Peer Review File

Manuscript Title: Antiviral signaling by a cyclic nucleotide activated CRISPR protease

Reviewer Comments & Author Rebuttals

Reviewer Reports on the Initial Version:

Referees' comments:

Referee #1:

Rouillon et al. characterize the mechanism of a recently-discovered CRISPR-Lon system using a combination of biochemical and structural approaches. Their data reveal that CRISPR-Lon is a protease that forms a complex with the small CRISPR-T protein from an adjacent gene within the locus. They demonstrate that CRISPR-Lon is activated by cA4, allowing it to proteolytically cleave CRISPR-T to release a small RNAse toxin. This is an impressive finding with important biological implications for both the CRISPR field and broader audience. It is a great candidate for Nature, provided the authors address one or more of my major concerns below. Overall, the structural and biochemical work has been expertly performed and the presentation is clear and concise. There are a few major gaps in the study that would really nail down a mechanism for this awesome system.

1. The authors determined a structure of CRISPR-Lon in the absence of any substrates, revealing the cOA pocket and protease active site. It remains unclear how cA4 activates the protease, since within the structure, the active site is already exposed, and the tentative substrate path has been outlined (Fig. 1). Why would it require cOA for activation? Is there a conformational change induced by cOA binding? While asking for more structures would be outside the scope of this study, the authors should at least include some experiments to provide a structural explanation. They point to other studies that have used SAXS, etc. Options such as these could be explored.

2. The failure to show convincing RNase activity of CRISPR-T is a very weak point of the manuscript. Any mechanistic insights into how cleavage of a flexible domain activates nuclease activity? Based on the AlphaFold2 (AF2) model and mutagenic data, the MazF-like and DUF2080-like domains are separated by a flexible linker and likely do not contact each other. How is the DUF2080 domain able to auto-inhibit the MazF-like domain? Based on Suppl. Fig. 10, cA4 does not appear to be necessary for activating the RNase activity of CRISPR-T, so perhaps the hypothesis is incorrect. There is also a contaminating nuclease that is making it even more difficult to interpret. The current data does not show that CRISPR-T has cytotoxic effects, thus undermining a major aspect of the presented hypothesis. Additional biochemistry is essential to support this important hypothesis.

3. Some CARF domain-containing proteins degrade cA4 resulting in an acute activation. Is this the case with CRISPR-Lon?

Minor comments:

1. CRISPR-T is in not introduced in the Summary.

2. The inclusion of the AF2 section not necessary. AF2 is amazing and most biologists (not just structural biologists) are aware of its merits and shortcomings. I would remove this section and focus on the incredibly interesting biology that this exciting manuscript contains instead!

David Taylor

Referee #2:

In this study, Rouillon et al. describe the structure and function of a CRISPR-associated Lon protease. CRISPR-Lon is a component of some type III CRISPR systems, which are complex CRISPR systems that elicit a multi-pronged defense. Part of this response is the generation of a secondary messenger, cOA. cOA is known to activate downstream effectors, through interaction with their CARD domains, that mediate the immune response.

Previous bioinformatics analysis had suggested that CRISPR-Lon was a membrane-associated protein that contained a SAVED domain. However, the authors show through a structure-function analysis (X-ray crystallography) that CRISPR-Lon is a soluble monomer and that its SAVED domain binds cA4. Through a genome analysis they also characterize a novel gene, which they call CRISPR-T, specifically showing that CRISPR-Lon forms a complex with CRISPR-T and that after activation by cA4 CRISPR-Lon cleaves CRISPR-T. Cleavage releases a 23-kD fragment (CRISPR-T23) that is predicted to have a structure very similar to the MazF toxin.

The novelty of this work is in the discovery of a cOA activated protease, the observation that the SAVED domain can bind cOA, and in the characterization of the CRISPR-Lon/CRISPR-T system. The work is expertly performed, the conclusions are justified given the data, and the manuscript is well written. However, the story feels preliminary in the context of a Nature publication, where I would expect a more complete picture of how this system functions. Providing genetic/cell-based data for the system or determining what the target of CRISPR-T23 is would go some way to fulfilling this.

Referee #3:

In response to phage infection, several types of defence system produce secondary signalling molecules (typically nucleoside based) that activate effector proteins in trans. The activated effector proteins then elicit the direct infection response (e.g. degrading phage RNA or inducing host dormancy). Examples of phage defences that use secondary signalling include CBASS (cyclic-oligonucleotide-based anti-phage signalling systems; Cohen 2019, Nature), Pycsar (Tal 2021, Cell; Ofir 2021, Nature), and the topic of this paper, CRISPR-Cas (Kazlauskiene 2017, Science; Niewoehner 2017, Nature). Recently, there has been a lot of interest in characterising the function of these effector proteins. To date, all effectors reported are directly activated by secondary signal (e.g. cOA) binding.

Here, the authors report the discovery of a signalling cascade involving the indirect post-

translational activation of a (hypothesised) effector protein. The study demonstrates that cOA signal produced by a type III CRISPR-Cas system activates a protease termed CRISPR-Lon. This is the first (to my knowledge) report demonstrating activity of a cOA-activated protease within the context of phage defence (despite earlier bioinformatic predictions of their occurrence). CRISPR-Lon contains a Lon-like protease domain and a cOA-sensing (named SAVED) domain. Until now, the function of CRISPR-Lon (formerly CARF-Lon) remained unknown — the key question being what is the target for the Lon domain and how does this lead to a phage infection response? First, the authors solved the structure of CRISPR-Lon and confirmed the presence of Lon and SAVED domains. The authors predicted the protease target to be a MazF-like protein encoded upstream of the CRISPR-Lon gene, which they named CRISPR-Toxin (CRISPR-T). Through a series of in vitro experiments with recombinant CRISPR-Lon and CRISPR-T, the authors demonstrate that CRISPR-Lon forms a complex with full-length CRISPR-T. Upon cOA binding by CRISPR-Lon, the CRISPR-T peptide is cleaved to release a ~23 kDa N-terminal fragment (CRISPR-T23) that contains a MazF-like RNase domain. It is hypothesised that CRISPR-T23 represents the activated effector protein for the system. However, nuclease activity of CRISPR-T23 was not detected, nor was any data presented to demonstrate CRISPR-T23 was essential for phage defence. No non-specific RNase activity was detected (as expected for MazF-like proteins); hence, the authors are left to speculate that the mechanism of action is via specific cleavage of rRNA, tRNA, or mRNA. Despite the lack of functional data for CRISPR-T23, the authors propose an interesting and plausible model for the overall activity and role of the CRISPR-Lon/CRISPR-T signalling cascade as distinct from other effector nucleases that are directly activated by the signalling molecule (typically cOAs).

The study itself appears robust and data are presented to support almost all conclusions, aside from the bona fide role of CRISPR-T23 as a toxin. The major outstanding gap is that a nuclease activity of the CRISPR-T toxin was not demonstrated. Additionally, there are no conclusive data that CRISPR-T is indeed a genuine toxin, rather than having some non-toxic (to the host) role in cleaving phage-specific RNAs, for example. MazF homologs have previously been shown to have specificity for triplets. Are all possible triplets included in the tested RNAs? Does activation of the system in vivo lead to cell death or dormancy (for example, testing expression of CRISPR-Lon and CRISPR-T in E. coli then addition of cOAs)? Or perhaps expression of CRISPR-T23 alone to demonstrate toxicity? However, determining the targets of toxins from TA systems is usually non-trivial (for many TA systems the targets remain unknown after many years), and the lack of these additional insights does not necessarily detract from the advances made by the paper. Studies to examine the CRISPR-T23 mechanism would be interesting to warrant publication in the future. Overall, the manuscript was interesting and generally well-presented (a few comments to improve communication of information are included below) and will surely be the seed for future studies (nicely covered in the Discussion) of this intriguing effector response.

Additional comments:

Substantially more detail should be added to the Methods to describe the mass spectrometry. For example, if a trypsin digest was performed, were any non-tryptic fragments identified that might correlate with the cleavage site?

For SEC-MALS, both the light scattering and a signal giving the protein concentration are required.

Typically, the latter is either measured by refractive index or UV absorption. Please indicate which was used. Also, were the mixed proteins preincubated (to allow complex formation) prior to loading on the SEC column?

X-ray crystallography: Presumably hanging/sitting-drop vapour diffusion experiments were performed for the screens. Please provide details to aid others in future endeavours. Temperature is also missing.

Addition of the sequences for the codon-optimised expression constructs would be useful (to enable replication or future studies of the system).

What method was used to determine protein concentrations?

L374: Is “phenix.autosolve” actually phenix.autosol?

L16: Cas9- -> Cas9

L30: CRISRP-T -> CRISPR-T

L57: “attempted to catalog” -> cataloged

L90: Please check whether “onnorineus” should be “onnurineus”

L114-115: Adding some context as to the roles of Cap4, Can1 and Cap5 in other systems might be helpful for readers (e.g. Cap4 being a SAVED domain effector nuclease found in CBASS systems). A generic statement grouping all three proteins mentioned into cOA-activated effector proteins from defence systems should suffice.

L258, “but no other CARF proteins such as the RNases Csm6 or Csx1”: I checked a couple of the genomes presented in Fig. S6 and there does appear to be several CARF domain proteins (likely effectors) encoded in these genomes (often clustered with the CRISPR-Cas loci). For example (shows a predicted CARF-PDDEXK protein encoded within the same cas locus as the CRISPR-Lon/T): https://padloc.otago.ac.nz/padloc/refseq/a97b2ce2802d9b9686bb6ea35a4cf169e2d4861978e64df5e1f56dc82b617f5b

L269-297 & 317: Is 1:50 correct, or should it be 50:1?

I had a couple of thoughts regarding the terminology of CRISPR-Lon and CRISPR-T and robustness of the nomenclature with future potential discoveries. Is CRISPR- the most appropriate prefix (rather than Cas- or SAVED-)? For CRISPR-T, have you looked for other potential families (non-MazF) of Lon-activated toxins, with potentially different mechanisms of toxicity? What would future analogous toxins be called? It is also possible that analogous CRISPR-Lon/T systems might be activated by other signal-producing defence systems, such as CBASS, so a more generic descriptor might be appropriate (e.g. SAVED-Lon). I cannot think of any obviously better nomenclature than what you currently have but it might be worth some thought.

Fig. 2 and Fig. S6: The gene arrangements 1/2 (coloured red in Fig. 2)/CRISPR-Lon appeared similar to the 1/8/CRISPR-Lon operons, so I checked whether 8 encodes a similar protein CRISPR-T and this appears to be the case (#8 from T. geofontis run through hhpred reveals an N-terminal MazF-like domain). Perhaps it's worth looking at a little closer.

Fig. 2 and Fig. S9: Additional labelling of the lanes would be informative. That is, lane 3 clearly doesn't contain CRISPR-Lon (due to the absence of the band), but this is not indicated in the labelling scheme.

Fig. S1: A) alone does not contain sufficient information to conclude CRISPR-Lon elutes as a monomer. Perhaps add an overlay showing the elution profile of protein standards? The light scattering data in Fig. 3 do support the monomer state, so an alternative would be to change the title of Fig. S1.

Fig. S2: B) is too small to read. C) Please define pLDDT (and provide information for readers not familiar with AlphaFold to interpret the meaning of this metric). D) appears to be a superposition of two structures, but these are not explained.

Fig. S7: Please define pLDDT. It seems a little odd to show MazF from the MazEF structure with the RNA from the MazF structure (but not showing the corresponding MazF). Perhaps separate figures showing 1) MazEF compared with CRISPR-T and 2) MazF/ssRNA with CRISPR-T?

Fig. S8: A) alone does not contain sufficient information to conclude CRISPR-T elutes as a monomer (Fig. 3 does contain supporting data). Perhaps add an overlay showing the elution profile of protein standards? The earlier peak also contains a large amount of CRISPR-T (based on (B)). Is this a multimer or is the peak at the void volume (aggregate)? A comment on this would be useful (protein standards for the SEC will also help resolve this).

Fig. S9: Indication of the A194 residue would be informative (especially since the numbering is based on the recombinant protein with His-TEV tag).

Author Rebuttals to Initial Comments:

We thank the referees for their overall very enthusiastic feedback. Following extensive further studies and a significant breakthrough in the elucidation of the system, we can now answer the two key questions that were posed by all three referees:

1) How is CalpL (formerly CRISPR-Lon) activated by cA4 binding?
2) What is the function of the $CalpT_{23}$ protein after cleavage by CalpL?

You'll find that based on the suggestion of Referee #3, we have renamed the proteins as follows:
CRISPR-Lon: CalpL (Calp stands for CRISPR associated Lon protease, L for Lon protease)
CRISPR-T: CalpT (T for target)
The newly discovered σ-factor in the operon is called CalpS (S for sigma)

With two new crystal structures, SAXS- and biochemical data, we are able to provide a convincing answer to question 1), as described below and in the revised manuscript.

Our substantial efforts to answer question 2) took a completely unexpected turn. We had followed several approaches to test for a nuclease activity, for example by Illumina sequencing of random RNA libraries but could not detect any nuclease activity of the toxin CalpT. We then thought that the third conserved gene in the operon, which has similarities to ECF-σ factors might be of importance. We cloned and purified this protein and found that CalpT and CalpS form a heterodimer, and together with CalpL form a heterotrimeric complex. Furthermore, we show that the sigma factor CalpS on its own copurified with *E. coli* RNA polymerase. This finding directly links the detection of CRISPR induced cA_4 to cellular transcription and we propose that CalpT/S act like an anti-σ-factor/sigma factor pair, with CalpT being under control of CalpL.

Referees' comments:
Referee #1:

Rouillon et al. characterize the mechanism of a recently-discovered CRISPR-Lon system using a combination of biochemical and structural approaches. Their data reveal that CRISPR-Lon is a protease that forms a complex with the small CRISPR-T protein from an adjacent gene within the locus. They demonstrate that CRISPR-Lon is activated by cA4, allowing it to proteolytically cleave CRISPR-T to release a small RNAse toxin. This is an impressive finding with important biological implications for both the CRISPR field and broader audience. It is a great candidate for Nature, provided the authors address one or more of my major concerns below. Overall, the structural and biochemical work has been expertly performed and the presentation is clear and concise. There are a few major gaps in the study that would really nail down a mechanism for this awesome system.
Thank you very much for your positive evaluation of our manuscript.

1. The authors determined a structure of CRISPR-Lon in the absence of any substrates, revealing the cOA pocket and protease active site. It remains unclear how cA4 activates the protease, since within the structure, the active site is already exposed, and the tentative substrate path has been outlined (Fig. 1). Why would it require cOA for activation? Is there a conformational change induced by cOA binding? While asking for more structures would be

outside the scope of this study, the authors should at least include some experiments to provide a structural explanation. They point to other studies that have used SAXS, etc. Options such as these could be explored.
We have conducted a large number of experiments to get more insight into this important question:
1) We have determined the crystal structure of the CalpL/CalpT$_{10}$ complex. Indeed, the CalpT binds to the N-terminal domain of CalpL and we can see electron density almost up to the protease cleavage site – the last residue is disordered. Our SPR results show that the CalpL/T complex is very strong, and hence, we believe, that the two proteins will always be tied together inside the bacterium. The crystal structure shows that the distance between the cleavage site and the protease active site is too large for the reaction to take place – hence, an activation mechanism is needed (see below)

2) We have determined the crystal structure of the CalpL/cA$_4$ complex. It reveals the binding mode of cA$_4$ and that no major conformational changes are induced when the activator binds to the SAVED domain.

4) We have conducted SAXS and DLS experiments on the CalpL protein in the presence and absence of cA$_4$. Strikingly and matching to observation to SAVED proteins from the CBASS system, the addition of cA$_4$ induces oligomerization of CRISPR-Lon. The SAXS- and DLS data and the surface charge distribution of CalpL are consistent with a model where the protein forms stacks, with a cA$_4$ molecule sandwiched between two CalpL molecules. In line with this, we found that mutations on the backside of the SAVED domain and thus far away from the protease active site impair the cleavage activity of CalpL.
We thus propose that the stack formation leads to a conformational change of the CRISPR-Lon-T complex that induces protease cleavage.

5) By conducting protease assays with mixtures of preformed CalpL/-T complexes (protease dead and cleavage site mutans) we could show that the cleavage reaction proceeds in trans.

2. The failure to show convincing RNase activity of CRISPR-T is a very weak point of the manuscript.

Any mechanistic insights into how cleavage of a flexible domain activates nuclease activity? Based on the AlphaFold2 (AF2) model and mutagenic data, the MazF-like and DUF2080-like domains are separated by a flexible linker and likely do not contact each other. How is the DUF2080 domain able to auto-inhibit the MazF-like domain?
Based on Suppl. Fig. 10, cA4 does not appear to be necessary for activating the RNase activity of CRISPR-T, so perhaps the hypothesis is incorrect. There is also a contaminating nuclease that is making it even more difficult to interpret. The current data does not show that CRISPR-T has cytotoxic effects, thus undermining a major aspect of the presented hypothesis. Additional biochemistry is essential to support this important hypothesis.

To provide a conclusive check for activity on single stranded RNAs, we have constructed random RNA libraries and look for depletion of target sequencing via Illumina deep sequencing. As a control, we incubated our library with MazF and found that sequences containing the ACA target sequence were indeed depleted. No such activity was found for

CalpT. Also, various truncated constructs that simulated the cleaved variant of CalpT did not show any effect on growth of *E. coli* cells. As described in our revised manuscript, we conclude that CalpT does in fact not have a nuclease activity but rather acts as an anti-σ-factor to CalpS.

3. Some CARF domain-containing proteins degrade cA4 resulting in an acute activation. Is this the case with CRISPR-Lon?
We added the following sentence to the manuscript in the Discussion section:
„Some CARF domain proteins are known auto deactivate by degrading cOA species REF (PMID: 33986148), but this has not been observed for SAVED domains and has not yet been investigated for CalpL.“

Minor comments:

1. CRISPR-T is in not introduced in the Summary.
We are now mentioning all proteins of the signaling cascade in the summary.

2. The inclusion of the AF2 section not necessary. AF2 is amazing and most biologists (not just structural biologists) are aware of its merits and shortcomings. I would remove this section and focus on the incredibly interesting biology that this exciting manuscript contains instead!
True! We removed the AF2 section.

Referee #2:

In this study, Rouillon et al. describe the structure and function of a CRISPR-associated Lon protease. CRISPR-Lon is a component of some type III CRISPR systems, which are complex CRISPR systems that elicit a multi-pronged defense. Part of this response is the generation of a secondary messenger, cOA. cOA is known to activate downstream effectors, through interaction with their CARD domains, that mediate the immune response.

Previous bioinformatics analysis had suggested that CRISPR-Lon was a membrane-associated protein that contained a SAVED domain. However, the authors show through a structure-function analysis (X-ray crystallography) that CRISPR-Lon is a soluble monomer and that its SAVED domain binds cA4. Through a genome analysis they also characterize a novel gene, which they call CRISPR-T, specifically showing that CRISPR-Lon forms a complex with CRISPR-T and that after activation by cA4 CRISPR-Lon cleaves CRISPR-T. Cleavage releases a 23-kD fragment (CRISPR-T23) that is predicted to have a structure very similar to the MazF toxin.

The novelty of this work is in the discovery of a cOA activated protease, the observation that the SAVED domain can bind cOA, and in the characterization of the CRISPR-Lon/CRISPR-T system. The work is expertly performed, the conclusions are justified given the data, and the manuscript is well written. However, the story feels preliminary in the context of a Nature publication, where I would expect a more complete picture of how this system functions.

Providing genetic/cell-based data for the system or determining what the target of CRISPR-T23 is would go some way to fulfilling this.

We thank the referee for his overall very positive evaluation of our work. We hope that the new insights provided in the revised version will convince the referee that the story is suitable for a publication in Nature.

Referee #3:

In response to phage infection, several types of defence system produce secondary signalling molecules (typically nucleoside based) that activate effector proteins in trans. The activated effector proteins then elicit the direct infection response (e.g. degrading phage RNA or inducing host dormancy). Examples of phage defences that use secondary signalling include CBASS (cyclic-oligonucleotide-based anti-phage signalling systems; Cohen 2019, Nature), Pycsar (Tal 2021, Cell; Ofir 2021, Nature), and the topic of this paper, CRISPR-Cas (Kazlauskiene 2017, Science; Niewoehner 2017, Nature). Recently, there has been a lot of interest in characterising the function of these effector proteins. To date, all effectors reported are directly activated by secondary signal (e.g. cOA) binding.

Here, the authors report the discovery of a signalling cascade involving the indirect post-translational activation of a (hypothesised) effector protein. The study demonstrates that cOA signal produced by a type III CRISPR-Cas system activates a protease termed CRISPR-Lon. This is the first (to my knowledge) report demonstrating activity of a cOA-activated protease within the context of phage defence (despite earlier bioinformatic predictions of their occurrence). CRISPR-Lon contains a Lon-like protease domain and a cOA-sensing (named SAVED) domain. Until now, the function of CRISPR-Lon (formerly CARF-Lon) remained unknown — the key question being what is the target for the Lon domain and how does this lead to a phage infection response? First, the authors solved the structure of CRISPR-Lon and confirmed the presence of Lon and SAVED domains. The authors predicted the protease target to be a MazF-like protein encoded upstream of the CRISPR-Lon gene, which they named CRISPR-Toxin
(CRISPR-T). Through a series of in vitro experiments with recombinant CRISPR-Lon and CRISPR-T, the authors demonstrate that CRISPR-Lon forms a complex with full-length CRISPR-T. Upon cOA binding by CRISPR-Lon, the CRISPR-T peptide is cleaved to release a ~23 kDa N-terminal fragment (CRISPR-T23) that contains a MazF-like RNase domain. It is hypothesised that CRISPR-T23 represents the activated effector protein for the system.
Thank you for your positive evaluation of our work.

However, nuclease activity of CRISPR-T23 was not detected, nor was any data presented to demonstrate CRISPR-T23 was essential for phage defence. No non-specific RNase activity was detected (as expected for MazF-like proteins); hence, the authors are left to speculate that the mechanism of action is via specific cleavage of rRNA, tRNA, or mRNA.
Despite the lack of functional data for CRISPR-T23, the authors propose an interesting and plausible model for the overall activity and role of the CRISPR-Lon/CRISPR-T signalling cascade as distinct from other effector
nucleases that are directly activated by the signalling molecule (typically cOAs).
The study itself appears robust and data are presented to support almost all conclusions, aside from the bona fide role of CRISPR-T23 as a toxin. The major outstanding gap is that a nuclease activity of the CRISPR-T toxin was not demonstrated.

Additionally, there are no conclusive data that CRISPR-T is indeed a genuine toxin, rather than having some non-toxic (to the host) role in cleaving phage-specific RNAs, for example.

The referee is indeed right. Despite extensive trials we could not detect a nuclease activity for the "toxin" CalpT. Instead, we found that CalpT literally links the cA_4 sensor CalpL to the transcription machinery of the cell and has stunning functional parallels to anti-σ-factors.

MazF homologs have previously been shown to have specificity for triplets. Are all possible triplets included in the tested RNAs? Does activation of the system in vivo lead to cell death or dormancy (for example, testing expression of CRISPR-Lon and CRISPR-T in E. coli then addition of cOAs)?
The referee raises a very important point. We have constructed a random RNA library with 10 consecutive bases. The library was tested with MazF and we could indeed find depletion of the ACA target sequence from our library by Illumina deep sequencing. We did however not find any activity of CalpT (formerly CRISPR-T). The data are shown in Extended Data Figure 8CD.

Or perhaps expression of CRISPR-T23 alone to demonstrate toxicity?
We tried to express different truncated constructs of CRISPR-T (1-130, 1-172, 1-194), all of which were tolerated by the cell. Our new data showing the tight interaction with CalpS sheds new light on its function.

However, determining the targets of toxins from TA systems is usually non-trivial (for many TA systems the targets remain unknown after many years), and the lack of these additional insights does not necessarily detract from the advances made by the paper.
Indeed, and thank you!

Studies to examine the CRISPR-T23 mechanism would be interesting to warrant publication in the future. Overall, the manuscript was interesting and generally well-presented (a few comments to improve communication of information are included below) and will surely be the seed for future studies (nicely covered in the Discussion) of this intriguing effector response.

Additional comments:

Substantially more detail should be added to the Methods to describe the mass spectrometry. For example, if a trypsin digest was performed, were any non-tryptic fragments identified that might correlate with the cleavage site?
We added more details to the corresponding section of Methods. We did not observe any non-tryptic fragments that correlated with the cleavage site. This is now mentioned in the manuscript.

For SEC-MALS, both the light scattering and a signal giving the protein concentration are required. Typically, the latter is either measured by refractive index or UV absorption. Please indicate which was used. Also, were the mixed proteins preincubated (to allow complex formation) prior to loading on the SEC column?
Refractive index was used. The proteins were incubated for 60 min before injection. This information has been added to the methods section.

X-ray crystallography: Presumably hanging/sitting-drop vapour diffusion experiments were

performed for the screens. Please provide details to aid others in future endeavours. Temperature is also missing.
Done.
"Pure CalpL protein was concentrated to 20 mg/ml and crystallized at 20 °C using a Gryphon pipetting robot (Art Robbins) and commercial crystallization screens (Molecular Dimensions) using sitting drop plates. Hexagonal crystals appeared after one day in condition D7 of the JCSG+ screen. Several rounds of optimization in sitting- and hanging drop plates were performed to achieve well-diffracting crystals. The final crystallization condition was 0.1 M Tris-Cl pH 8.0, 38.8 % PEG 400, 0.29 M Li_2SO_4. The SeMet derivative (see above) was crystallized under similar conditions and yielded identical crystals. The crystals were harvested without further cryo-protection and a diffraction dataset was recorded at beamline P13 (λ = 0.9795) operated by EMBL Hamburg at the PETRA III storage ring (DESY, Hamburg, Germany) [51]. The diffraction data were automatically processed with XDS [52]. The structure was solved using phenix.autosol and refined with phenix.refine [53]. Further model building was performed in Coot [54] and figures were prepared with PyMOL (www.pymol.org). The geometry of the model was checked with MolProbity [55]."

Addition of the sequences for the codon-optimised expression constructs would be useful (to enable replication or future studies of the system).
We will upload the sequences (.gb files) as supplementary data.

What method was used to determine protein concentrations?
Throughout the study, we used UV280 measurements to determine protein concentrations.

L374: Is "phenix.autosolve" actually phenix.autosol?
Yes. We changed it.

L16: Cas9- -> Cas9
Done.

L30: CRISRP-T -> CRISPR-T
Thanks

L57: "attempted to catalog" -> cataloged
Changed.

L90: Please check whether "onnorineus" should be "onnurineus"
Thanks for catching this. We changed it.

L114-115: Adding some context as to the roles of Cap4, Can1 and Cap5 in other systems might be helpful for readers (e.g. Cap4 being a SAVED domain effector nuclease found in CBASS systems). A generic statement grouping all three proteins mentioned into cOA-activated effector proteins from defence systems should suffice.
Done.

L258, "but no other CARF proteins such as the RNases Csm6 or Csx1": I checked a couple of the genomes presented in Fig. S6 and there does appear to be several CARF domain

proteins (likely effectors) encoded in these genomes (often clustered with the CRISPR-Cas loci). For example (shows a predicted CARF-PDDEXK protein encoded within the same cas locus as the CRISPR-Lon/T):
https://padloc.otago.ac.nz/padloc/refseq/a97b2ce2802d9b9686bb6ea35a4cf169e2d4861978e64df5e1f56dc82b617f5b
The referee is right there are indeed Csx1-like proteins in the Webflags result. We have amended the sentence in the manuscript and mention this in the discussion.

L269-297 & 317: Is 1:50 correct, or should it be 50:1?
Thanks for catching this! It is actually 20:1 (w/w).

I had a couple of thoughts regarding the terminology of CRISPR-Lon and CRISPR-T and robustness of the nomenclature with future potential discoveries. Is CRISPR- the most appropriate prefix (rather than Cas- or SAVED-)? For CRISPR-T, have you looked for other potential families (non-MazF) of Lon-activated toxins, with potentially different mechanisms of toxicity? What would future analogous toxins be called? It is also possible that analogous CRISPR-Lon/T systems might be activated by other signal-producing defence systems, such as CBASS, so a more generic descriptor might be appropriate (e.g. SAVED-Lon). I cannot think of any obviously better nomenclature than what you currently have but it might be worth some thought.
After a lot of thought, and as mentioned at the beginning of this document, we renamed the system to CalpL, CalpT, CalpS. Calp stands for “CRISPR associated Lon Protease”. L, T, S would be short for Lon, Target, Sigma-factor, respectively.

Fig. 2 and Fig. S6: The gene arrangements 1/2 (coloured red in Fig. 2)/CRISPR-Lon appeared similar to the 1/8/CRISPR-Lon operons, so I checked whether 8 encodes a similar protein CRISPR-T and this appears to be the case (#8 from T. geofontis run through hhpred reveals an N-terminal MazF-like domain). Perhaps it's worth looking at a little closer.
We have noticed the CalpT (CRISPR-T) homolog and will look at it in future studies.

Fig. 2 and Fig. S9: Additional labelling of the lanes would be informative. That is, lane 3 clearly doesn’t contain CRISPR-Lon (due to the absence of the band), but this is not indicated in the labelling scheme.
Done.

Fig. S1: A) alone does not contain sufficient information to conclude CRISPR-Lon elutes as a monomer. Perhaps add an overlay showing the elution profile of protein standards? The light scattering data in Fig. 3 do support the monomer state, so an alternative would be to change the title of Fig. S1.
The referee is right. We changed the title of the figure.

Fig. S2: B) is too small to read. C) Please define pLDDT (and provide information for readers not familiar with AlphaFold to interpret the meaning of this metric). D) appears to be a superposition of two structures, but these are not explained.
This figure has changed a lot, due to the new data that was added. B) is now Extended Data Figure 1B and we have improved the readability of the figure.

Fig. S7: Please define pLDDT. It seems a little odd to show MazF from the MazEF structure with the RNA from the MazF structure (but not showing the corresponding MazF). Perhaps separate figures showing 1) MazEF compared with CRISPR-T and 2) MazF/ssRNA with CRISPR-T?
We agree that the figure was a bit convoluted. The Figure is now part of Extended Data Figure 4 and we improved it according to the referees' suggestions.

Fig. S8: A) alone does not contain sufficient information to conclude CRISPR-T elutes as a monomer (Fig. 3 does contain supporting data). Perhaps add an overlay showing the elution profile of protein standards? The earlier peak also contains a large amount of CRISPR-T (based on (B)). Is this a multimer or is the peak at the void volume (aggregate)? A comment on this would be useful (protein standards for the SEC will also help resolve this).
Due to its peculiar structure, isolated CalpT is not well suited for size determination by SEC. Hence, we have removed the "monomer" statement from the figure (now Extended Data Figure 4c) and referenced the MALS data in Figure 3. We have now also labelled the void peak as aggregates.

Fig. S9: Indication of the A194 residue would be informative (especially since the numbering is based on the recombinant protein with His-TEV tag).
The numbering was fixed and the cleavage site is now indicated. (now Extended Data Figure 4d)

Reviewer Reports on the First Revision:

Referees' comments:

Referee #1:

The authors have addressed all of my comments more than satisfactorily. Their new discovery of oligomeric state changes significantly strengthens the story. I strongly support publication in Nature. Well done!

David Taylor

Referee #2:

The authors were very receptive to the previous reviewers' comments and now provide a wealth of additional data. Excitingly, they now provide an activity for CalpT (formerly CRISPR-T), finding that it isn't a nuclease but in fact an anti-sigma-factor. This work will be of broad interest, and I recommend publication.

Referee #3:

Firstly, well done to the authors on their efforts to unravel the mystery of the CRISPR-T protein and extensive experimental work undertaken to address the review comments. The research team have been on an interesting journey with this project. Never ignore the operon (#NITO).

In the previous submission, the major outstanding question arising from the study was the role of the CalpT23 product (formerly designated CRISPR-T23), with the authors proposing a function as a CRISPR/Lon-activated toxin. In the revised manuscript, new experimental data revealed that CalpT23 is unlikely to function as a toxin. In the initial parts of the revised manuscript, the authors strengthen the mechanistic understanding and model for how CalpL is activated by the cyclic oligoadenylate signal (specifically cOA4) to proteolytically cleave CalpT into CalpT23 and CalpT10. They provide new SAXS data to support a cA4-induced CalpL/T oligomerisation, activation and trans-cleavage model, which is further supported by an elegant heterodimer/complex (WT and mutant proteins) cleavage experiment.

Next, the authors introduce a major twist (since the first manuscript), where a third protein, designated CalpS, is implicated in the Calp response mechanism. CalpS appears to be an anti-sigma factor and the authors demonstrate binding of CalpS to the E. coli RNA polymerase (Fig. 4). CalpS also binds the N-terminal MazF-like domain of CalpT (including to CalpL-bound CalpT). Importantly, CalpT-bound CalpS does not bind RNAP. Based on these findings, the authors propose a model whereby a stable CalpL-CalpT-CalpS complex exists in uninfected cells. Upon phage infection and recognition by the type III-B CRISPR-Cas system, cA4 is produced by target-RNA–bound CmrCas10. The cA4 binds CalpL, inducing oligomerisation of the Calp complex and leading to trans-cleavage of

CalpT, resulting in liberation of the CalpT23-CalpS subcomplex. The fate of liberated CalpT23-CalpS and subsequent mechanism for RNAP inhibition is speculated on but not demonstrated experimentally. The authors speculate that subsequent breakdown of CalpT23 (facilitated by host proteases, as demonstrated for other anti-sigma–sigma factor pairs) releases CalpS, which is then free to bind and inhibit RNAP.

The lack of experimental testing the hypothesised final piece of the mechanism leaves the manuscript somewhat open-ended. In the earlier version of the manuscript, where the function of CRISPR-T23 was speculated, the lack of experimental evidence demonstrating toxicity of CRISPR-T23 was highlighted as a major limitation, prompting the further investigations (described above) that revealed the role of CalpS. Demonstration that cA4 addition to cells possessing CalpT-CalpT-CalpS results in an accumulation of RNAP-bound CalpS and associated changes in host transcription would strengthen the present manuscript. However, the lack of experimental data directly linking CalpT23-S liberation to a transcriptional response (or phage defense phenotype) does not compromise the earlier parts of the study, which represent a substantial body of work to uncover an interesting and, to my knowledge, novel CRISPR immune response pathway/signaling cascade that will be of general interest. As is, the study will likely stimulate multiple new lines of research (some of which are covered in the Discussion), and there appears to be potential for multiple biotechnology applications (not directly discussed in the manuscript), signifying potential for widespread impact.

In terms of addressing reviewer comments from the first submission, the authors have satisfactorily addressed the major (addressing toxicity of T23) and minor comments.

Minor comments in the revision:

L98: No longer refers to Fig. 1b.*
*There are many instances where the figure references no longer point to the correct figures/panels.

L120: homologies -> similarity

L141: 10 kDa -> later, this is stated as 8.7 kDa. If so, should CalpT10 be CalpT9?

L159: I agree with the likely cleavage site, but the statement here is incorrect (“extended exactly to the identified cleavage site”). The HVA preceding the cleavage site was not detected (coloured black). Perhaps rephrase.

L349: A194 -> A195?

L357: S196 -> S196?

Author Rebuttals to First Revision:

Point-by-point reply to referees' comments (2nd round):

We would like to thank all three referees for their very constructive comments and the recognition of our work.

Referee #1:

The authors have addressed all of my comments more than satisfactorily. Their new discovery of oligomeric state changes significantly strengthens the story. I strongly support publication in Nature. Well done!

David Taylor

Thank you very much!

Referee #2:

The authors were very receptive to the previous reviewers' comments and now provide a wealth of additional data. Excitingly, they now provide an activity for CalpT (formerly CRISPR-T), finding that it isn't a nuclease but in fact an anti-sigma-factor. This work will be of broad interest, and I recommend publication.

Thank you very much!

Referee #3:

Firstly, well done to the authors on their efforts to unravel the mystery of the CRISPR-T protein and extensive experimental work undertaken to address the review comments. The research team have been on an interesting journey with this project.

Never ignore the operon (#NITO).
Indeed!

In the previous submission, the major outstanding question arising from the study was the role of the CalpT23 product (formerly designated CRISPR-T23), with the authors proposing a function as a CRISPR/Lon-activated toxin. In the revised manuscript, new experimental data revealed that CalpT23 is unlikely to function as a toxin. In the initial parts of the revised manuscript, the authors strengthen the mechanistic understanding and model for how CalpL is activated by the cyclic oligoadenylate signal (specifically cOA4) to proteolytically cleave CalpT into CalpT23 and CalpT10. They provide new SAXS data to support a cA4-induced CalpL/T oligomerisation, activation and trans-cleavage model, which is further supported by an elegant heterodimer/complex (WT and mutant proteins) cleavage experiment.
Thanks
Next, the authors introduce a major twist (since the first manuscript), where a third protein,

designated CalpS, is implicated in the Calp response mechanism. CalpS appears to be an anti-sigma factor and the authors demonstrate binding of CalpS to the E. coli RNA polymerase (Fig. 4). CalpS also binds the N-terminal MazF-like domain of CalpT (including to CalpL-bound CalpT). Importantly, CalpT-bound CalpS does not bind RNAP. Based on these findings, the authors propose a model whereby a stable CalpL-CalpT-CalpS complex exists in uninfected cells. Upon phage infection and recognition by the type III-B CRISPR-Cas system, cA4 is produced by target-RNA–bound CmrCas10. The cA4 binds CalpL, inducing oligomerisation of the Calp complex and leading to trans-cleavage of CalpT, resulting in liberation of the CalpT23-CalpS subcomplex. The fate of liberated CalpT23-CalpS and subsequent mechanism for RNAP inhibition is speculated on but not demonstrated experimentally. The authors speculate that subsequent breakdown of CalpT23 (facilitated by host proteases, as demonstrated for other anti-sigma–sigma factor pairs) releases CalpS, which is then free to bind and inhibit RNAP.

The lack of experimental testing the hypothesised final piece of the mechanism leaves the manuscript somewhat open-ended. In the earlier version of the manuscript, where the function of CRISPR-T23 was speculated, the lack of experimental evidence demonstrating toxicity of CRISPR-T23 was highlighted as a major limitation, prompting the further investigations (described above) that revealed the role of CalpS. Demonstration that cA4 addition to cells possessing CalpT-CalpT-CalpS results in an accumulation of RNAP-bound CalpS and associated changes in host transcription would strengthen the present manuscript.

In our manuscript, we have changed all speculations that sounded too strong around that point. We believe we are now fairly staying in line with (i) our experimental data and (ii) a recently published article

(i) We have experimentally shown that CalpS binds to RNAP, as expected for a sigma factor. Thus, we feel that our conclusion that CalpS acts as a sigma factor and hence influences transcription is a not too far-fetched and we think that we have carefully discussed the limitations of our experimental data.

(ii) We have included a very recent study published this month showing the release of a similar sigma factor from a different CRISPR subtype. The protease involved is not a cOA-activated Lon protease but a caspase-like protease that is part of the effector complex. In their article, they don't show the binding of the sigma factor to the RNA polymerase but they provide proof of binding to a specific DNA sequence. They found three promotors containing this sequence; one of them promote the expression of the CRISPR proteins needed for adaptation.

Of course, we agree that such data would have made the story complete and we are also very curious about the role of CalpS and its molecular and cellular downside effects.
As we would like to carry out such experiments in *Sulfurihydrogenibium*, we hope the referee agrees that the amount of experimentation to really "nail this down" merits a separate follow up study, as indicated by the referee in the next paragraph.

However, the lack of experimental data directly linking CalpT23-S liberation to a transcriptional response (or phage defense phenotype) does not compromise the earlier parts of the study, which represent a substantial body of work to uncover an interesting

and, to my knowledge, novel CRISPR immune response pathway/signaling cascade that will be of general interest. As is, the
study will likely stimulate multiple new lines of research (some of which are covered in the Discussion), and there appears to be potential for multiple biotechnology applications (not directly discussed in the manuscript), signifying potential for widespread impact.
Thank you! See also comment above.

In terms of addressing reviewer comments from the first submission, the authors have satisfactorily addressed the major (addressing toxicity of T23) and minor comments.

Minor comments in the revision:

L98: No longer refers to Fig. 1b.*
*There are many instances where the figure references no longer point to the correct figures/panels.
Thank you! We have carefully checked the references to the figures/panels again.

L120: homologies -> similarity
Done.

L141: 10 kDa -> later, this is stated as 8.7 kDa. If so, should CalpT10 be CalpT9?
We decided to leave it as CalpT10

L159: I agree with the likely cleavage site, but the statement here is incorrect (“extended exactly to the identified cleavage site”). The HVA preceding the cleavage site was not detected (coloured black). Perhaps rephrase.
We changed it to “almost exactly”.

L349: A194 -> A195?
Fixed.

L357: S196 -> S196?
Fixed.