NgAgo-enhanced homologous recombination in *E. coli* is mediated by DNA endonuclease activity

Kok Zhi Lee[1], Michael A. Mechikoff[1], Archana Kikla[2], Arren Liu[2], Paula Pandolfi[2], Frederick S. Gimble[3,4], and Kevin V. Solomon[1,2*]

[1]Department of Agricultural and Biological Engineering, Purdue University, West Lafayette, IN 47906, USA.

[2]Department of Biological Sciences, Purdue University, West Lafayette, IN 47906, USA.

[3]Purdue University Interdisciplinary Life Science Program (PULSe), Purdue University, West Lafayette, IN, 47906, USA.

[4]Department of Biochemistry, Purdue University, West Lafayette, IN, 47906, USA.

* To whom correspondence should be addressed. Tel: +1 765-494-1134; Fax: +1 765-496-1115; Email: kvs@purdue.edu

ABSTRACT

Prokaryotic Argonautes (pAgos) have been proposed as more flexible tools for gene-editing as they do not require sequence motifs adjacent to their targets for function, unlike popular CRISPR/Cas systems. One promising pAgo candidate, from the halophilic archaeon *Natronobacterium gregoryi* (NgAgo), however, has been subject to intense debate regarding its potential in eukaryotic systems. Here, we revisit this enzyme and characterize its function in prokaryotes. NgAgo expresses poorly in non-halophilic hosts with the majority of protein being insoluble and inactive even after refolding. However, we report that the soluble fraction does indeed act as a DNA endonuclease. Structural homology modelling revealed that NgAgo shares canonical domains with other catalytically active pAgos but also contains a previously unrecognized single stranded DNA binding domain (repA). Both repA and the canonical PIWI domain participate in DNA cleavage activities. We also found that these endonuclease activities are essential for enhanced NgAgo-guided homologous recombination, or gene-editing, in *E. coli*. Collectively, our results provide insight into the poorly characterized NgAgo for subsequent gene-editing tool development and sheds new light on seemingly contradictory reports.

INTRODUCTION

Long prokaryotic Argonaute proteins (pAgos) are programmable endonucleases that have recently been proposed as flexible tools for genome editing[1]. Like Cas9-based gene editing strategies, single-stranded nucleic acids bind to pAgos and enhance pAgo cleavage of complementary target nucleic sequences, enabling DNA repair and editing. However, pAgos have the distinct advantage of not requiring a protospacer adjacent motif (PAM) for function[2-5], which means that pAgos are not limited to targets flanked by PAM sites and can potentially cut any DNA target regardless of composition. Despite this potential, no pAgo has been developed that rivals the simplicity and function of Cas9-based strategies.

Long pAgos are predicted to serve as a form of adaptive defense against invading nucleic acids such as phage/viral DNA and RNA[6,7]. With a single-stranded DNA and/or RNA as a guide, long pAgos cleave complementary target DNA, RNA, or both via the conserved catalytic tetrad, DEDX[1]. To create a double-stranded DNA break, long pAgos require two guides. Target recognition and cleavage is enabled by four canonical domains[3]: N (N-terminal), PAZ (PIWI-Argonaute-Zwille), MID (middle), and PIWI (P element-induced wimpy testis). The N-terminal domain is essential in target cleavage[8,9] and dissociation of cleaved strands[9,10], though the detailed mechanism remains poorly understood. The MID domain interacts with the 5'-end of the guide[11] and promotes binding of the guide to its target nucleic acids[12]. The PAZ domain interacts with the 3' end of a guide[13-16], protecting it from degradation[17]. The PIWI domain plays a pivotal role in nucleic acid cleavage via the conserved catalytic tetrad, DEDX (D: aspartate, E: glutamate, X: histidine, aspartate or asparagine)[6]. Despite the presence of these canonical domains in all long pAgos, currently characterized pAgos including TtAgo[2], MpAgo[5], PfAgo[18] and MjAgo[3,19] work at very high temperatures (>55 °C)[2,3,5,18], making them infeasible for gene editing in common mesophilic organisms.

The halophilic Argonaute from *Natronobacterium gregoryi* (NgAgo) was recently put forth as a promising candidate for pAgo-mediated gene editing as it is believed to operate at mesophilic (~37°C) temperatures[20]. However, these claims have since been refuted due to an inability to demonstrate *in vitro* DNA cleavage or to replicate these findings in a number of eukaryotic hosts [21-25]. NgAgo expression is poor, presumably due to its halophilic characteristics that make low salt expression challenging[26,27]. Thus, all published *in vitro* cleavage assays have relied on refolded protein[21-25], which may be non-functional, resulting in the inconclusive results. Nonetheless, recent work by Fu and colleagues demonstrated that NgAgo may still have potential as a gene editor for prokaryotic hosts. While the authors were able to confirm that gene-editing was mediated by homologous recombination via RecA, which physically associated with NgAgo in an unanticipated manner, the specific role of NgAgo remained unclear. Here, we demonstrate that NgAgo is indeed a DNA endonuclease by identifying a catalytic mutant that is required for DNA cleavage, and provide evidence that this activity is essential for NgAgo-mediated gene editing via homologous recombination repair.

MATERIAL AND METHODS

Strains and plasmids

E. coli strains and plasmids used in this study are listed in Table 1. Cloning was carried out according to standard practices[28] with primers, template, and purpose listed in Supplementary Table 1. Plasmids were maintained in *E. coli* DH5α. NgAgo variants (wildtype, D663A/D738A, N-del, and repA with GST or His tag) that were used for *in vitro* activity assays were cloned into a IPTG-inducible T7 plasmid via the pET32a-GST-ELP64 (provided by Professor Xin Ge, University of California, Riverside).

To test the homologous recombination ability of NgAgo, we cloned pTKDP-KanR-mNeonGreen-hph for recombineering and made p15-KanR-PtetRed as our donor plasmid with inducible lambda-red recombinase (Table 1).

NgAgo expression and purificatio

All GST-NgAgo or His-NgAgo variants were transformed into BL21 (DE3) electrocompetent cells and were plated on agar plates containing ampicillin (100 µg/ml). A single colony was inoculated in LB with ampicillin for 16 hours and then cultured in 100 ml of LB containing ampicillin. Expression was induced with IPTG at 0.1 mM final concentration when the culture OD600 reached 0.5. After 4 hours incubation at 37 °C or 22 °C overnight, cells were collected by centrifuge 7500 rpm at 4 °C for 5 minutes. The cell pellet was resuspended in TN buffer (10 mM Tris and 100mM NaCl, pH 7.5) and lysed via sonication at a medium power setting (~50 W) in 10 s intervals, with intervening 10 s incubations on ice to reduce heat denaturation. Cell lysates were then clarified at 12000 rpm at 4 °C for 30 minutes. The supernatant was collected as a soluble protein fraction. Both soluble and insoluble (cell pellet) fractions were purified via His-IDA nickel column (Clontech Laboratories, Mountain View, CA. Cat. No: 635657) according to the manufacturer instructions. Insoluble NgAgo protein was refolded on the column after denaturation with guanidium chloride according to manufacturer instructions. GST-tagged NgAgo variants were purified by glutathione agarose (Thermo Fisher Scientific, Waltham, MA. Cat. No: 16100) according to the manufacturer protocol.

In vitro activity assay

For the reloading protocol, five micrograms of purified NgAgo were mixed with one microgram total of phosphorylated single-stranded DNA (P-ssDNA) targeting mNeonGreen (Supplementary Table 2) and incubated at 55 °C for an hour. 200-300 ng of substrate plasmid DNA (pNCS-mNeonGreen) was then added to the sample. The final volume of the reaction was 50 µl (working concentration: 20 mM Tris-Cl, 300 mM KCl, 500 µM $MgCl_2$, and 2 mM DTT). The sample was then incubated at 37 °C for three hours. 0.8 units of Proteinase K (NEB, Ipswich, MA. Cat. No: P8107S) were added to the sample to digest the protein for 5 minutes at 37 °C. The nucleic acids were then cleaned up by the DNA Clean & Concentrator™-5 kit (Zymo Research, Irvine, CA. Cat. No: D4003T) according to manufacturer instructions and mixed with 6X loading dye containing SDS (Thermo Fisher S, Waltham, MA. Cat. No: R1151) before gel electrophoresis.

The gel containing Sybrsafe (Thermo Fisher S, Waltham, MA. Cat. No: S33102) was visualized under a blue light (Azure Biosystems, Dublin, CA. Azure c400).

For our standard protocol, we incubated the same amount of guides and proteins at 37 °C for 30 minutes, and added the same amount of plasmid DNA (p15-KanR or pBSI-SceI(E/H)[31]) with 50 ul final volume (working concentration: 20mM Tris-Cl, 300mM NaCl, 250 uM $MgCl_2$, and 2mM DTT). The samples were incubated at 37 °C for an hour before Proteinase K treatment. The rest of the procedure is the same as the reloading protocol.

As positive controls for nicked and linearized DNA, we digested plasmid pBSI-SceI(E/H) with I-SceI or a K223I I-SceI mutant[31], generating linearized and nicked DNA, respectively. We tested five micrograms of each NgAgo variant with pBSI-SceI(E/H) and conducted electrophoresis to check the plasmid conformation. To exclude the possibility of band shift due to DNA binding, we treated the samples with 0.8 units of proteinase K and used a gel loading dye with SDS when running on a gel.

Electrophoretic mobility shift assay (EMSA)

Five microgram of purified N-del and repA were incubated with one microgram of mNeonGreen ssDNA guide in 50ul in buffer (working concentration: 20 mM Tris-Cl, 300 mM KCl, 500 µM $MgCl_2$, and 2 mM DTT) at 37°C for an hour and treated with 0.8 units proteinase K for 5 minutes if needed before running with 20% TBE gel with 0.5X TBE buffer. Gels were stained with Sybr Gold (Thermo Fisher Scientific, Waltham, MA. Cat. No: S11494) before visualizing under a green fluorescent channel (Azure Biosystems, Dublin, CA. Azure c400). Positional marker 10/60 ladder (Coralville, IA. Cat. No: 51-05-15-01) was used in the EMSA assay.

Gene-editing assay

MG1655 (DE3) *atpI*::KanR-mNeonGreen was transformed with pET-GST-NgAgo-His (to induce DNA cleavage) and p15-KanR-PtetRed (for lambda-red recombinase expression and to provide donor DNA for repair) and made electrocompetent. Electrocompetent cells were transformed with either no guides or one microgram total of FW, RV, both guides and incubated in Miller LB with ampicillin, chloramphenicol, and IPTG for an hour. These cultures were then diluted ten-fold in Miller LB containing ampicillin (working concentration: 100 µg/ml), chloramphenicol (working concentration: 25 µg/ml), IPTG (working concentration: 0.1mM), and anhydrotetracycline (aTc) (working concentration: 50 µg/ml), incubated for 2 hours before plating with and without kanamycin. Colony forming units (CFU) were counted after 16-20 hours incubation at 37 °C. The unguided control was normalized to 100% and guided-treatments were normalized to the unguided control.

Phyre 2 and HHpred analysis

NgAgo protein (IMG/M Gene ID: 2510572918) was analyzed via Phyre 2[34] with normal mode on 2018 November 19. The normal mode pipeline involves detecting sequence homologues, predicting secondary structure and disorder, constructing a hidden Markov model (HMM), scanning produced HMM against library of HMMs of proteins with experimentally solved structures, constructing 3D models of NgAgo, modelling insertions/deletions, modelling of amino acid sidechains, submission of the top model, and transmembrane helix and topology prediction[34]. NgAgo was analyzed via HHpred[35] (https://toolkit.tuebingen.mpg.de/#/tools/hhpred) on 2018 November 27. The parameters for HHpred are HHblits=>uniclust30_2018_08 for multiple sequence alignment (MSA) generation method, 3 for maximal number of MSA generation steps, 1e-3 for E-value incl. threshold for MSA generation, 0% for minimum sequence identity of MSA hits with query, 20% for minimum coverage of MSA hits, during_alignment for secondary structure scoring, local for alignment mode, off for realign with MAC, 0.3 for MAC realignment threshold, 250 for number of target sequences, and 20% for minimum probability in hit list.

Phylogenetic analysis

BLAST was used to compare NgAgo protein sequence with all the isolates in the database via the IMG/M server (https://img.jgi.doe.gov/). Representative full-length Argonautes with a repA domain were used to represent each species. Selected pAgos with repA domains and some well-characterized pAgos were compared, and the midpoint rooted tree was generated via the server http://www.genome.jp/tools-bin/ete with unaligned input type, mafft_default aligner, no alignment cleaner, no model tester, and fasttree_default Tree builder parameters. The nwk output file was then used for phylogenetic tree generation in R with ggtree package.

RESULTS

NgAgo has canonical N-terminal, PIWI, MID, and PAZ domains, and a putative single stranded DNA binding (repA) domain.

Given the ongoing debate of the function of NgAgo, we analysed its sequence (IMG/M Gene ID: 2510572918) with Phyre 2[34] and HHpred[36] to predict its structure based on characterized structural homologs. Phyre 2 and HHpred analyses found with high confidence (probability = 100%) that NgAgo shares structural features with catalytically active pAgos and eukaryotic Agos (eAgos) including archaeal MjAgo, bacterial TtAgo, and eukaryotic hAgo2 (Supplementary Table 3 and 4). Since MjAgo is the only characterized pAgo from Archaea, we used it as a template for comparative modelling. The predicted NgAgo structure is similar to the crystal structure of MjAgo, consisting of canonical N-terminal, PAZ, MID, and PIWI domains (100% probability in both Phyre 2 and HHpred) (Fig. 1a and b). However, the N-terminal domain of NgAgo is truncated, relative to MjAgo, potentially suggesting a novel mechanism for strand displacement and binding due to the N-terminal domain's role in pAgo targeted cleavage.

Structural analysis also identified an uncharacterized oligonucleotide/oligosaccharide-binding (OB) fold domain between residues 13-102 of NgAgo that is known to bind single-stranded DNA in eukaryotes and prokaryotes[37] (Fig. 1b). This OB domain has recently been identified as a new feature of pAgos[38]. As repA proteins were the most common matches on both Phyre 2 and HHpred, we will refer to this OB domain as repA (Supplementary Tables 5 and 6). While the repA domain is absent in all characterized pAgos, at least 12 Ago homologs from various species deposited on IMG/M (https://img.jgi.doe.gov/) share this domain. Phylogenetic analysis showed that all the repA-containing pAgos were from halophilic Archaea forming a clade that is distinct from that of the current well-characterized pAgos (Fig. 1c). This monophyletic group of repA-containing pAgos may represent a new class of pAgos that is currently unrecognized in the literature[39]. Moreover, its unique presence within halophiles suggests that the repA domain may be required for function in high salt environments, potentially replacing the role of the canonical N-terminal domain, which was then truncated through evolution.

Our analysis of NgAgo also confirmed the presence of a conserved catalytic tetrad, DEDX (X: H, D or N)[6], which is critical for nucleic acid cleavage by the PIWI domain of Argonautes. The catalytic tetrad (D663, E704, D738, and D863) of NgAgo aligns well with those from other catalytically active pAgos, including MjAgo[3], PfAgo[18], MpAgo[5], and TtAgo[2] (Fig. 1d). Moreover, structural alignment of NgAgo and MjAgo display good colocalization of the catatytic tetrad, except for E704, suggesting that NgAgo may have similar nucleic acid cleavage activity (Fig. 1e).

Soluble but not refolded NgAgo exhibits random DNA cleavage activity *in vitro*

As halophilic proteins tend to be insoluble when expressed in a low-salt environment due to their sequence adaptations[26,27,40], we first optimized expression conditions to obtain more soluble NgAgo protein (Supplementary Fig. 1). We purified wildtype NgAgo (Fig. 2a) from both the soluble and insoluble fractions to test for guide-dependent DNA cleavage using 5'P-ssDNA as guides. Insoluble NgAgo was refolded during purification using a previously published method[41]. Our results showed that purified NgAgo from the soluble cell lysate fraction (sNgAgo) nicks plasmid DNA and genomic DNA, independent of guide (Fig. 2b and supplementary Fig. 2e), as evidenced by the presence of the nicked and linearized plasmid. However, purified refolded NgAgo from the insoluble lysate fraction (rNgAgo) has little or no activity on DNA (Fig. 2c), consistent with a study by Ye and colleagues[41]. We hypothesized that NgAgo generates random guides in the host via DNA chopping[42], which co-purifies with NgAgo leading to apparent guide-independent activity *in vitro*. While we were able to confirm the presence of these random copurified guides (Fig. 2d), we were unable to displace them with incubation at high temperature (55 °C) and reload with our target guides (reloading protocol). Subsequent testing had similar guide-independent cleavage activity with no evidence of increased linearized plasmid (Supplementary Fig. 3). As refolded NgAgo had no cleavage activity, we used soluble NgAgo to study its function *in vitro* unless otherwise stated.

Previous studies have demonstrated that TtAgo can obtain random guides from the expression plasmid DNA via DNA chopping[2]. Thus, the observed guide-independent cleavage may indeed be guide-dependent as a result of chopping and subsequent guide loading with homologous DNA, which cannot be easily displaced as demonstrated in Fig 2c. To examine this hypothesis, we completed the *in vitro* cleavage assay with a 'related' plasmid, pNCS-mNeonpGreen (Supplementary Fig. 2a), and an 'unrelated' plasmid, p15-KanR (Supplementary Fig. 2c). The unrelated plasmid, p15-KanR, shares no DNA homology with the NgAgo expression plasmid while the related plasmid, pNCS-mNeonGreen, has the same ampicillin resistance gene. NgAgo cleaved both related and unrelated plasmids independent of guide (Supplementary Fig. 2b and 2d), suggesting that the guide-independent cleavage activity of our purified NgAgo does not rely on pre-loaded DNA. These results confirmed that NgAgo has guide-independent cleavage activity *in vitro*, sharing similar properties with bacterial TtAgo[42] and archaeal MjAgo[19].

RepA and PIWI domains are responsible for NgAgo DNA cleavage

As NgAgo cuts plasmids independent of guide, we used this activity to identify which domains are responsible for DNA cleavage. Since *in silico* analysis identified an uncharacterized repA domain, we constructed a repA mutant (residues 1-102) and a repA-deletion (residues 105-887, referred to as N-del) (Fig. 2a) to examine whether repA is required for NgAgo function. We also constructed double mutants, D663A/D738A, containing mutations at putative active site residues (this double mutant corresponds to the catalytic double mutant, D478A/D546A, of TtAgo[2] that loses all cleavage activities[2,42]) in the full-length protein and N-del (Fig. 2a). *In vitro* cleavage assays with repA confirm that it nicks and cleaves plasmid DNA, as evidenced by open-circular and linearized plasmid (Fig. 2e). Although the repA domain is able to bind to ssDNA as demonstrated on an electrophoretic mobility shift assay (EMSA) (Fig. 2c), the mechanism by which it cuts DNA remains unknown.

Our cleavage assays with NgAgo mutants suggest that multiple domains are involved in NgAgo activity (Fig. 2e). An N-del truncation mutant that lacks the repA domain displays cleavage activity. Similarly, D663A/D863A mutants containing mutations in the canonically catalytic PIWI domain maintain similar guide-independent nicking and cleaving activity relative to wildtype. Thus, repA and PIWI domains appear to both cut DNA independently from one another and can complement the loss of function from the other. Indeed, mutants containing combined mutations (N-del/D663A/D863A) lose all ability to nick/linearize plasmids (Fig. 2e), suggesting that the nicking/cleaving activities of N-del is dependent on the putative catalytic tetrad within the PIWI domain (Fig. 1d and 1e). Collectively, our work shows that NgAgo is a DNA endonuclease, dependent on the function of its repA and PIWI domains.

repA and PIWI domains are essential for programmable DNA editing

Since we have shown that NgAgo can cleave DNA, and since work from other groups indicated the protein is active *in vivo*[43], we asked if NgAgo can be repurposed as a guided gene-editing tool in *E.coli*. We chose

E.coli instead of mammalian cells as our model because *E.coli* lacks histones, which are known to inhibit pAgo activity[19]. To test for NgAgo gene editing activity, we created an MG1655 (DE3) strain harbouring a cassette composed of a *kanR* gene and a *mNeonGreen* gene lacking an RBS and promoter, flanked by two double terminators (Fig. 3a). This arrangement prevents any KanR/mNeonGreen expression from transcription read-through and translation from upstream and downstream genes. Since DNA breaks in *E.coli* are lethal, only correct recombinants will survive on kanamycin plates when provided with donor plasmid, which harbors a truncated *mNeonGreen*, a constitutive promoter, an RBS and a truncated *kanR* (Fig. 3a). We then demonstrated that ssDNA could survive long enough to form a complex with NgAgo before degradation (Supplementary Fig. 4). Wildtype NgAgo increased homologous recombination efficiency 107%, 82%, and 31% when provided with FW, RV, and both guides, respectively, compared with an unguided control (Fig. 3b), demonstrating that guide-dependent NgAgo activity can enhance gene editing.

Given that the PIWI domain is essential for guide-dependent cleavage activity in other studied pAgos[2,5,18], we tested its essentiality for homologous recombination in NgAgo. The PIWI mutant, D663A/D738A, of NgAgo demonstrated a statistically significant enhancement in homologous recombination; however, this was roughly half of what was seen in the wildtype protein (43% above no guide controls). The PIWI mutant displayed no significant enhancement of recombination with the FW or both guides (Fig. 3b). While the mechanism behind this pattern is unclear, these data suggest that the PIWI domain is not essential for guide-dependent cleavage activity of NgAgo.

Additionally, as the repA domain is not common amongst pAgos, we tested if it was required for DNA targeting activity. The N-del mutant of NgAgo lacking the repA domain displayed only an 11% enhancement in homologous recombination above unguided controls in the presence of the RV guide only (Fig. 3b). Nonetheless, this is consistent with a mechanism in which repA also plays a role in guide-dependent cleavage activity. Consistent with *in vitro* results, an N-del/D663A/D738A catalytic mutant showed no increase in gene editing activity in the presence of FW, RV, or both guides compared to an unguided control. Thus, the DNA endonuclease activity mediated by the repA and PIWI domains is essential for enhanced homologous recombination and gene editing.

DISCUSSION

NgAgo may represent a new class of mesophilic pAgos

To our knowledge, NgAgo is the first studied pAgo with an uncharacterized repA domain, which indeed binds to single-stranded DNA (Fig. 2f). Surprisingly, we found that repA alone contributes to DNA cleavage activity (Fig. 2e). Moreover, repA aids the PIWI domain in NgAgo targeted DNA cleavage as homologous recombination is reduced in N-del mutants relative to wildtype (Fig. 3). Interestingly, all repA domain-containing pAgos are from halophilic Archaea mesophiles, suggesting that the repA domain may be

required for pAgos to function in high-salt environments. Given that *Natronobacterium gregoryi*, the native host of NgAgo, is a halophile, the protein must have evolved ways to maintain protein-DNA interactions for catalysis in high salt environments where many electrostatic interactions are reduced. As demonstrated by Hunt and co-workers, single-stranded binding (SSB) protein enhances TtAgo activity[44]; repA at the N-terminus of NgAgo may be involved in the cleaving process without recruiting SSB protein. Moreover, as the N-terminal domain of pAgos is essential for target cleavage[6], repA may supplant its role resulting in the truncated N-terminal domain of NgAgo. Further research, however, is needed to clarify the function of this repA domain.

NgAgo is a DNA-guided DNA endonuclease

Although previous studies demonstrated that refolded NgAgo does not cut DNA *in vitro*[41,44], consistent with our findings, we establish that soluble NgAgo can, in fact, cleave DNA *in vitro*. That is, refolded NgAgo may not be fully functional. As we showed that an N-del/D663A/D738A catalytic mutant lacks DNA cleaving activity (Fig. 2e), the catalytic activity is unlikely to be the result of sample contamination. However, we are unable to demonstrate unequivocal guide-dependent cleavage with both double-stranded DNA target and single-stranded DNA target *in vitro* (data not shown). This may be due to inefficient guide loading, as we observe that N-del co-purifies guides (Fig. 2c).

NgAgo can be repurposed as a DNA editing tool

Our results provide supporting evidence to encourage the development of NgAgo for gene-editing. When provided with homologous target regions, NgAgo can aid in homologous recombination. Much like other pAgos, the PIWI domain participates in DNA editing as shown here and by Fu *et al*. Moreover, without repA, PIWI mutants of NgAgo exhibit reduced cleavage activity with a concomitant reduction in homologous recombination efficiency. Both the repA deletion and the PIWI mutation (N-del/D663A/D738) are needed to fully abolish catalytic and gene editing functions. Thus, in the presence of both functional domains, NgAgo can effectively enhance homologous recombination by inducing a double stranded break at a targeted region. Despite the programmable DNA-cleaving ability of NgAgo, there remains several challenges to its development as a robust tool for gene-editing applications: high off-target activity or guide independent cleavage, poor expression, and potentially low activity in eukaryotic hosts. Nonetheless, further insight may lead to protein engineering strategies to overcome these hurdles and develop NgAgo as a robust tool for gene-editing.

Conclusion

Based on the above findings, we conclude that NgAgo is a novel DNA endonuclease that belongs to an unrecognized class of pAgos defined by a characteristic repA domain. NgAgo cleaves DNA through both a well-conserved catalytic tetrad in PIWI and through a novel uncharacterised repA domain. This cleavage

activity is essential to enhancing gene-editing efficiency in prokaryotes. Despite the challenges of NgAgo, our work provides insight into poorly characterized NgAgo for subsequent gene-editing tool development, and sheds new light on seemingly contradictory reports.

FUNDING

This research was supported by the startup funds from the Colleges of Engineering and Agriculture, and the USDA National Institute of Food and Agriculture (Hatch Multistate Project S1041).

Acknowledgment

We are grateful to Dr. Xin Ge (University of California, Riverside) and Dr. Kristala J. Prather (Massachusetts Institute of Technology) for providing pET32a-GST-ELP64 plasmid and MG1655 (DE3), respectively. We also thank Dr. Mathew Tantama (Purdue University) for providing pBAD-mTagBFP2 plasmid.

CONFLICT OF INTEREST

K.V.S., K.Z.L., and M.A.M. have filed a patent related to this work.

Author contributions

K.V.S. and K.Z.L. designed the experiments. K.Z.L., M.A.M., A.K., A.L., and P.P. conducted and analyzed the experiments. K.V.S., F.G., and K.Z.L. supervised research and experimental design. K.V.S, K.Z.L, M.A.M, and F.G. wrote the manuscript.

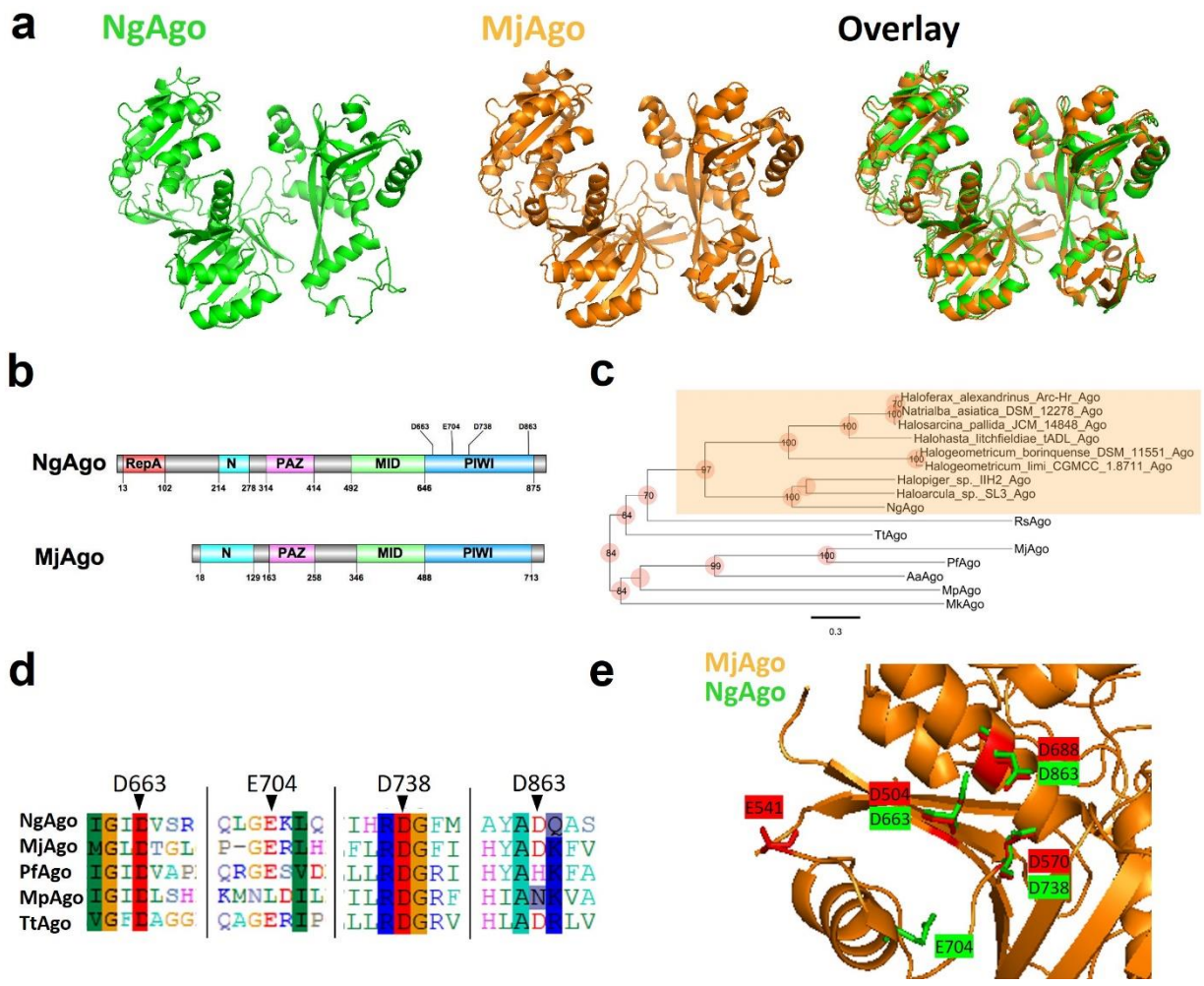
References

1. Hegge, J. W., Swarts, D. C. & van der Oost, J. Prokaryotic Argonaute proteins: novel genome-editing tools? *Nature Reviews Microbiology* **16**, 5 (2018).
2. Swarts, D. C. *et al.* DNA-guided DNA interference by a prokaryotic Argonaute. *Nature* **507**, 258-261 (2014).
3. Willkomm, S. *et al.* Structural and mechanistic insights into an archaeal DNA-guided Argonaute protein. *Nature Microbiology* **2**, 17035 (2017).
4. Enghiad, B. & Zhao, H. Programmable DNA-guided artificial restriction enzymes. *ACS synthetic biology* **6**, 752-757 (2017).
5. Kaya, E. *et al.* A bacterial Argonaute with noncanonical guide RNA specificity. *Proceedings of the National Academy of Sciences* **113**, 4057-4062 (2016).
6. Swarts, D. C. *et al.* The evolutionary journey of Argonaute proteins. *Nature structural & molecular biology* **21**, 743-753 (2014).
7. Koonin, E. V. Evolution of RNA-and DNA-guided antivirus defense systems in prokaryotes and eukaryotes: common ancestry vs convergence. *Biology direct* **12**, 5 (2017).
8. Hauptmann, J. *et al.* Turning catalytically inactive human Argonaute proteins into active slicer enzymes. *Nature Structural and Molecular Biology* **20**, 814 (2013).
9. Faehnle, C. R., Elkayam, E., Haase, A. D., Hannon, G. J. & Joshua-Tor, L. The making of a slicer: activation of human Argonaute-1. *Cell reports* **3**, 1901-1909 (2013).
10. Kwak, P. B. & Tomari, Y. The N domain of Argonaute drives duplex unwinding during RISC assembly. *Nature structural & molecular biology* **19**, 145 (2012).
11. Ma, J.-B. *et al.* Structural basis for 5′-end-specific recognition of guide RNA by the A. fulgidus Piwi protein. *Nature* **434**, 666 (2005).
12. Künne, T., Swarts, D. C. & Brouns, S. J. J. Planting the seed: target recognition of short guide RNAs. *Trends in microbiology* **22**, 74-83 (2014).
13. Lingel, A., Simon, B., Izaurralde, E. & Sattler, M. Nucleic acid 3′-end recognition by the Argonaute2 PAZ domain. *Nature Structural and Molecular Biology* **11**, 576 (2004).
14. Ma, J.-B., Ye, K. & Patel, D. J. Structural basis for overhang-specific small interfering RNA recognition by the PAZ domain. *nature* **429**, 318 (2004).
15. Sheng, G. *et al.* Structure-based cleavage mechanism of Thermus thermophilus Argonaute DNA guide strand-mediated DNA target cleavage. *Proceedings of the National Academy of Sciences* **111**, 652-657 (2014).
16. Wang, Y. *et al.* Structure of an argonaute silencing complex with a seed-containing guide DNA and target RNA duplex. *nature* **456**, 921 (2008).
17. Hur, J. K., Zinchenko, M. K., Djuranovic, S. & Green, R. Regulation of Argonaute slicer activity by guide RNA 3'end interactions with the N-terminal lobe. *Journal of Biological Chemistry*, jbc-M112 (2013).

18. Swarts, D. C. *et al.* Argonaute of the archaeon Pyrococcus furiosus is a DNA-guided nuclease that targets cognate DNA. *Nucleic acids research* **43**, 5120-5129 (2015).
19. Zander, A. *et al.* Guide-independent DNA cleavage by archaeal Argonaute from Methanocaldococcus jannaschii. *Nature Microbiology* **2**, 17034 (2017).
20. Cyranoski, D. Authors retract controversial NgAgo gene-editing study. *Nature News*, doi:10.1038/nature.2017.22412 (2017).
21. Javidi-Parsijani, P. *et al.* No evidence of genome editing activity from Natronobacterium gregoryi Argonaute (NgAgo) in human cells. *Plos One* **12**, 14, doi:10.1371/journal.pone.0177444 (2017).
22. Wu, Z. *et al.* NgAgo-gDNA system efficiently suppresses hepatitis B virus replication through accelerating decay of pregenomic RNA. *Antiviral Research* (2017).
23. Burgess, S. *et al.* Questions about NgAgo. *Protein & Cell* **7**, 913-915, doi:10.1007/s13238-016-0343-9 (2016).
24. Khin, N. C., Lowe, J. L., Jensen, L. M. & Burgio, G. No evidence for genome editing in mouse zygotes and HEK293T human cell line using the DNA-guided Natronobacterium gregoryi Argonaute (NgAgo). *PloS one* **12**, e0178768 (2017).
25. Qin, Y. Y., Wang, Y. M. & Liu, D. NgAgo-based fabp11a gene knockdown causes eye developmental defects in zebrafish. *Cell Research* **26**, 1349-1352, doi:10.1038/cr.2016.134 (2016).
26. Elcock, A. H. & McCammon, J. A. Electrostatic contributions to the stability of halophilic proteins. *Journal of molecular biology* **280**, 731-748 (1998).
27. Tadeo, X. *et al.* Structural basis for the aminoacid composition of proteins from halophilic archea. *PLoS biology* **7**, e1000257 (2009).
28. Sambrook, J., Fritsch, E. F. & Maniatis, T. *Molecular cloning: a laboratory manual*. (Cold spring harbor laboratory press, 1989).
29. Wood, W. B. Host specificity of DNA produced by Escherichia coli: bacterial mutations affecting the restriction and modification of DNA. *Journal of molecular biology* **16**, 118-IN113 (1966).
30. Tseng, H.-C., Martin, C. H., Nielsen, D. R. & Prather, K. L. J. Metabolic engineering of Escherichia coli for enhanced production of (R)-and (S)-3-hydroxybutyrate. *Applied and environmental microbiology* **75**, 3137-3145 (2009).
31. Niu, Y., Tenney, K., Li, H. & Gimble, F. S. Engineering variants of the I-SceI homing endonuclease with strand-specific and site-specific DNA-nicking activity. *Journal of molecular biology* **382**, 188-202 (2008).
32. Rhodius, V. A. *et al.* Design of orthogonal genetic switches based on a crosstalk map of σs, anti-σs, and promoters. *Molecular systems biology* **9**, 702 (2013).
33. Reisch, C. R. & Prather, K. L. J. The no-SCAR (Scarless Cas9 Assisted Recombineering) system for genome editing in Escherichia coli. *Scientific reports* **5**, 15096 (2015).
34. Kelley, L. A., Mezulis, S., Yates, C. M., Wass, M. N. & Sternberg, M. J. E. The Phyre2 web portal for protein modeling, prediction and analysis. *Nature protocols* **10**, 845-858 (2015).

35. Zimmermann, L. *et al.* A completely Reimplemented MPI bioinformatics toolkit with a new HHpred server at its Core. *Journal of molecular biology* **430**, 2237-2243 (2018).
36. Söding, J., Biegert, A. & Lupas, A. N. The HHpred interactive server for protein homology detection and structure prediction. *Nucleic acids research* **33**, W244-W248 (2005).
37. Flynn, R. L. & Zou, L. Oligonucleotide/oligosaccharide-binding fold proteins: a growing family of genome guardians. *Critical reviews in biochemistry and molecular biology* **45**, 266-275 (2010).
38. Ryazansky, S., Kulbachinskiy, A. & Aravin, A. The expanded universe of prokaryotic Argonaute proteins. *bioRxiv*, 366930 (2018).
39. Ryazansky, S., Kulbachinskiy, A. & Aravin, A. A. The expanded universe of prokaryotic Argonaute proteins. *mBio* **9**, e01935-01918 (2018).
40. Müller-Santos, M. *et al.* First evidence for the salt-dependent folding and activity of an esterase from the halophilic archaea Haloarcula marismortui. *Biochimica et Biophysica Acta (BBA)-Molecular and Cell Biology of Lipids* **1791**, 719-729 (2009).
41. Sunghyeok, Y. *et al.* DNA-dependent RNA cleavage by the Natronobacterium gregoryi Argonaute. *BioRxiv*, 101923 (2017).
42. Swarts, D. C. *et al.* Autonomous Generation and Loading of DNA Guides by Bacterial Argonaute. *Molecular Cell* **65**, 985-998 (2017).
43. Fu, L. *et al.* The prokaryotic Argonaute proteins enhance homology sequence-directed recombination in bacteria. *Nucleic acids research* (2019).
44. Hunt, E. A., Evans Jr, T. C. & Tanner, N. A. Single-stranded binding proteins and helicase enhance the activity of prokaryotic argonautes in vitro. *PloS one* **13**, e0203073 (2018).

411 **FIGURES**



412

413

414 **Figure 1 | NgAgo belongs to a distinct clade of pAgos with a catalytic DEDX tetrad and novel repA**
415 **domain. a,** Phyre 2 simulation 3D structure based on MjAgo structure (PDB: 5G5T). NgAgo structure is
416 similar to the MjAgo structure except for the N-terminal domain. **b**, Domain architecture of NgAgo based on
417 Phyre2 and HHpred reveals that NgAgo has an uncharacterized repA domain, a truncated N-terminal
418 domain, a MID domain, and a PIWI domain. **c**, Phylogenetic analysis of repA-containing pAgos (orange
419 shaded) found from BLASTP against all isolates via JGI-IMG portal and other characterized pAgos. **d**, The
420 catalytic tetrad of NgAgo is conserved with catalytically active pAgos including MjAgo, PfAgo, MpAgo, and
421 TtAgo in sequence alignment. **e**, All residues of the catalytic tetrad (D663, E704, D738, and D863) DEDD,
422 except E704 are structurally colocalized with the catalytic tetrad of MjAgo (D504, E541, D570, and D688).

423

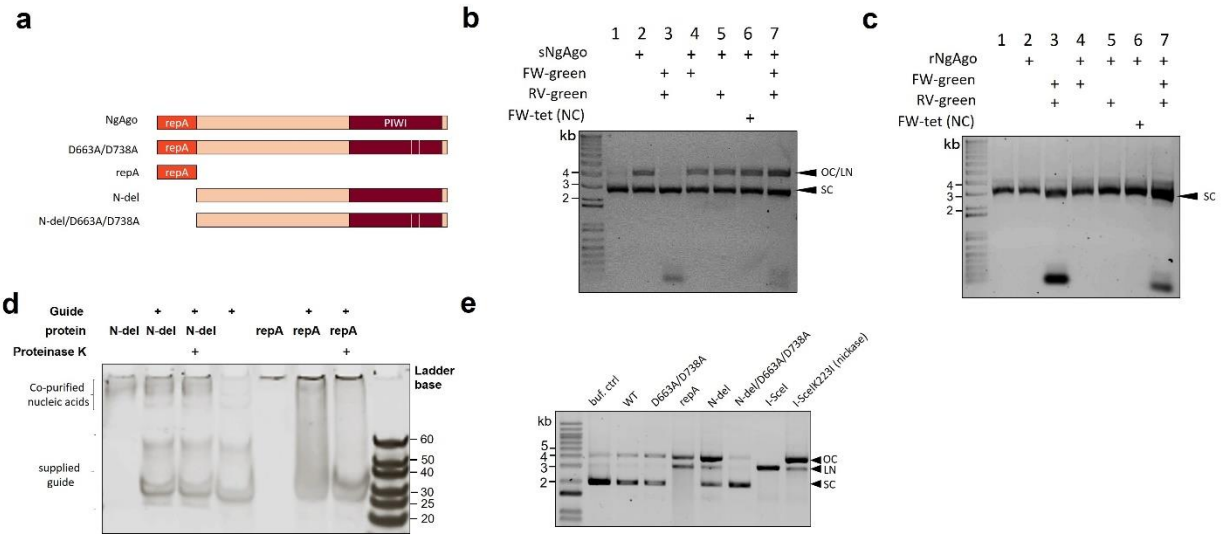


Figure 2 | Soluble NgAgo variants nick and cut plasmid DNA *in vitro* via repA and D663/D738 mutations in the PIWI domain. a, NgAgo variants used in the *in vitro* assay to identify which domain is essential for nicking and cleaving activity. **b**, Soluble NgAgo (sNgAgo) nicks and cuts plasmid DNA regardless of the presence of guide DNA. **c**, Refolded NgAgo, rNgAgo, has no effect on plasmid DNA. **d**, Electrophoretic mobility shift assay (EMSA) of N-del and repA domain with guides. N-del copurifies with nucleic acids and does not bind (shift) supplied guide. repA does not copurifiy with nucleic acid and readily binds and shifts supplied guide, confirming its single-stranded DNA binding ability. **e**, Plasmids were treated with NgAgo variants for an hour before analysis on an agarose gel. Wildtype and D663A/D738A incompletely nicks plasmids DNA while repA and N-del nick and cleave plasmids DNA. N-del/D663A/D738A loses the ability to nick and cleave. I-SceI and I-SceI K223I are used as positive cleavage and nicking controls, respectively. OC, open circular; LN, linear; SC, supercoiled.

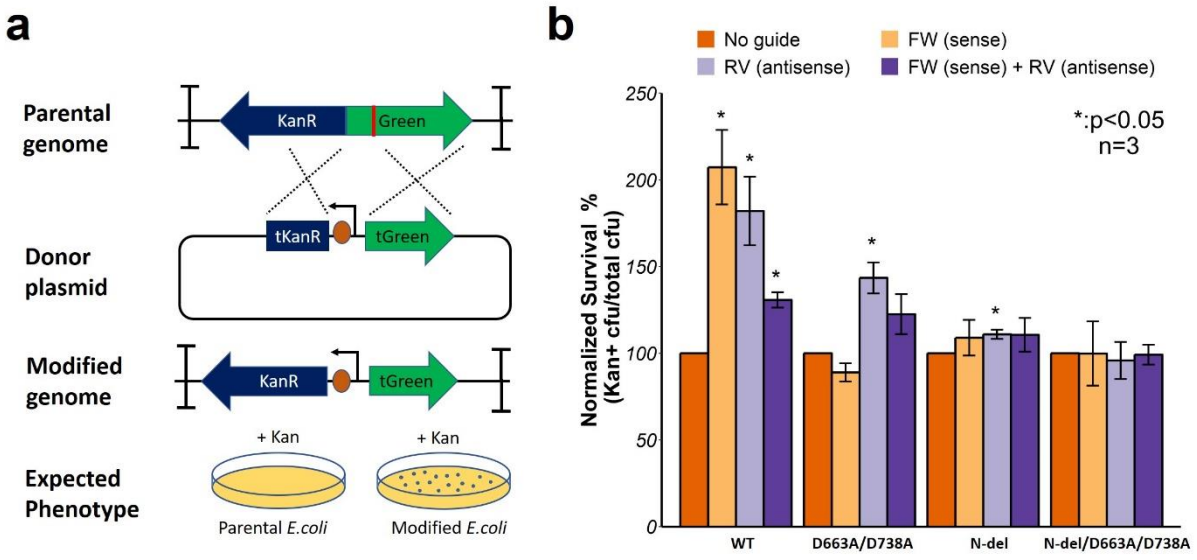


Figure 3 | NgAgo enhances gene-editing via λ-red-mediated homologous recombination in *E.coli*. a, Design of gene-editing assay in MG1655 (DE3). *KanR* and *mNeonGreen* (Green) cassette without promoter and RBS, flanked by two double terminators, is integrated in MG1655 (DE3). Donor plasmid with truncated *mNeonGreen* (tGreen) encodes a nonfunctional truncated *KanR* (tKanR). Guide was transformed to target the *mNeonGreen* (red line). After successful gene editing, modified genome has a functional KanR cassette, enabling survival in Kan selective plate. **b**, NgAgo variants enhance gene editing efficiency with ~1 microgram of guide(s) relative to an unguided control. Error bars are the standard errors generated from three replicates. Statistically significant results are indicated with * (p-value< 0.05, paired t-test)

448 **Table 1. Strains and Plasmids**

Name	Relevant genotype	Vector backbone	Plasmid origin	Source
Strains				
BL21 (DE3)	F– ompT gal dcm lon hsdSB(rB–mB–) λ (DE3) [lacI lacUV5-T7p07 ind1 sam7 nin5]) [malB+]K-12(λS)			[29]
MG1655 (DE3)	K-12 F– λ– ilvG– rfb-50 rph-1 (DE3)			[30]
MG1655 (DE3) *atpI*::KanR-mNeonGreen	K-12 F– λ– ilvG– rfb-50 rph-1 (DE3) *atpI*::KanR-mNeonGreen			This study
Plasmids				
pBSI-SceI(E/H)	*bla*		ColE1 derivative	[31]
pET32a-GST-ELP64	*bla,* lacI, P_{T7}-GST-ELP64			Professor Xin Ge (University of California, Riverside)
pTKDP-hph	*bla*, *hph*, sacB		pMB1	[32]
pCas9-CR4	*cat*, P_{Tet}-Cas9		p15A	[33]
pET-GST-Ago-His	*bla*, lacI, P_{T7}-GST-NgAgo-His	pET32a-GST-ELP64	pBR322	This study
pET32a-His-Ago	*bla*, lacI, P_{T7}-GST-NgAgo-His	pET32a-GST-ELP64	pBR322	This study
pET32a-His-repA	*bla*, lacI, P_{T7}-His-repA	pET32a-GST-ELP64	pBR322	This study
pET-GST-N-del-His	*bla*, lacI, P_{T7}-GST-N-del-His	pET32a-GST-ELP64	pBR322	This study
pET-GST-N-del/D663A/D738A-His	*bla*, lacI, P_{T7}-GST- N-del/D663A/D738A -His	pET32a-GST-ELP64	pBR322	This study
pTKDP-KanR-mNeonGreen-hph	*bla*, *hph*, KanR-mNeonGreen	pTKDP-hph	pMB1	This study
p15-KanR-PtetRed	*cat*, KanR-mNeonGreen, P_{Tet}-gam-beta-exo	pCas9-CR4	p15A	This study
pET32-BFP	*Amp,* lacI, P_{T7}-BFP	pET32a-GST-ELP64 and pBAD-mTagBFP2	pBR322	This study

449