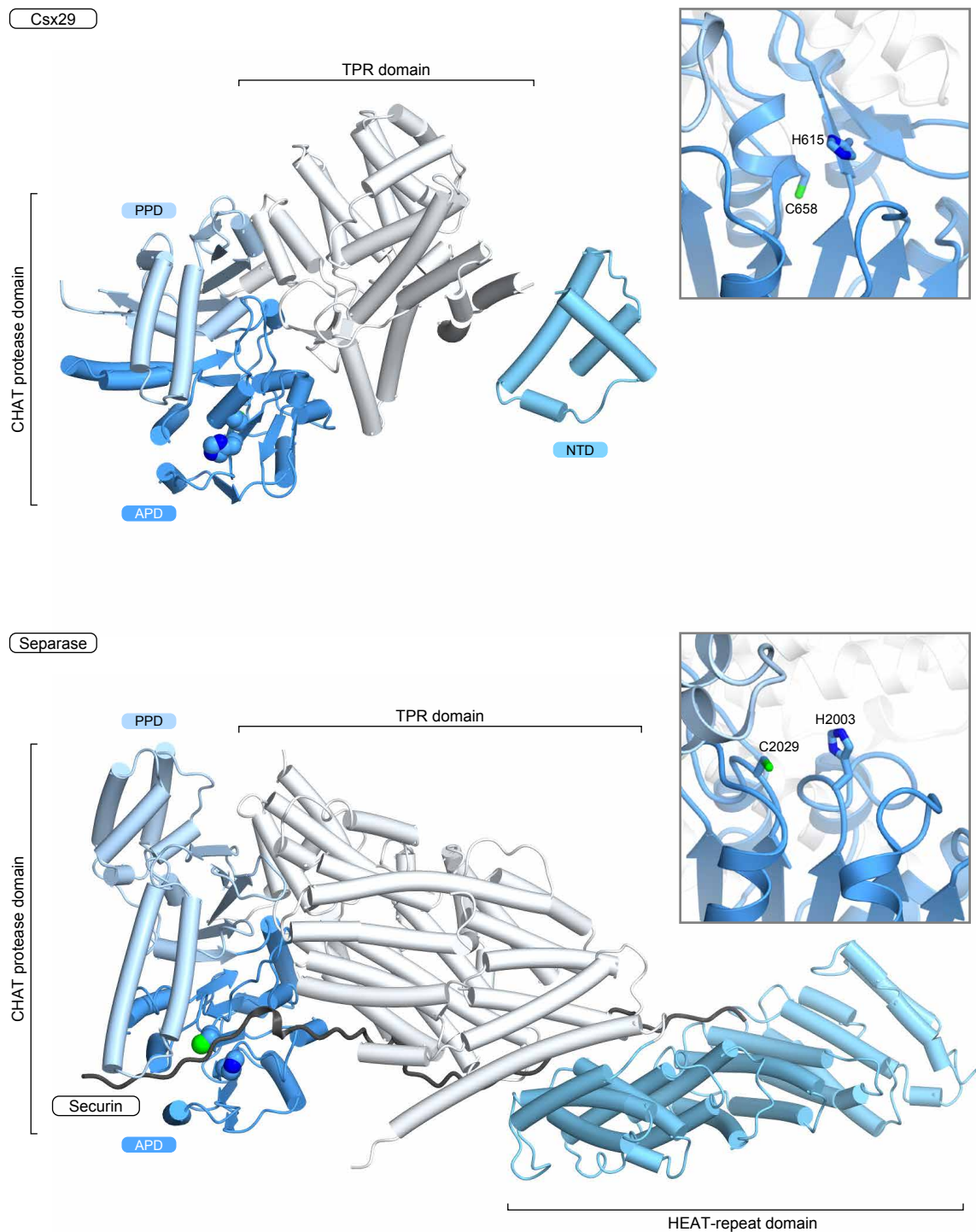


Figure S5

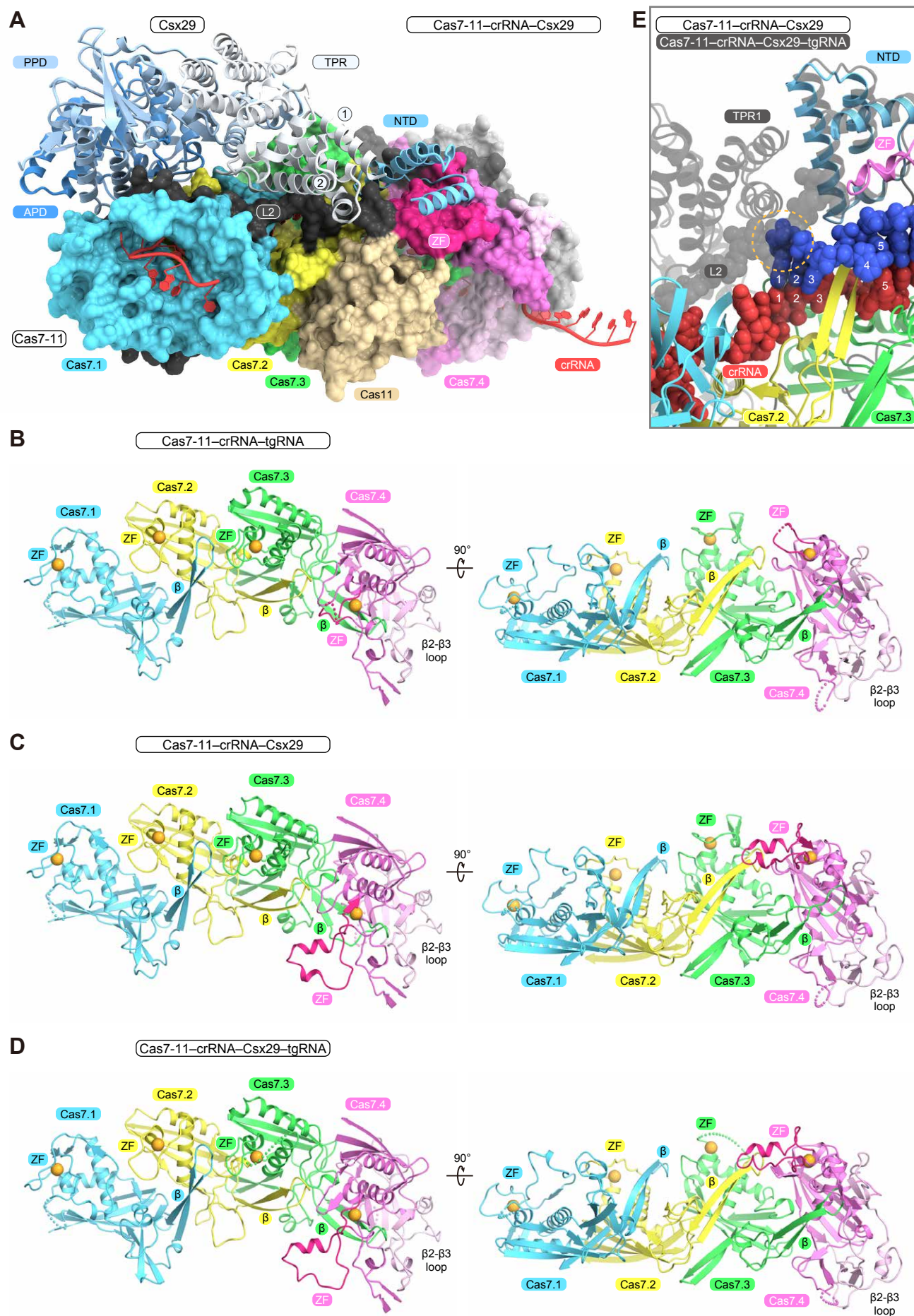


Figure S6

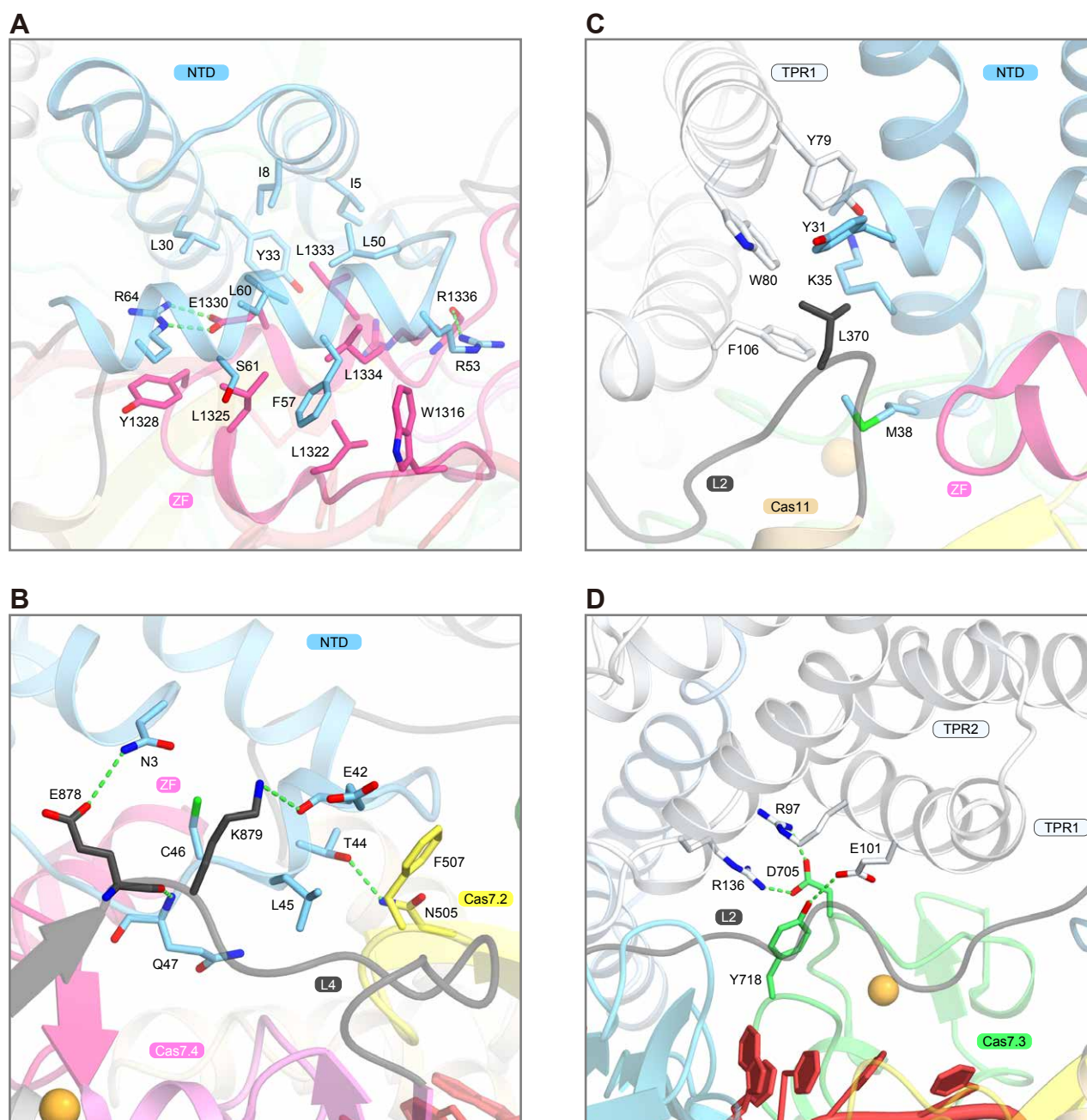


Figure S7

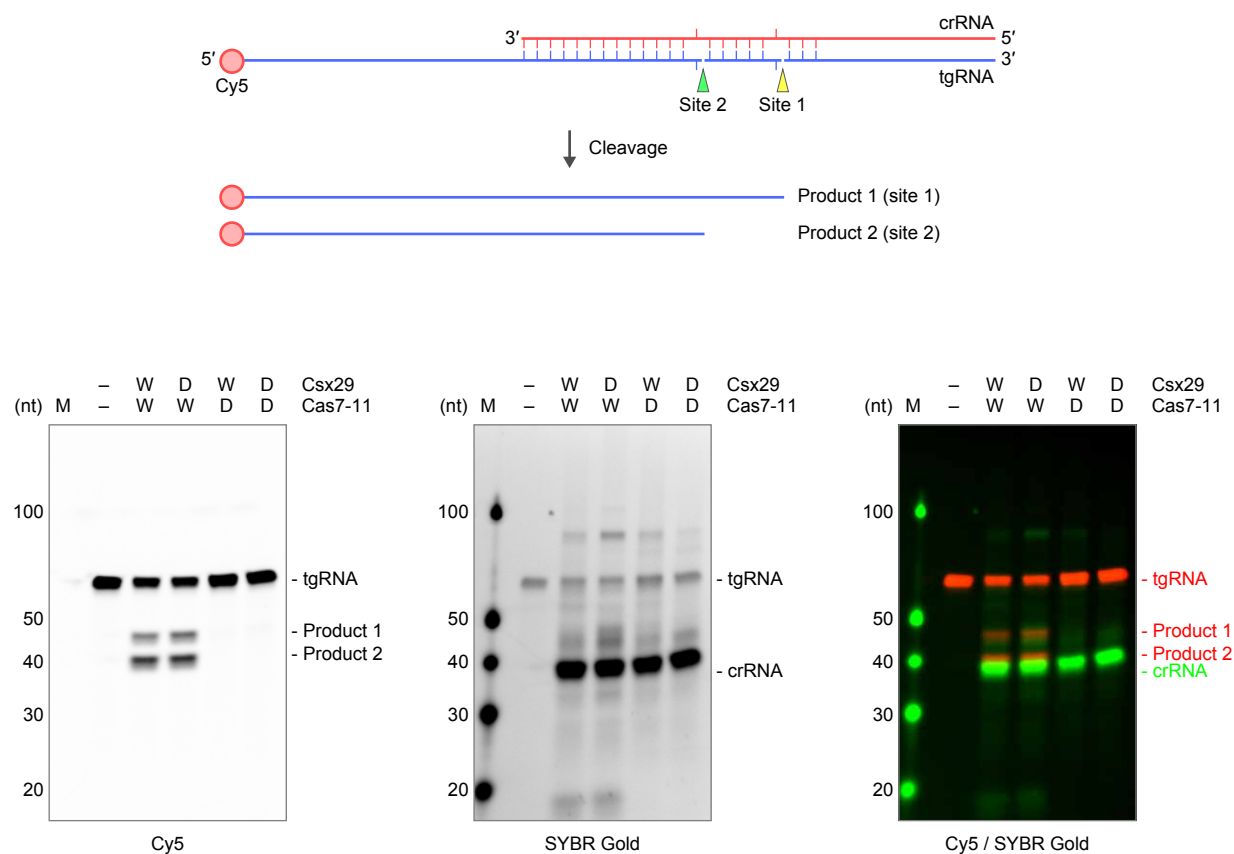


Figure S8

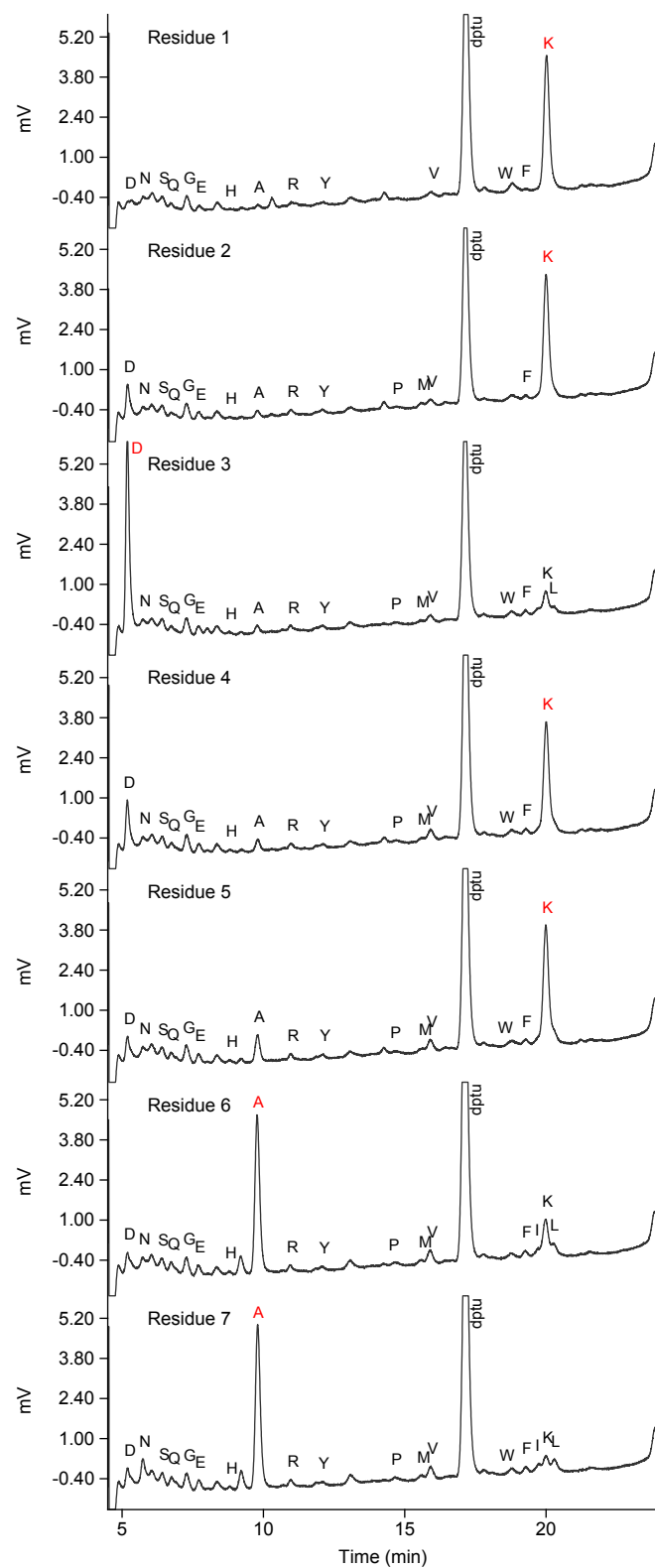


Figure S9

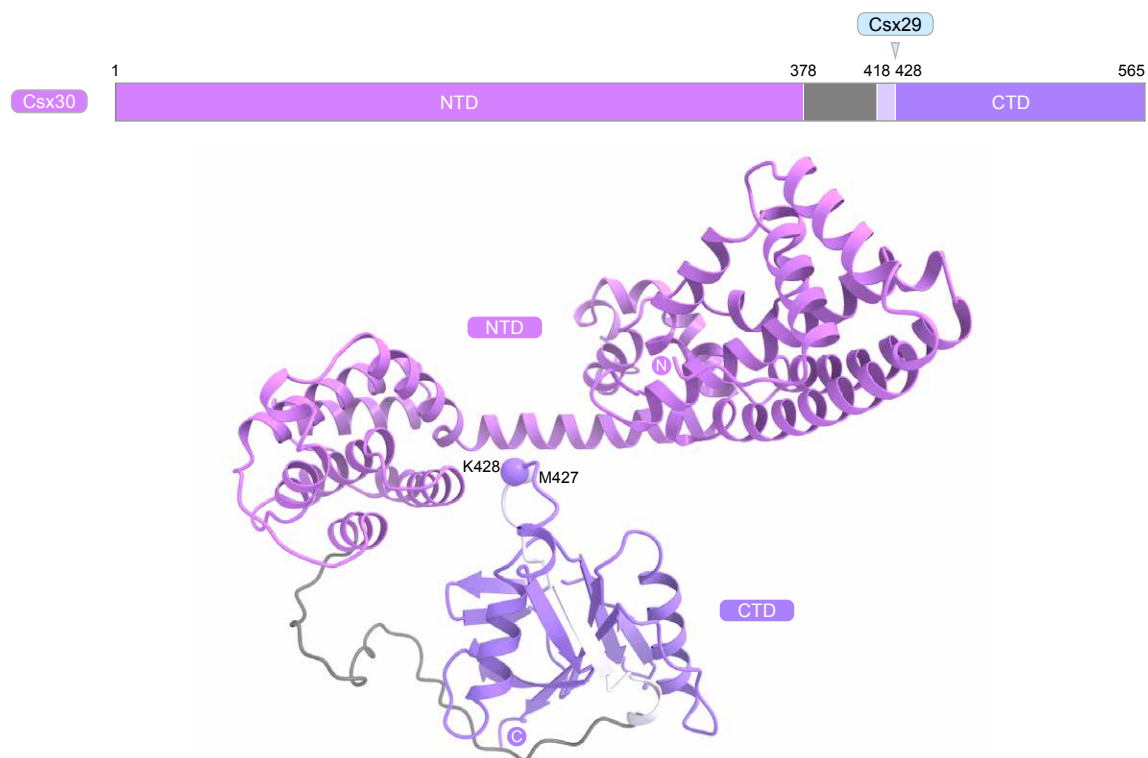


Figure S10

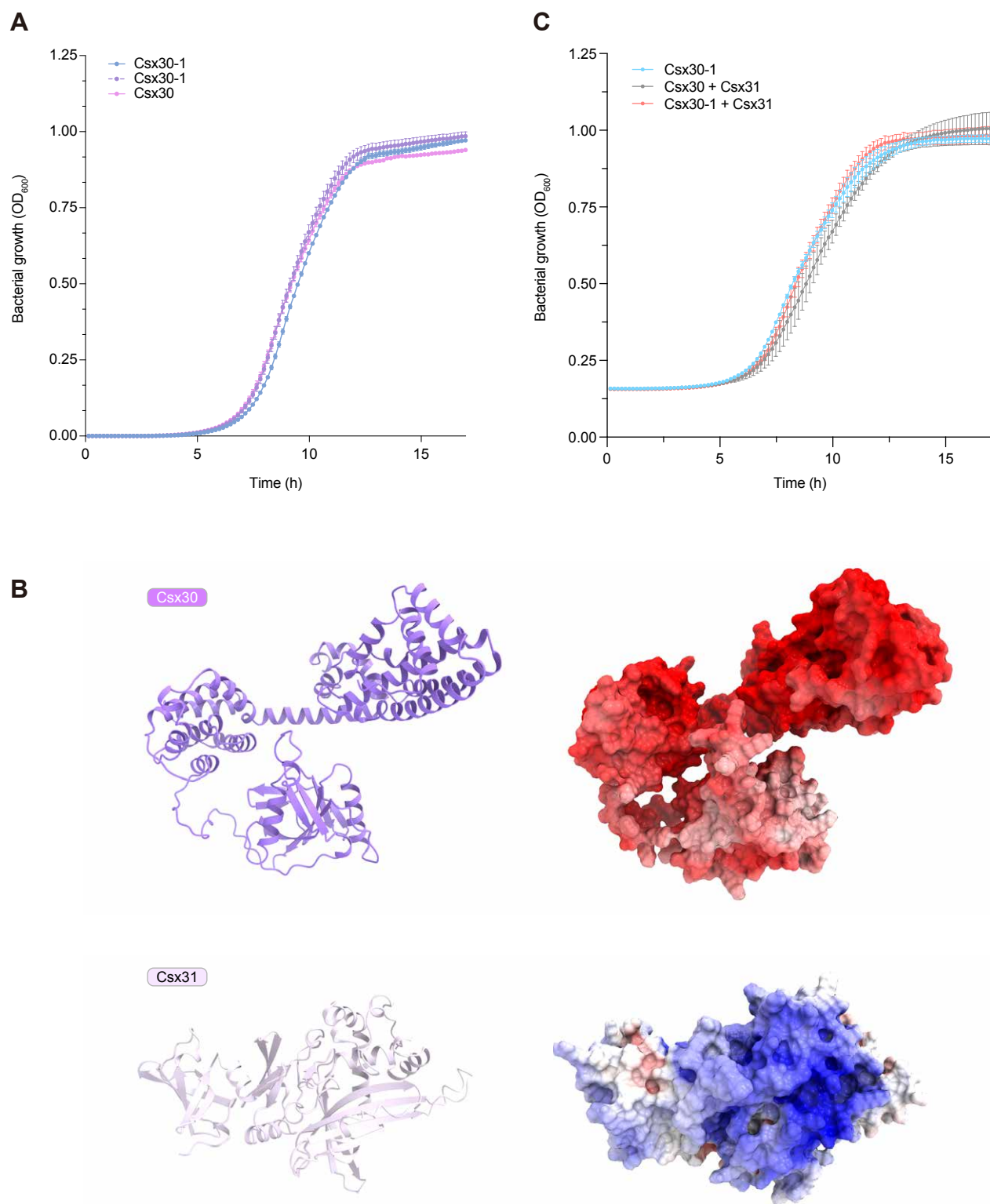


Figure S11

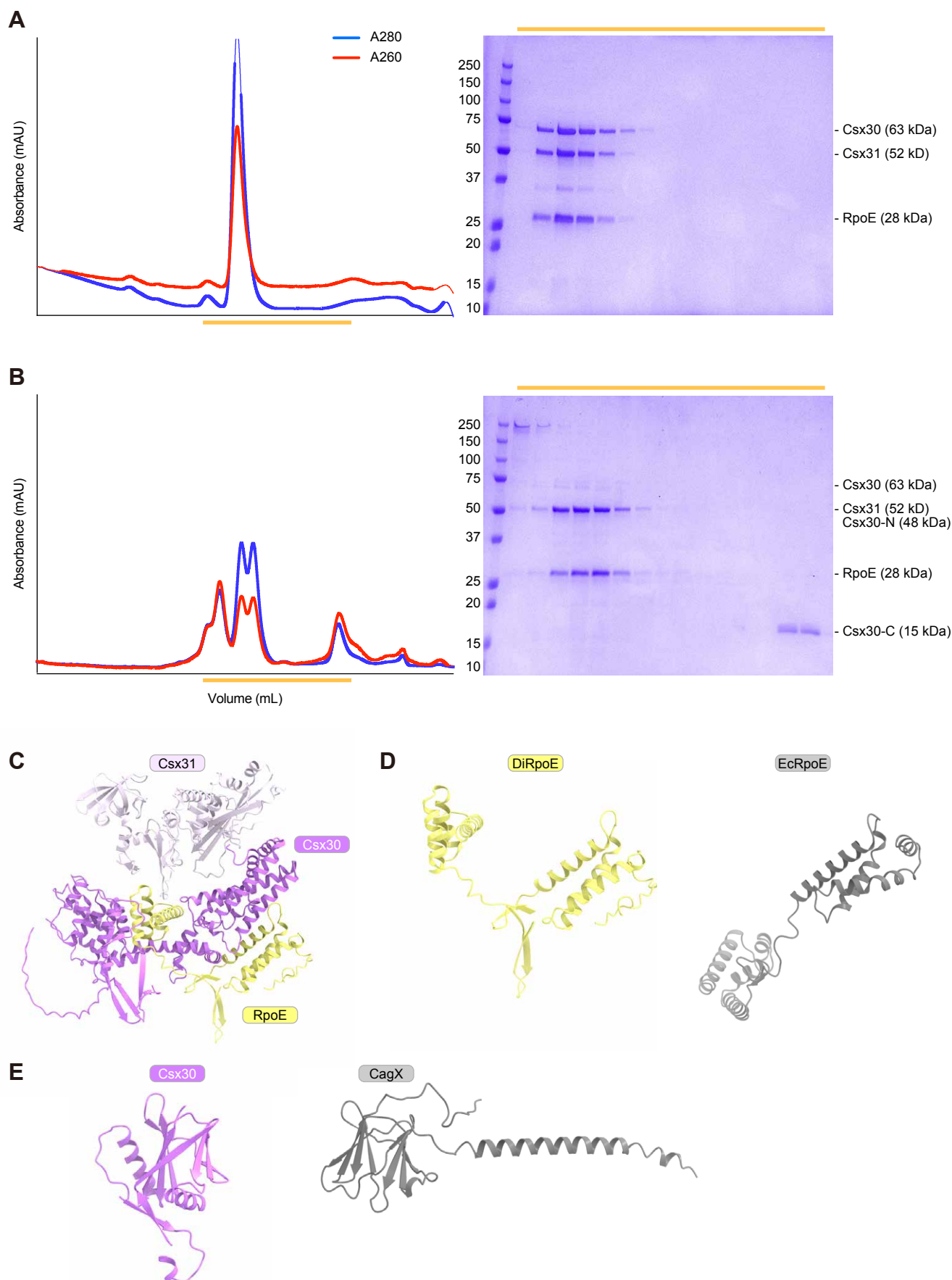


Figure S12

MCC7212117.1 -----MLLPKNQKRNFTNM--VQWGHKTHLSAGVP-ETEQQCLKEVDCRTLI---NSNVLDLDFSSIEEINKKVALAISAINDALITGTD 63
KHE91660.1 -----MYEHDYIGAI--LEWGHKTRILS--P-HIKYREDIEEYCKPFLLESNEDELCDLLRAIGELKQVESAISITFGDILKIVTST 77
NUN21992.1 -----MQPVMNMQRGFCNVL--LEWGHKTHS--P-QITSYLSLEIEYCKPFI--HDMRLDFLHSIEELKQVDTAVSILTSLELVST 80
MBC6927463.1 -----MQLTDKSRNELFSAL--LEWGHKTHS--P-EIVQDISEIBENCPEFI--KSDFTNFLNSVSEKIRKQVFAISLFTNIELVITAD 80
KAA0249750.1 -----MLSD--P-EIVQDISEIBENCPEFI--KSDFTNFLNSVSEKIRKQVFAISLFTNIELVITAD 58
WP_207678646.1 -----MFRDYKNRKFEDFLYSPTDIL--INWGRAHLIR--P-----EVLQETEDENWFLK--KKSDFTRGLLETGTASLEKTDENALNHLFGQETDLDL-- 82
MCF8114645.1 -----MSKEYTVDYAHLSSYSGKIDLLSTIAWATFFSWH-----NKYITKLTLEWQSVL--KNDDHPKLDLDDGEMEKVVSITLALITGP--MTLTI 83
WP_124327587.1* -----MNTTYYNTTIDAL--LEWGVVYFQK--E-DFSEFLDNLEIAYISDA-----GDSLKDELSGVEKLVLGKISAAVIFGCEAVIGTTP 76
QOY56907.1 -----MSDDISYNTIGTDIL--LEWGVVYFQK--E-APFEFLSDIEEALS--GESLDTHLKQGTARIAAAEAVSARDLIGASDITGTT 77
MBF0120740.1 -----MNTNDKNTIIDIV--IQWGIKFMQK--E-KFPNVLSSLEDLDFSSI--DNLESEIKSGTSFTIQQSHDATVILSQPEILGVT 77
WP_045212335.1 -----MIPNELHISTINFL--EWGRLSFRD--HIQQLLEHLHGRVDFSY--EKLGLERHLDLILIKKALALDFNFGFOLLOWT 76
WP_207261537.1 -----MTAFNLNMHKAETALQWVAFPAKSL--LEWGHKTHS--P-QITSYLSLEIEYCKPFI--HDMRLDFLHSIEELKQVDTAVSILTSLELVST 80
WP_071544145.1 -----MTASLNMHKVEALMKHAFSALRRF--LEFGKGLIN--V-----PGDEPLPEAADDFEPLAGMLAQAGDAFAGDITGLIAS 75
WP_027366343.1 -----MSMSYEPQPDQPPRYSMATDIL--IVLGRMGLIG--Q-GSAPEREDLRHGLDAAL--AADPADLARGGEALGRAVEALASLCGDESEFLQD 87
WP_014258638.1 -----MSMSYEPQPDQPPRYSTADLL--IVLGRMGLIG--Q-GSAPEREDLRHGLDAAL--AADPADLARGGEALGRAVEALASLCGDESEFLQD 87
WP_124329907.1 -----MSTENEFHRTENTLFSCELDIL--VQWGLRLIM--Q-DNDAFLQELGQWYAL--NAVSPQDPGLDFEGKTEAALALHGRITLVGMDM 84
OEU46499.1 -----MEEVLYSPLDITL--LEWGHKTHS--P-DHKEYLSLLEEGVGAS--SPADIGNELPIGVNTICAEASGEMTSLKPKDLILATP 77
QOY56028.1 -----MMASDTKEFHSQHLELL--MKWGEHHLQ--P-GKNGLLQDLAKTWKTL--DPSFSDSLSDLEILGKVKVMALEQLGN--WLTSTI 77

MCC7212117.1 EEGIHWELEBGTAFMWNVYMLSHYLIAPKK--EASKFIKIDLETMSNIIISWIE-DGDFSPHETLVINNVYRHQLSQIPVWQYILFVYIYSCYN 180
KHE91660.1 YEEHLHKGAEFCETKRNIVLMSSHYLICIRDK--RYNEILQELDNVFSRITLWIE-BGDFSPHETLVINNVYRHQLSQIPVWQYILFVYIYSCYN 180
NUN21992.1 DTFPLHNDSESLITIRVNIVLMHAFHIAPIYRD--AVYKAFSIHKLDEIFSFITSWIE-BGAVSPHETLVINNVYRHQLSQIPVWQYILFVYIYSCYN 180
MBC6927463.1 EDLSLHNDSEVSFIAPRVNIVLMHAFHIAPIYRD--AVYKAFSIHKLDEIFSFITSWIE-BGAVSPHETLVINNVYRHQLSQIPVWQYILFVYIYSCYN 180
KAA0249750.1 EDLSLHNDSEVSFIAPRVNIVLMHAFHIAPIYRD--AVYKAFSIHKLDEIFSFITSWIE-BGAVSPHETLVINNVYRHQLSQIPVWQYILFVYIYSCYN 180
WP_207678646.1 YDMDIAQCTALEIILDNVLMSSHYLICIRDK--RYNEILQELDNVFSRITLWIE-BGDFSPHETLVINNVYRHQLSQIPVWQYILFVYIYSCYN 180
MCF8114645.1 QRMDSEHGLPLQVLDQAVLSLHSGQGLCPFL--DRLAQKIYDIDKDSIVL--NDPVPYHRIPIINWWRBELRDELFDIAPVILFVYIYSCYN 180
WP_124327587.1* ENEMAYCESEFSLITICAVWLSADRVVRKQDA--SLADSLIAKLDEAINRVAKLY-ADNLSPHETLVINNVYRHQLSQIPVWQYILFVYIYSCYN 180
QOY56907.1 DDEYVYKEDAFKIKINLVLYVYHDTLLKQNE--NLRIDSVLEDDTLKINEFTY-ESGFSALDNLNLSHRIEIGISGYSYLVFVYIYSCYN 180
MBF0120740.1 DDEYVYKEDAFKIKINLVLYVYHDTLLKQNE--NLRIDSVLEDDTLKINEFTY-ESGFSALDNLNLSHRIEIGISGYSYLVFVYIYSCYN 180
WP_045212335.1 DSDSKASLEAFDRIDNLVFFMAERLLAPKDE--KMAELLQKTTCYQVTSFF--SHDFSALBFTFHYKRALEVIRPQYILFVYIYSCYN 180
WP_207261537.1 GDABI--HPIPLAANLGNVYACANRDLTAPADH-----AQDIAMEDVFEISDDIEDLFFGRIIPCNLEBELLGQIPQENHFLFTVTECVQD 163
WP_071544145.1 EDSEI--HPIPLAANLGNVYACANRDLTAPADH-----ARDAAMKDVCLSDDDIGLFFGRIIPCNLEBELLGQIPQENHFLFTVTECVQD 163
WP_027366343.1 GDEYVYKEDAFKIKINLVLYVYHDTLLKQNE--NLRIDSVLEDDTLKINEFTY-ESGFSALDNLNLSHRIEIGISGYSYLVFVYIYSCYN 180
GDEYVYKEDAFKIKINLVLYVYHDTLLKQNE--NLRIDSVLEDDTLKINEFTY-ESGFSALDNLNLSHRIEIGISGYSYLVFVYIYSCYN 180
WP_124329907.1 DTEADQMTAMGTIDQVILFVYACDA--FPDAG--QFPKAFQEKVRELLKQVIT--DRPFGRIIPCNLEBELLGQIPQENHFLFTVTECVQD 163
OEU46499.1 DSDIWFEDGGLFALFVFLFASL--EYSTWMSHESVDITETVYKQVDEALQVSDWLFAGEBFSALDNLNLSHRIEIGISGYSYLVFVYIYSCYN 180
QOY56028.1 HRDEPEFQSALRILGSIIFLVSEEC--FPGSA--DLDFPHGKAENILHKKIT--DQETPLGRIIPCNLEBELLGQIPQENHFLFTVTECVQD 163

MCC7212117.1 KDILVEHIDENFTDVLFTG--KEHISALLEPIAEH--SFDGRIKCLFSIMKHEY--ILHKSILMKAIKRSALTNRWLGVSNSLRDIMPQIEKADV 272
KHE91660.1 DDTVGIHIDENFTDVLFTG--KEHISALLEPIAEH--SFDGRIKCLFSIMKHEY--ILHKSILMKAIKRSALTNRWLGVSNSLRDIMPQIEKADV 272
NUN21992.1 ENILETHILYFDDFLSKW--KKLAKKIFVEHLVEH--MHDKRIKCLFSIMKHEY--ILHKSILMKAIKRSALTNRWLGVSNSLRDIMPQIEKADV 272
MBC6927463.1 ENILETHILYFDDFLSKW--KKLAKKIFVEHLVEH--MHDKRIKCLFSIMKHEY--ILHKSILMKAIKRSALTNRWLGVSNSLRDIMPQIEKADV 272
KAA0249750.1 ENILETHILYFDDFLSKW--KKLAKKIFVEHLVEH--MHDKRIKCLFSIMKHEY--ILHKSILMKAIKRSALTNRWLGVSNSLRDIMPQIEKADV 272
WP_207678646.1 PETLEIIGHKWDIENKQLEMIETDSVLDGCTLYV--LTAIKTHVLLNHSHQES--RTAEVLSQVKKDFKQLQLLKMSDAKYFVSDEVEDK 269
MCF8114645.1 ENILETHILYFDDFLSKW--KKLAKKIFVEHLVEH--MHDKRIKCLFSIMKHEY--ILHKSILMKAIKRSALTNRWLGVSNSLRDIMPQIEKADV 272
WP_124327587.1* ENILETHILYFDDFLSKW--KKLAKKIFVEHLVEH--MHDKRIKCLFSIMKHEY--ILHKSILMKAIKRSALTNRWLGVSNSLRDIMPQIEKADV 272
QOY56907.1 SGILNIDAVHYHEFAKK--RNDINKLPEIKKINLYLSSESSQILKKYMEHN--TIFLSINQALTKHWAFLWEASKSGEVKRIEIRGI 267
MBF0120740.1 SGILNIDAVHYHEFAKK--RNDINKLPEIKKINLYLSSESSQILKKYMEHN--TIFLSINQALTKHWAFLWEASKSGEVKRIEIRGI 267
WP_045212335.1 SDTFLDAENING--PDNLPRIKDEPLFYAPDQ--QVLLKNVYKQES--LIARKFVKLLDTHNSLRWLFVEINLKLIPMVEVKA 259
WP_207261537.1 IDSLERHIDTWDLEW--DKTAPGDDABALVY--LTAIKTHVLLNHSHQES--RTAEVLSQVKKDFKQLQLLKMSDAKYFVSDEVEDK 269
WP_071544145.1 IDSLERHIDTWDLEW--DKTAPGDDABALVY--LTAIKTHVLLNHSHQES--RTAEVLSQVKKDFKQLQLLKMSDAKYFVSDEVEDK 269
WP_027366343.1 EDALERHAAPESARAH--VGADELDADLMAH--LQISAPPLAREPMR--RIMAGVYVAFAGHSLGMLRLAREANERFVTSVLENG 264
WP_014258638.1 EDALERHAAPESARAH--VGADELDADLMAH--LQISAPPLAREPMR--RIMAGVYVAFAGHSLGMLRLAREANERFVTSVLENG 264
WP_124329907.1 PDALDAHAHWDALRG--QANGMGIPOPTLGR--MADNPGAPPLARWOKA--RQRQHSKKAAGRLPGRLLALCESRSQVFPNFGGEAD 266
OEU46499.1 EFLDLEHIDENFTLQ--ETAPETLPEGIHLVELYMAERH--KILARVDEEH--RLSRGQAVVSILALRLFTVSDLSMKYITFVPEVAGL 268
QOY56028.1 ENFLDLAHWDIEDLGR--EASALGLNIATLNAH--LTAIKTHVLLNHSHQES--RTAEVLSQVKKDFKQLQLLKMSDAKYFVSDEVEDK 269

MCC7212117.1 VRTVNNLIKGVLT-----SDVTNDRKAYWLFSAFCGPGGLDDKQRNLNLEKVEIKV-KIDTSDIV-GDMANVLK--KHW-YDSNC--EDSQIVSIS 359
KHE91660.1 VRTVNNLIKGVLT-----SDVTNDRKAYWLFSAFCGPGGLDDKQRNLNLEKVEIKV-KIDTSDIV-GDMANVLK--KHW-YDSNC--EDSQIVSIS 359
NUN21992.1 IIVCVSLILDAIL-----EIFSEAEKIYWTFLPALCGPSLNGKQLEFLDTVENMIS-KIDIMIEA-RDKGELKT--KWL-FEGKI--EDSYLIKIS 356
MBC6927463.1 VVVSFKLIEDA-----KIAFEADEICWKFLLAFCGPNLDDKQRLSLLKQVEERIK-KIDIHIS--ENENILKT--KSW-FEGNC--DNALIKIS 353
KAA0249750.1 VVVSFKLIEDA-----KIAFEADEICWKFLLAFCGPNLDDKQRLSLLKQVEERIK-KIDIHIS--ENENILKT--KSW-FEGNC--DNALIKIS 353
WP_207678646.1 VYASCKNIDTRDPLDPREVLNPLRSKEDRTERLFLAFCGPGGLTDEKRIDLPGRVENLID-RDIFSCKI--SDSPLKIDQWKQGDG--QKLLSIDIV 363
MCF8114645.1 QAVAVQAIYSK-----PMFKEYHRIELFLAFCGPGPNLDDKQRLSLLKQVEERIK-KIDIHIS--ENENILKT--KSW-FEGNC--DNALIKIS 353
WP_124327587.1* HQVAVRAVFSQ-----TNLSPAERLAVAIAGACTPEISEDRRLLEILDCERVC-EIEAPTGD-DTSVRVIK--KAL-ADHRV--RHEIPAESL 346
QOY56907.1 ERISIRGILAQ-----KPQIQDQGLVAVTACAPAEISEDRRLLEILDCERVC-EIEAPTGD-DTSVRVIK--KAL-ADHRV--RHEIPAESL 346
MBF0120740.1 ENISIRAVLSD-----IPETNSERVIATAGSCFAPDISBDRVKTLLKIELRIK-VLSLEGND-DLISNGIK--TNF-ASMKI--SANEARAL 351
WP_045212335.1 VELSGLLQGS-----STMTQORFALGLSALPADDIABERKILQECNIFD-RKETSPPNPAQRVQV--QVQL--SLEBAVQAT 344
WP_207261537.1 MDHLVSWFRM-----PAVSGAVALVLV-----VGIVPGTQV-----GK--DGQRMPIVMT----- 287
WP_071544145.1 MDHLVSWFRM-----PAVSGAVALVLV-----VGIVPGTQV-----GK--DGQRMPIVMT----- 287
WP_027366343.1 AAAALGANTR-----PPASEGERHEHRLFLYALCGPFLAARLRLDLEFQVAFILR-DPDAAEQR--PGSPLAVYAAW-MRGEV--ADQQLSATA 346
WP_014258638.1 AAAAMHANTR-----PPASEGERHEHRLFLYALCGPFLAARLRLDLEFQVAFILR-DPDAAEQR--PGSPLAVYAAW-MRGEV--ADQQLSATA 346
WP_124329907.1 LALACLTLDV-----PVRSETGRAELFLAFCAPFPDARYISLFSETERFLK-RIPPSALP--SGSLMESVYRW-AHDLR--SDRELIRKA 348
OEU46499.1 SATAFHVISNA-----PRTPADQALVLQTFGCGPQLTDEQREAFSWEHEHDAIDYTGIL--MTDSMLGRATW-RAGRTGTEQTFADYV 355
QOY56028.1 HAACITLTA-----PIFSEAEIRIGHLFTAFCSPPFDDQRDLRLSIVEDFLK-QPDCAHD-----GIGR--QW-HEGRM--SDEKISRVA 338

MCC7212117.1 FDTLTRMEENARMQTREYDE--DPKELNNTLNT-----LLKKPIPSIPFEDSW-----IERISEWINRVAVQ-APALAW 428
KHE91660.1 FETWLGKMRK-----ANEI-DTESFNAHLET-----LLRGVPGLTKFLEEDV-----AKEPI-LEGSCTIDL----- 407
NUN21992.1 FDTWKNLIEEKASRMPTREYDA--APDELKNAINN-----LVKKEPIQVSEIRIIRFESLET-----MREKR-KSKHQSEFYNIPLHKV 432
MBC6927463.1 FDTWKNMEEKAINMHAVEFDE--SPEKLWNAIME-----LQKGRMGASNSVEKYINDFVNGCR-----QVWQI-FKSTVDEPSVLYGARA 431
KAA0249750.1 FDTWKNMEEKAINMHAVEFDE--SPEKLWNAIME-----LQKGRMGASNSVEKYINDFVNGCR-----QVWQI-FKSTVDEPSVLYGARA 431
WP_207678646.1 LESWLDNLTVQA--KRVAVFPVN-ENISFLDTVNR-----IMTDAIERTVTLTGTE-----WINFNCASELYTEA-- 427
MCF8114645.1 LKLMESDLKAA--DQIRSFQF--EDGRFNAVY-----VQMPLOTIVPHEKAFPNKIDIPQIFPIGWHDRPSE-VSMIYQTCFALPALS 432
WP_124327587.1* VLSMTEQIAGSTDPDKRMDELVLRLMSDMVIT--LSVDKRAAQSTEDQVKKQKGIIP-----FPFVQ-IANDVEVYQAVGV-- 427
QOY56907.1 LDFWFGQME-----SLVSDERQ--DMAELVGLLV-----QEPGQDGGKTFPSAED-----ILKMI-IGAIICQRIRWELSF 434
MBF0120740.1 CDWFMTQLERAGTVNEEK--DMIMLIEDILE-----ERKEKDRKKTVPDTYNTRTF-----FLLAI-IINKYKSLRIIPTM-- 421
WP_045212335.1 LDYVTEQLNVVQAEARNRKAAMA-GLVRSIISVEDSDGVEKISRQRPASSVGVKDKPKSWLEKLKAW-----LGNPILVITAAVAFILVTSVQF 433
WP_207261537.1 -----EMLVYQDGLAV-----RGGEETKPD-----VRKPG----- 312
WP_071544145.1 -----EMLVYQDGLAV-----RGGEETKPD-----VRKPG----- 312
WP_027366343.1 FAVWDLARLARA--QAGQGAEV-TARQPDHVR-----EAAQPIRKSEAGEPTSPFIARLQTLATL-----LPTPA-LRAATAGLIVAVLAGS 427
WP_014258638.1 FAVWDLARLARA--QAGQGAEV-TARQPDHVR-----EAAQPIRKSEAGEPTSPFIARLQTLATL-----LPTPA-LRAATAGLIVAVLAGS 427
WP_124329907.1 FEWEGRLGRN--EYTA-PEILTRAVEK--VRNAPVGHILRSPOGVANVAVSAYV--YIAT--MLVAVAAACILALALAF 423
OEU46499.1 FSDWLEKLSAAPINTGSPHE--TINSFLEGMMMAECERKLMSVAFPRGSYGVIEPDEIFDRLVGLSQNPIKRVGFFRSE-RFQPLRLQPAVALGA 452
QOY56028.1 FRQWGNLALAR--KRSPRAITY-SASSFWDAIET-----LRAAGISARKENGMSQIQKISLSELA-----FSPSK-IAATATACLVIFIII 419

MCC7212117.1 LSEDEERTTE
KHE91660.1 --DDQCASTQ
NUN21992.1 AANAGKEEAG
MBC6927463.1 IAAASEKSVQ
KAA0249750.1 IAAASEKSVQ
WP_207678646.1 --AKSAGPLS
MCF8114645.1 YGDHEESKK
WP_124327587.1* --KDKKAAN
QOY56907.1 PMGETDDEEK
MBF0120740.1 --SDEEII
WP_045212335.1 LLRQEDQMHI
WP_207261537.1 -----
WP_071544145.1 ALTMHLSSTP
WP_027366343.1 VLLTMHLSAP
WP_014258638.1 FTIYFPAPSV
WP_124329907.1 LQTFEQDRME
OEU46499.1 YTRQDYEQQLQ
QOY56028.1

Figure S13

DUF4384 (pfam14326) family

WP_045212235.1 NFEMRVYQVLFILTRS---DEAEVTVAESGAVLSKRCFQILFTETRLWYVYLHQDAGGQTTKYLEGTHHKDSTCFEDKRSRSCFTADSGSEVFVILVY 97
 WP_014258638.1 TAPIGTLRLLAAGEEGQFRGGAPRIIEQSGEVMLSGRRQVRFTLPPBAYVFFRDAQGEVSTLQGRVEAG-PRVLFSPGESYVDEEGSSRVVILV 99
 WP_027366343.1 TAFMGTLRLLAAGEAGQFRGGAPRIIAQSGETMLSGRRQVRFTLPPBAYVFFRDAQGEVSTLQGRVEAG-PRVLFSPGESYVDEEGSSRVVILV 99
 WP_207261537.1 -----DIPPEGGAIQIRFTLDBAYITLYHQSDDKRELLDPSKLEKCDHVFDPKSKFMRHGRKFGEDBYIIM- 69
 WP_071544145.1 -----DAPLEGGAIQIRFTLDBAYITLYHGRKPHFEDPSRGLKGC-PRVLFSPGESYVDEEGSSRVVILV 69

WP_045212235.1 FDNVAVDGFARITQSMNTDILQDPVFEQEIIVTFPMACQEIFSISYK 145
 WP_014258638.1 LAKEQKLEDEAL----HGEV---EAIERHRCVSUREIASHK 139
 WP_027366343.1 LAKEQKLEDEAL----HGEV---EAIERHRCVSUREIASHK 139
 WP_207261537.1 ASERALNKIDQTNMNVNKTNO-----PGTHIELEL--- 100
 WP_071544145.1 ASREPLSPADIEKPLADEALGGA-----AKATTHIFCS--- 102

Family 1

KHE91660.1 KEQFSGKNNWGLTILFIREESISMPDPDN----IRASSEDYKRLMKTIDKVEDN--YWGTCFADKKNLVFFKRTITYFLGIEGSKGVREAVIG 93
 NUN21992.1 TEEIDGKNNHYYSILFYDRGANEHPPPE--EVSGCTAVKQKKLWVLGNIKGSCVAGGICPKEDWQIK-EIPQPKKRIFETESKEGSRNAVIG 97
 KAA0249750.1 PRPEIDANNHILLSLKPNPKGVILLSSSLALKRVVLGEGVEDYKRIWNYLEHTKND--YWGCFITNDKRFQ-VASVQIENRILAKTR--LYKRAIIG 95
 MBC6927463.1 PRPEIDANNHILLSLKPNPKGVILLSSSLALKRVVLGEGVEDYKRIWNYLEHTKND--YWGCFITNDKRFQ-VASVQIENRILAKTR--LYKRAIIG 95
 MCC7212117.1 QLLIDGKNNHYYSILFYDRGANEHPPPE--EVSGCTAVKQKKLWVLGNIKGSCVAGGICPKEDWQIK-EIPQPKKRIFETESKEGSRNAVIG 94

KHE91660.1 ISNSEDDEIFVNRKRS-----VSITEKGQRTYFVDAAGSRIAEVVDYTKQKPLNVVLLIKYTYEE--- 155
 NUN21992.1 ISNSEDDEIFVNRKRS-----VSITEKGQRTYFVDAAGSRIAEVVDYTKQKPLNVVLLIKYTYEE--- 155
 KAA0249750.1 VSEKRVLEDETCDF-----PAVIFSGKGP-----LK-----DS--LVKKVLLVLSIG----- 137
 MBC6927463.1 VSEKRVLEDETCDF-----PAVIFSGKGP-----LK-----DS--LVKKVLLVLSIG----- 137
 MCC7212117.1 VSEKRVLEDETCDF-----PAVIFSGKGP-----LK-----DS--LVKKVLLVLSIG----- 137

Family 2 (csx30)

WP_124327587.1* DSKVKFPSLIEGCSGDGKRLLEDTD-----AAA-N---HRQLFSILKAGKLN---SAFFIQSDGEWVESEKPTMEDN--RILHPS-HHSSGV 84
 OQY56907.1 --TVTIGVNRGSEHQEADVAILIASVSVFPFMEQISHEILTALLERLNSG-RETFYQVLLKNFSGHSTFPDKPEEIVD--LEPDTIIVPDEGG 94
 MBF0120740.1 EETFEPTIRHGTCHIKRENSVILLNSIQNKSD-----FDFIQFNNKKDLKLYINVISMSKQKWIILKGTG-KTYKPSIEPCD---SD-IGF 88

WP_124327587.1* WILDIGSMILRQSVKCVKDALNK--MTGSAKTKKPKMIMWVVTLPQEG 130
 OQY56907.1 LLSDTKGMSEKSAEAMKMSLNKTKGPGDESKYSSSTHLLVLEI---- 138
 MBF0120740.1 LILNSKKDINKTKTEKLLALANKSEKLSDED-IDECTHISYYIK--- 132

Family 3 (single representative)

WP_207678646.1
 KESVASQDIKSLNLKKNIFEDDLYIEDDSLKIRDOSPLQQLDNIEYWNWGYEPLDGDMEINYSDAEEDGDTSPFEIGKDKKEYKTIVVVCIAATDEEIVNTVSKINSKEEPLPEEKNTIWLIIYTKK

Family 4 (single representative)

MCF8114645.1
 EIKFPTPNPMVLSPIPTDKLLNETGELNELRSNYDLSFGAIAINGSGEINITEPKFTGPVKRKKIDIDKDAIVLIIAVSGDKQLKAALEQFRNISSKSDPLFVIENVFVIIYYDKKE

Family 5

OQY56028.1 QVPHVPMKSEDAVEKAAVPESQSLRREVKKKDAMSLPGKAERMERFSRAEVS--EQTPPPALAPAPAVBKQKSLPSEKRIKRVV-NIKSTVSGGK 97
 WP_124329907.1 -----SEPKPFVKKIVFPATSPAPALH-----LPSEKRM-FKRRDFVSGF--DGT 45

OQY56028.1 GDDIDQAEIRIPPKADGLAPLPATEFTFSTQGFHEDSVARESLHFKTWLKDIDHLHFAILSSDAGSAYLEDAYENFYAVVGMANGNSGVHVEI 197
 WP_124329907.1 EDDIAY-----TTP--TGAUPLPFL--LH-----SDRTAEERIMAILLPIGMAKMKAPICRGVILGKLSAGFHMIRVHPI 119

OQY56028.1 LSRNIVEEKHETSSSEIMTSNRRLALKFSI 229
 WP_124329907.1 FRDRVMEERQIDGCSPPVPRVAVVLLAGEAS 151

Family 6 (single representative)

OEU46499.1
 EPLPELHGNLLVIDLMPDEGGGRVYLLLPQHEIKSDESDKLEQYLKCFKTFGKYWQAGIIFGANLPQTWKPDLSRTPEPITQYVPKQTAAIILGLCSDEELLEKAIALVCEGKECTEESILWILYRPA
 RKHAELE

Figure S14