nature research

Peer Review File

Manuscript Title: DNA targeting and interference by a bacterial Argonaute nuclease

Editorial Notes:

Redactions – Mention of other journals

This document only contains reviewer comments, rebuttal and decision letters for versions considered at *Nature*. Mentions of the other journal have been redacted.

Reviewer Comments & Author Rebuttals

Reviewer Reports on the Initial Version:

Referee #1 (Remarks to the Author):

This manuscript investigates the generation of DNA associated with the Clostridium butyricum Argonaute protein when it is expressed in E. coli. Argonaute proteins, best characterized for their role in eukaryotic RNA interference, are found ubiquitously in archaea and in bacteria, where they interact with small DNA molecules (smDNA). This work examines the role of the nuclease activity of CbAgo and the genomic derivation of the smDNA associated with it. By genetic analysis, they implicate the double-strand break repair proteins in the generation of smDNA. Over-represented in smDNA is DNA from multiple copy number elements, including plasmids and transposons, as well as sites of double-strand breaks. By examining persistence of plasmid DNA and single-strand DNA bacteriophage titers, they show that bacterial Ago can protect cells from extragenomic infectious elements.

All in all, this is an important demonstration of the role of this group of proteins and an intriguing mechanism for smDNA generation. The data are sound and well-presented. I have some issues with some of the language used and there are few things that would improve and clarify the reasoning in the manuscript.

Line 26 and throughout. It took me some time to realize that the protein is being expressed in E. coli and the analysis entirely done in this organism. It should be "bacterial Argonaute nuclease CbAgo expressed in Escherichia coli." This should also be mentioned when introducing the results, for instance line 60 "purified it from E. coli cells".

Line 28 and throughout. This is misleading. I would not use the term "cooperation" and the generation of smDNA guides does not require double-strand break repair machinery, since smDNA is generated fine in recBCD and recA mutants. Cooperation implies that the processes occur together or are more efficient in concert, neither of which are demonstrated here. The double-strand break machinery merely influences the composition of the guide DNA.

Line 70 "restricted" evokes restriction/modification systems, a better term would be "bounded" or "flanked

Line 77 and Fig. 1: It is hard to see the offset of the single peak in tus- strains relative to TerC

Line 85 and Fig. 1. I would very much like to see a larger scale representation of the spread of smDNA enrichment relative to the boundaries of plasmid homology region, close to lacI or araC. This is an important point and this would emphasize it. The spreading phenomenon is different than the very targeted type 2 R/M or CRISPR

Line 89 and throughout: "DNA interference" is not being measured here and it's not clear what "interference between the plasmid and chromosomal loci" means. Do you mean "against"?

Fig. 1 legend: please state that the lower panel in parts b, c, d is dAgo (it's only said in the legend for b)

Line 115: how sure are we that the mutant is indeed catalytically inactive, since much of the interpretation keys on this? An experimental demonstration would be nice, or at least some reasoning/reference that makes us more confident about this fact.

Line 132: the enrichment of repetitive/higher copy DNA seems like it would require some amplification step, which maybe you should elaborate on. It would be easy enough to investigate plasmid copy number on smDNA.

Line 187: Do you mean persistence of DNA ends in recA-? Expressing gam protein (phage-encoded inhibitors of RecBCD) would be interesting.

Line 221: the lower efficiency is hard to see in Fig. 3
Line 225: How fast does smDNA turn over? (just inerested, not a critique)

Line 242: I wouldn't say the effect is "massive" , substantial, maybe

Line 248: At least for transposons, free ends are rare, transposition rates low. Is this system especially sensitive to conjugal plasmids, since they transmit via single-strand DNA?

Line 256: It is important to mention here that the preferential use of plasmid sequences occurs even when CbAgo is expressed from the chromosome. I missed this the first time through, and thought that the effect could be in cis from the expression locus

Line 279: You are not directly measuring substrate processing by the nuclease, this is misleading

Line 282: again, not necessarily cooperating

Line 289: not clear to me why smaller replicons are more vulnerable. The lack of Chi sites could have the opposite effect, by promoting complete degradation. Is there any evidence that ssDNA is more vulnerable to these nucleases? Interesting because you saw the effect on ssDNA phage: would you see this for lambda?

Line 314: a potential universal defense system, one example does not make a universe

Line 547: does "negative colonies" mean plaques?

Fig. 4E: I'm a bit worried that potential toxicity of the expressed protein if it is plasmid-encoded. Is CbAgo expressed from the genome here? That would be the best experiment.

Figure extended 7: what does "long-A" etc refer to?

Referee #2 (Remarks to the Author):

Kuzmenko et al. explore the mechanism of guide generation for prokaryotic Argounaut nucleases. This nucleases have been identified bioinformatically about 15 years ago, but both their mechanism of function and biological significance is still relatively unknown. In the first part of the manuscript (Figs. 1 and 2) it is clearly demonstrated that CbAgo is preferentially loaded with guides from DSB sites as well as DNA sequences that are present in multiple copies within the cell.

In both cases the DSB repair machinery is involved, as guide hot-spots are delimited by chi sites and largely eliminated in the absence of the RecBCD machinery. This is an important results that significantly advances our understanding of pAgo mechanism.

The second part of the paper explores the biological significance of these findings. Here the data shown is quite limited. Figure 3 shows that sequences from which the CbAgo guides are extracted from are subject to CbAgo cleavage, which is a somewhat expected result given the well known DNA-guided nuclease activity of this enzyme. In my opinion, the main issue to investigate is the consequence of this self-targeting for the bacterial population. And Fig. 4 attempts to demonstrate that CbAgo provides defense against mobile genetic elements. While the data is convincing for plasmids, the results for M13 infections are a confusing, as the nuclease dead version of CbAgo still provides a significant level of protection.

Therefore, the paper will benefit from more experiments to clarify this second part:

- What portion of the cells in the culture acquire self-targeting guides? Is it possible to do single cell DNA seq to find out?
- Is self-targeting toxic to the host? A competition assay of cells with and without CbAgo should answer this.
- Does self-targeting results in DNA repair of the break? Is this repair mutagenic in a way that it could eliminate, say, the I-SceI site to prevent further cleavage and generate "escapers" of CbAgo toxicity (if there is such toxicity)?
- M13 is in many ways a very special phage. Does CbAgo provide protection against the most common E. coli phages such as lambda, T7 and P1?
- Direct evidence of the presence of M13-derived guides in CbAgo following infection should be provided.

Other concerns:
- It seems that the extraction of guides from homologous sequences is related to their recombination. It is highly likely that recombination of these identical sequences is resolved generating some degree of DSBs that are used as a source of guides. This process will involve RecBCD also. In the present text, the reason behind the enhanced guide abundance from duplicated/repetitive sequences is largely left undiscussed.
- Define Cb in the abstract.
- Line 60. The authors need to explicitly mention that CbAgo is expressed in E. coli. Since this is not the natural host for this nuclease, the discussion should address why the authors think that the results of their study are most likely not affected by this experimental set up.
- Lines 88-89: it is impossible to know the order of events from these results. The first guide could come from the chromosome and lead to the cleavage of the plasmid, which will then begin a positive feedback loop of cleavage and guide extraction at both loci.
- Fig. 2b is reference before 2a. Fig. 4e is before 4d.
- Lines 177-178: the functions of RecB, C and D should be briefly explained (helicase/nuclease for B and D; chi recognition for C). Can the results be explained by these different functions? The conclusion that "different components… have independent roles" is not very useful for the reader. It should also be explained that in the absence of RecB and D the complex will not form and the results of the mutant strain lacking these components are not a reflection of RecC activity, but of a situation where there is no RecBCD complex (that is not the case in the absence of RecC, where RecBD can still recognize the DSB and degrade it). On this topic, have the authors tested mutants in the active sites of RecB and/or RecD? These mutations allow more accurate conclusions as they allow complex formation, just without different activities.
- Line 180. The "single peak is hard to see. A zoomed-in panel could help.
- Line 209. Which host? Not E. coli.
- Line 279-280: it is not clear that the guides are preferentially acquired from foreign DNA. Fig. 4a shows that there are 10-fold more guides from the plasmid than what would be predicted by chance. But this is probably true also of the ter sites in the chromosome. Did the authors compare

the total amount of reads for chromosomal vs plasmid guides? Are these different when normalized by genome size and by copy number? The only comparison I can make from the presented data is in Fig. 1a, where many of the araC reads must come from the plasmid but still the terC peak is much higher.
- Line 288, point (ii). It is not clear to me why replication intermediates are "located in smaller … replicons". Every replicon regardless of its size has replication intermediates. Also, it may be better to clarify if the authors are referring to replication intermediates that lead to DSBs.
- Line 290, point (iii). While it is true that phages lack chi sites, they all have RecBCD inhibitors (lambda gamma gene from the Red system for example). This should be discussed, as it could be an evolutionary response of phages to prevent pAgos from obtaining guides from their genomes.
- Fig. 2a: label each RecBCD subunit with its letter.
- Fig. 3. The title is confusing. CbAgo does not attack DSBs; guides are taken from DSB sites during their repair by the RecBCD machinery and then, once the DSB is fixed, cleaved again by the CbAgo molecules that where loaded with those guides. Also, if this is the model, it should be also included in Extended Fig. 6.

Referee #3 (Remarks to the Author):

Kuzmenko et al. present a study of prokaryotic Argonaute (pAgo) and study DNA interference in a bacterium, with targeting of chromosomal, plasmid and phage DNA. Specifically, using genetics and deep-coverage DNA sequencing techniques, the authors map the Clostridium butyricum pAgo guide DNAs on genomic areas of interest and implicate ter sites, DSBs and DNA repair machinery in pAgo targeting. Their primary conclusion is that pAgo targeting protects bacteria from invasive DNA such as plasmids and phages. This is a generally timely and interesting topic.

The key conclusion of the manuscript, as highlighted in the title, abstract and throughout the manuscript is that pAgo-mediated DNA interference provides protection against invasive DNA, notably phage, plasmid and IS elements. Yet, the large majority of the work performed focuses on chromosomal self-targeting and seems to indicate that there is broad and non-toxic targeting of various chromosomal locations, especially at ter, rRNA operons and IS elements loci. If the DNA targeting is occurring and efficient and does enable phage DNA and plasmid DNA and thus chromosomal DNA interference, and if DNA cleavage is the outcome then it must be shown, characterized, quantitatively measured and mechanistically and functionally explained. Putatively, the nuclease-enabled pAgo cleaves phage and plasmid DNA, leading to protection against invasive foreign genetic elements, so this cleavage activity must be shown and documented on phage and plasmid DNA. The conclusion that pAgo protects bacteria from invaders DNA is not fully documented nor supported by the data provided. For instance, the authors claim to have "demonstrated that it induces cleavage of multicopy genetic elements, including plasmids, transposons and repetitive chromosomal loci" and that "the mechanism … endows CbAgo with capacity to eliminate plasmid DNA and fight phage infection", but neither is proven since no DNA cleavage data is provided. Likewise, it is not clear that the "results show that foreign DNA … is a preferential substrate for processing by CbAgo".

The authors repeatedly show that CbAgo smDNA occur throughout the chromosome, with deep(er) coverage at ter sites. Is there any indication that DNA cleavage occurs in the chromosome, or not? When targeting IS elements (including two families, namely IS1 and IS3 with copies throughout the genome), is there any indication that DNA is actually cleaved? The authors do show that the catalytically inactive "deactivated" dCbAgo has loss of smDNA biogenesis, but it is still "active" from a functional standpoint (e.g. provides plasmid loss and phage titer reduction, and can retain the ability to target DSBs, albeit with lower efficiency). The authors must discuss whether DNA cleavage is the driver of the process, or not. If DNA is driving the observed effect, then the authors must explain how the deactivated version dCbAgo provides interference in the absence of cleavage.

Mechanistically, clearance of plasmid by loss over passages in the absence of selective pressure is not the same as cleavage of plasmid following transformation under selective pressure. To claim plasmid interference, the latter should be determined. The former does show "quantitative loss of plasmid" over time in some conditions, but this may not constitute plasmid "attack" per se. Likewise, reduction of phage titer by CbAgo is certainly noteworthy, but this was performed over a short timeframe, and the impact of the "active" CbAgo is better but in some ways comparable to that of the "inactive" dCbAgo, begging the question as to what is exactly happening, and that the actual impact is. Phage interference should be measured at the DNA level, and possibly at the RNA level too. Critically, phage biology encompassing plaque size, burst size, and efficiency of plaquing should be investigated and quantitatively determined. Multiple phages and various conditions over more extended periods of time should be assessed (at least 16-24 hrs). The authors likely must determine what happens to phage DNA and phage RNA over time when targeted by both CbAgo and dCbAgo, given the relative effect of both on phage titer (Figure 4F).

Unfortunately, the authors extensively conclude with a "remarkable" comparison between pAgo an CRISPR-Cas immune systems, and make a series of misguided statements about their mechanistic commonalities and the potential to repurpose pAgos as molecular machines to manipulate genomes. Invoking the DNA repair machinery in the context of Cas-mediated interference is not appropriate, nor is the mention of "similar principles for differentiation between self and nonself" given the conspicuous absence of a PAM. Concluding with a hyperbolic comparison to the most impactful and disruptive technological development in two decades should be re-considered. This is particularly important in light of prior attempts by several groups to exploit pAgos for genome manipulation, and the generated results.

Editorial comments:

- The authors should consider additional panels (perhaps an additional row) for zooming in on rRNA and IS hotspots from Figure 1D (see supplemental data and an even higher genomic resolution version thereof).
- Some of the verbiage may need to be toned down and re-aligned with what the data shows (e.g. "attacks" vs. locates or targets DSBs; or "fights" vs. interferes with or impacts plasmids and phages).
- Data in Figure 4E and 4F seem to be the most important and valuable when assessing the overall narrative, yet they represent merely one third of a quarter of the figures (e.g. approximately one twelfth of the data presented). The authors should consider performing additional experiments (at least more phages) and show time-courses (for both plasmids and phages), and compare with and without selection (for plasmids), and RNAseq and DNAseq data (how does plasmid targeting impact plasmid DNA coverage overall and at targeted loci; how does phage DNA and/or RNA get impacted by CbAgo and even dCbAgo?)
- A subset of supp Fig1 should be dedicated to plasmid targeting and another to phage targeting.
- A zoom in of Supp Fig3 should be used as an additional panel for Figure 1.
- As the authors invoke CRISPR, they may want to consider using the approach of Supp Fig7 and investigate whether there is a pattern of co-occurrence of pAgo and CRISPR-Cas immune systems.

Author Rebuttals to Initial Comments:

We are very grateful to all three reviewers for their valuable criticism and suggestions. Before going to the point-by-point response to their comments, we would like to outline the scope and the major findings of our study. We aimed to understand the functions of CbAgo in cell defense against invader DNA, discover the specificity of DNA targeting by CbAgo *in vivo* and reveal molecular mechanisms underlying this specificity. The **principal results** are summarized below.

1) We show that CbAgo fights selfish genetic elements that invade bacteria (Fig. 4). This is the first report of the role of pAgos in cell defense against phages. The results firmly establish pAgos as a genome defense system in bacteria.

2) We show that CbAgo targets homologous regions and induces degradation of multicopy sequences present in the cell (Fig. 1). This process of 'genuine' DNA interference, not depending on any RNA intermediates, can participate in the whole range of genetic processes in bacteria including elimination of invader and repetitive DNA elements, DNA repair and recombination.

3) We describe the molecular mechanism that generates small guide DNAs used by CbAgo (Fig. 2 and Fig. 5). These guides are produced at the sites of double-strand breaks, depending on the catalytic activity of CbAgo and the cellular RecBCD helicase-nuclease. This mechanism ensures that the guides are enriched in sequences corresponding to foreign genetic elements. Importantly, CRISPR-Cas systems were also shown to rely on RecBCD (AddAB) during spacer acquisition.

4) We show that CbAgo can greatly stimulate DNA degradation at double-strand breaks in the genomic DNA (Fig. 3).

5) We reveal association of catalytically active pAgos with CRISPR-Cas systems in bacterial genomes (Fig. 5), suggesting cooperation between these two systems.

The following new results have been included in the manuscript in response to the reviewers' comments.

1) Effects of CbAgo on phages T7 and P1. CbAgo does not inhibit T7 infection, possibly because this phage encodes several factors that suppress host systems, including an inhibitor of RecBCD. At the same time, CbAgo provides very efficient (up to four orders of magnitude) defense against P1, including prevention of culture lysis, decrease in phage titers and impaired plaque formation.

2) Analysis of the effects of CbAgo on the integrity of plasmid and genomic DNA using unbiased genome sequencing (DNA seq). The data show that the plasmids are evenly targeted by CbAgo, in agreement with the uniform pattern of smDNA processing. At the same time, CbAgo expression has no dramatic effects on the genomic DNA coverage, even at the sites of preferential smDNA processing, suggesting that these sites are efficiently repaired. Thus, analysis of smDNAs associated with CbAgo allows visualization of DSBs that cannot be detected by genomic DNA sequencing. We would like to emphasize that there are currently no conventional methods to map double-strand breaks in the bacterial chromosome with high precision, and CbAgo presents a unique instrument for such analysis.

3) Analysis of the effects of CbAgo on bacterial growth. The data show that mutant strains with inactivated RecBCD function grow slower in the presence of CbAgo, while the growth kinetics of the wild-type strain is almost unaffected by CbAgo. These observations are in full agreement with the proposed role of RecBCD in the repair of DNA breaks formed upon CbAgo expression.

4) Analysis of the time course of plasmid loss in the presence of CbAgo, showing that some plasmids are lost after just a single passage.

5) Analysis of processing of large foreign replicons by CbAgo, in a model strain containing an F' plasmid, a ~200 kb derivative of the conjugation F factor that contains a large genomic insertion. We demonstrate that the F factor is strongly preferred by CbAgo over the chromosome, due to the presence of a large 100 kb region lacking Chi sites, but that smDNA

processing is stopped in the genomic part containing Chi. The results further confirm the proposed model of foreign DNA recognition based on the combined action of CbAgo and the cellular DSB repair machinery.

6) Analysis of the relative distributions of pAgo genes and CRISPR-Cas systems in bacteria and archaea. We have revealed strong association of catalytically active pAgos – but not other pAgo variants – with CRISPR-Cas.

Below, we provide our point-by-point response (shown in blue) to the comments of each reviewer.

Referee #1 (Remarks to the Author):

This manuscript investigates the generation of DNA associated with the Clostridium butyricum Argonaute protein when it is expressed in E. coli. Argonaute proteins, best characterized for their role in eukaryotic RNA interference, are found ubiquitously in archaea and in bacteria, where they interact with small DNA molecules (smDNA). This work examines the role of the nuclease activity of CbAgo and the genomic derivation of the smDNA associated with it. By genetic analysis, they implicate the double-strand break repair proteins in the generation of smDNA. Over-represented in smDNA is DNA from multiple copy number elements, including plasmids and transposons, as well as sites of double-strand breaks. By examining persistence of plasmid DNA and single-strand DNA bacteriophage titers, they show that bacterial Ago can protect cells from extragenomic infectious elements.

All in all, this is an important demonstration of the role of this group of proteins and an intriguing mechanism for smDNA generation. The data are sound and well-presented. I have some issues with some of the language used and there are few things that would improve and clarify the reasoning in the manuscript.

Line 26 and throughout. It took me some time to realize that the protein is being expressed in E. coli and the analysis entirely done in this organism. It should be "bacterial Argonaute nuclease CbAgo expressed in Escherichia coli." This should also be mentioned when introducing the results, for instance line 60 "purified it from E. coli cells".

We have indicated in the revised manuscript that CbAgo was expressed and analyzed in the *E. coli* system. As we also argue in our response to the comments of reviewer 2, this system can be used to investigate the functional properties of pAgo proteins, since CbAgo can efficiently process genomic and invader DNA and is active against foreign genetic elements in *E. coli*. Importantly, previous studies revealed that pAgos spread among non-related bacterial and archaeal species through horizontal gene transfer suggesting that pAgos are naturally adapted to function in 'heterologous' systems. We discuss this in the revised manuscript (second paragraph of Discussion).

Line 28 and throughout. This is misleading. I would not use the term "cooperation" and the generation of smDNA guides does not require double-strand break repair machinery, since smDNA is generated fine in recBCD and recA mutants. Cooperation implies that the processes occur together or are more efficient in concert, neither of which are demonstrated here. The double-strand break machinery merely influences the composition of the guide DNA.

We agree that the generation of smDNA guides per se does not require double-strand break repair machinery. However, the specificity of smDNA guides bound by CbAgo targeting strongly depends on the presence of the functional RecBCD, and RecA. We therefore re-write the abstract to say that "Loading of CbAgo with locus-specific small DNA guides requires cooperation between its intrinsic endonuclease activity and the cellular double-strand break repair machinery." (lines 31-33)

Line 70 "restricted" evokes restriction/modification systems, a better term would be "bounded" or "flanked

We modified this sentence as suggested: "The smDNA hotspots are bounded by the Ter sites from the outside" (lines 71-72)

Line 77 and Fig. 1: It is hard to see the offset of the single peak in tus- strains relative to TerC

Thank you for this note. We have indicated the exact positions of TerA and TerC for all strains in the revised Extended data Fig. 2. The centre of the single peak in the tus- strain is located just on

the left of terC. We also showed strand-specific distributions of smDNAs for all strains shown in this Figure (with most important regions from these plots shown in Fig. 2 and Extended data Fig. 4).

Line 85 and Fig. 1. I would very much like to see a larger scale representation of the spread of smDNA enrichment relative to the boundaries of plasmid homology region, close to lacI or araC. This is an important point and this would emphasize it. The spreading phenomenon is different than the very targeted type 2 R/M or CRISPR

We have included larger scale presentations of both regions in Extended data Fig. 4C, where we indicate the exact boudaries of *lacI* and *araC*. We would like to note that in Fig. 1C the relative size of araC (vertical blue line) is shown to scale, and the whole peak corresponds to the surrounding genomic regions (araC-mapped smDNAs are exluded from the plot, since they map to both chromosomal and plasmid DNA). At first glance, this may seem different from type 2 R/M or CRISPR but it should be noted that no direct experiments on the extent of genomic DNA processing (by cellular nucleases including RecBCD) following site-specific cleavage was performed for these systems, and the actual patterns may be not so different but may be masked by efficient DSB repair by cellular systems. In fact, extensive DNA processing is observed under certain conditions for the highly specific palindrome-dependent cleavage (Hasan et al., 2018), and also for the CRISPR-Cas adaptation – in the latter case, new spacers are inserted with high efficiency from the vast region surrounding the site of I-SceI cleavage (Modell et al., 2017). The use of CbAgo thus allows to visualize DSB processing in unprecedented detail, and can be applied for analysis of DNA targeting by other systems including R/M and CRISPR-Cas (see also our response to the reviewer 3).

Line 89 and throughout: "DNA interference" is not being measured here and it's not clear what "interference between the plasmid and chromosomal loci" means. Do you mean "against"?

We use the term "DNA interference" to indicate that the presence of the two (or more) homologous DNA regions in the same cell triggers processing of these sequences and generation of smDNAs, depending on CbAgo. We also introduce the term DNA interference in the Abstract. Extensive processing of chromosomal loci flanking the sequences with homology to the plasmid strongly suggests that there is DNA interference between these loci, as we argue in the manuscript. So, we would prefer to keep the term "DNA interference" in the manuscript. At the same time, we acknowledge that this process does not necessarily lead to loss of genomic DNA (which would be absolutely detrimental for cell survival). Therefore, we used more accurate wording in this section about DNA interference, e.g. by saying "The presence of plasmids therefore leads to generation of smDNA guides from cognate chromosomal loci and flanking sequences" instead of "CbAgo can therefore induce DNA interference between plasmid and chromosomal regions independently of its genetic location."

Fig. 1 legend: please state that the lower panel in parts b, c, d is dAgo (it's only said in the legend for b)

We have indicated this in the figure legend. We have also improved labelling of the wild-type and dCbAgo in corresponding panels in Fig.1.

Line 115: how sure are we that the mutant is indeed catalytically inactive, since much of the interpretation keys on this? An experimental demonstration would be nice, or at least some reasoning/reference that makes us more confident about this fact.

We thank the reviewer for this note. The properties of the mutant protein were extensively studied in our previous work (Kuzmenko et al., [REDACTED] 2019). It was shown that it is completely inactive and cannot cleave either single-stranded or double-stranded DNA substrates. We have indicated this in the revised manuscript: "dCbAgo (catalytically dead CbAgo) that contains substitutions of two out of four catalytic residues in the active site. Unlike the wild-type protein,

dCbAgo loaded with cognate DNA guides is incapable of cleaving either single-stranded or double-stranded DNA in vitro." (lines 110-113)

Line 132: the enrichment of repetitive/higher copy DNA seems like it would require some amplification step, which maybe you should elaborate on. It would be easy enough to investigate plasmid copy number on smDNA.

Indeed, the observed enrichment of smDNAs with repetitive sequences above the arithmetical copy number implies some amplification step. For chromosomal regions, this enrichment is seen directly from chromosomal smDNA distribution (e.g. Fig. 1 or Extended data Fig. 3). For plasmids, the level of this enrichment can be estimated after normalization of smDNAs by the plasmid copy numbers relative to the chromosome, as suggested by the reviewer and as we did in our study (Fig. 4a).

We discuss a possible mechanism of signal amplification in the model shown in Fig. 5 (formerly Extended data Fig.7) and in the text: "CbAgo induces DNA interference between multicopy sequences present in the cell that likely underlies its functional activity against invading genetic elements… Loading of locus-specific guides into pAgo increases the concentration of effector complexes and induces DNA interference leading to further processing of new smDNAs from the target locus, amplification of guide DNAs and effective destruction of the target"

Line 187: Do you mean persistence of DNA ends in recA-? Expressing gam protein (phage-encoded inhibitors of RecBCD) would be interesting.

We propose that the increased production of smDNAs in the *recA*- strain may indeed result from the persistence of unrepaired DNA ends extensively processed by RecBCD. This results in increased genomic DNA degradation and highly efficient generation of small guide DNAs captured by CbAgo. Indeed, the phenomenon of "reckless DNA degradation" has been known since long ago, and it was shown that mutations in RecBCD can supress the extensive processing of genomic DNA in recA- strains (e.g. Capaldo et al., 1975). We indicate this in the revised manuscript: "Thus, RecA guards the genome from excessive processing by CbAgo and RecBCD, a phenomenon known as 'reckless' DNA degradation due to persistence of DNA ends and their continuous degradation by RecBCD in recA-minus strains" (lines 198-201).

We plan to perform additional experiments to understand the details of RecBCD-dependent DNA processing in bacterial strains expressing pAgos, including analysis of additional mutations in individual components of RecBCD and expression of the Gam protein. While these experiments go beyond the scope of this initial publication, we thank the reviewer for this suggestion.

Line 221: the lower efficiency is hard to see in Fig. 3

The peak of smDNAs at the DSB locus is indeed smaller in the case of catalytically inactive dCbAgo. We indicated the percent of smDNA reads mapping around the DSB (in a fixed area around the DSB) in each strain the revised Fig. 3b.

Line 225: How fast does smDNA turn over? (just interested, not a critique)

We investigated the kinetic parameters for *in vitro* cleavage of ssDNA by CbAgo in our previous publication (Kuzmenko et al., 2019). It was shown that the cleavage half-time was ~17 and ~90 min for single-turnover and multiple-turnover reactions, respectively. Therefore, dissociation of the cleavage products is likely the rate-limiting step in these reactions. We have not measured the rates of guide exchange for CbAgo. The complexes of guide-Ago complexes were shown to be quite stable for other studied pAgos (RsAgo from *Rhodobacter sphaeroides* and MpAgo from *Marinitoga piezophila*), with half-lives of dozen minutes and even hours. However, as we have shown for RsAgo in another publication (Liu, Esyunina et al., [REDACTED] 2018), rapid guide dissociation can be triggered by interactions with mismatched target molecules. So, the actual rates of smDNA exchange *in vivo* may be hard to determine.

Line 242: I wouldn't say the effect is "massive", substantial, maybe

We have replaced "massive" with "substantial" in the revised manuscript.

Line 248: At least for transposons, free ends are rare, transposition rates low. Is this system especially sensitive to conjugal plasmids, since they transmit via single-strand DNA?

In the revised manuscript, we removed 'often' from the corresponding phrase, it now says that "CbAgo may target mobile genetic elements, such transposons, plasmids and phages, which have multicopy nature and form free DNA ends in their life-cycle".

Indeed, conjugation plasmids, and also DNA substrates recognized by natural transformation systems, might be particularly sensitive to pAgos because they form single-stranded intermediates during transmission. Extensive transposition of up-regulated transposon elements may also result in their preferable processing by CbAgo. Other cellular processes in the cell connected to formation of single-stranded regions – including DNA repair and SOS-response under stress conditions – may also produce ssDNA processed by CbAgo. Thus, determination of the whole range of *in vivo* DNA targets recognized by CbAgo may be an important subject for further studies.

During revision, we analyzed generation of smDNA guides from a conjugative F' factor (though in the absence of conjugation). We demonstrated that it is highly preferred by CbAgo over chromosomal DNA. We showed that this is explained by the presence of a large (100 kb) region lacking Chi sites. These results further confirm our proposed model on the joint action of RecBCD and CbAgo in foreign DNA processing. The new data are presented in the Extended data Fig. 10 and described in the text (lines 280-289).

Line 256: It is important to mention here that the preferential use of plasmid sequences occurs even when CbAgo is expressed from the chromosome. I missed this the first time through, and thought that the effect could be in cis from the expression locus

We have indicated this in the revised manuscript and in the Figure legend. All experiments on plasmid loss presented in Fig. 4c and Extended data Fig. 11 were also performed with a genome-encoded CbAgo (see also our response to the comment related to line 132 above).

Line 279: You are not directly measuring substrate processing by the nuclease, this is misleading

We have modified the sentence to say that "foreign DNA – plasmids, transposons and phages – is a preferential substrate for CbAgo and, possibly, other pAgos".

Line 282: again, not necessarily cooperating

We have changed the wording to say "CbAgo relies on the cellular DNA break repair machinery for smDNA loading and targets vulnerable DNA regions containing breaks and free DNA ends that are formed during DNA replication, repair, and interference."

Yet, we would like to note that the joint action of RecBCD and CbAgo may be considered as cooperation since RecBCD determines the specific loci of preferential smDNA processing (depending on the presence of Chi sites), and RecB/RecD/RecA are strongly required for DNA interference between homologous loci. On the other hand, the catalytic activity of CbAgo likely increases the intensity of RecBCD-dependent DNA processing at the sites of engineered DSBs, since more smDNAs are loaded in the Chi-dependent way in the case of active CbAgo in comparison with dCbAgo, resulting in a deeper decrease in the genomic DNA content at the site of DSB (Fig. 3b and 3c).

Line 289: not clear to me why smaller replicons are more vulnerable. The lack of Chi sites could have the opposite effect, by promoting complete degradation. Is there any evidence that ssDNA is more vulnerable to these nucleases? Interesting because you saw the effect on ssDNA phage: would you see this for lambda?

We propose that smaller replicons may be more vulnerable because they can be completely degraded by RecBCD and other nucleases after DSB formation during replication, while much longer chromosomal DNA can be repaired via homologous recombination. And this process can indeed be greatly influenced by the presence of Chi sites, which are often lacking in foreign DNA. We indicate this in the revised manuscript: "…the ability of pAgo to target double-strand breaks and replication intermediates, which can form in multiple copies of plasmids and phages; such small replicons can then be completely degraded by cellular nucleases before their repair;" (lines 344-346).

In vitro, ssDNA is indeed a preferred substrate for CbAgo and several other tested pAgo nucleases (e.g. Kuzmenko et al., 2019, Swarts et al., 2014). Our results presented in the current manuscript also suggest that CbAgo may preferentially act on specific intermediates formed during DNA replication (since there is a preference for the lagging DNA strand) and DNA repair (since more guide DNAs are produced from the 3'-terminated strand at DSBs). However, both DNA strands are quite efficiently targeted on the whole-genomic level, indicating that there may be no specific requirement that the DNA targets should be single-stranded by their nature – probably because any DNA becomes single-stranded at a certain stage during replication or repair.

We have not performed any experiments with phage lambda but we shall consider it for testing in further studies. For lambda, these experiments might be especially tricky because it encodes a RecBCD inhibitor (Gam) and its own recombination system but lacks Chi-sites (which makes infection very inefficient in the absence of Gam). So, the experiments may require the use of Chi-containing lambda lacking Gam. At the same time, we have tested two additional phages, T7 and P1. While CbAgo had no effect on the infection with T7, it strongly inhibited replication of phage P1 (even more efficiently than for M13). We included these results in the revised manuscript (Fig. 4F-G and corresponding section in the text). So, CbAgo can target both single-stranded and double-stranded phages.

Line 314: a potential universal defense system, one example does not make a universe

We replaced it with "as an ancient defence system" to avoid confusion. Our meaning is that Ago proteins can govern guide-dependent genome defence in all three domains of life, but of course it is not true for every species and depends on individual genomic features and the presence of Ago proteins themselves.

Line 547: does "negative colonies" mean plaques?

We replaced "negative colonies" with "phage plaques".

Fig. 4E: I'm a bit worried that potential toxicity of the expressed protein if it is plasmid-encoded. Is CbAgo expressed from the genome here? That would be the best experiment.

CbAgo was expressed from the genome in all experiments shown in Fig. 4. We indicate this in the figure legend. Importantly, the preference of CbAgo for plasmids is observed independently of whether it is expressed from plasmid or genomic locations – see the data for smDNA distribution from Fig. 4A and extended data Fig. 9. We have clearly indicated this in the revised manuscript.

Figure extended 7: what does "long-A" etc refer to?

We thank the reviewer for this question. We have explained designations of different groups of pAgos in the revised manuscript (both in the Fig. 5 legend in Discussion and in the extended Fig. 12, former Fig. 7). Long-A is one of three major phylogenetic groups of pAgos revealed in our previous analysis (Ryazansky et al., 2018). This, and only this, group mainly contains catalytically active pAgos. During revision, we also performed additional analysis of the distribution of CRISPR-Cas systems in prokaryotic genomes and found that they are specifically enriched in bacteria containing long-A pAgos, which implies their functional interaction (see also our response to Reviewer 3).

Referee #2 (Remarks to the Author):

Kuzmenko et al. explore the mechanism of guide generation for prokaryotic Argounaut nucleases. This nucleases have been identified bioinformatically about 15 years ago, but both their mechanism of function and biological significance is still relatively unknown. In the first part of the manuscript (Figs. 1 and 2) it is clearly demonstrated that CbAgo is preferentially loaded with guides from DSB sites as well as DNA sequences that are present in multiple copies within the cell. In both cases the DSB repair machinery is involved, as guide hot-spots are delimited by chi sites and largely eliminated in the absence of the RecBCD machinery. This is an important results that significantly advances our understanding of pAgo mechanism.

The second part of the paper explores the biological significance of these findings. Here the data shown is quite limited. Figure 3 shows that sequences from which the CbAgo guides are extracted from are subject to CbAgo cleavage, which is a somewhat expected result given the well known DNA-guided nuclease activity of this enzyme. In my opinion, the main issue to investigate is the consequence of this self-targeting for the bacterial population.

We thank the reviewer for these comments. The main conclusion that we derive from the experiments in Fig. 3 is that CbAgo is actively involved in processing of double-strand breaks, which explains the mechanism of extensive DNA degradation by the combined action of CbAgo and RecBCD. In addition, the ability of pAgos loaded with specific guide DNAs to cleave corresponding DNA targets *in vivo* has long been implied but was never demonstrated previously. So, these data provide the first direct evidence that pAgos can target specific genomic loci *in vivo* and intensify DNA degradation at these regions, probably through repetitive rounds of DNA breakage and repair.

We describe the problem of self-targeting, in comparison with CRISPR-Cas, in a separate paragraph in the revised manuscript (4th para in section 'Role of the DSB repair machinery') (lines 202-213). In the case of genomic DNA, even persistent breaks can be effectively repaired by the cellular DSB repair systems (as demonstrated, for example, for palindrome-induced DSBs, Fig. 3). Indeed, genomic DNA sequencing for strains with functional DSBR demonstrates that expression of CbAgo does not decreased genomic DNA coverage in the target regions (Extended data Fig. 6). Accordingly, this self-targeting has only minor effect on bacterial growth in the case of wild-type strains, but the growth defect can indeed be observed in strains with defects in the DSB repair system. We show corresponding growth curves for the wild-type and mutant strains in the revised manuscript (Extended data Fig. 5).

And Fig. 4 attempts to demonstrate that CbAgo provides defense against mobile genetic elements. While the data is convincing for plasmids, the results for M13 infections are a confusing, as the nuclease dead version of CbAgo still provides a significant level of protection.

Therefore, the paper will benefit from more experiments to clarify this second part:

In the revised manuscript, we provide additional data for plasmids, including more plasmid variants with different number of passages (Extended data Fig. 11) and analysis of plasmid DNA coverage depending on the expression of CbAgo (Fig. 4 and Extended data Fig. 9).

For M13, the nuclease dead version of CbAgo still provides some level of defence but it is significantly lower than in the case of active CbAgo (in comparison, no effect of dCbAgo is seen in the case of plasmids). This can possibly be explained by the ability of dCbAgo to bind single-stranded M13 DNA, when loaded with corresponding guides. Indeed, we observed loading of guide DNAs into dCbAgo in our smDNA sequencing experiments (e.g. Fig. 1B, 2B, 3B). In the case of M13, simple DNA binding by dCbAgo-guide complexes may slow down phage replication. Importantly, more than two thirds of all pAgo proteins are predicted to lack catalytic activity, because of substitutions in the catalytic center in the PIWI domain (Ryazansky et al., 2018). Thus, the demonstration that a catalytically inactive pAgo variant can have some inhibitory effects on

phage infection is important for further understanding of the functions of various groups of pAgos in prokaryotic cells. We indicate this in the revised manuscript: "It is plausible that dCbAgo might provide modest protection against M13 by guide-dependent binding to its single-stranded DNA even if it cannot cleave it. Notably, many pAgos lack endonuclease activity due to natural substitutions in their active sites1-3, suggesting that cleavage of target DNA is not essential for their function" (lines 326-329).

Furthermore, we have performed experiments with two more phages, T7 and P1, and demonstrated that the P1 infection is strongly inhibited by CbAgo, while catalytically inactive dCbAgo has no effect (see below).

Overall, the results with both plasmids and phages support the proposed model of CbAgo action (Fig. 5). Below, we describe specific experiments, including additional data, that allow us to make firm conclusions about the biological role of CbAgo.

- What portion of the cells in the culture acquire self-targeting guides? Is it possible to do single cell DNA seq to find out?

This is very difficult to address experimentally, especially for smDNA guides associated with CbAgo since this would require analysis of protein-DNA complexes in single cells. However, given that CbAgo lacks strict replicon-specificity (similarly to CRISPR-Cas systems during the adaptation step) it is highly likely that self-targeting occurs in most cells. For the specific case of induced DSBs, significant degradation of genomic DNA in the presence of CbAgo is clearly visible (as a decreased DNA coverage upon expression of CbAgo, Fig. 3c), suggesting that most cells are indeed targeted. Furthermore, CbAgo expression slows down bacterial cell growth in strains with mutations in DSB repair (Extended data Fig. 5 in the revised manuscript) – this effect may be explained by genomic targeting by CbAgo (see next comment).

- Is self-targeting toxic to the host? A competition assay of cells with and without CbAgo should answer this.

In the revised manuscript, we present growth curves for bacterial strains with and without CbAgo, performed under the same conditions as experiments on smDNA acquisition by CbAgo (Extended data Fig.5). We show that expression of CbAgo results in a certain delay in cell growth in comparison with corresponding control strain (panel A). This effect becomes stronger in a minimal medium (panel B) and in the case of strains with defects in the DSB repair system (*recA*- and *recB-recD*-) but not in the strain without Tus protein (*tus*-). Therefore, self-targeting by CbAgo may indeed be toxic to the host, especially when DSB repair is disrupted ('Role of the DSB repair machinery', 4th paragraph).

- Does self-targeting results in DNA repair of the break? Is this repair mutagenic in a way that it could eliminate, say, the I-SceI site to prevent further cleavage and generate "escapers" of CbAgo toxicity (if there is such toxicity)?

Our data suggest that self-targeting indeed results in DNA repair of the break. In the revised manuscript, we present data on genomic sequencing of bacterial strains expressing CbAgo, which reveal no decrease in DNA coverage at the sites targeted by CbAgo, including the *araC* locus strongly preferred by CbAgo (Extended data Fig. 6). Whether this repair is mutagenic may be an interesting point for further research, which likely goes beyond the scope of this manuscript.

- M13 is in many ways a very special phage. Does CbAgo provide protection against the most common E. coli phages such as lambda, T7 and P1?

M13 is a good model to study the effects of 'naïve' DNA targeting by pAgos since it does not encode specific inhibitors of cellular defence systems. It is indeed unusual in that it is a 'chronic' phage without lytic cycle, which makes it somewhat similar to plasmid targets. We indicate this in

the revised text: "expression of CbAgo protected bacteria from phage M13, a chronic phage with a circular ssDNA genome" (lines 304-306).

In the revised manuscript, we provide additional data for phages T7 and P1. We show that CbAgo does not provide protection against T7 (Extended data Table 1). This can possibly be explained by the action of inhibitory proteins encoded in the T7 genome. Such proteins include gp5.9 that inhibits RecBCD (Pacumbaba and Center, 1975; Bobay et al., 2013) and Ocr that inhibits various DNA binding proteins (Walkinshaw et al., [REDACTED] 2002; Isaev et al., [REDACTED] 2020). In future studies, it would be interesting to test the effects of these and other phage proteins on pAgos.

At the same time, CbAgo provided strong protection against P1: it prevented lysis of infected bacterial culture, dramatically decreased phage titers and prevented plaque formation in a plating assay (Fig. 4E-G). We have described all results obtained with various phages in a separate section in the revised manuscript ('CbAgo provides protection against phages').

We also indicate in Discussion that many phages could counteract the action of pAgos and other defense systems by encoding inhibitors of RecBCD (such as phage lambda Gam protein), which allow them to replicate despite the absence of Chi sites, while encoding their own recombination systems. Indeed, recent analysis of new bacterial defence systems with a panel of phages showed that most systems are active against only a subset of phages, and only one of four tested systems had considerable activity against T7 while two were completely inactive (Doron et al., [REDACTED] 2018). We indicate this in the revised manuscript: "Many phages, such as T7 or lambda, encode inhibitors of RecBCD which may thus help them to cope with multiple defense systems in host bacteria" (lines 386-388).

Overall, the experiments with phages M13 and P1 firmly demonstrate that CbAgo can inhibit multiplication of at least two different types of phages. Further experiments are definitely warranted to establish the spectrum of antiphage activities for various pAgo proteins, and we plan to perform them in our future studies.

- Direct evidence of the presence of M13-derived guides in CbAgo following infection should be provided.

We present this evidence in the Extended data Fig. 9c. We have isolated and sequenced CbAgo-associated smDNAs in a strain infected with M13 and found smDNAs derived from both strands of the M13 genome – the whole M13 genome is covered (with a slight preference for the minus strand). The results suggest that M13 is targeted during its replication within cells.

Other concerns:
- It seems that the extraction of guides from homologous sequences is related to their recombination. It is highly likely that recombination of these identical sequences is resolved generating some degree of DSBs that are used as a source of guides. This process will involve RecBCD also. In the present text, the reason behind the enhanced guide abundance from duplicated/repetitive sequences is largely left undiscussed.

At present, we have no data suggesting that generation of guide DNAs from homologous sequences may be related to their recombination or, more specifically, that recombination may be the main driver of this process, as proposed. As mentioned above, similarly efficient generation of guide DNAs is observed for various plasmids independently of the presence of regions of homology with the genome (Fig. 4a and Extended data Fig. 9A,B). We indicate this in the revised text: "SmDNA guides were strongly enriched in plasmid sequences independently of whether CbAgo itself was expressed from plasmid or chromosome and whether the plasmid had any homology to chromosomal DNA or not" (lines 267-269). Furthermore, the *araC* and rDNA peaks are well observed even when the most common pathway of recombination is disrupted (in the *recC*-strain).

Overall, our data suggest that plasmid DNA, and not recombination intermediates, is the primary source of guide DNAs during DNA interference between plasmid and chromosomal DNA. However, a possible interplay between DSB formation and recombination during DNA interference certainly deserves further analysis, as we indicate in the revised manuscript: "CbAgo-dependent DNA processing may potentially play roles in other genomic processes such as elimination of repetitive genetic elements, DNA repair and recombination, response to DNA damage and suicidal systems" (lines 367-369).

We explain a likely mechanism leading to the observed targeting of repetitive sequences in Discussion (first paragraph, Fig. 5A). The presence of multiple copies of the same sequence increases the load of guide DNAs corresponding to this sequence into CbAgo, and thus increases the effective concentration of locus-specific guide-CbAgo complexes and the likelihood of repeated cleavage of the same target sequence, leading to amplification of the signal.

- Define Cb in the abstract.

We have indicated the species name for CbAgo (*Clostridium butyricum*) in the abstract.

- Line 60. The authors need to explicitly mention that CbAgo is expressed in E. coli. Since this is not the natural host for this nuclease, the discussion should address why the authors think that the results of their study are most likely not affected by this experimental set up.

We have explicitly mentioned that CbAgo was investigated in *E. coli* in the revised manuscript, and discussed why pAgo proteins can likely function in heterologous systems. Previous analyses performed by us (Ryazansky et al., 2018) and others (Swarts et al., 2014, Makarova et al. 2009) demonstrated that pAgos are spread between prokaryotic species via horizontal gene transfer, with no correlation between their phylogeny and their host species. Furthermore, pAgos from various parts of their phylogenetic tree can be associated with various types of cellular recombination systems (Extended data Fig. 12) suggesting that their activity does not strictly depend on species-specific cellular cofactors. Indeed, CbAgo is active against plasmids and phages in *E. coli*, can be efficiently loaded with guide DNA *in vivo* and can process invader and chromosomal DNA. A similar approach with the use of convenient heterological hosts was previously used for analysis of other defence systems encoded by various prokaryotic species (e.g. Doron et al., [REDACTED] 2018; Ofir et al., [REDACTED] 2018), as we indicate in the manuscript: "Similarly, other bacterial defence systems were shown to be active in heterologous systems" (lines 359-360).

- Lines 88-89: it is impossible to know the order of events from these results. The first guide could come from the chromosome and lead to the cleavage of the plasmid, which will then begin a positive feedback loop of cleavage and guide extraction at both loci.

This scenario seems very unlikely, since plasmid DNA is present at a higher copy number, is processed with much higher efficiency than the chromosome, and is evenly covered with smDNAs, without strong peaks at *araC*, *lacI* or any other sites (Fig. 4a, Extended data Fig. 9). Furthermore, such preference is observed for various plasmids independently of the presence of regions of homology to the chromosome, and plasmids with and without such regions of homology are similarly eliminated from the cells containing CbAgo.

- Fig. 2b is reference before 2a. Fig. 4e is before 4d.

We have corrected the order of references (please note that Fig. 4 was rearranged significantly).

- Lines 177-178: the functions of RecB, C and D should be briefly explained (helicase/nuclease for B and D; chi recognition for C). Can the results be explained by these different functions? The conclusion that "different components… have independent roles" is not very useful for the reader. It should also be explained that in the absence of RecB and D the complex will not form and the results of the mutant strain lacking these components are not a reflection of RecC activity, but of a

situation where there is no RecBCD complex (that is not the case in the absence of RecC, where RecBD can still recognize the DSB and degrade it).

We have explained the functions of individual subunits of RecBCD in the revised manuscript: "The RecBCD complex (RecB, 3'-5' helicase and nuclease; RecD, 5'-3' helicase; RecC, Chi site recognition) recognizes Chi in the 3'-terminated strand during DSB processing" (lines 147-149). We have also indicated that knockouts of both *recB*/*recD* and *recC* disrupt the whole RecBCD complex. *RecC* mutants are as defective in the RecBCD activity (including Chi site recognition and nuclease activity) as *recB* (Ref. 23, Chaudhury and Smith, 1984), so RecB and RecD cannot degrade DSBs. However, they still assist CbAgo in DNA interference, possibly by unwinding dsDNA substrates.

On this topic, have the authors tested mutants in the active sites of RecB and/or RecD? These mutations allow more accurate conclusions as they allow complex formation, just without different activities.

We thank the reviewer for this note. We have not tested such mutants yet this may be an interesting point for further studies.

- Line 180. The "single peak is hard to see. A zoomed-in panel could help.

The single peak corresponds to the whole ter region between 1 and 2 Mb, with its highest point located close to dif/TerC. We added the reference to the Extended data Fig. 2, in which this peak is better visible for the *recBrecD*- and *recC*- strains in the smaller scale whole-genome diagrams.

- Line 209. Which host? Not E. coli.

The host is indeed *E. coli*, and the SbcCD nuclease (the homolog of eukaryotic Mre11-Rad50) is normally involved in processing of noncanonical DNA structures and DNA repair. We indicate this in the revised text.

- Line 279-280: it is not clear that the guides are preferentially acquired from foreign DNA. Fig. 4a shows that there are 10-fold more guides from the plasmid than what would be predicted by chance. But this is probably true also of the ter sites in the chromosome. Did the authors compare the total amount of reads for chromosomal vs plasmid guides? Are these different when normalized by genome size and by copy number? The only comparison I can make from the presented data is in Fig. 1a, where many of the araC reads must come from the plasmid but still the terC peak is much higher.

The data in Fig. 4a are for the total read counts mapping to the chromosomal and plasmid DNA, and they were already normalized by the genome/plasmid size and by plasmid copy numbers (we indicate this in the Figure legend and Materials and Methods). So, there is a clear enrichment for plasmid-derived smDNAs even including the *ter* and *araC* peaks in the chromosome. Calculations for genomic sequences excluding *araC* and *ter* reveal an even higher enrichment for plasmid DNA, from 4.5 to 18-fold.

We would like to note that all smDNA reads corresponding to the *araC* locus visible in Fig. 1 come from genomic DNA, not the plasmid. In this figure, the *araC* gene is much smaller than the peak width (the vertical blue line in Fig. 1C corresponds to the actual length of *araC*), and only unique sequences were used for mapping, excluding multimappers (such as *araC* which is present both in the plasmid and in the genome). We indicate this in the figure legend: "(d) SmDNA peak in the *ara* locus (only unique chromosomal reads are shown)." The enlarged peaks around *araC* and *lacI* are shown in Fig. 2D and Extended data Fig. 4B; in these panels, the size of the gene in the center of the peaks can be clearly visible.

- Line 288, point (ii). It is not clear to me why replication intermediates are "located in smaller … replicons". Every replicon regardless of its size has replication intermediates. Also, it may be better to clarify if the authors are referring to replication intermediates that lead to DSBs.

We thank the reviewer for this note. Two factors may contribute to the increased processing of plasmids and phages: (1) the higher number of replication intermediates because of the multicopy nature of these replicons; (2) the small size of these replicons, which together with the absence of Chi sites makes them particularly vulnerable to RecBCD and other nucleases. We are not referring exclusively to DSBs as replication intermediates because our data also show enrichment of smDNAs for the lagging DNA strand in both replichores (Fig. 2F). We have clarified this point in the revised manuscript: "Preferential targeting of foreign DNA by CbAgo can be explained by combination of several mechanisms: … (ii) the ability of pAgo to target double-strand breaks and replication intermediates, which can form in multiple copies of plasmids and phages; such small replicons can then be completely degraded by cellular nucleases before their repair" (lines 341-350).

During revision, we also analyzed smDNA processing for a much larger replicon, an F' factor containing a large genomic insert (~100 kb) in addition to the native sequence of the F factor (also ~100 kb) (lines 280-289, Extended data Fig. 10). In comparison with multi-copy plasmids, it is maintained as a single copy episome. Nevertheless, it is highly preferred over the chromosome (>10-fold smDNA enrichment after normalization by the genome size). Importantly, smDNAs are preferentially generated from the F factor sequence lacking Chi sites, and their numbers drop several fold for the chromosomally-derived part of this replicon, upon encountering Chi-sites. These results further support our proposed model of the recognition of foreign replicons, depending on their multicopy nature and the absence of Chi sites.

- Line 290, point (iii). While it is true that phages lack chi sites, they all have RecBCD inhibitors (lambda gamma gene from the Red system for example). This should be discussed, as it could be an evolutionary response of phages to prevent pAgos from obtaining guides from their genomes.

We thank the reviewer for this note. We mentioned the 'arms race' between phages and host defense systems in the revised manuscript: "Many phages, such as T7 or lambda, encode inhibitors of RecBCD which may thus help them to cope with multiple defence systems in host bacteria" (lines 386-388). Since RecBCD or AddAB also play important roles in spacer acquisition by CRISPR-Cas systems (and in the action of restriction/modification systems), inhibition of RecBCD may be a general mechanism to counteract various defence systems by phages.

- Fig. 2a: label each RecBCD subunit with its letter.

We have labelled all RecBCD subunits in the revised figure.

- Fig. 3. The title is confusing. CbAgo does not attack DSBs; guides are taken from DSB sites during their repair by the RecBCD machinery and then, once the DSB is fixed, cleaved again by the CbAgo molecules that where loaded with those guides. Also, if this is the model, it should be also included in Extended Fig. 6.

In according to the reviewer's suggestion, we changed the title to "CbAgo targets DSBs". We also made changes to Fig. 5 with the model (formed Extended Fig. 6) to emphasize that RecBCD and CbAgo are involved in DNA processing.

Referee #3 (Remarks to the Author):

Kuzmenko et al. present a study of prokaryotic Argonaute (pAgo) and study DNA interference in a bacterium, with targeting of chromosomal, plasmid and phage DNA. Specifically, using genetics and deep-coverage DNA sequencing techniques, the authors map the Clostridium butyricum pAgo guide DNAs on genomic areas of interest and implicate ter sites, DSBs and DNA repair machinery in pAgo targeting. Their primary conclusion is that pAgo targeting protects bacteria from invasive DNA such as plasmids and phages. This is a generally timely and interesting topic.

The key conclusion of the manuscript, as highlighted in the title, abstract and throughout the manuscript is that pAgo-mediated DNA interference provides protection against invasive DNA, notably phage, plasmid and IS elements. Yet, the large majority of the work performed focuses on chromosomal self-targeting and seems to indicate that there is broad and non-toxic targeting of various chromosomal locations, especially at ter, rRNA operons and IS elements loci. If the DNA targeting is occurring and efficient and does enable phage DNA and plasmid DNA and thus chromosomal DNA interference, and if DNA cleavage is the outcome then it must be shown, characterized, quantitatively measured and mechanistically and functionally explained. Putatively, the nuclease-enabled pAgo cleaves phage and plasmid DNA, leading to protection against invasive foreign genetic elements, so this cleavage activity must be shown and documented on phage and plasmid DNA.

We thank the reviewer for raising this discussion. Most issues raised by the reviewer can be answered based on the data present in the revised manuscript (including new results on genomic DNA sequencing). First of all, we would like to outline the logic of experiments that has led us to our conclusions (we have also made necessary modifications in the manuscript text):

- CbAgo is an active DNA nuclease; its activity towards single-stranded and double-stranded DNA has been well documented *in vitro* (Kuzmenko et al., 2019; Hegge et al., 2019) (previously published data; lines 52-54 in Introduction);

- CbAgo is loaded with guide DNAs corresponding to specific genomic regions (Fig. 1) and invader DNA (plasmid and phage) (Extended data Fig. 9);

- this loading results from DNA cleavage, since small DNA fragments in the cell can be produced only by cleavage, and since their site-specific production is dependent on both the catalytic activity of CbAgo and on RecBCD (Fig. 1 and Fig. 2);

- smDNA processing clearly involves formation of double-strand breaks in the case of both genomic and plasmid DNAs; this is the key conclusion that follows from the pattern of smDNA processing in the presence of RecBCD and Chi sites (Fig. 2 for the chromosome, Fig. 4B for plasmids, Extended data Fig. 10 for the F' factor): only in the case of DSB formation there can be any dependence of smDNA distribution on Chi sites since they are only recognized during processing of DSBs by RecBCD;

- when the cleavage site is localised in a specific genomic region (with engineered DSBs), CbAgo significantly amplifies DNA degradation in this region, as seen from both a huge increase in smDNA production and a significant loss of genomic DNA at the site of DSB (Fig. 3B and C);

- for most genomic loci, the cleavage sites are efficiently repaired (genomic DNA sequencing data, Extended data Fig. 6);

- in contrast, foreign DNA is extensively degraded due to the high copy numbers and the absence of Chi sites (resulting in signal amplification, as depicted in Fig. 5A); plasmid and phage DNA degradation is uniform, as seen from both smDNA and genomic DNA sequencing data (Extended data Fig. 9);

- as a result, plasmids are eliminated (Fig. 4C, Extended data Fig. 11) and phage multiplication is inhibited (Fig. 4E-G);

Currently, there are no commonly used techniques for the highly sensitive detection of DSBs in prokaryotic cells. The two methods available to date – RecA-ChIP and marker frequency analysis (HTS of genomic DNA) - suffer from low sensitivity and their resolution is limited to dozens kilobases. In comparison, analysis of smDNAs associated with CbAgo allows much more precise detection of DSBs, as illustrated by comparison of panels B (for smDNAs) and C (for genomic DNA) for engineered DSBs in Fig. 3. Furthermore, since the catalytically dead version of CbAgo (dCbAgo) is still loaded with smDNAs in vivo, this allows to analyze the ongoing DSB processing in the cell unperturbed by the catalytic activity of CbAgo (this does not mean though that this activity is not required for DNA interference and anti-plasmid/phage defence, see below).

CbAgo thus gives a much-needed instrument for analysis of the formation and repair of DSBs in bacterial cells. We anticipate that it can be used in numerous studies of DNA processing in both prokaryotic and eukaryotic systems. We included this note in the revised manuscript (end of third paragraph in Discussion): "Highly specific targeting of DSBs by pAgos can potentially be used for genomics applications, in particular as an instrument to study the genome architecture and DNA processing in both prokaryotic and eukaryotic systems."

The conclusion that pAgo protects bacteria from invaders DNA is not fully documented nor supported by the data provided. For instance, the authors claim to have "demonstrated that it induces cleavage of multicopy genetic elements, including plasmids, transposons and repetitive chromosomal loci" and that "the mechanism … endows CbAgo with capacity to eliminate plasmid DNA and fight phage infection", but neither is proven since no DNA cleavage data is provided. Likewise, it is not clear that the "results show that foreign DNA … is a preferential substrate for processing by CbAgo".

As we argue above and indicate in the manuscript, the observed dependence of smDNA production on both the catalytic activity of CbAgo and RecBCD/Chi sites strongly suggests that double-stranded breaks are indeed formed in the target regions, and they can be visualized by analysis of smDNAs associated with CbAgo: "The observed dependence of smDNA production on Chi sites and RecBCD strongly implies the formation of DSBs in the chromosomal target regions during DNA interference" (lines 202-203). In the revised manuscript, we also show data on genomic and plasmid DNA sequencing in bacterial strains expressing CbAgo (Extended data Fig. 6 and 9). As can be seen, this method shows some decrease in the plasmid DNA coverage (even in the presence of selective antibiotics) but does not reveal any plasmid regions specifically targeted by CbAgo (plasmid DNA is uniformly covered with sequencing reads). This agrees well with the observed distribution of smDNAs along the plasmid sequences (Extended data Fig. 9).

Furthermore, the preferences of CbAgo toward certain genomic regions or foreign replicons can clearly be seen from the enrichment of smDNAs associated with CbAgo with sequences corresponding to these regions. In particular, we detect at least 10-fold preference for multicopy plasmids over the chromosome (after normalization for the length and copy number) (Fig. 4A) and a single-copy F' factor (Extended data Fig. 10).

The authors repeatedly show that CbAgo smDNA occur throughout the chromosome, with deep(er) coverage at ter sites. Is there any indication that DNA cleavage occurs in the chromosome, or not? When targeting IS elements (including two families, namely IS1 and IS3 with copies throughout the genome), is there any indication that DNA is actually cleaved?

As we argue above, the dependence of smDNA distribution on Chi sites and RecBCD makes it clear that they are produced from double-strand breaks. We explain this in more detail in the revised manuscript. In the revised manuscript, we show the patterns of DNA processing not only around ter sites, but also around other loci targeted by CbAgo, araC, rDNA and IS-elements. In all cases, the distribution of smDNAs is clearly dependent on Chi sites (Fig. 2B and Extended data Fig. 4).

The authors do show that the catalytically inactive "deactivated" dCbAgo has loss of smDNA biogenesis, but it is still "active" from a functional standpoint (e.g. provides plasmid loss and phage titer reduction, and can retain the ability to target DSBs, albeit with lower efficiency). The authors must discuss whether DNA cleavage is the driver of the process, or not. If DNA is driving the observed effect, then the authors must explain how the deactivated version dCbAgo provides interference in the absence of cleavage.

We have explained this point better in the revised manuscript. Note, that the inactive dCbAgo actually does not cause plasmid loss and the strain with its expression is indistinguishable from a control strain without CbAgo in this assay (Fig. 4C). For DSBs, dCbAgo is loaded with smDNAs, which are likely produced by RecBCD, but its expression does not increase DNA degradation around the break, in contrast to the active CbAgo (Fig. 3C).

In the revised manuscript, we include new results showing that CbAgo also provides strong defense against dsDNA phage P1 (new section 'CbAgo provides protection against phages', Fig. 4E-G). In contrast, dCbAgo is almost inactive (there is only a week protective effect at low MOI), suggesting that the catalytic activity of CbAgo is essential for protection.

For phage M13, we observe a protective effect of dCbAgo, but it is significantly lower than in the case of active CbAgo. We speculate that this may be explained by the single-stranded nature of the M13 genome, so even the catalytically inactive dCbAgo can bind to it and interfere with replication. Similarly, a naturally inactive RsAgo from *Rhodobacter sphaeroides* can bind its target DNA *in vivo*. Importantly, many other pAgos are catalytically inactive because of substitutions in their active sites, and this observation suggests that they may still provide defence against some invader DNAs: "In contrast, dCbAgo had a smaller than wild-type CbAgo but still significant effect on infection with phage M13 (Fig. 4d). It is plausible that dCbAgo might provide modest protection against M13 by guide-dependent binding to its single-stranded DNA even if it cannot cleave it. Notably, many pAgos lack endonuclease activity due to natural substitutions in their active sites1-3, suggesting that cleavage of target DNA is not essential for their function." (lines 324-329).

Mechanistically, clearance of plasmid by loss over passages in the absence of selective pressure is not the same as cleavage of plasmid following transformation under selective pressure. To claim plasmid interference, the latter should be determined. The former does show "quantitative loss of plasmid" over time in some conditions, but this may not constitute plasmid "attack" per se.

In the manuscript, we provide data on small DNA and genomic DNA sequencing for plasmids under selective pressure. We show that plasmid DNA coverage copy number is slightly decreased (~1.2-1.5-fold) in the presence of CbAgo (Extended data Fig. 9, stationary phase). Furthermore, efficient loading of smDNAs into CbAgo is observed for plasmids under these conditions (Extended data Fig. 9A,B), resulting in DNA interference with chromosomal regions (araC or lacI), depending on the catalytic activity of CbAgo (Fig. 1 and Extended data Fig. 2), suggesting that the plasmids are actively degraded. However, it is difficult to estimate the actual effect of CbAgo on plasmid elimination in the presence of selective antibiotics (which kill the cell if the plasmid is lost). This is why we performed experiments on plasmid elimination in the absence of selective pressure: "The experiment was performed in the absence of selection for plasmid genes, to allow elimination of plasmids" (lines 291-292).

Likewise, reduction of phage titer by CbAgo is certainly noteworthy, but this was performed over a short timeframe, and the impact of the "active" CbAgo is better but in some ways comparable to that of the "inactive" dCbAgo, begging the question as to what is exactly happening, and that the actual impact is. Phage interference should be measured at the DNA level, and possibly at the RNA level too. Critically, phage biology encompassing plaque size, burst size, and efficiency of plaquing should be investigated and quantitatively determined. Multiple phages and various conditions over more
extended periods of time should be assessed (at least 16-24 hrs). The authors likely must

determine what happens to phage DNA and phage RNA over time when targeted by both CbAgo and dCbAgo, given the relative effect of both on phage titer (Figure 4F).

In the revised manuscript, we provide additional data for phages T7 and P1. For T7, no strong effects of CbAgo on phage titer could be detected, possibly because T7 encodes an inhibitor of RecBCD (gp5.9) and an Ocr protein that can inhibit various DNA-targeting systems. For P1, we observed strong protection against infection. As suggested, we performed experiments with different multiplicities of infection and for different time points (up to 8 hours of infection) - similar times are used in other studies of antiphage defence systems (e.g. DISARM - Ofir et al., 2018 [REDACTED]; BREX - Gordeeva et al., 2018 [REDACTED]). It was observed that CbAgo completely prevents lysis of bacterial culture at low MOI (0.1), supports steady culture growth at intermediate MOI (1) and significantly delays culture lysis at high MOI (5). In comparison, cells without CbAgo or cells with inactive dCbAgo are highly sensitive to P1 at all tested MOI (Fig. 4E,F). P1 also could not form normal plaques on the strain with expression of wild-type CbAgo (Fig. 4G).

We note that the small DNA and genomic DNA sequencing experiments demonstrate that the whole sequences of plasmid and phage DNA are uniformly covered, with the exception of the F' factor containing Chi sites that change the pattern of DNA processing. This suggests that gene-specific effects of CbAgo transcription (i.e. on the RNA level) are highly unlikely. Therefore, the RNA seq experiments could add little to our understanding of the mechanisms of DNA targeting by CbAgo. It may be interesting to uncover potential effects of CbAgo on gene expression on the whole genome level, however, these experiments seem to go beyond the scope of this study.

Clearly more experiments with different pAgos and different phages can reveal the whole spectrum of pAgo specificities toward different groups of foreign DNA elements. However, the main point of this study is to demonstrate the phenomenon of DNA interference *in vivo* and determine the primary mechanism of specific targeting of foreign DNA by CbAgo. We plan to investigate this phenomenon in more detail in our future studies.

Unfortunately, the authors extensively conclude with a "remarkable" comparison between pAgo an CRISPR-Cas immune systems, and make a series of misguided statements about their mechanistic commonalities and the potential to repurpose pAgos as molecular machines to manipulate genomes. Invoking the DNA repair machinery in the context of Cas-mediated interference is not appropriate, nor is the mention of "similar principles for differentiation between self and nonself" given the conspicuous absence of a PAM. Concluding with a hyperbolic comparison to the most impactful and disruptive technological development in two decades should be re-considered. This is particularly important in light of prior attempts by several groups to exploit pAgos for genome manipulation, and the generated results.

We thank the reviewer for this comment. It seems that the meaning of this comparison was not sufficiently clear in the original submission. Of course, there are big differences between whole-genome sampling of chromosomal and invader DNA by CbAgo (even if it has strong preferences for certain loci) and highly specific DNA cleavage by CRISPR-Cas effector complexes, targeted by guide RNA encoded in the CRISPR loci. However, there are clear parallels between generation of guide molecules for CbAgo and spacer acquisition by CRISPR-Cas (during adaptation, not interference). Similarly to CbAgo, for both type I and II CRISPR-Cas systems spacer acquisition was shown to be dependent on RecBCD/AddAB – which may help in preferential selection of foreign DNA even at the adaptation step – and also results in specific targeting of the Ter region and double-strand breaks (Levy et al., 2015, Modell et al., 2017). We clarify this in the revised manuscript (lines 381-386 in the last paragraph).

While the CRISPR-Cas technology has been indeed the most impactful development in this century, this does not exclude the potential use of pAgos as an instrument for analysis of genome architecture and DNA processing, as we discuss in our manuscript. Previous attempts of genome editing with a few pAgos were indeed unsuccessful but the vast diversity of pAgo proteins (Ryazansky et al., 2018) gives hope for their use in various applications (e.g. Hegge et al.,

[REDACTED] 2018). Specifically, our study shows that CbAgo can detect with high sensitivity the regions of DSBs in the genome, and we plan to develop this analysis in the future: "Highly specific targeting of DSBs by pAgos can potentially be used for genomics applications, in particular as an instrument to study the genome architecture and DNA processing in both prokaryotic and eukaryotic systems" (lines 372-374).

Reviewer Reports on the First Revision:

Referee #1 (Remarks to the Author):

I am happy with the revised manuscript and it addressed my concerns.

Referee #2 (Remarks to the Author):

The efforts by the authors to improve the paper are very much appreciated. The new version is indeed better, however I still find difficult to draw solid conclusions about the function of pAgo in vivo. The authors made many assumptions to interpret the data as a clear indication that pAgo protects bacteria from invaders, but I find the data confusing and the arguments weak.

Referee #3 (Remarks to the Author):

Kuzmenko et al. present a revised version of their study of prokaryotic argonaute (pAgo) with targeting of chromosomal, plasmid and phage DNA. Using genetics and DNA sequencing techniques, the authors convincingly map the Clostridium butyricum pAgo guide DNAs on genomic areas of interest and implicate ter sites, DSBs and DNA repair machinery in pAgo targeting. Despite some additional data, the primary conclusion that pAgo targeting protects bacteria from invasive DNA such as plasmids and phages still seems overly conclusive and not aligned with the bulk of the data presented throughout the paper (which is still mostly consisting of chromosomal targeting, and is still relatively shallow on the phage side). The added data on phage is noted, but the large majority of the results is still focusing on chromosomal and plasmid targeting and the narrative should be framed and revised accordingly.

The phage data is interesting, intriguing and still preliminary. The authors should certainly feature this in the closing section of the manuscript (the data is limited and comes after 3 multi-panel figures setting the stage for chromosomal and plasmid targeting). It is indeed "the first report of pAgos in cell defense against phages" (P1 results are interesting and noteworthy), but this remains a minor portion of the data presented, and should accordingly not be the key conclusion presented, especially in light of the actual effect on some phages tested (e.g. T7 infection). This also applies to the "suggested cooperation" between CRISPR-Cas systems and pAgos, with no data actually shown to establish a bona fide cooperation (merely a genomic co-occurrence pattern for which statistical significance is not provided). Rather than "association" and/or "cooperation", the authors present some genomic co-occurrence patterns which is not convincing and does not reveal actual interaction between the two.

As previously mentioned, Data in Figure 4E and 4F (now expanded with more panels) seem to be the most important and valuable when assessing the overall narrative, yet they represent merely one half of a quarter of the data figures (e.g. approximately one eighth of the data presented, though it is more than the one twelfth initially submitted). The new data does provide support in the case of P1, but not T7, and the impact on some phage counts is still limited (supplemental

table 1). While this warrants discussion and featuring in the manuscript, I still believe the narrative should more properly reflect the data presented and focus on chromosome and plasmid targeting rather than phages. Specifically, the first section (CbAgo binds small DNAs from ter sites) covers chromosomal targeting; as does the second section (DNA interference between multicopy sites), covering chromosomal regions; with the third section (role of the catalytic activity of CbAgo) covering araC, rDNA and IS elements; the fourth section (asymmetry in smDNA processing) covering ter, rDNA, ara and implicated Chi sites; the fifth section covering the role of the DSB repair machinery (RecBCD) and again encompassing multicopy sequences and replication termination sites; the sixth section covers DSB processing by CbAgo. Eventually, getting to the sixth section, plasmids are encompassed, and eventually in the seventh section, phage protection is covered, with no impact on T7, some impact on M13 and a nice effect on P1. This 32-lines section is still relatively short, and cross-references results presented in figure 4d-g. This overall relative text bias is consistent with the figures relative focus, with figure 1 covering chromosomal regions (ter, rRNA, ara, IS); figure 2 showing RecBCD cooperation, relative to chi sites; figure 3 showing DSBs targeting; and eventually 4 panels of figure 4 presenting phage data.

The abstract stated that "CbAgo… induces DNA degradation at double-stranded breaks in target DNA", and I believe this still was not shown for phage DNA. Likewise, the abstract states that "The mechanism of guide generation ensures that small DNA guides are enriched in sequences that target foreign DNA and endows CbAgo with capacity to eliminate plasmids and suppress phage infection", but the direct link to small DNA guides targeting phages is still (mostly) missing (one panel in one supp. Figure).

For figure 5B, the authors should clarify what CRISPR-Cas system types and/or subtypes are implicated, and perhaps investigate more what is happening with the long-A type (any instances where pAgo and CRISPR-Cas systems are genetically linked?, any insights into what is happening with the Clostridium butyricum CRISPR-Cas system?, any specific CRISPR-Cas type or subtype more frequently associate with long-A pAgos? Is there a way for the authors to delve into the long-A portion of extended data 12 and add a CRISPR ring to show potential occurrence patterns of interest? Besides the occurrence of CRISPR-Cas systems in bacteria, there are heavily enriched in archaea, so any insights from archaeal long-A pAgo co-occurrence with CRISPR-Cas systems?

It is unclear why the authors included M13 data in supp fig 9 (panel C) rather than P1, for which the effect is relatively limited.

Author Rebuttals to First Revision:

We are very grateful to the referees for reviewing our manuscript and for the comments. In the revised manuscript, we have reframed the text to shift the focus on DNA interference, and to make more accurate conclusions about possible functions of pAgos as a defense system in bacteria. In particular, we have changed the title, the first paragraph, and corresponding places in Results and Discussion. We are confident in our results that CbAgo provides defense against two different phages in the *E. coli* host, but agree that the molecular mechanisms underlying this effect as well as the general role of pAgos in anti-phage defense indeed require further investigation. We have changed the Discussion accordingly, starting from the mechanism of DNA interference and concluding with possible functions of pAgos *in vivo*. Furthermore, we provide additional analysis of the distribution of different types of CRISPR-Cas systems in bacterial genomes, and show that Type I and Type III systems are strongly enriched in the same genomes as active pAgos. Our point-by-point response to the comments of referee #3 is provided below.

Referee #3 (Remarks to the Author) (our response is shown in blue)

Kuzmenko et al. present a revised version of their study of prokaryotic argonaute (pAgo) with targeting of chromosomal, plasmid and phage DNA. Using genetics and DNA sequencing techniques, the authors convincingly map the Clostridium butyricum pAgo guide DNAs on genomic areas of interest and implicate ter sites, DSBs and DNA repair machinery in pAgo targeting. Despite some additional data, the primary conclusion that pAgo targeting protects bacteria from invasive DNA such as plasmids and phages still seems overly conclusive and not aligned with the bulk of the data presented throughout the paper (which is still mostly consisting of chromosomal targeting, and is still relatively shallow on the phage side). The added data on phage is noted, but the large majority of the results is still focusing on chromosomal and plasmid targeting and the narrative should be framed and revised accordingly.

The phage data is interesting, intriguing and still preliminary. The authors should certainly feature this in the closing section of the manuscript (the data is limited and comes after 3 multi-panel figures setting the stage for chromosomal and plasmid targeting). It is indeed "the first report of pAgos in cell defense against phages" (P1 results are interesting and noteworthy), but this remains a minor portion of the data presented, and should accordingly not be the key conclusion presented, especially in light of the actual effect on some phages tested (e.g. T7 infection).

As previously mentioned, Data in Figure 4E and 4F (now expanded with more panels) seem to be the most important and valuable when assessing the overall narrative, yet they represent merely one half of a quarter of the data figures (e.g. approximately one eighth of the data presented, though it is more than the one twelfth initially submitted). The new data does provide support in the case of P1, but not T7, and the impact on some phage counts is still limited (supplemental table 1). While this warrants discussion and featuring in the manuscript, I still believe the narrative should more properly reflect the data presented and focus on chromosome and plasmid targeting rather than phages. Specifically, the first section (CbAgo binds small DNAs from ter sites) covers chromosomal targeting; as does the second section (DNA interference between multicopy sites), covering chromosomal regions; with the third section (role of the catalytic activity of CbAgo) covering araC, rDNA and IS elements; the fourth section (asymmetry in smDNA processing) covering ter, rDNA, ara and implicated Chi sites; the fifth section covering the role of the DSB repair machinery (RecBCD) and again encompassing multicopy sequences and replication termination sites; the sixth section covers DSB processing by CbAgo. Eventually, getting to the sixth section, plasmids are encompassed, and eventually in the seventh section, phage protection is covered, with no impact on T7, some impact on M13 and a nice effect on P1. This 32-lines section is still relatively short, and cross-references results presented in figure 4d-g. This overall relative text bias is consistent with the figures relative focus, with figure 1 covering chromosomal regions (ter, rRNA, ara, IS); figure 2 showing RecBCD cooperation, relative to chi sites; figure 3 showing DSBs targeting; and eventually 4 panels of figure 4 presenting phage data.

We have reframed the text to make more emphasis on the mechanism of DNA interference, starting from the title and the abstract, and going through all the sections of Results, as listed by the referee. In the last part of Results, we present data showing that CbAgo indeed provides defense against plasmids and phages. For two of three analyzed phages, the effects are strong (up to four orders of magnitude) and highly statistically significant, and for P1 they are observed using several different approaches. So, we are confident that CbAgo indeed protects *E. coli* cells from phages. The fact that this protection is not universal is not unusual, since many other previously studied prokaryotic defense systems provide protection against only a subset of phages (sometimes only against one in a dozen or so, *e.g.* Doron et al., [REDACTED] 2018). At the same time, we agree that the molecular details of this protection deserve further investigation, including analysis of more pAgos, more phages, and possibly a more detailed analysis of targeting of phage genomic DNA by pAgos, especially for phages with long genomes and for phages encoding various anti-defense functions. We have therefore modified Discussion accordingly, starting from DNA

interference and chromosomal targeting by CbAgo, and concluding with discussion of possible functions of pAgos in cell defense and other cellular processes.

The abstract stated that "CbAgo… induces DNA degradation at double-stranded breaks in target DNA", and I believe this still was not shown for phage DNA. Likewise, the abstract states that "The mechanism of guide generation ensures that small DNA guides are enriched in sequences that target foreign DNA and endows CbAgo with capacity to eliminate plasmids and suppress phage infection", but the direct link to small DNA guides targeting phages is still (mostly) missing (one panel in one supp. Figure).

We have modified the abstract to avoid confusion. It is now said that "CbAgo targets multicopy genetic elements and suppresses propagation of plasmids and infection by phages. CbAgo induces DNA interference between homologous sequences and triggers DNA degradation at double-strand breaks in the target DNA."

The last sentence is replaced completely, the new version is: "These results identify molecular mechanisms that generate guides for DNA interference and suggest common principles of recognition of foreign nucleic acids by prokaryotic defense systems."

This also applies to the "suggested cooperation" between CRISPR-Cas systems and pAgos, with no data actually shown to establish a bona fide cooperation (merely a genomic co-occurrence pattern for which statistical significance is not provided). Rather than "association" and/or "cooperation", the authors present some genomic co-occurrence patterns which is not convincing and does not reveal actual interaction between the two.

We have replaced "association" and "cooperation" with "co-occurrence" in the Discussion and Fig. 4 legend. We show strong enrichment of genomes encoding active pAgo nucleases with CRISPR-Cas systems, in sharp contrast to genomes without pAgos or with inactive pAgos. This may indeed suggest some kind of cooperation, but this requires further investigation. We indicate this in the revised manuscript: "The ability of pAgos to target DSBs and their co-occurrence with CRISPR-Cas suggests that the two systems may cooperate in foreign DNA targeting, which deserves further investigation."

For figure 5B, the authors should clarify what CRISPR-Cas system types and/or subtypes are implicated, and perhaps investigate more what is happening with the long-A type (any instances where pAgo and CRISPR-Cas systems are genetically linked?, any insights into what is happening with the Clostridium butyricum CRISPR-Cas system?, any specific CRISPR-Cas type or subtype more frequently associate with long-A pAgos? Is there a way for the authors to delve into the long-A portion of extended data 12 and add a CRISPR ring to show potential occurrence patterns of interest? Besides the occurrence of CRISPR-Cas systems in bacteria, there are heavily enriched in archaea, so any insights from archaeal long-A pAgo co-occurrence with CRISPR-Cas systems?

In the revised manuscript, we provide more data on co-occurrence of pAgos with various CRISPR-Cas systems. We show that Type I and Type III CRISPR-Cas systems and most their subtypes are enriched in the genomes with active pAgos, while Type II systems are somewhat depleted. These data are included in the revised Fig. 4c and Extended data Fig. 10b. We have also added information about distribution of CRISPR systems to the phylogenetic tree in Extended data (Fig. 10a), and a miniature version of this tree is also shown in Fig. 4a. We also indicate in the text that *C. butyricum* contains a type I-B CRISPR-Cas system (located on a chromid) together with CbAgo (located on the chromosome).

We have not analyzed genetic linkages of pAgos and CRISPR-Cas, which would be an important point for further studies. However, in our analysis we used only completely assembled genomes, to be confident that both systems are indeed encoded in the same genome. Because of this reason, there is only a very limited number of archaeal pAgos co-encoded with CRISPR-Cas in our dataset: 27 active pAgos are found in 456 available complete genomes; 20 (74%) of them also

encode CRISPR-Cas systems. However, due to the low number of genomes, this enrichment may not be significant, so we do not consider them separately.

It is unclear why the authors included M13 data in supp fig 9 (panel C) rather than P1, for which the effect is relatively limited.

The data for M13 were already included in the first version of manuscript. They clearly show that the whole genome of M13 is fully covered with smDNAs, for both genomic and minus strands, indicating that it is targeted during its replication cycle. M13 is a chronic phage which makes it technically easier to analyze CbAgo-bound smDNAs during infection. For P1, this is more technically challenging due to lysis of bacterial cultures during infection. Another practical reason was that preparation and sequencing of smDNA libraries had to be halted during the pandemy (and it requires the use of radioactive isotopes which have not been available for several months already).

Reviewer Reports on the Second Revision:

Referee #3 (Remarks to the Author):

Kuzmenko et al. present another revised version of their study of prokaryotic argonaute (pAgo) with targeting of chromosomal, plasmid and phage DNA. The re-framed and re-focused narrative on Clostridium butyricum pAgo DNA interference is more comprehensive and supported by the presented data, and altogether a more convincing study in the current version.

Changes to the title, abstract, discussion and throughout the text are noted and strengthen the manuscript, while allowing the maintenance of intriguing aspects of in vivo function and phage targeting. Besides the edits to the manuscript, the rebuttal is also noted, and in some cases, the current verbiage and key conclusions are improved and even more insightful (e.g. "co-occurrence" of CRISPR and pAgo; future areas of research and need for mechanistic determination).

Minor comments:
- Check the title and consider either the singular "DNA targeting by a pAgo nuclease during DNA interference", or the general "DNA targeting by pAgo nucleases during DNA interference", or the more simplistic "DNA targeting and interference by pAgo"
- I am unsure the word "cooperation" is best in the next to last sentence in the abstract and perhaps "mechanism" or "interaction" or other alternatives may be more descriptive.
- Please ensure a different set of continuously numbered references should be used in the Methods References section (for which the formatting for refs 47-51 seems slightly different).
- Ensure the "code availability" statement is consistent with Nature editorial policy, rather than being posted online.
- Ensure the fonts used throughout figures are large enough for the readership (a few small fonts may seem on the fringe in figures 1 and 2).
- Check whether active/non-active or active/in-active is best for figure 4a

Author Rebuttals to Second Revision:

Response to referee #3

We are grateful to the referee for the additional suggestions. Our response below is shown in blue.

Minor comments:
- Check the title and consider either the singular "DNA targeting by a pAgo nuclease during DNA interference", or the general "DNA targeting by pAgo nucleases during DNA interference", or the more simplistic "DNA targeting and interference by pAgo"

We have changed the title to "DNA targeting and interference by a bacterial Argonaute nuclease".

- I am unsure the word "cooperation" is best in the next to last sentence in the abstract and perhaps "mechanism" or "interaction" or other alternatives may be more descriptive.

We have replaced "cooperation" with "interplay".

- Please ensure a different set of continuously numbered references should be used in the Methods References section (for which the formatting for refs 47-51 seems slightly different).

We have updated the References section in Methods; in accordance to the Nature format, continuous numbering is used for references in both main text and Methods.

- Ensure the "code availability" statement is consistent with Nature editorial policy, rather than being posted online.

We have uploaded the code used for data analysis in a public repository (https://github.com/AntKuzmenko/CbAgo_DNAi.git) and provided a corresponding statement in the manuscript.

- Ensure the fonts used throughout figures are large enough for the readership (a few small fonts may seem on the fringe in figures 1 and 2).

We have checked the fonts in all the figures in accordance with Nature guidelines (the final font size should be no less than 5 points).

- Check whether active/non-active or active/in-active is best for figure 4a

We have replaced non-active with inactive in Fig. 4a.