In vitro reconstitution of Cascade-mediated CRISPR immunity in Streptococcus thermophilus

Tomas Sinkunas, Giedrius Gasiunas, Sakharam P Waghmare, Mark J Dickman, Rodolphe Barrangou, Philippe Horvath, and Virginijus Siksnys

Corresponding author: Virginijus Siksnys, Institute of Biotechnology, Vilnius University

Review timeline:

Submission date: 26 August 2012
Editorial Decision: 01 October 2012
Revision received: 11 November 2012
Editorial Decision: 27 November 2012
Author Correspondence: 29 November 2012
Accepted: 29 November 2012

Transaction Report:

(Note: With the exception of the correction of typographical or spelling errors that could be a source of ambiguity, letters and reports are not edited. The original formatting of letters and referee reports may not be reflected in this compilation.)

1st Editorial Decision 01 October 2012

Thank you for submitting your manuscript for consideration by the EMBO Journal. It has now been seen by three referees whose comments are shown below.

As you will see from the reports, while referees #2 and #3 are generally positive about your manuscript, referee #1 raises concerns about the novelty of the reported findings relative to a recently published paper addressing the Cascade complex in E. coli. In light of these concerns I consulted with the other referees, and they both disagreed with referee #1 and found your manuscript to provide sufficient novelty over the literature to be within the scope of The EMBO Journal. However, they both emphasized that for the revised version, you will need to discuss your findings extensively in the context of the E. coli study and to highlight similarities and differences between the two Cascade systems. In particular, the difference between the E. coli and S.thermophilus system with regard to binding of supercoiled DNA should be addressed further.

Given the referees' positive recommendations, we offer you the opportunity to submit a revised version of the manuscript, addressing the comments of all three reviewers. I should add that it is EMBO Journal policy to allow only a single round of revision, and acceptance or rejection of your manuscript will therefore depend on the completeness of your responses to the full satisfaction of all three referees in this revised version. Please do not hesitate to contact me if you have questions related to the review process and the requests made by the referees.
When preparing your letter of response to the referees' comments, please bear in mind that this will form part of the Review Process File, and will therefore be available online to the community. For more details on our Transparent Editorial Process, please visit our website: http://www.nature.com/emboj/about/process.html

We generally allow three months as standard revision time. As a matter of policy, competing manuscripts published during this period will not negatively impact on our assessment of the conceptual advance presented by your study. However, we request that you contact the editor as soon as possible upon publication of any related work, to discuss how to proceed. Should you foresee a problem in meeting this three-month deadline, please let us know in advance and we may be able to grant an extension.

Thank you for the opportunity to consider your work for publication. I look forward to your revision.

Yours sincerely,

Editor
The EMBO Journal

REFEREE REPORTS:

Referee #1 (General Remarks):

This manuscript describes biochemical characterization of the Type I-E CRISPR interference complex from S. thermophilus, which is analogous in composition and activity to the previously studied Type I-E Cascade complex from E. coli. Data are presented showing that just as for the Cascade complex, the S. thermophilus complex assembles from 5 proteins and a single CRISPR RNA (crRNA), and recruits a separate endonuclease, Cas3, which cleaves dsDNA bearing a sequence complementary to the crRNA. As previously described, a short sequence motif (PAM), in this case a single nucleotide, adjacent to the targeted DNA sequence, is required for target cleavage. Although the experiments are solid and the results appear sound, there is little that is new here given the extensive prior biochemical, genetic and structural work published for the highly analogous E. coli Type I-E system. Overall, this work is better suited to a more specialized journal given that it provides little additional insight into the composition or function of this CRISPR-associated complex.

Other comments:

1. References need to be cross-checked with the text; for example, Young et al (2012), cited on p. 4, is not listed in the References.

2. In Fig. 4, it looks like the majority of the DNA remains intact after a 1-hour incubation with an excess of St-Cascade+Cas3. Does this imply that some additional factor(s) are necessary for the complex to turn over? Is the remaining DNA cleaved upon longer incubation times? Also, is there a role for supercoiling in the cleavage reaction efficiency, as has been suggested for the E. coli Cascade system?

Referee #2 (General Remarks):

Sinkunas et al, describe isolation and characterization of Streptococcus thermophiles Cascade complex. Authors were able to purify the St-Cascade complex assembled in E coli by using affinity, size exclusion and heparin affinity chromatography. The bound crRNA was isolated from the complex and identified by mass spectrometry. Using the purified Cascade complex, authors successfully demonstrated that, together with St Cas3, the St-Cascade complex binds to and cleaves
proto-spacer DNA. They further addressed the requirement for PAM sequences in proto-spacer binding, R-loop formation, and target DNA cleavage. Experiments were designed and executed beautifully and the findings are significant. The most significant finding of this work is the demonstration that of St Cascade cleaves DNA in a Cas3 and ATP dependent manner. Authors did not stop at this discovery and further moved on to detailed characterization of DNA sequences required for Cascade binding and Cas3 cleavage. The results are convincing that the PAM sequence plays an important role in these processes.

This is a rare paper that I can't find much criticism for. Well done.

Referee #3 (General Remarks):

Sinkunas et al. have carried out an impressive functional in vitro study of the St-CRISPR4-Cas system of Streptococcus thermophilus (St), which belongs to CRISPR type I-E. This type of CRISPR/cas system was previously studied in E. coli where a multi-subunit RNP complex termed Cascade was shown to bind to invading target DNA and interfere with it via the accessory Cas3 protein.

Here, Sinkunas et al. achieved to reconstitute in vitro the corresponding S. thermophilus system to show that the Cas proteins of this model organism assemble into a Cascade complex and recruit Cas3 to form a functional effector-complex that then cleaves target DNA. Using these biochemical assays, they are able to break down the last step of CRISPR mediated target DNA interference into target site recognition and cleavage stages, which are executed by St-Cascade and St-Cas3. They determine experimentally an usually short (1 nucleotide) PAM motif that serves as the Cascade binding side in the proto-spacer; such a promiscuous PAM like motif argues for the presence of a second more stringent PAM requirement for the spacer acquisition step which is perhaps mediated by either Cas1 or 2.

The authors reveal that ATPase and nuclease activities of St-Cas3 are induced by St-Cascade binding to the proto-spacer dsDNA. Furthermore they show that St-Cas3 initially cuts both DNA strands in the proto-spacer and is translocating along the non-target strand that is further cleaved off. The unidirectional translocation of Cas3 may contribute to the selection of a specific DNA strand from which new spacers are subsequently acquired and suggests a functional link between interference and acquisition. Their findings are summarized in a model for CRISPR-mediated interference mechanism.

This is an important study that reveals novel mechanistic details of the different steps of DNA interference by CRISPR-Cas (and sufficiently transcends the previous E. coli work to meet the novelty criteria of EMBO Journal). The paper is very well written and presents well-conceived and well-conducted experiments. All the figures are presented in high quality format.

Specific comments:

1. In general, the CRISPR systems that target dsDNA in an ATP-independent manner require an energy source for strand opening. Westra et al., 2012 showed a preference of Cascade from type I-E Escherichia coli strain K12 for negative supercoiled target DNA since most of the energy needed can be supplied by the free energy of supercoiling. This observation that DNA binding by the crRNA-guided Cascade complex is constrained by strict topological requirements of the target DNA is therefore likely to be a general characteristic for all mesophilic CRISPR/Cas systems that target dsDNA. However in the St-CRISPR4-Cas TypeI system studied here, the authors show that crRNA guided Cascade binds to the matching DNA strand of both linear oligonucleotide DNA duplex (Fig. 4AB) and plasmid DNA (Fig. 3). The authors should elaborate more on this obvious discrepancy between the typeI-E systems of E. coli and Streptococcus.

2. The authors conclude from their experiments that R-loop formation resulting from crRNA-guided Cascade binding to the target DNA serves to load the St Cas3 protein. St-Cas3 binding to the ssDNA present in the R-loop triggers ATP/helicase activity which presumably contributes to remodelling of the Cascade complex. Cascade remodelling makes both strands available for St-Cas3...
induced cleavage. However, an recent impressive study by Sashital et al. 2012 (Molecular Cell, 46(5):606-15) showed for E. coli that after Cascade-mediated R loop formation, the Cse1 subunit (casA) recruits Cas3, which catalyzes nicking of target DNA through its HD-nuclease domain. The authors must comment on this observation and should cite the paper.

Minor points:

3. To help the reader understand the function of Cascade proteins, the authors should discuss which of them is involved in PAM recognition. Include the properties of the recognition motif provided by loop L1 of CasA present in S. thermophilus and highlight structural or sequence differences.

4. Page 3, second para, type II system: Add reference to the article by Deltcheva et al., 2011 in Nature. In addition, note that the system can either target RNA or DNA as shown for S. epidermidis by Marraffini and Sontheimer, 2010.

5. Page 3, last para: When describing the presence of other CRISPR type II and III systems in S. thermophilus, the authors may want to comment why acquisition has not been observed for type I.

6. Page 5, first para: The authors conclude from the band intensity in the stained SDS gel that the stoichiometry of the S. thermophilus Cascade proteins is the same as in E. coli. However, since some proteins get better stained then others independent of their copy numbers, this statement needs to be taken with a grain of salt.

7. A characterization of co-purified crRNAs reveals a 7-nucleotide handle. However, all other investigated type I CRISPR-Cas systems showed that endoribonucleases Cas6e or Cas6f cleave precursor crRNA leaving an 8-nucleotide handle. So, why would it be a 7-nucleotide handle for the Cas6e nuclease of S. thermophilus?

8. Page 7, first para "St-Cascade binding to...": Mention the recently reported presence of a seed sequence that enhances target recognition as observed for typeI-E of E. coli (Semenova et al., 2011) and typeI-F in Pseudomonas aeruginosa (Wiedenheft et al., 2011).

1st Revision - authors' response 11 November 2012

Response to Reviewers’ Comments

Referee #1 (General Remarks):

Other comments:

1. References need to be cross-checked with the text; for example, Young et al (2012), cited on p. 4, is not listed in the References.

Answer: reference Young et al (2012) is included in the reference list in the revised manuscript.

2. In Fig. 4, it looks like the majority of the DNA remains intact after a 1-hour incubation with an excess of St-Cascade+Cas3. Does this imply that some additional factor(s) are necessary for the complex to turn over? Is the remaining DNA cleaved upon longer incubation times? Also, is there a role for supercoiling in the cleavage reaction efficiency, as has been suggested for the E. coli Cascade system?

Answer: The majority of the substrate DNA remains intact after 1-hour incubation with an excess of St-Cascade+Cas3 when ATP is missing in the reaction mixture. In the presence of ATP only ~20 % of oligoduplex or plasmid substrate is cleavage-resistant even at longer incubation times. It is possible that St-Cascade binding to the proto-spacer yields a fraction of unproductive complexes that hamper St-Cas3 binding or, alternatively, a fraction of St-Cas3 is cleavage-defective. Furthermore, gel shift data indicate that an effector complex remains bound to the cleaved DNA...
(Fig. S11) and this might affect a turnover rate under the multiple turnover conditions. In principle, we cannot exclude the possibility that some cellular components can displace Cascade and promote turnover, however, currently we are not aware of such factors. A role for supercoiling in the cleavage reaction efficiency has been recently suggested for the *E. coli* Cascade system (Westra et al., 2012). To test whether DNA supercoiling affects DNA cleavage rate in the St-Cascade/Cas3 system, we monitored St-Cascade/Cas3 cleavage of supercoiled and linear DNA. Data provided in Figure S12 of the SI section demonstrate that cleavage rates of supercoiled and linear DNA are very similar (Figure S12). These results do not exclude the role of supercoiling in the Cascade and target DNA interactions but indicate that supercoiling does not limit the cleavage reaction rate for St-Cascade/Cas3 system. Further explanatory sentences have been added to the Results and Discussion sections in the revised manuscript.

Referee #3 (General Remarks):

1. **In general, the CRISPR systems that target dsDNA in an ATP-independent manner require an energy source for strand opening.** Westra et al., 2012 showed a preference of Cascade from type I-E *Escherichia coli* strain K12 for negative supercoiled target DNA since most of the energy needed can be supplied by the free energy of supercoiling. This observation that DNA binding by the crRNA-guided Cascade complex is constrained by strict topological requirements of the target DNA is therefore likely to be a general characteristic for all mesophilic CRISPR/Cas systems that target dsDNA. However in the St-CRISPR4-Cas Type I system studied here, the authors show that crRNA guided Cascade binds to the matching DNA strand of both linear oligonucleotide DNA duplex (Fig. 4AB) and plasmid DNA (Fig. 3). The authors should elaborate more on this obvious discrepancy between the typeI-E systems of *E. coli* and Streptococcus.

   **Answer:** We appreciate the reviewer’s concern and have performed an additional experiment which shows that degradation rates of supercoiled and linear substrates by St-Cascade/Cas3 are very similar (Figure S12). These results do not exclude the role of supercoiling for the Cascade and target DNA interactions but indicate that supercoiling does not limit the cleavage reaction rate for St-Cascade/Cas3 system. Explanatory sentences have been added to Results and Discussion sections in the revised manuscript.

2. **The authors conclude from their experiments that R-loop formation resulting from crRNA-guided Cascade binding to the target DNA serves to load the St Cas3 protein.** St-Cas3 binding to the ssDNA present in the R-loop triggers ATP/helicase activity which presumably contributes to remodelling of the Cascade complex. Cascade remodelling makes both strands available for StCas3 induced cleavage. However, an recent impressive study by Sashital et al. 2012 (Molecular Cell, 46(5):606-15) showed for *E. coli* that after Cascade-mediated R loop formation, the Cse1 subunit (casA) recruits Cas3, which catalyzes nicking of target DNA through its HD-nuclease domain. The authors must comment on this observation and should cite the paper.

   **Answer:** We thank the reviewer for this update and added the reference and relevant text to the manuscript.

Minor points:

3. To help the reader understand the function of Cascade proteins, the authors should discuss which of them is involved in PAM recognition. Include the properties of the recognition motif provided by loop L1 of CasA present in *S. thermophilus* and highlight structural or sequence differences.

   **Answer:** Following reviewer’s suggestion we added an explanatory sentence on p. 11 and included a reference to the paper by Sashital et al. 2012 which demonstrates that the L1 loop of CasA protein in the Cascade complex of *E. coli* is involved in PAM recognition. Sequence comparison suggests that a similar loop may be present in the CasA of St-Cascade, however we were unable to locate a putative PAM recognition motif in the predicted L1 loop of *S. thermophilus* CasA protein.

4. **Page 3, second para, type II system:** Add reference to the article by Deltcheva et al., 2011 in *Nature*. In addition, note that the system can either target RNA or DNA as shown for *S. epidermidis* by Marraffini and Sontheimer, 2010.
Answer: The text was modified and references added per the reviewer’s comment.

5. Page 3, last para: When describing the presence of other CRISPR type II and III systems in *S. thermophilus*, the authors may want to comment why acquisition has not been observed for type I.

Answer: There is no universal definition of CRISPR activity in the literature, as it may be defined by either the ability to acquire novel spacers, or the ability of existing spacers to provide interference against homologous sequences. Because the two Type II systems in *S. thermophilus* are so highly active, these two loci would likely have to be inactivated or deleted to assess whether the Type I system is active in terms of its ability to acquire spacers against invasive elements. This is clearly beyond the scope of this manuscript. Conversely, we attempted to reconstruct activity as defined by interference *in vitro*. Similarly to the CRISPR system present in *E. coli* (strain K12), “natural” spacer acquisition has not been observed for the CRISPR4 system of *S. thermophilus* upon phage challenge, though some Cas proteins have been detected, notably following phage infection (Young et al. 2012). In this work we have focused on the functional characterization of the interference complex of St-CRISPR4-Cas system and provide experimental evidence that it is active *in vitro*. The spacer acquisition mechanism is beyond the scope of this manuscript. We can only speculate that the spacer acquisition machinery of CRISPR4 is repressed by some unknown factor(s), or that some of its components (Cas1, Cas2, leader…) are inactivated, or that the activity of the Type II loci is superseding that of the Type I system.

6. Page 5, first para: The authors conclude from the band intensity in the stained SDS gel that the stoichiometry of the *S. thermophilus* Cascade proteins is the same as in *E. coli*. However, since some proteins get better stained then others independent of their copy numbers, this statement needs to be taken with a grain of salt.

Answer: We agree with reviewer’s comment that stoichiometry of St-Cascade suggested by the SDS-gel analysis should be taken with caution. In a strict sense, SDS gel analysis of St-Cascade shows that: i) all 5 CasABCDE proteins of the *S. thermophilus* CRISPR4 system are present in the complex, suggesting that they may assemble into a St-Cascade complex similar to that of *E. coli* ii) CasC protein is likely present in St-Cascade in multiple copies if its staining efficiency is similar to that of other proteins iii) mass spectrometry analysis of the St-Cascade tryptic digest suggests that the CasC protein, is the most abundant protein present in St-Cascade similar to the Ec-Cascade.

7. A characterization of co-purified crRNAs reveals a 7-nucleotide handle. However, all other investigated type I CRISPR-Cas systems showed that endoribonucleases Cas6e or Cas6f cleave precursor crRNA leaving an 8-nucleotide handle. So, why would it be a 7-nucleotide handle for the Cas6e nuclease of *S. thermophilus*?

Answer: In the seminal Science paper (Brouns et al. 2008) 29-nt repeat and 32-nt spacer sequences were considered for the *E. coli* CRISPR-Cas system. However, recent experimental data (Goren et al., 2012; Swarts et al., 2012) on spacer incorporation show that the last nucleotide in the repeat is sometimes degenerated and most likely originates from the spacer, therefore the conserved repeat sequence (named duplcon by Goren et al., 2012) in the *E. coli* Type I system is 28 nt and the spacer sequence is 33 nt. In St-CRISPR4-Cas system, the first 28 nt of the repeat are conserved while the 29th nt is either C or T. Therefore, we postulate that in the St-CRISPR4-Cas system the repeat length is 28 nt and the spacer sequence is 33 nt. The processed crRNAs in the *E. coli* and *S. thermophilus* systems have the same length (61 nt) suggesting that in the precursor crRNA the cleavage position of endoribonucleases Cas6e is conserved in both systems and located at the 21st nucleotide within the repeat sequence. In this case, crRNAs are made of a 7-nt 5’ handle, a 33-nt spacer, and a 21-nt 3’ handle. An explanatory paragraph has been added to Discussion section.

8. Page 7, first para "St-Cascade binding to...": Mention the recently reported presence of a seed sequence that enhances target recognition as observed for typeI-E of *E. coli* (Semenova et al., 2011) and typeI-F in *Pseudomonas aeruginosa* (Wiedenheft et al., 2011).

Answer: We have modified the text accordingly.
Thank you for submitting your revised manuscript for our consideration.

It has now been seen once more by one of the original referees (see comments below), and I am happy to inform you that there are no further objections towards publication in The EMBO Journal.

However, before we will able to send you a formal letter of acceptance, there are just a few minor things I need to ask you for: While the manuscript figure legends are generally very informative, I noticed that the values indicated by the error bars in Fig. 3a are not specified in the legend. I would therefore ask you to add this information to the text and send the updated file to me by email. We will then upload it in house.

In addition, we now encourage the publication of source data, particularly for electrophoretic gels and blots, with the aim of making primary data more accessible and transparent to the reader. Would you be willing to provide files comprising the original, uncropped and unprocessed scans of all gels used in the figures? We would need 1 file per figure (which can be a composite of source data from several panels) in jpg, gif or PDF format, sent to the editor by email as "Source data files". The gels should be labelled with the appropriate figure/panel number, and should have molecular weight markers; further annotation would clearly be useful but is not essential. These files will be published online with the article as a supplementary "Source Data". Please let me know if you have any questions about this policy.

Thank you for letting us consider your manuscript. Pending the additional information mentioned above we should then be able to swiftly proceed with formal acceptance and production of the manuscript!

Yours sincerely,

Editor
The EMBO Journal

REFEREE REPORT

Referee #3

The revisions by Sinkunas et al. have satisfactorily addressed my previous criticism, and the additional experiment in Figure S12 has improved the manuscript further.

Author Correspondence

Thank you for e-mail and I am happy to hear that there are no further objections towards publication of our manuscript.

Attached please find Figure 3 legend corrected following your suggestion (changes are marked in yellow). The source files for the Figures are also attached.