

**DEVELOPMENT OF AN ASSAY TO IDENTIFY AND QUANTIFY  
ENDONUCLEASE ACTIVITY**

by

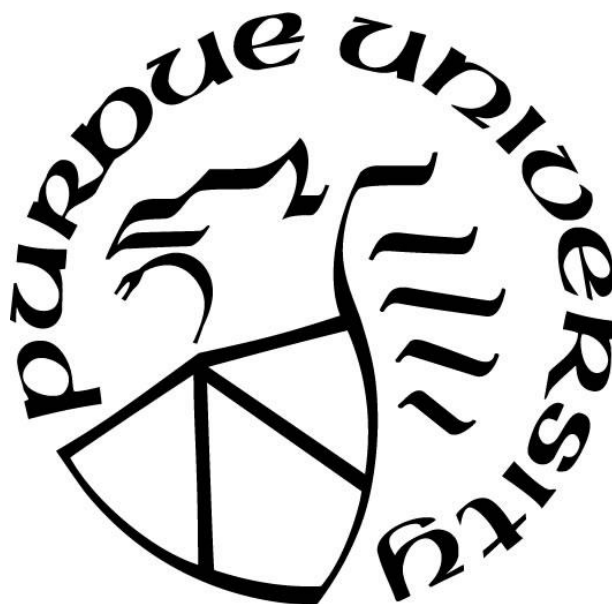
**Michael A. Mechikoff**

**A Thesis**

*Submitted to the Faculty of Purdue University*

*In Partial Fulfillment of the Requirements for the degree of*

**Master of Science**



School of Agricultural and Biological Engineering

West Lafayette, Indiana

December 2019

**THE PURDUE UNIVERSITY GRADUATE SCHOOL**  
**STATEMENT OF COMMITTEE APPROVAL**

**Dr. Kevin V. Solomon, Chair**

School of Agricultural and Biological Engineering

**Dr. Nathan Mosier**

School of Agricultural and Biological Engineering

**Dr. Kari Clase**

School of Agricultural and Biological Engineering

**Approved by:**

Dr. Nathan Mosier

*Dedicated to my family and girlfriend whose love and support were unfaltering. This thesis is also dedicated to my lab mate and friend, Kok Zhi Lee, for his patience and help with educating me on valuable lab techniques.*

## ACKNOWLEDGEMENTS

I thank my advisor and mentor, Dr. Kevin V. Solomon, for his guidance and patience with this project. Without his expertise and dedication to my work, this project would have never occurred. I am extremely grateful to have studied under him and will always consider him a valuable mentor.

I also thank my lab mates for their valued help in the lab, which contributed greatly to the success of this project. Specifically, I would like to thank Kok Zhi Lee, Dr. Jacob Englaender, Ethan Hillman, and Casey Hooker.

I thank those who contributed resources for my project. I thank Dr. Frederick Gimble for his generous donation of the pEndoSceWT and pEndoSce-D44S plasmids. I thank Dr. Kristala Prather for her generosity in gifting our lab the MG1655 (DE3) *E. coli* strain. I thank Dr. Laszlo Csonka for his guidance on selection systems and his knowledge of Tetracycline resistance in bacteria, as well as his gift of Tet<sup>r</sup> plasmids. I also thank the Ralph W. and Grace M. Showalter Research Trust and Colleges of Engineering and Agriculture Purdue Research Foundation for their funding of this research.

Finally, I thank the United States Air Force Academy and the Air Force Institute of Technology for providing me with a scholarship to attend graduate school.

# TABLE OF CONTENTS

LIST OF TABLES .....	7
LIST OF FIGURES .....	8
ABBREVIATIONS .....	10
ABSTRACT.....	11
1. INTRODUCTION .....	12
1.1 Motivation.....	12
1.2 Plasmid Vectors .....	12
1.3 Chromosomal Integrations.....	14
1.4 Programmable Endonucleases .....	16
1.5 Directed Evolution and Selection Screens.....	18
1.6 Project Objectives .....	21
1.7 Assumptions.....	21
2. A SENSITIVE <i>IN VIVO</i> ENDONUCLEASE ACTIVITY ASSAY .....	22
2.1 Background .....	22
2.2 Materials and methods .....	22
2.2.1 Strain Construction .....	24
2.2.2 Plasmid construction.....	24
2.2.3 I-SceI lethality assay.....	26
2.2.4 Nuclease activity assay .....	26
2.2.5 Statistical analysis.....	27
2.3 Results.....	27
2.3.1 Design and construction of selection system.....	27
2.3.2 Optimization of selection system sensitivity .....	27
2.3.3 Validation of targeted endonuclease activity.....	31
2.4 Discussion .....	34
3. CONCLUSIONS AND FUTURE DIRECTIONS .....	40
3.1 Summary of Current Progress.....	40
3.2 Modification of the system to rank commercially available plasmids .....	41
3.3 Application of NgAgo for Future Work .....	42

APPENDIX A.....	44
REFERENCES .....	47
PUBLICATION.....	52

## LIST OF TABLES

<b>Table 1.1</b> Generalized comparison of various genome engineering tools. Table and legend modified from original source <sup>24</sup> under CC BY 4.0 License. ....	17
<b>Table 2.1</b> Strains and plasmids.....	25
<b>Table A1.</b> Oligonucleotides used. Restriction sites are in upper case.....	44
<b>Table A2.</b> Values for paired Student's T-Test comparing endonucleases when induced (total rescue) and uninduced (total death).....	44

## LIST OF FIGURES

**Figure 1.1.** Plasmid inheritance with respect to incompatibility groups, assuming copy number for each plasmid is 2. .... 13

**Figure 1.2.** Chromosomal integration. Target DNA (chromosome) is cut using an endonuclease. The break is repaired using a donor piece of DNA which has sequences on either end that are homologous to the target DNA. Cellular machinery then uses homologous recombination to repair the break using the donor DNA. .... 15

**Figure 1.3.** An example of directed evolution with comparison to natural evolution. The inner cycle indicates the 3 stages of the directed evolution cycle with the natural process being mimicked in brackets. The outer circle demonstrates steps in a typical experiment. The red symbols indicate functional variants, the pale symbols indicate variants with reduced function. Reproduced without edit under the Creative Commons license 4.0. <sup>38</sup> ..... 19

**Figure 2.1. Design of selection system.** a) Overview of selection system. b) Plasmid designs used in the selection system. The lethal plasmid encodes for the homing endonuclease, I-SceI, which targets a modified *E. coli* MG1655 (DE3) genome. I-SceI is under the control of the arabinose-inducible promoter, araBAD, and the lethal plasmid has an AmpR selectable marker and a ColE1 origin of replication (~20 copies). The rescue plasmid encodes the endonuclease of choice under the control of a P<sub>Tet</sub> promoter, inducible with aTc. The rescue plasmid also has corresponding endonuclease guides targeting major origins of replication, under the control of a rhamnose-inducible rhamBAD promoter. The rescue plasmid has a KanR selectable marker and a ColA origin of replication (~20-40 copies). c) The exponential nature of the *in vivo* system (blue line) amplifies the signal of endonuclease activity compared to linear, *in vitro* activity assays (orange line). Small differences in endonuclease activity will be magnified in an *in vivo* assay. Cells containing the I-SceI plasmid (I-SceI<sup>+</sup>) will die off, leaving cells without the I-SceI plasmid (I-SceI<sup>-</sup>) to live. Endonuclease activity is linked to cell survival because the endonuclease targets and cures the I-SceI plasmid, resulting in I-SceI<sup>-</sup> cells. .... 29

**Figure 2.2. I-SceI is lethal in engineered KS 165.** a) *E. coli* MG1655 (DE3) *nth::tetA* (KS 165) was transformed with the wild type I-SceI plasmid, pEndoSceWT, or the catalytic mutant, pEndoSceD44S. Cultures were grown with either 0 or 10mM arabinose for induction of wild type and mutant I-SceI. All cultures were able to form healthy colonies after 16 hours at 37°C although those with the induced wild type I-SceI plasmid were observably smaller, suggesting I-SceI on the pEndoSceWT plasmid induces some cell death via a DSB in the host genome. **b-c)** Cells harbouring the pEndoSceWT plasmid were allowed to grow for 24 hours (b) or 4 hours (c), followed by a 4 hour I-SceI induction period. Cultures were grown with and without ampicillin to test plasmid retention in the presence (or absence) of a selective pressure. Cells grown for 24 hours without the selective pressure lost some of the plasmid whereas cells grown for 4 hours without selective pressure did not lose the plasmid. **d)** Cells harbouring the higher copy number plasmid, pColE1-I-SceI, were allowed to grow for 4 hours, followed by a 4 hour I-SceI induction period. pColE1-I-SceI was able to reduce cellular growth by 84.1%. .... 33

**Figure 2.3. Endonuclease activity recovers growth.** a) Cell growth was recovered using wild type SpCas9 enzyme. b) Comparison of commonly used endonucleases. Each endonuclease was



given four hours of induction and targeted toward the lethal plasmid, after which I-SceI of the lethal plasmid was induced to cause cell death. Each endonuclease significantly rescued cell growth (unpaired t-test,  $p < 0.05$ , standard error,  $n=3$ )..... 35

**Figure 3.1.** Modified with permission<sup>30</sup>. NgAgo cuts DNA *in vitro*. Soluble NgAgo (sNgAgo) nicks and cuts plasmid DNA nonspecifically without the presence of guide (FW-green and RV-green), as indicated by the open circular and linear DNA (OC/LN). Without the presence of NgAgo, plasmid DNA remains in its supercoiled form (SC). ..... 42

**Figure A1.** Timeline for endonuclease activity assay. .... 45

**Figure A2.** eSpCas9 was used to rescue cell growth in our endonuclease activity assay. Cas refers to the endonuclease. + and – indicate induced and uninduced conditions, respectively. Data displayed is the mean of 3 replicates, error represented as standard error. .... 45

**Figure A3.** xCas9 was used to rescue cell growth in our endonuclease activity assay. Cas refers to the endonuclease. + and – indicate induced and uninduced conditions, respectively. Data displayed is the mean of 3 replicates, error represented as standard error. .... 46

## ABBREVIATIONS

aTc – Anhydrotetracycline

Cas – CRISPR associated proteins

CRISPR – Clustered, regularly interspaced, palindromic repeats

ddH<sub>2</sub>O – double distilled water

DE3 – prophage derived from bacteriophage carrying the T7 RNA polymerase gene

DSB – Double stranded break

*E. coli* – *Escherichia coli*

eSpCas9 – enhanced specificity Cas9

Kb – Kilobases

KS 165 – *E. coli* strain MG1655 (DE3) with I-SceI recognition site in the genome

LB – Luria-Bertani medium

MCS – Multiple cloning site

mM – millimolar

NgAgo – Argonaute from *Natronobacterium gregoryi*

NiCl<sub>2</sub> – Nickel chloride

OD 600 – Optical density taken at wavelength 600nm

pAgo – Prokaryotic Argonaute

PAM – Protospacer adjacent motif

SpCas9 – Cas9 enzyme from *Streptococcus pyogenes*

TALEN – Transcription activator-like effector nuclease

Tc – Tetracycline

Tc<sup>r</sup> – Tetracycline resistance

tRNA – Transfer ribonucleic acid

xCas9 – derivative of Cas9 with broader PAM site recognition

ZFN – Zinc-finger nuclease

## ABSTRACT

Synthetic biology reprograms organisms to perform non-native functions for beneficial reasons. An important practice in synthetic biology is the ability to edit DNA to change a base pair, disrupt a gene, or insert a new DNA sequence. DNA edits are commonly made with the help of homologous recombination, which inserts new DNA flanked by sequences homologous to the target region. To increase homologous recombination efficiency, a double stranded break is needed in the middle of the target sequence. Common methods to induce double stranded breaks use nucleases, enzymes that cleave ribonucleotides (DNA and RNA). The most common nucleases are restriction enzymes, which recognize a short, fixed, palindromic DNA sequence (4-8 base pairs). Because of the short and fixed nature of the recognition sites, restriction enzymes do not make good candidates to edit large chromosomal DNA. Alternatively, scientists have turned to programmable endonucleases which recognize user-defined DNA sequences, often times much larger than the recognition sites of restriction enzymes (15-25 base pairs). Programmable endonucleases such as CRISPR-based systems and prokaryotic Argonautes are found throughout the prokaryotic kingdom and may differ significantly in activity and specificity. To compare activity levels among endonuclease enzymes, activity assays are needed. These assays must clearly delineate dynamic activity levels of different endonucleases and work with a wide variety of enzymes. Ideally, the activity assay will also function as a positive selection screen, allowing modifications to the enzymes via directed evolution. Here, we develop an *in vivo* assay for programmable endonuclease activity that can also serve as a positive selection screen using two plasmids, a lethal plasmid to cause cell death and a rescue plasmid to rescue cell growth. The lethal plasmid houses the homing endonuclease, I-SceI, which causes a deadly double-stranded break at an 18 base pair sequence inserted into an engineered *E. coli* genome. The rescue plasmid encodes for a chosen endonuclease designed to target and cleave the lethal plasmid, thereby preventing cell death. With this, cell growth is directly linked to programmable endonuclease activity. Three endonucleases were tested, SpCas9, eSpCas9, and xCas9, displaying recovered growth of 49.3%, 26.1%, and 16.4% respectively. These values translate to kinetic enzymatic activity and are congruent with current literature findings as reported values find WT-SpCas9 to have the fastest kinetics cleaving around 95% of substrate within 15 seconds, followed closely by eSpCas9 cleaving 75% of substrate within 15 seconds and finally trailed by xCas9 cleaving 20% of substrate in about 30 seconds. The differences between each endonuclease's activity is exacerbated in our *in vivo* system when compared to similar *in vitro* methods with much lower resolution. Therefore, slight differences in activity between endonucleases within the first few minutes in an *in vitro* assay may be a few percentages different whereas in our *in vivo* assay, these differences in activity result in a more amplified signal. With the ability to display the dynamic response of enzymes, this assay can be used to compare activity levels between endonucleases, give insight into their kinetics, and serve as a positive selection screen for use in directed evolution applications.

# 1. INTRODUCTION

## 1.1 Motivation

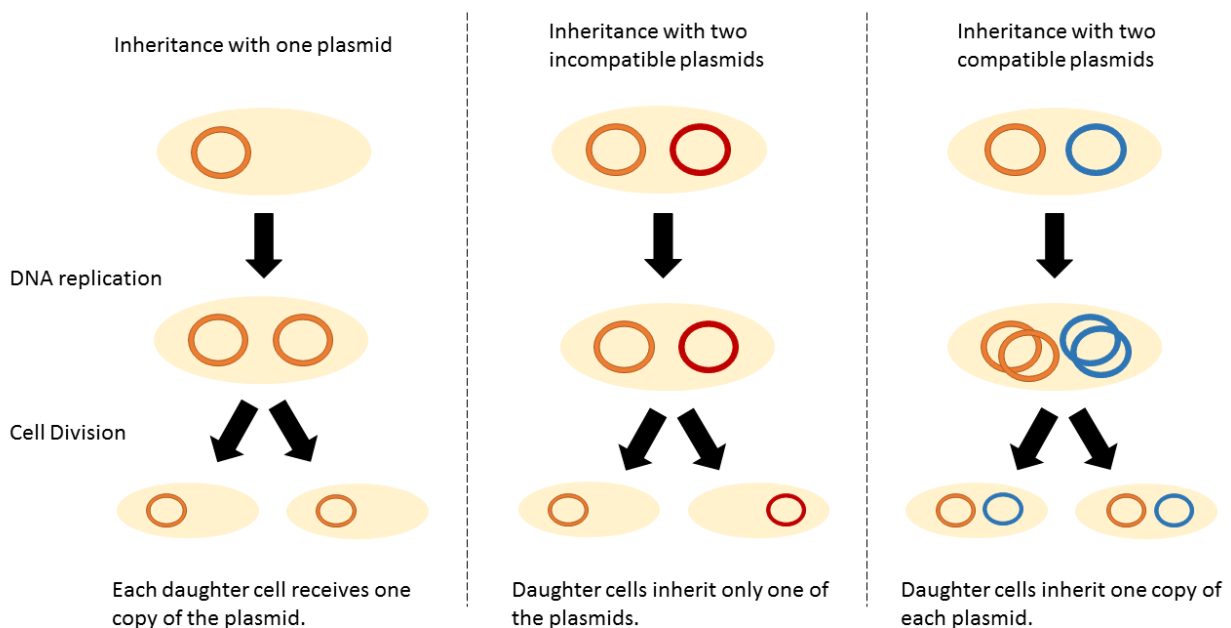
Synthetic biology is a field of biology that engineers organisms to have new abilities, thereby redesigning them for useful purposes<sup>1</sup>. Through synthetic biology techniques, researchers have been able to revolutionize aspects of agriculture, biofuels, and medicine. For example, scientists have created pest-resistant plants that increase crop yields and improve the quality of the crop<sup>2</sup>, they have produced biodiesels from a number of different hosts<sup>3</sup>, and they have produced and harvested human insulin in bacteria for use in diabetic patients,<sup>4</sup> which has been widely studied and improved over the years<sup>5,6</sup>. An important technique used in synthetic biology is the ability to edit DNA using recombinant DNA technology. This technology introduces a break in the target DNA which can then be recombined with a donor, often foreign, piece of DNA. The organism then treats this foreign DNA as its own and produces the protein for which the DNA codes for. However, creating this break can be tricky and depending upon the type of DNA one is working with, the technique may be different. In the *Escherichia coli* bacterium, there are two main types of DNA, chromosomal and plasmid DNA. Chromosomal DNA is much larger in sequence and contains almost all of the genes necessary for an organism to live. Plasmid DNA, on the other hand, is much smaller in terms of nucleotide length, often less than 1500 kilobases (Kb) naturally<sup>7</sup> and often around 3Kb in a laboratory setting<sup>8</sup>. Working with either plasmid or chromosomal DNA will dictate the techniques used to produce recombinant DNA.

## 1.2 Plasmid Vectors

Plasmid vectors are extremely common for producing recombinant DNA in *E. coli* for a few reasons. First, plasmids have been shortened to around 3Kb, which is drastically shorter than those found naturally in *E. coli*<sup>8</sup>. Because of this, plasmid DNA is easier to purify and many plasmids have been designed specifically for molecular cloning purposes. For example, most plasmids incorporate a multiple cloning site (MCS) which includes a large number of different restriction sites, recognized by restriction enzymes<sup>9</sup>. Restriction enzymes are nucleases which recognize a short, palindromic sequence, and induce a double stranded break (DSB) at the restriction site. Many restriction enzymes will make a sticky end for their DSB, which leaves a

single-stranded oligonucleotide overhang. Cutting the plasmid and insert DNA with the same restriction enzymes ensures the sticky ends are complimentary and allows for accurate cloning of insert DNA. Moreover, the design of the plasmid can ensure only desired restriction sites are present and therefore the restriction enzyme will not cut at unintended locations, preventing the integration of insert DNA into undesired locations.

Second, plasmids replicate independently from the host chromosome. When the cell divides, each daughter cell usually gets a copy of chromosomal DNA and a copy of plasmid DNA. This independent replication sparked vast interest in studying cellular replication machinery, leading to the discovery and development of new origins of replication. These origins of replication are crucial for the independent replication of plasmid DNA as they are each governed by a set of replication machinery different from the machinery used to replicate the chromosome<sup>9</sup>. However, origins of replication fall in groups, and each group is governed by its own replication machinery<sup>10</sup>. The replication machinery must be different in order for the plasmids to be stably inherited together<sup>10</sup>. Therefore, to have more than one type of plasmid in the cell, the origins of replication must be in different incompatibility groups (**Figure 1.1**).

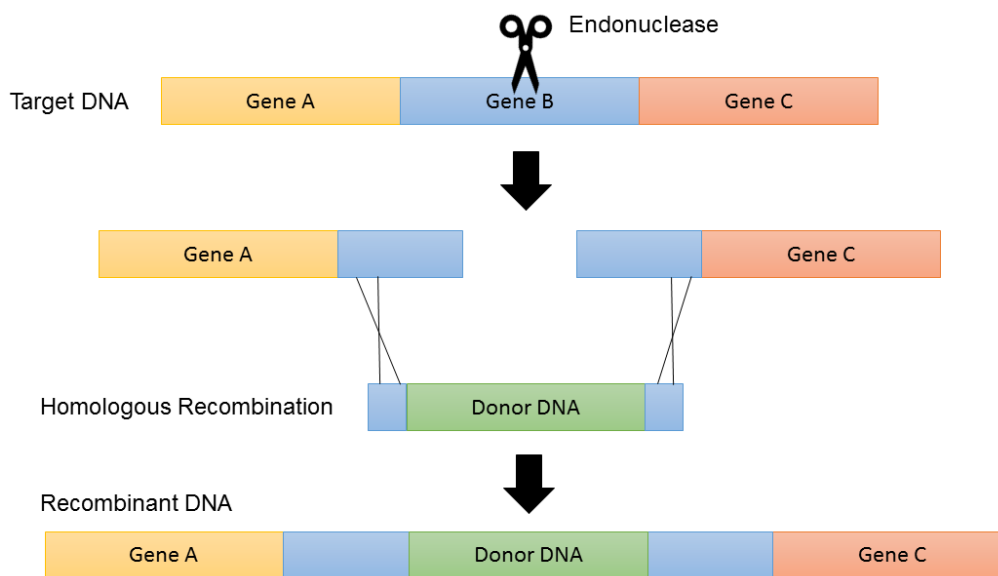


**Figure 1.1.** Plasmid inheritance with respect to incompatibility groups, assuming copy number for each plasmid is 2.

However, during cellular replication, plasmids are not always divided and inherited equally<sup>11</sup>. To ensure each daughter cell ends up with a plasmid, the plasmid must have a selectable marker which allows for artificial selection of cells harboring the plasmid. One of the most common types of selectable markers in bacteria is the use of antibiotic resistance genes<sup>12</sup>. Inserting an antibiotic resistance gene into a plasmid allows the scientist to select for cells that have the plasmid by simply growing the plasmid in the presence of that antibiotic. Cells not harboring the plasmid will die off, while those with the plasmid will live. Though plasmid vectors are extremely useful, industrial applications which grow cultures in large quantities would require a large supply of antibiotics which is expensive. Moreover, each plasmid needs its own selective marker which drastically increases the cost for large batches of culture harboring more than one plasmid. Therefore, alternative approaches incorporating recombinant DNA technology are sought after.

### **1.3 Chromosomal Integrations**

To avoid plasmid inheritance issues, researchers can integrate the donor DNA into the host chromosome. Because the chromosome is duplicated and each daughter cell receives one copy of the chromosome, scientists can ensure the donor DNA is inherited into each daughter cell and propagated with each cell division. Typical chromosomal integrations work by creating a DSB and repairing this break with a donor piece of DNA via homologous recombination (**Figure 1.2**).



**Figure 1.2.** Chromosomal integration. Target DNA (chromosome) is cut using an endonuclease. The break is repaired using a donor piece of DNA which has sequences on either end that are homologous to the target DNA. Cellular machinery then uses homologous recombination to repair the break using the donor DNA.

Cloning a gene into the host chromosome though, is not as simple as a plasmid. Researchers do not have the luxury of using restriction enzymes due to their small recognition sites. These recognition sites, by chance, will show up every so often in the very large sequences of chromosomes. So, if a restriction enzyme is used to cleave the 4.6 mega-base pair sequence of the *E. coli* genome<sup>13</sup>, and the restriction site appears every 10,000 base pairs by chance, roughly 463 fragments will be generated, each of which could bind to the sticky end of the donor DNA. With this many possibilities, finding the correctly assembled clone would be a daunting task. Thus, researchers have turned to different nucleases for chromosomal integrations. One approach is to use a homing endonuclease, which has a much larger recognition site<sup>14</sup>. These sites can be 18 base pairs or longer and are absent in many bacterial genomes. However, these homing endonucleases still require a specific sequence for cleavage that is not user-defined, which necessitates the engineering of the genome to incorporate the long recognition site. These limitations have led researchers to find new approaches to make chromosomal edits.

## 1.4 Programmable Endonucleases

One prominent solution to increasing the ability to make precise chromosomal edits is the use of programmable endonucleases. These enzymes do not have canonical sequences they cleave like restriction enzymes do, but rather can be programmed or targeted to cleave a given sequence. This unique ability allows these enzymes to access a much larger area of the genome and gives the user more control over regions they edit. Scientists have been attempting to discover new endonucleases for decades as these enzymes pose to be some of the most powerful tools for genetic engineering<sup>15</sup>. Some have attempted to synthetically create new endonucleases<sup>16</sup>, while others have relied on nature to provide them. Current popular programmable endonucleases include transcription activator-like effector nucleases (TALENs)<sup>17</sup>, Zinc Finger Nucleases (ZFNs)<sup>18</sup>, clustered, regularly interspaced, short, palindromic repeats (CRISPR)/CRISPR-associated (Cas) systems<sup>19</sup>, and prokaryotic Argonautes (pAgos). Each endonuclease has its advantages and limitations (**Table 1.1**). TALENs work by fusing a transcription activator-like effector DNA binding domain to a DNA-cleaving nuclease<sup>20</sup>. The binding domain recognizes a single nucleotide, and the combination of the binding domain and cleavage domain allows the scientist to direct the nuclease to a given location and cleave a specified DNA sequence. Similarly, ZFNs work by fusing a zinc finger domain with a nuclease. The zinc finger domain recognizes nucleotide triplets<sup>21</sup> and can be designed to bind specific DNA sequences while the nuclease cleaves the DNA, again allowing the scientist to specify the location of DNA cleavage. CRISPR systems use Cas enzymes to first bind guide RNA which then directs the enzyme to a specific DNA location<sup>19</sup>. The Cas enzyme then cleaves the DNA at that location. pAgos use two 5'-phosphorylated DNA guides<sup>22</sup> which targets the enzyme to a specified location and also induces a conformational change<sup>23</sup>, allowing the enzyme to cleave the target DNA.



**Table 1.1** Generalized comparison of various genome engineering tools. Table and legend modified from original source<sup>24</sup> under CC BY 4.0 License.

<b>Nuclease platform</b>	<b>ZFN</b>	<b>TALEN</b>	<b>CRISPR/Cas</b>	<b>pAgo</b>
<b>Source</b>	Bacteria, Eukaryotes	Bacteria ( <i>Xanthomonas</i> sp.)	Bacteria ( <i>Streptococcus</i> sp.)	Archaea, Bacteria
<b>Number of component(s)</b>	2	2	1–2 (depends)	2
<b>Type of recognition</b>	Protein-DNA	Protein-DNA	RNA-DNA	RNA-DNA, DNA-DNA
<b>Recognition site (bp)</b>	18–36	24–40	17–23	13–25
<b>Double strand break pattern</b>	Staggered cut (4–5 nt, 5' overhang)	Staggered cut (Heterogeneous overhangs)	SpCas9 creates blunt ends	User-defined
<b>Function</b>	Nuclease, Nickase	Nuclease, Nickase	Nuclease, Nickase	Nickase
<b>Best suited for</b>	Gene knockout, Transcriptional regulation	Gene knockout, Transcriptional regulation	Gene knockout, Transcriptional regulation, Base editing	Gene knockout, Gene knock-in
<b>Ease of design</b>	Difficult	Moderate	Easy	Easy
<b>Dimerization required</b>	Yes	Yes	No	Depends <sup>a</sup>
<b>Ease of generating large scale libraries</b>	Laborious	Moderately laborious	Easy	Easy
<b>Specificity</b>	Low–Moderate	Moderate	Low–Moderate	Low–Moderate
<b>Improved/other versions</b>	AZP-SNase	Tev-mTALEN	Cpf1, eSpCas9, xCas9	Many
<b>Cost (USD)</b>	5–10,000	< 1000	< 100	<100
<b>Targeting constraints</b>	Non-guanosine rich sequence hard to target	5' targeted base must be thymine for each TALEN monomer	PAM sequence must follow target site	Depends <sup>b</sup>
<b>Methylation sensitive</b>	Yes	Yes	No	No
<b>First use in human cells</b>	2003	2011	2013	2016 <sup>c</sup>

\*All information for ZFN, TALEN, and CRISPR/Cas in this table is from original source.<sup>24</sup>

\*\*CRISPR/Cas information is specific for the most widely used Cas9 nuclease, SpCas9, from *Streptococcus pyogenes*.

\*\*\*pAgo data is for long pAgos involved in nucleic acid-guided host defense mechanisms. Data attempts to generalize for all pAgos though it could be inaccurate for some pAgos due to their diverse nature.

For pAgo data: Source<sup>23</sup>, Number of components<sup>23,25,26</sup>, Type of recognition<sup>23,25</sup>, Recognition site<sup>26,27</sup>, Double strand break pattern<sup>23,28</sup>, Function<sup>22,26,29,30</sup>, Best suited for<sup>30,31</sup>, Ease of generating large scale libraries<sup>26</sup>, Specificity<sup>23,29-31</sup>, Methylation sensitive<sup>28</sup>.

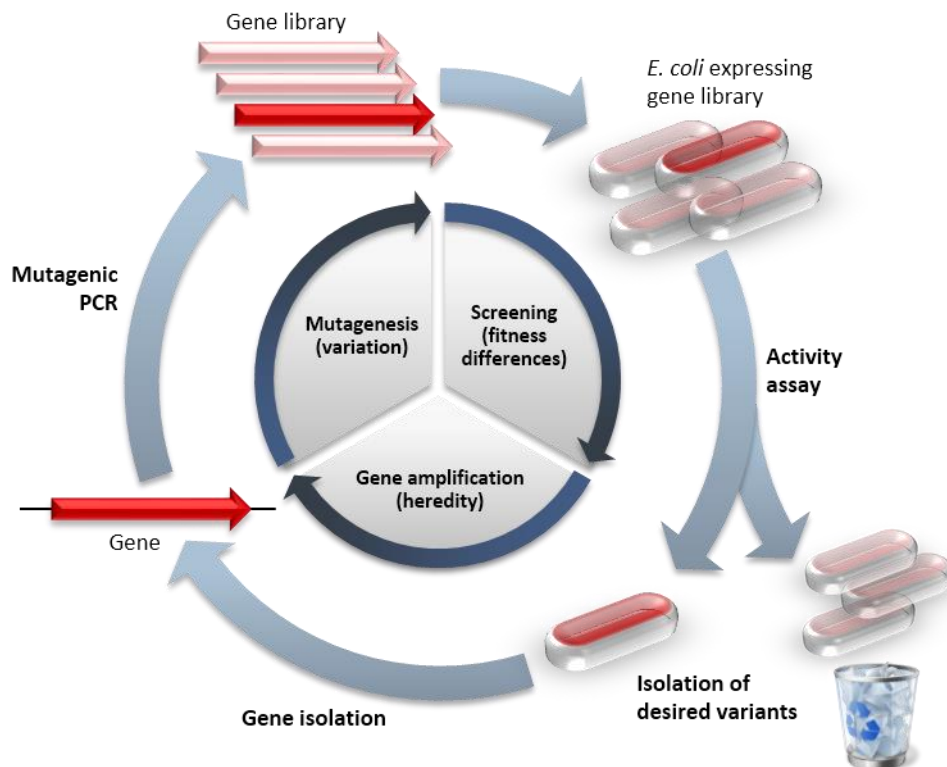
- a. Some pAgos act as nickases which would require two proteins to create a DSB, while other pAgo nucleases can act as monomers to create a DSB.
- b. Some pAgos cleave DNA and RNA<sup>26</sup>, while others only cleave DNA<sup>29,30</sup>. Some pAgos display bias toward target sequences<sup>26</sup>.
- c. NgAgo was used to edit genes in human cell lines<sup>32</sup>, though this study is highly disputed and has since been retracted.

Although these powerful tools have widely expanded genomic editing abilities, they are not without limitation. ZFNs and TALENs require difficult and timely engineering of the protein complex to properly bind and cleave target sequences. Moreover, ZFN and TALEN nucleases only function as dimers<sup>21</sup>. Therefore, two proteins must be designed, one that recognizes the sequence upstream of the target and one that recognizes the sequence downstream of the target. Cas enzymes, while easier to work with, require protospacer adjacent motifs (PAM) for site recognition<sup>19</sup>. Therefore, Cas enzymes cannot bind to a sequence unless it is adjacent to a PAM site. pAgos, on the other hand, do not need to be specifically engineered for each target nor do they require a PAM site to bind and cleave DNA. However, most characterized pAgos are thermophilic and thus do not function well in temperatures more suitable for biotechnology applications<sup>26,27,29,33</sup>. Because of these limitations, scientists are still busy trying to discover new endonucleases. However, this discovery process requires an assay that clearly delineates endonuclease activity, even at lower levels. To my knowledge, until now there lacked a simple and effective assay that quickly identifies endonuclease activity.

## 1.5 Directed Evolution and Selection Screens

An alternative approach to finding new endonucleases is to modify and further develop existing endonucleases. A powerful and common method for engineering proteins is through a process called directed evolution. With this process, scientists have been able to increase protein stability in non-native environments<sup>34</sup>, increase binding affinity of antibodies<sup>35</sup>, increase substrate specificity of enzymes<sup>36</sup>, among a number of other applications. Directed evolution works by first adding mutations to a DNA sequence, and after expression in a host, screening or selecting for the mutated protein with desired characteristics (**Figure 1.3**). More desirable proteins are recovered

while less desirable ones are discarded. Then, the recovered proteins can be put through the cycle again, and eventually, a protein with the ideal phenotype/function will emerge<sup>37</sup>.



**Figure 1.3.** An example of directed evolution with comparison to natural evolution. The inner cycle indicates the 3 stages of the directed evolution cycle with the natural process being mimicked in brackets. The outer circle demonstrates steps in a typical experiment. The red symbols indicate functional variants, the pale symbols indicate variants with reduced function. Reproduced without edit under the Creative Commons license 4.0.<sup>38</sup>

While inducing mutations in DNA is rudimentary today, the selection process for recovering enzymes with these mutations is more difficult. Some systems screen for activity using a reporter gene, such as a fluorescent protein, while others make use of nuclear magnetic resonance (NMR) technology or high-performance liquid chromatography (HPLC)<sup>37</sup>. Each of these approaches have their benefits, though they all require screening of hundreds to thousands of colonies. In contrast, screening approaches that link enzyme activity to cell survival have been developed to decrease the total amount of cells needed to be screened. These techniques are called positive selection systems.

Positive selection systems are commonly used to select for mutant proteins in a wide variety of applications. Some argue that these systems may be the best option for screening large libraries of mutants<sup>39</sup>, leading to their frenzied development. In 1998, Reyrat *et. al* proposed a list of counter selectable markers that may make good candidates for selection strategies<sup>40</sup>, and some have been explored extensively. For example, one selection system employs the *tetA* gene which confers tetracycline resistance (Tc<sup>r</sup>) to the cells. The system utilizes the consequences of *tetA* causing hypersensitivity to lipophilic chelating agents, such as fusaric acid<sup>41</sup>. Bochner and colleagues used fusaric acid and quinaldic acid to eliminate cells containing Tc<sup>r</sup> transposons. This same technique was improved a year later by Maloy and Nunn by optimizing the medium to better suit *E. coli* species<sup>42</sup>. Years later, Ryu *et. al* used a similar system to select for successful recombinants in *E. coli*. However, they used NiCl<sub>2</sub> instead of fusaric acid or quinaldic acid to select for properly edited genomes via MAGE, ZFNs, or CRISPR<sup>43</sup>. Although useful, this system requires fine tuning of Tc and NiCl<sub>2</sub> concentrations for each host strain<sup>44</sup>. Moreover, this fine tuning may change as the plasmids within the system change, requiring the user to try different concentrations of each chemical for each plasmid tested.

In 2002, Gruen and colleagues created a positive selection system specifically designed to highlight endonuclease activity<sup>45</sup>. Their system was designed to use two plasmids, one hosting the toxin, Barnase, which kills the cell when induced, and the other hosting the endonuclease of interest. The endonuclease was targeted to the Barnase plasmid and cell survival was linked to successful endonuclease activity. While a powerful tool for endonuclease development, the system uses suppressor tRNAs to control expression and display activity, and it requires the use of two inducer chemicals to induce cell death. Additionally, this system requires the endonuclease to be incorporated as an N-terminal fusion to the suppressor tRNA, which can complicate interactions between DNA and protein.

In direct response to Gruen's Barnase system, Chen *et al.* used a similar two plasmid strategy using a different toxin, CcdB<sup>46</sup>. In this system, the suppressor tRNAs were removed and the requirement of the endonuclease to be fused to the suppressor tRNA was void. This modification made the system simpler because only one protein needed expression to cause cell death rather than two. Guo and colleagues were able to successfully use Chen's system for directed evolution of Zinc

Finger Nucleases<sup>47</sup>, proving the feasibility of the system. However, Chen's system uses CcdB, which is extremely toxic to cells even at low levels<sup>46</sup>. Therefore, even with a tightly regulated promoter, background or leaky expression of CcdB can cause retarded growth and result in much lower cell densities. Moreover, the stringency of the system can result in missed detection of less efficient endonucleases.

## 1.6 Project Objectives

This project is aimed at **creating a new system** that has two primary objectives. First, the system must be able to serve as a preliminary screen to test for endonuclease activity. Ideally, organisms will be collected from nature and screened for different endonucleases. To do this, the organisms' genomes will be sequenced and alignments will be run against known endonuclease classes. Potential hits will then be tested in the system for endonuclease activity. Second, the system must be able to serve as a positive selection system for directed evolution applications. Enzymes with endonuclease activity will be put through rounds of directed evolution to increase their endonuclease activity. The system must permit the recovery of these enzymes if successful endonuclease activity is detected.

## 1.7 Assumptions

In order for this research to be conducted, some assumptions regarding the basic functionality of biological parts as well as scientific processes must be made.

The assumptions for this research project are as follows:

1. Cell death is due to I-SceI cleavage of the host genome, and not due to mutation.
2. Rescued cells are due to functional endonuclease activity and not due to mutation in either the genome or I-SceI gene.
3. Cell growth is pure and not due to contamination. Suspected contamination will be tested.
4. miRNA or siRNA are not interacting with the I-SceI gene, recognition site, or endonuclease of interest.

## 2. A SENSITIVE *IN VIVO* ENDONUCLEASE ACTIVITY ASSAY

This chapter is modified from a paper in preparation by Michael A. Mechikoff, Kok Zhi Lee, Paula Pandolfi, Kevin Fitzgerald, Ethan Hillman, and Kevin V. Solomon.

### 2.1 Background

Programmable endonucleases are important enzymes for modern biotechnology that can be targeted to cleave specific DNA sequences, enabling sequence modification. CRISPR/Cas9 from *Streptococcus pyogenes* (SpCas9) is the most widely used and studied programmable endonucleases due to its ease of expression and high activity in a wide variety of hosts. However, its use is restricted to regions adjacent to a defined NGG protospacer adjacent motif (PAM), which it modifies with variable efficiency dictated by the target sequence. As a result, different variants of SpCas9 have been engineered to optimize its properties. For example, xCas9, an engineered variant of SpCas9, has been developed to recognize a diversity of PAMs including NG, GAA and GAT, as opposed to NGG<sup>48</sup>. Another variant, eSpCas9, was developed with enhanced specificity, eliminating the off-target activity of the wildtype SpCas9<sup>49</sup>. While the availability of these SpCas9 variants greatly expand our ability to edit varying regions with increasing amounts of specificity, the rates, and thus efficiencies, of these enzymes are challenging to measure.

Current methods of assessing activity rely on *in vitro* characterization, which involves protein expression, purification, and activity assays. This method of *in vitro* characterization is laborious, time-consuming, and not suitable for high-throughput characterization of variants made via directed evolution. In addition, *in vitro* characterization excludes endonucleases that are difficult to be isolated with high purity and yield. All these limitations slow down the process of characterization of endonuclease variants and newly found endonucleases, impeding the development of programmable endonucleases.

*In vivo* characterization, however, overcomes the shortcomings of *in vitro* characterization by linking cell phenotypes to endonuclease activity, which allows users to rapidly characterize better performing endonucleases variants without tedious protein purification. Ideally, an endonuclease

activity assay *in vivo* would display low levels of endonuclease activity and limit the possibility for false negatives. The system should also link a visible phenotype to endonuclease activity, allowing for quick identification of activity level. Preferably, the assay would be quick and give some insight into the kinetics of the enzymes tested. Moreover, an *in vivo* system can more easily capture the dynamic state of the enzyme compared to an *in vitro* system. Specifically, quicker acting or more efficient enzymes would cause an earlier phenotypic difference. Linking enzymatic activity to cell survival amplifies signal from the assay, as bacterial cell growth is exponential, better highlighting kinetics differences between enzymes. In contrast, *in vitro* systems require constant monitoring to capture enzyme kinetics and, depending upon the turnover rate of the enzyme and amount of substrate, may only be able to identify the steady state response of the enzyme, which is likely different *in vivo*.

Current *in vivo* assays are not ideal to assess endonuclease enzymes due to their extreme sensitivity to environmental parameters. These assays rely on toxins encoded on a target plasmid that is cured by endonuclease activity to rescue growth. Thus, this type of system links cell survival to nuclease activity. However, common toxins used, CcdB<sup>46,50,51</sup> and barnase<sup>45</sup>, are highly lethal and even low levels of leaky expression can cause cell death<sup>45,46</sup>. Tuning the toxins expression to extremely low levels is possible but difficult to achieve. Moreover, the extreme toxicity of CcdB and Barnase make the systems hard to tune to measure lower levels of endonuclease activity. Endonucleases with less activity may not sufficiently cleave the toxin-encoding plasmid, and the small amount of highly lethal toxins will kill the cell. Therefore, this system falsely reports a lack of endonuclease activity when activity levels are low. For the discovery and comparison of endonucleases that have differing levels of activity, a less toxic, yet effective and tunable system is desired.

Here, we develop a novel, *in vivo* endonuclease activity assay in *E. coli* that links cell survival to programmable endonuclease activity. Endonuclease activity rescues a lethal phenotype induced by the homing endonuclease I-SceI. As the assay output is relative growth, small differences in endonuclease activity are amplified by exponential cell growth for ease in detection. Moreover, an *in vivo* assay of this nature is faster than conventional *in vitro* assays, which requires lengthy protein expression and purification before evaluation. As a proof of concept, we validated this

assay with wildtype SpCas9, xCas9, and eSpCas9, demonstrating its versatility to work with an array of enzymes and rapidly quantify activity.

## 2.2 Materials and methods

### 2.2.1 Growth Conditions

For each experiment, KS 165 (MG1655, DE3) *E. coli* cells were grown in Luria-Bertani medium (LB) at 37° C at 250 RPM unless otherwise noted. Super Optimal broth with Catabolite repression (SOC)<sup>52</sup> was used to recover cells after transformation. Antibiotics were introduced at 100 ug/ul for ampicillin, 50 ug/ul for kanamycin, and 10 ug/ul for tetracycline. The inducers, arabinose, aTc, and Rhamnose, were added for a final concentration of 10mM, 200 ng/uL, and 0.2% respectively. Tetracycline, kanamycin, aTc, glycerol and LB are from Fisher Bioreagents, Fairlawn, NJ. Rhamnose is from Sigma-Aldrich, St. Louis, MO. Ampicillin and arabinose are from Acros Organics, New Jersey.

### 2.2.2 Strain Construction

The *in vivo* endonuclease activity assays were conducted in *E. coli* strain KS165, which contains a DE3 cassette for T7 induction of endonuclease expression and an I-SceI recognition target integrated within the chromosome (Table 1). *E. coli* MG1655 (DE3) was a gift from Prof. Kristala Prather (MIT Chemical Engineering, Cambridge, MA). To generate KS165, I-SceI recognition sites were integrated in the genome of MG1655 (DE3) using a standard recombineering protocol<sup>53</sup>. *tetA* from pTKS/CS was amplified with I-SceI recognition sites and the PCR product was integrated at the *nth* locus. Tetracycline resistant clones were isolated on Luria-Bertani (LB) agar plates supplemented with tetracycline. Colonies were then checked via colony PCR and Sanger sequenced to confirm correct integration at the *nth* locus. Strain information can be found in **Table 2.1**.

### 2.2.3 Plasmid construction

All plasmid information can be found in **Table 2.1**. The I-SceI lethal plasmid, pColEI-ISceI, was constructed by amplifying I-SceI from pEndoSceWT<sup>50</sup> (a gift from Prof. Frederick Gimble, Purdue



Biochemistry) and cloned into pBAD-mTagBFP2 (a gift from Prof. Mathew Tantama) at the BglII and EcoRI restriction sites (**Appendix A, Table A1a**). An inactive I-SceI catalytic mutant, pColEI-ISceI-D44S, was constructed with the same primers and restriction sites but used pEndoSce-D44S from Frederick Gimble as a template. pFREE was purchased from Addgene (Addgene plasmid # 92050)<sup>54</sup>. The pFREE-xCas9 and pFREE-eSpCas9 plasmids were constructed by amplifying the pFREE backbone and the Cas9 mutants from pxCas9CR4 and pJSC114, respectively (**Appendix A, Table A1c,d**). SalI and SpeI (BcuI) restriction sites were added to the ends of the backbone and insert (**Appendix A, Table 1b**). pJSC114 (Addgene plasmid # 101215)<sup>55</sup> and pxCas9CR4 (Addgene plasmid # 111656) were purchased from Addgene. All plasmid constructs were verified via Sanger Sequencing.

**Table 2.1** Strains and plasmids.

<b>Name</b>	<b>Relevant Phenotype</b>	<b>Plasmid Origin of Replication</b>	<b>Source</b>
<b>Strain</b>			
<b>KS 165</b>	$\Delta$ endA, $\Delta$ recA, TetA, I-SceI recognition site	N/A	This study
<b>Plasmids</b>			
<b>pEndoSceWT</b>	AmpR, AraC, P <sub>BAD</sub> I-SceI	p15A	(Doyon, 2006) <sup>50</sup>
<b>pEndoSce-D44S</b>	AmpR, AraC, P <sub>BAD</sub> I-SceI	p15A	(Doyon, 2006) <sup>50</sup>
<b>PColEI-ISceI</b>	AmpR, AraC, P <sub>BAD</sub> I-SceI	ColE1	This study
<b>pColEI-ISceI-D44S</b>	AmpR, AraC, P <sub>BAD</sub> I-SceI	ColE1	This study
<b>pFREE</b>	KanR, P <sub>rhambAD</sub> guides, TetR, P <sub>Tet</sub> Cas9	ColA	(Lauritsen, 2017) <sup>54</sup>
<b>pFREE-xCas9</b>	KanR, P <sub>rhambAD</sub> guides, TetR, P <sub>Tet</sub> xCas9	ColA	(Hu, 2018)* <sup>48</sup>
<b>pFREE-eSpCas9</b>	KanR, P <sub>rhambAD</sub> guides, TetR, P <sub>Tet</sub> eSpCas9	ColA	(Slaymaker, 2016)* <sup>49</sup>

\*Source refers to the insert.

#### 2.2.4 I-SceI lethality assay

50ng of pColEI-ISceI or pEndoSceWT were electroporated into *E. coli* MG1655 (DE3) *nth::tetA* (denoted KS 165) competent cells. KS 165 was made competent by growing to OD<sub>600</sub> 0.5 and washing with chilled ddH<sub>2</sub>O twice and once with chilled 10% glycerol, then resuspended in 10% glycerol. Transformed cells were recovered in SOC for one hour at 37°C shaking, then plated on LB plates with ampicillin. After 16 hours, a single colony was picked and inoculated into 3mL of LB with and without ampicillin and allowed to grow for 4 hours at 37°C shaking. After 4 hours, the cultures were diluted 100x into LB and ampicillin, with and without 10mM arabinose (inducer for I-SceI). The cultures were allowed to grow for 4 hours at 37°C shaking and OD<sub>600</sub> was taken to determine growth. The experiment was done in triplicate.

#### 2.2.5 Nuclease activity assay

The pColEI-ISceI plasmid was transformed into KS 165 competent cells and recovered in 1 mL SOC for one hour at 37°C shaking. The cells were then plated on LB with ampicillin and grown for 16 hours. A single colony was picked and inoculated in LB with ampicillin and allowed to grow to OD 0.5. These cells were then made competent by washing twice with chilled ddH<sub>2</sub>O and once with chilled 10% glycerol. Each Cas9 variant (pFREE, pFREE-xCas9, pFREE-eSpCas9) was electroporated into the pColEI-ISceI/KS 165 competent cells and recovered in SOC for one hour at 37°C shaking. The cells were then plated on LB with ampicillin and allowed to grow for 16 hours at 37°C. A single colony was picked and grown in 3mL of LB with ampicillin, kanamycin, and tetracycline for 3 hours at 37°C shaking. Cultures were diluted 100x into LB with 100μL 10% rhamnose and 10 μL of aTc (inducers for guide and Cas enzyme), and allowed to grow for 4 hours at 37°C shaking. Cultures were then diluted 100x again into LB with 10mM arabinose (inducer for I-SceI) and allowed to grow for 4 hours at 37°C shaking. OD 600 readings were taken before and after this final 4-hour growth period (**Appendix A, Figure A1**). For each trial with inducer, a trial without inducer was run as a negative control.

### 2.2.6 Statistical analysis

Each experiment was replicated in triplicate. Comparison of recovered growth between induced and uninduced endonucleases were done using unpaired t-tests. Data shown are the mean +/- standard deviation, except for Fig. 3b which reports mean +/- standard error.

## 2.3 Results

### 2.3.1 Design and construction of selection system

Our positive selection system links targeted DNA endonuclease activity to cell survival in a quantitative way via a two-plasmid system in a modified *E. coli* MG1655 (DE3) host strain (**Figure 2.1a**). The first plasmid, a lethal plasmid, encodes a homing endonuclease that creates a lethal double-stranded DNA break (DSB) at a target site introduced in the chromosome of our modified strain (KS 165). The enzyme, I-SceI, targets a large recognition site, TAGGGATAACAGGGTAAT, which was integrated at the *nth* locus via standard recombineering approaches<sup>53</sup>. The size of this recognition site mitigates accidental cleavage of the host chromosome due to the absence of similar, slightly mismatched sequences. Chromosomal double-stranded DNA breaks are inefficiently repaired in *E. coli* inhibiting cell replication and growth<sup>51</sup>. Thus, we hypothesize that I-SceI will generate a lethal double-stranded DNA break that inhibits cell growth only when induced with arabinose. To prevent unintended cell death via basal expression of I-SceI, the tightly controlled, arabinose-inducible P<sub>araBAD</sub> promoter<sup>22,56</sup> was chosen to regulate expression of I-SceI (**Figure 2.1b**).

The second plasmid serves as a rescue plasmid that encodes a programmable endonuclease that is targeted to the lethal plasmid, linearizing and curing it. The rescue plasmid is derived from a pFREE backbone<sup>54</sup>, which expresses programmable endonucleases under the control of a P<sub>Tet</sub> promoter that is induced by anhydrotetracycline (aTc) (**Figure 2.1b**). Programmable endonuclease cleavage of the lethal plasmid is targeted via rhamnose-inducible guides for the ColE1 origin in the lethal plasmid<sup>57</sup>. Upon successful cleavage, the linearized lethal plasmid is rapidly degraded, rescuing growth<sup>58</sup>. Partial cleavage of the lethal plasmid would allow for fewer cells to escape cell death resulting in slower apparent growth. As cell growth is exponential, however, this growth based output amplifies small differences in cleavage activity to give an exponential correlation

between endonuclease activity and culture optical densities at later time points (**Figure 2.1c**). While signal amplification increases the sensitivity for detection of activity in low activity endonucleases, it can also make it difficult to discriminate between endonucleases with relatively high levels of activity due to more rapid signal saturation. As growth rates are finite, there will be an upper bound to the endonuclease activity that can be detected. Thus, the rescue plasmid was designed to also be self-curing by targeting its ColA origin with a separate rhamnose-inducible guide, limiting the amount of endonuclease that is expressed. This self-curing ability maintains assay sensitivity at low activity levels (poor endonuclease activity leads to longer expression times for endonuclease), while at high endonuclease activity expression is short allowing for detection of a wider range of activities.

**Figure 2.1. Design of selection system.** a) Overview of selection system. b) Plasmid designs used in the selection system. The lethal plasmid encodes for the homing endonuclease, I-SceI, which targets a modified *E. coli* MG1655 (DE3) genome. I-SceI is under the control of the arabinose-inducible promoter, araBAD, and the lethal plasmid has an AmpR selectable marker and a ColE1 origin of replication (~20 copies). The rescue plasmid encodes the endonuclease of choice under the control of a P<sub>Tet</sub> promoter, inducible with aTc. The rescue plasmid also has corresponding endonuclease guides targeting major origins of replication, under the control of a rhamnose-inducible rhamBAD promoter. The rescue plasmid has a KanR selectable marker and a ColA origin of replication (~20-40 copies). c) The exponential nature of the *in vivo* system (blue line) amplifies the signal of endonuclease activity compared to linear, *in vitro* activity assays (orange line). Small differences in endonuclease activity will be magnified in an *in vivo* assay. Cells containing the I-SceI plasmid (I-SceI<sup>+</sup>) will die off, leaving cells without the I-SceI plasmid (I-SceI<sup>-</sup>) to live. Endonuclease activity is linked to cell survival because the endonuclease targets and cures the I-SceI plasmid, resulting in I-SceI<sup>-</sup> cells.



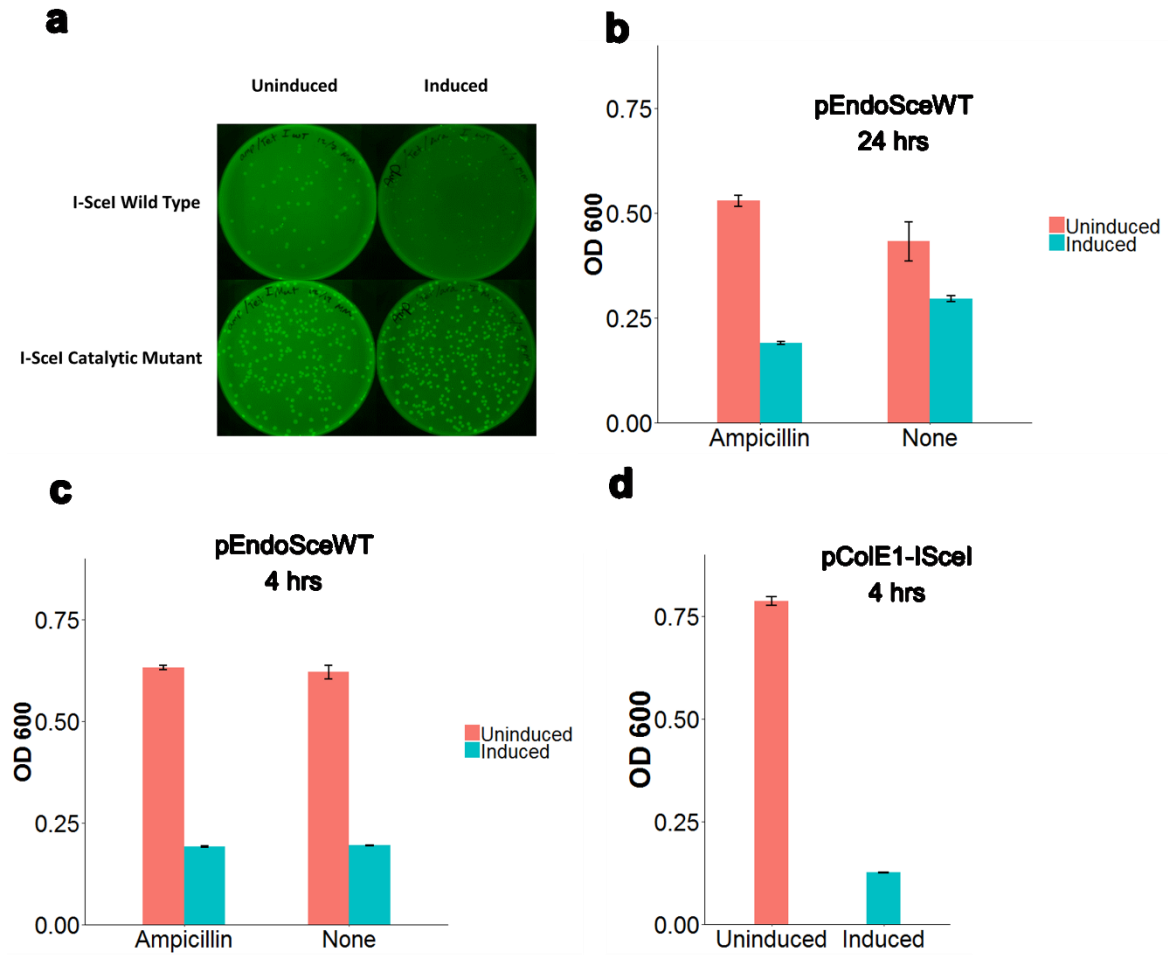
### 2.3.2 Optimization of selection system sensitivity

We first validated the ability of I-SceI to be conditionally lethal using a pACYC vector with a low copy p15A origin of replication for the lethal plasmid. KS165 cells containing plasmid pEndoSceWT or pEndoSceD44S were plated on LB agar and induced with arabinose. As I-SceI was only induced on the plate with arabinose, cells experienced a brief period of growth before sufficient I-SceI had been expressed to arrest growth (**Figure 2.2a**). This resulted in a noticeable phenotypic difference in colony size between the induced and uninduced wild type I-SceI plates, indicating some I-SceI activity. Negative controls with a catalytically inactive mutant of I-SceI resulted in no phenotypic difference in cell size between the induced and uninduced cells. That is, I-SceI was indeed conditionally lethal in our cells.

To improve parallelizability and ease of detection we tested conditional lethality in liquid media by measuring the optical density of growing cultures (**Figure 2.2b**). We grew cultures for 24 h, to simulate potential endonuclease induction and expression from the rescue plasmid before diluting the cultures, and inducing I-SceI for 4 h before measuring the optical density. In the presence of ampicillin, cell numbers (or OD 600) were reduced by 64.3% by I-SceI from pEndoSceWT, confirming the conditional lethality observed on plate-based assays. When fully implemented, the activity assay rescues cell growth by curing the I-SceI plasmid and its selection marker. That is, the cell will lose its ampicillin resistance. Thus, to prevent systematic bias in assay output, conditional lethality must also function in the absence of any antibiotic. We tested the pEndoSceWT plasmid's ability to reduce cell growth in the absence of antibiotic for 24 hours and found it only reduced growth by 31.8% (**Figure 2.2b**). Because the pEndoSceWT plasmid has a low copy, p15A origin of replication ( $\sim 10$ -12 copies/cell)<sup>10</sup>, we hypothesized that the cells were spontaneously curing themselves of the plasmid over the 24 hour time period in the absence of any selection pressure. We then decided to decrease the growth time from 24 hours to 4 hours, which should allow for ample expression and cleavage via any tested endonucleases, and test whether the cells would retain the conditionally lethal phenotype (**Figure 2.2c**). The cultures with and without selective pressure (ampicillin) were able to reduce cell growth to a similar extent, 69.7% and 68.6%, respectively. This result suggests that a 4 hour growth period prior to I-SceI induction is sufficient to preserve plasmid retention within the cells.

To increase the sensitivity of the assay we increased the copy number of the lethal plasmid. However, this copy number must be lower than that of the rescue plasmid to ensure sufficient endonuclease for activity detection. As the rescue plasmid has a ColA origin of replication which generates 20-40 plasmid copies per cell<sup>10</sup>, we chose a pET vector with a ColE1 origin of replication (~15-20 copies/cell)<sup>10</sup>, which is compatible with ColA<sup>10,59</sup>. The generated pColE1-ISceI lethal plasmid significantly reduced cell growth by 84.1% (**Figure 2.2d**), increasing the sensitivity of the system due to the higher copy number, as designed.





**Figure 2.2. I-SceI is lethal in engineered KS 165.** **a)** *E. coli* MG1655 (DE3) *nth::tetA* (KS 165) was transformed with the wild type I-SceI plasmid, pEndoSceWT, or the catalytic mutant, pEndoSceD44S. Cultures were grown with either 0 or 10mM arabinose for induction of wild type and mutant I-SceI. All cultures were able to form healthy colonies after 16 hours at 37°C although those with the induced wild type I-SceI plasmid were observably smaller, suggesting I-SceI on the pEndoSceWT plasmid induces some cell death via a DSB in the host genome. **b-c)** Cells harbouring the pEndoSceWT plasmid were allowed to grow for 24 hours (b) or 4 hours (c), followed by a 4 hour I-SceI induction period. Cultures were grown with and without ampicillin to test plasmid retention in the presence (or absence) of a selective pressure. Cells grown for 24 hours without the selective pressure lost some of the plasmid whereas cells grown for 4 hours without selective pressure did not lose the plasmid. **d)** Cells harbouring the higher copy number plasmid, pColE1-I-SceI, were allowed to grow for 4 hours, followed by a 4 hour I-SceI induction period. pColE1-I-SceI was able to reduce cellular growth by 84.1%.

### 2.3.3 Validation of targeted endonuclease activity

The ability of the system to function as an activity assay for endonuclease activity was validated using the popular wild type SpCas9 enzyme<sup>19</sup>. Cells harbouring both the rescue and lethal plasmids were grown without any inducer to establish a baseline optical density without induced lethality. The cells were grown for four hours, diluted 100x, then grown for another four hours, resulting in an optical density of  $0.682 \pm 0.04$ . Expression of SpCas9 did not have a statistically significant effect on cell growth ( $OD = 0.666 \pm 0.04$ ). In the presence of uninduced rescue plasmid, I-SceI induction resulted in conditional lethality, reducing optical density by 68%, in agreement with our earlier studies. However, preexpression of SpCas9 targeted to the I-SceI plasmid resulted in optical densities of  $0.446 \pm 0.07$ , rescuing 49.3% of wildtype growth. This result validates the efficacy of the system to display endonuclease activity as cells with induced SpCas9 and I-SceI showed an increase in cellular growth compared to the growth of cells with only I-SceI induction. A final control was added which only included the host strain, void of any plasmid, to test the effect of metabolic burden due to plasmid maintenance and heterologous gene expression. These cultures grew to an optical density  $0.735 \pm 0.02$ , demonstrating no significant metabolic burden effects that would decrease growth (**Figure 2.3a, Appendix A, Figure A1,2**).

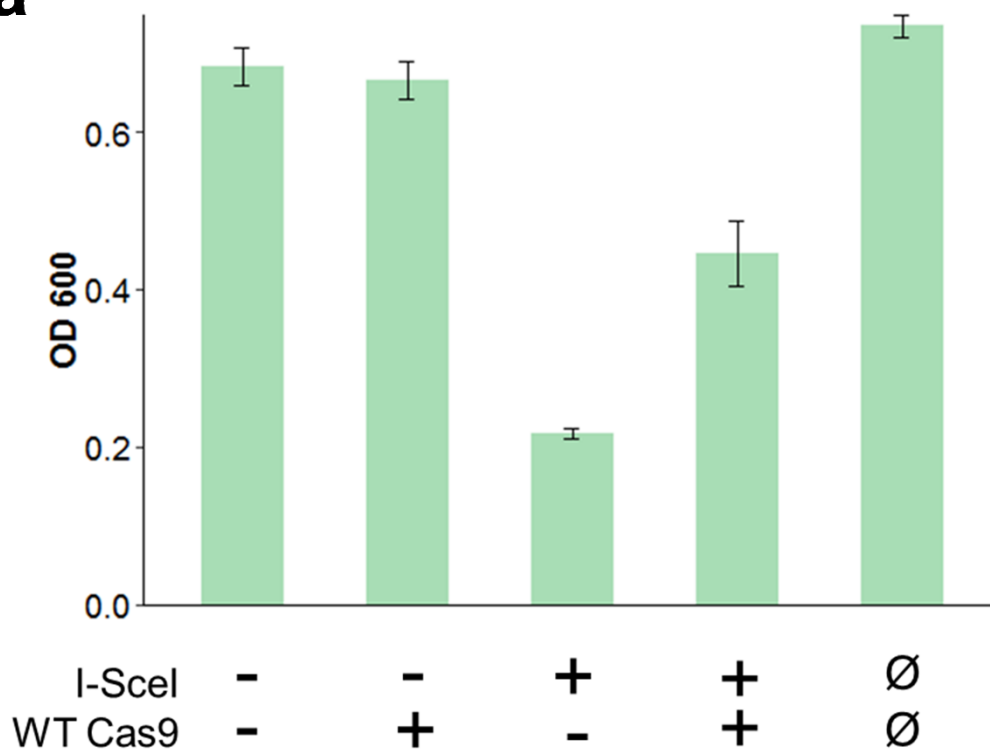
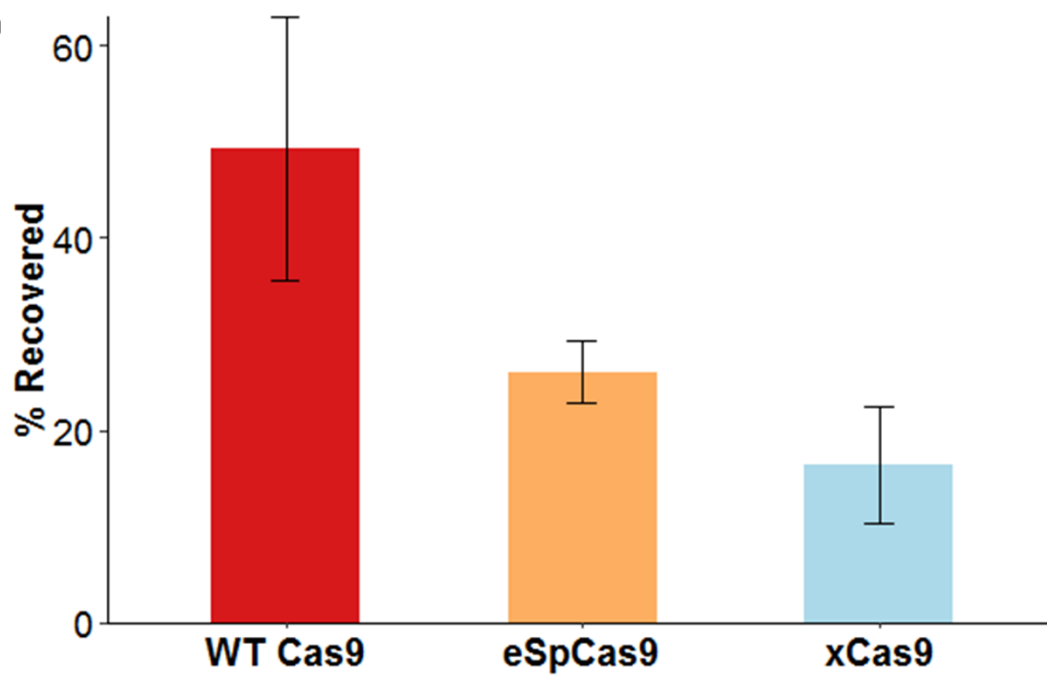
To compare and rank the activity of each endonuclease, growth recovery percentages were calculated. Endonuclease activity was quantified by dividing the difference between the uninduced control and the I-SceI/endonuclease induced treatments by the difference between the uninduced control and I-SceI-only induced control (**Equation 1**). By doing so, the ability of the endonuclease to prevent I-SceI-mediated cell death is calculated and can be used as a way to quantify endonuclease activity. Thus, higher growth recovery correlates to higher endonuclease activity while less growth recovery results in lower endonuclease activity (**Figure 2.1c**). With this system, and under the conditions tested, wild type Cas9 had the greatest growth recovery at 49.3%, followed by eSpCas9 at 26.1%, and xCas9 at 16.4%. (**Figure 2.3b**). Each endonuclease displayed activity by significantly recovering cell growth (unpaired t-test,  $p < 0.05$ ,  $n = 3$ ).

**Equation 1.** Equation to determine endonuclease activity.

**Percentage growth recovery**

$$= \frac{OD_{600}I\text{SceI}^- \text{Endonuclease}^- - OD_{600}I\text{SceI}^+ \text{Endonuclease}^+}{OD_{600}I\text{SceI}^- \text{Endonuclease}^- - OD_{600}I\text{SceI}^+ \text{Endonuclease}^-}$$

**Figure 2.3. Endonuclease activity recovers growth.** **a)** Cell growth was recovered using wild type SpCas9 enzyme. **b)** Comparison of commonly used endonucleases. Each endonuclease was given four hours of induction and targeted toward the lethal plasmid, after which I-SceI of the lethal plasmid was induced to cause cell death. Each endonuclease significantly rescued cell growth (unpaired t-test,  $p < 0.05$ , standard error,  $n=3$ ).

**a****b**

## 2.4 Discussion

We have developed an *in vivo* endonuclease activity assay designed to identify and compare enzymes for DNA-cleavage activity. Our system makes use of the homing endonuclease, I-SceI, which recognizes and cleaves an engineered site in a modified *E. coli* genome, causing cell death. I-SceI was chosen as a lethal factor due to its low toxicity when uninduced. In contrast, previous attempts to develop an endonuclease activity assay used the toxic proteins barnase and CcdB as their lethal factors. Basal expression of barnase in a low-copy plasmid with a p15A origin was reportedly lethal in *E. coli* cells, even under the tight regulation of the araBAD promoter<sup>45</sup>. Therefore, the authors introduced amber nonsense codons to control the toxicity of barnase. While successful, this technique necessitates the use of amber suppressor tRNAs for proper functionality, which requires another inducer. To combat this, a CcdB version of an endonuclease activity assay was created, nullifying the suppressor tRNA requirement. However, CcdB was also reported to be toxic under basal expression from the same araBAD promoter, though the copy number was much higher (100-300 copies per cell)<sup>46</sup>. Therefore, the authors engineered the ribosome binding site for the *ccdB* gene to thwart some of its toxicity. Unlike barnase and CcdB, we did not experience any overt toxicity issues employing I-SceI as our lethal factor.

Originally in a pACYC vector with a low copy p15A origin of replication, I-SceI was able to reduce cell growth by 64.3% after a 24 hour growth period, in the presence of a selective pressure to retain the plasmid (**Figure 2b**). However, the assay accounts for a growth period to induce the endonuclease, at which time selective pressures cannot be incorporated. After a 24 hour growth period without selective pressure, the cells were unable to fully retain the plasmid and less lethality was observed (**Figure 2b**). To mitigate this issue, we decreased the growth time to 4 hours and saw similar reductions in growth regardless of the presence of selection pressure (**Figure 2c**). We also sought to increase the sensitivity of the assay by increasing the plasmid copy number from 10-12 to 15-20 copies per cell. This new plasmid, pColE1-I-SceI, was able to reduce cell growth by 84.1% after a 4 hour growth period (**Figure 2d**). In the pColE1-I-SceI plasmid, I-SceI is under the control of the tightly regulated P<sub>BAD</sub> promoter. This strict regulation alleviated leaky expression and prevented unwanted cell death. To demonstrate this, uninduced cells harbouring the pColE1-I-SceI plasmid had similar growth rates to cells not harbouring the plasmid (**Figure 3a**).

With the lethal plasmid properly inducing cell death, we next had to design the rescue plasmid. Fortunately, Lauritsen and colleagues previously designed a plasmid denoted pFREE which was used for plasmid curing. We modified this plasmid by flanking the Cas9 enzyme with SalI and SpeI restriction sites to allow for easy swapping of the endonuclease. We chose the pFREE plasmid as a host for endonuclease expression because of its SpCas9 RNA guides and unique origin of replication. The guides are targeted toward a sequence shared by the vast majority of bacterial cloning and expression vectors<sup>54</sup>, allowing SpCas9 and derivatives to target a wide selection of plasmid vectors, including the pET vector of the lethal plasmid. Meanwhile, the origin, ColA, is in a class of its own allowing it to stably coexist with a wide variety of origins of replication, including ColE1. Critically, the copy number for ColA is higher than the copy number for ColE1, which allows for enough expression of endonuclease to sufficiently cleave each copy of lethal plasmid. If the rescue plasmid's copy number were lower than the lethal plasmid's, less efficient endonucleases would likely fail to cleave every copy of lethal plasmid, which would lead to cell death. By selecting a plasmid that has a higher copy number origin than the lethal plasmid, we ensure that we do not miss low levels of endonuclease activity that could result from novel endonuclease enzymes or imperfect conditions.

To confirm our system was robust with respect to a number of endonucleases and could distinguish between differing activity levels, we tested our system with three endonucleases, Cas9, xCas9, and eSpCas9. Of the three, wild-type Cas9 displayed the highest level of endonuclease activity, followed by eSpCas9 and xCas9. We hypothesize that the mutations made to develop eSpCas9 and xCas9 affect the kinetics of the enzymes, slowing their activity levels, congruent with published findings<sup>60</sup>. This idea is corroborated in literature as one study found eSpCas9 to cleave 75% of substrate within 15 seconds and 90% within 30 seconds using a substrate to guide/Cas9 molar ratio of 1:100<sup>55</sup>. Meanwhile wild-type Cas9 cleaved about 95% of substrate within the first 15 seconds. Another study found xCas9 to cleave around 20% of substrate at around 30 seconds and 60% at 5 minutes using a substrate to guide/Cas9 molar ratio of 1:2<sup>61</sup>. These findings support our results as wild-type Cas9 has the highest activity in each study, followed by eSpCas9 and xCas9. However, over time, these enzymes may reach the same steady state as wild-type Cas9. Therefore, *in vitro* assays require constant monitoring, especially at very short time intervals, to detect kinetic differences in enzymatic activity. Because our system is performed *in vivo* in liquid

culture, kinetics of enzymes play a large role in comparing endonuclease activity. Cellular growth is exponential and thus slight changes in kinetics will result in large differences in total growth over time. For example, the difference between eSpCas9 and wild-type Cas9 activity levels *in vitro* may only be a percentage or two after 1 minute, but would be far greater in an *in vivo* assay (**Figure 3b**). Therefore, differences in activity levels between endonucleases that would otherwise resolve themselves are amplified and obvious over time.

Moreover, *in vivo* activity assays require much less time than traditional *in vitro* methods. For example, *in vitro* assays require cell transformation and long growth periods followed by cell lysis, protein purification, protein quantification, activity assays, and finally gel electrophoresis. *In vivo* systems, on the other hand, negate cell lysis, protein purification/quantification, and gel electrophoresis. Instead, *in vivo* assays can skip from cell transformation to the activity assay directly. Typically, *in vivo* systems can be performed within 24 hours start to finish while *in vitro* methods take days.

With removal of the ColA Cas9 guide RNA from the pFREE plasmid, our system may also be used as a positive selection screen in directed evolution approaches to enhance enzymatic activity of endonucleases. With the pFREE plasmid no longer targeting itself, endonuclease activity will still be retained when the endonuclease cleaves the lethal plasmid and rescues cell growth. However, as more active mutants result in faster growth, those high performing variants will begin to dominate mixed cultures of endonuclease variants facilitating recovery, characterization and subsequent round of directed evolution

In conclusion, our novel endonuclease activity assay can identify and rank endonuclease activity. By using a tightly controlled two-plasmid system, our assay positively links cell growth to endonuclease activity and serves as a ranking system among different endonucleases. The *in vivo* nature of our system drastically decreases the time required to identify endonuclease activity compared to *in vitro* methods. Moreover, the endonuclease activity signal is amplified in our *in vivo* assay compared to *in vitro* methods, facilitating the clear discrimination between endonuclease activity levels.

### 3. CONCLUSIONS AND FUTURE DIRECTIONS

#### 3.1 Summary of Current Progress

During the span of this project, a novel system was created that can serve to detect endonuclease activity and function as a positive selection screen for directed evolution applications. This system built upon the successes of two previous positive selection screens whose foundations paved the way for this research<sup>45,46</sup>. The system incorporates two plasmids, a lethal plasmid that induces cell death and a rescue plasmid that rescues cell growth. With this system, endonuclease activity is positively linked to cell survival, and thus higher cell growth indicates more endonuclease activity.

Due to our desire to create a system that not only functions as a positive selection screen but also as a preliminary screen for endonuclease activity, we designed our system to display less-efficient endonuclease activity while retaining the stringency needed for directed evolution applications. Our lethal plasmid induces cell death using the homing endonuclease, I-SceI, which targets an 18 base pair recognition site, engineered into an *E. coli*, MG1655 (DE3) genome. Due to insufficient repair mechanisms found in the *E. coli* MG1655 (DE3) strain, the DSB caused by I-SceI becomes deadly and thus prevents cell growth. The I-SceI enzyme is under the tight control of the araBAD promoter, inducible with arabinose. This tight regulation helps mitigate the leaky expression of I-SceI and therefore prevents unwanted cell death. Using a lethal plasmid with a pACYC vector (10-12 copies per cell), our system was able to reduce cell growth by 69.7%. However, we sought to increase this number and cloned the I-SceI enzyme into a pET vector (15-20 copies). This increase in plasmid copy number helped reduce cell growth by 84.1%. Moreover, the system is designed so that selective markers cannot be used to retain plasmids. Using our pET plasmid, denoted pColE1-I-SceI, we do not see curing of this plasmid from dilution due to growth regardless of the presence of a selection pressure, indicating that the copy number and the system run time are a good combination.

To test the efficacy of the system to identify endonuclease activity, we tested three endonuclease enzymes, wild type-SpCas9, xCas9, and eSpCas9. Endonuclease activity was dictated by recovery of cellular growth. While each enzyme displayed activity levels that were significant ( $p < 0.05$ ),



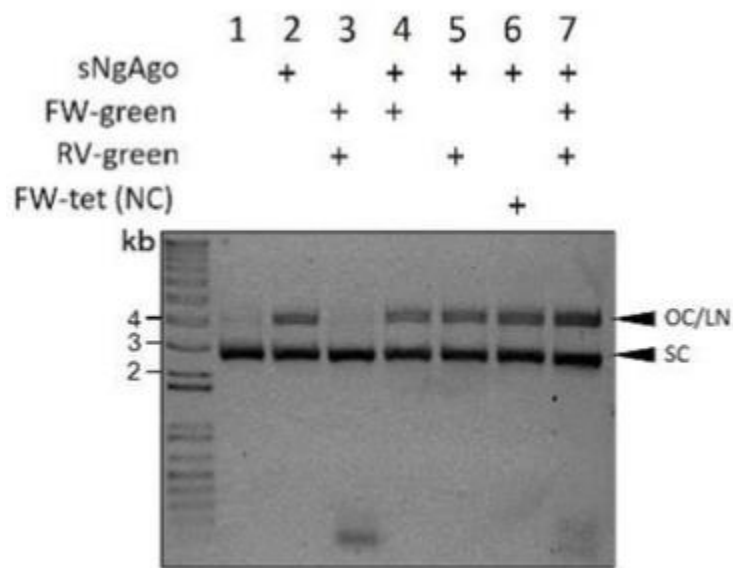
wild type-SpCas9 had the highest activity level with 49.3%, followed by eSpCas9 at 26.1%, and xCas9 at 16.4%. However, these enzymes were removed from their host vectors and placed into our rescue plasmid, which may not serve as the ideal vector for each endonuclease enzyme. A system that compares enzymes in their host vectors would be highly beneficial to scientists who want to commercially buy an endonuclease without having to clone it into a different expression vector.

### **3.2 Modification of the system to rank commercially available plasmids**

While our system can positively identify endonuclease activity, it removes external influence on endonuclease activity that may be different in non-identical host plasmids by expressing each endonuclease enzyme in the same host vector. Thus, the enzymes are tested in a commensurable environment which only allows for the comparison of the endonuclease enzymes, not the commercially available vectors. To compare commercially available plasmids, the lethal plasmid would need to be modified to be compatible with each commercial plasmid, which would take the place of the rescue plasmid. As discussed previously (section 1.1.1), the lethal plasmid and commercially available endonuclease plasmid would need to have different origins of replication and different selectable markers. However, the ColE1 origin and ampicillin resistance marker of the lethal plasmid are not uncommon. To accommodate for a wide selection of commercially available plasmids to serve as the rescue plasmid, the lethal plasmid would need to have a rarer origin of replication and a less common selectable marker. With these two modifications, the system should be able to test endonuclease activity using commercially available vectors. This would alleviate the need for researchers to clone the endonuclease into our rescue plasmid to test its activity. Moreover, it is likely that the commercial vectors housing the endonuclease enzymes have been optimized for their specific enzymes. Some enzymes may work better under higher or lower expression levels and under different conditions. Therefore, we are currently modifying our system to incorporate these changes, allowing us to test endonuclease activity of commercially available vectors.

### 3.3 Application of NgAgo for Future Work

One of our main goals for this project was creating a positive selection system for directed evolution applications of endonucleases. One endonuclease, the Argonaute from *Natronobacterium gregoryi* (NgAgo), has been studied extensively by our lab and others. Unlike other pAgos, NgAgo has been shown to work at mesophilic conditions, making it a better candidate as a gene-editing tool in conditions more relevant to biotechnology<sup>30,31</sup>. Moreover, a recent paper showed that NgAgo has gene-editing potential by enhancing homologous recombination<sup>31</sup> while our lab showed that it does this via endonuclease activity (**Figure 3.1**).



**Figure 3.1.** Modified with permission<sup>30</sup>. NgAgo cuts DNA *in vitro*. Soluble NgAgo (sNgAgo) nicks and cuts plasmid DNA nonspecifically without the presence of guide (FW-green and RV-green), as indicated by the open circular and linear DNA (OC/LN). Without the presence of NgAgo, plasmid DNA remains in its supercoiled form (SC).

However, NgAgo poses certain challenges as an effective gene-editing tool. For example, *Natronobacterium gregoryi* is halophilic and thus NgAgo expression in common, low-salt conditions is poor<sup>30</sup>, a common challenge among halophilic proteins<sup>62,63</sup>. Poor expression of NgAgo results in less soluble protein, and because only soluble NgAgo cuts DNA<sup>30</sup>, the ability for NgAgo to serve as a gene-editing tool is limited. Moreover, NgAgo displays some non-specific DNA cleavage<sup>30</sup>, further reducing its capacity to function in gene-editing applications. Non-specific cleavage, or off-target activity, is when the endonuclease cuts a piece of DNA at a location

other than its targeted site. Ideally, endonucleases will have limited to no off-target activity, giving the researcher precise control over where they cut DNA. To address these issues, our lab is applying directed evolution techniques to modify NgAgo, increasing its solubility and decreasing its off-target activity. The directed evolution cycles will utilize the positive selection system developed in this project.

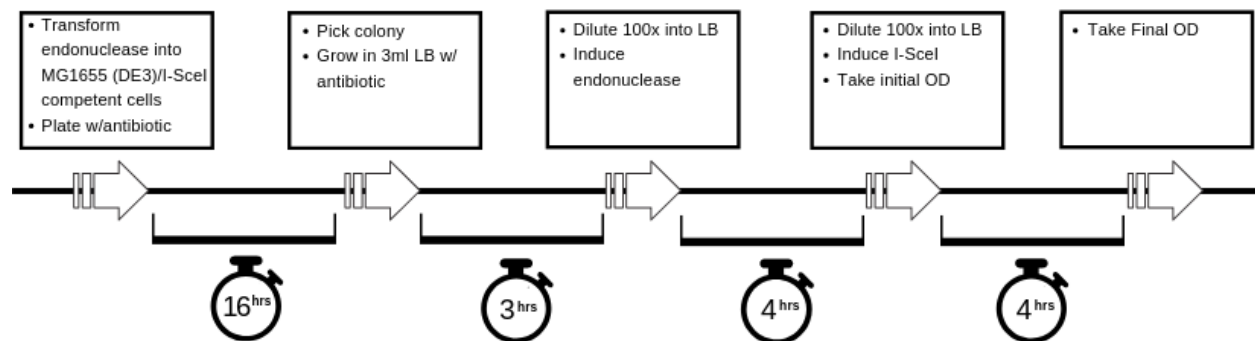
## APPENDIX A

**Table A1.** Oligonucleotides used. Restriction sites are in upper case.

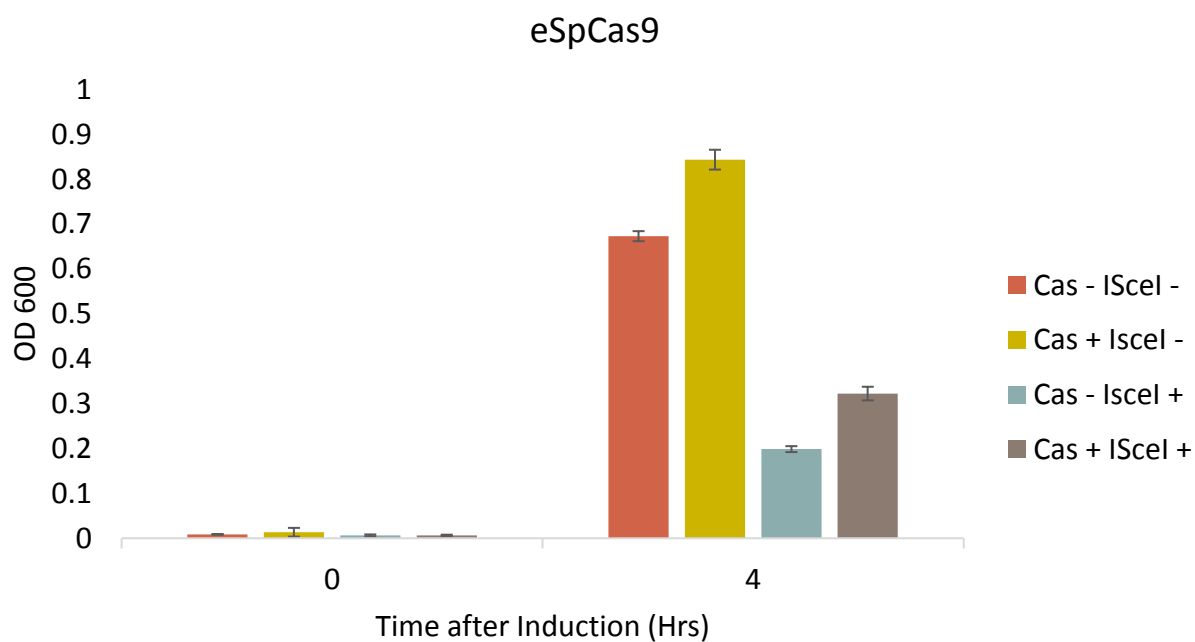
Name	Sequence 5' → 3'
<b>a pColE1-ISceI Construction</b>	
pColE1-ISceI Fwd	ttttagatctATGAAAaacatcaaaaaaaccaggtaatgaacctgg
pColE1-ISceI Rev	tttGAATTCttatttcaggaaagtttcggaggagatagtgttc
<b>c pFREE-xCas9 Construction</b>	
pFREE-xCas9 Backbone Fwd	tggtACTAGTgatcccatgttaccggtatccaag
pFREE-xCas9 Backbone Rev	tggtGTCGACctatcactgatagtgtcagttatttctatc
pFREE-xCas9 Insert Fwd	GTCGACagatactgagcacagaaggagatatacatatggataagaaa
pFREE-xCas9 Insert Rev	ACTAGTttagtcacctcctagctgactca
<b>d pFREE-eSpCas9</b>	
pFREE-eSpCas9 Backbone	tggtACTAGTgatcccatgttaccggtatccaag
pFREE-eSpCas9 Backbone	tggtGTCGACctatcactgatagtgtcagttatttctatc
pFREE-eSpCas9 Insert Fwd	GTCGACagatactgagcacagaaggagatatacatatggataagaaa
pFREE-eSpCas9 Insert Rev	ACTAGTttagtcacctcctagctgactca
<b>f KS 165 Construction</b>	
nth TetA Fwd	ctgctttccgctcaggcgaccgatgtcagtggttaataaggcgacggcgaata
nth TetA Rev	cggaaaatgtgcgtgtcgacagcaatagtcggccagccgaatgcagtggttc

**Table A2.** Values for paired Student's T-Test comparing endonucleases when induced (total rescue) and uninduced (total death).

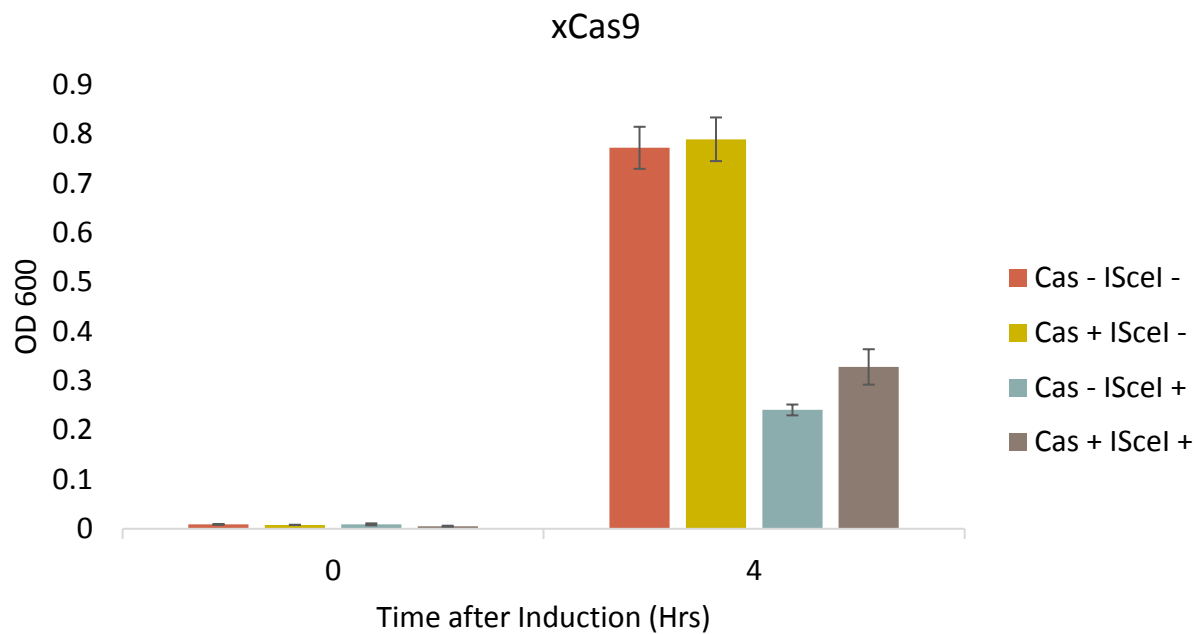
<i>Endonuclease</i>	<i>% Growth Recovered</i>	<i>Standard Error</i>	<i>Total Replicates</i>	<i>p value (one-tail)</i>
<b>WT SpCas9</b>	49.3	13.7	3	0.019
<b>xCas9</b>	16.4	10.5	3	0.013
<b>eSpCas9</b>	26.1	5.6	3	0.0049



**Figure A1.** Timeline for endonuclease activity assay.



**Figure A2.** eSpCas9 was used to rescue cell growth in our endonuclease activity assay. Cas refers to the endonuclease. + and – indicate induced and uninduced conditions, respectively. Data displayed is the mean of 3 replicates, error represented as standard error.



**Figure A3.** xCas9 was used to rescue cell growth in our endonuclease activity assay. Cas refers to the endonuclease. + and – indicate induced and uninduced conditions, respectively. Data displayed is the mean of 3 replicates, error represented as standard error.

## REFERENCES

1. Synthetic Biology. *National Human Genome Research Institute* Available at: <https://www.genome.gov/about-genomics/policy-issues/Synthetic-Biology>.
2. Hellmich, R. L. & Hellmich, K. A. Use and Impact of Bt Maize. *Nat. Educ. Knowl.* **3**, 4 (2012).
3. Jagadevan, S. *et al.* Recent developments in synthetic biology and metabolic engineering in microalgae towards biofuel production. *Biotechnol. Biofuels* **11**, 1–21 (2018).
4. Genentech. First Successful Laboratory Production of Human Insulin Announced. (1978).
5. Zieliński, M. *et al.* Expression and purification of recombinant human insulin from E. coli 20 strain. *Protein Expr. Purif.* **157**, 63–69 (2019).
6. Baeshen, N. A. *et al.* Cell factories for insulin production. *Microb. Cell Fact.* **13**, 1–9 (2014).
7. Griswold, A. Genome Packaging in Prokaryotes: the Circular Chromosome of E. coli. *Nat. Educ.* **1**, 57 (2008).
8. Lodish, H., Berk, A. & Zipursky, S. DNA Cloning with Plasmid Vectors. in *Molecular Cell Biology* (Freeman, W. H., 2000).
9. Leonard, A. C. & Mechali, M. DNA Replication Origins. *Cold Spring Harb. Perspect. Biol.* **5**, 1–17 (2013).
10. Tolia, N. H. & Joshua-Tor, L. Strategies for protein coexpression in Escherichia coli. *Nat. Methods* **3**, 55–64 (2006).
11. Million-Weaver, S. & Camps, M. Contribution submission to the International Conference on the Physics of Semiconductors 2012 Part : Type : Topic : Email : ICPS Poster only 6 . Topological insulators. *Plasmid* **27–36** (2014). doi:10.1016/j.plasmid.2014.07.002.Mechanisms
12. Summers, D. The kinetics of plasmid loss. *Trends Biotechnol.* **9**, 273–278 (1991).
13. Blattner, F. R. *et al.* The complete genome sequence of Escherichia coli K-12. *Science* (80-. ). **277**, 1453–1462 (1997).
14. 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. *J. Mol. Biol.* **382**, 188–202 (2008).
15. Roberts, R. J. How restriction enzymes became the workhorses of molecular biology. *Proc.*

- Natl. Acad. Sci. U. S. A.* **102**, 5905–5908 (2005).
16. Chevalier, B. S. *et al.* Design, activity, and structure of a highly specific artificial endonuclease. *Mol. Cell* **10**, 895–905 (2002).
  17. Miller, J. C. *et al.* A TALE nuclease architecture for efficient genome editing. *Nat. Biotechnol.* **29**, 143–150 (2011).
  18. Porteus, M. H. & Carroll, D. Gene targeting using zinc finger nucleases. *Nat. Biotechnol.* **23**, 967–973 (2005).
  19. Jinek, M. *et al.* A Programmable Dual-RNA – Guided DNA Endonuclease in Adaptive Bacterial Immunity. *Science* (80-. ). **337**, 816–822 (2012).
  20. Joung, J. K. & Sander, J. D. TALENs: a widely applicable technology for targeted genome editing. *Nat. Rev. Mol. Cell Biol.* **14**, 49–55 (2013).
  21. Gaj, T., Gersbach, C. A. & Barbas III, C. F. ZFN, TALEN and CRISPR/Cas-based methods for genome engineering. *Trends Biotechnol.* **31**, 397–405 (2013).
  22. Hegge, J. W. *et al.* DNA-guided DNA cleavage at moderate temperatures by *Clostridium butyricum* Argonaute. *Nucleic Acids Res.* **47**, 5809–5821 (2019).
  23. Hegge, J. W., Swarts, D. C. & Van der Oost, J. Prokaryotic Argonaute proteins: novel genome-editing tools? *Nat. Rev. Microbiol.* **16**, 5–11 (2017).
  24. Guha, T. K., Wai, A. & Hausner, G. Programmable Genome Editing Tools and their Regulation for Efficient Genome Engineering. *Comput. Struct. Biotechnol. J.* **15**, 146–160 (2017).
  25. Ryazansky, S., Kulbachinskiy, A. & Aravin, A. A. The Expanded Universe of Prokaryotic Argonaute Proteins. *MBio* **9**, 1–20 (2018).
  26. Swarts, D. C. *et al.* DNA-guided DNA interference by a prokaryotic Argonaute. *Nature* **507**, 258–261 (2014).
  27. Willkomm, S. *et al.* Structural and mechanistic insights into an archaeal DNA-guided Argonaute protein. *Nat. Microbiol.* **2**, (2017).
  28. Enghiad, B. & Zhao, H. Programmable DNA-Guided Artificial Restriction Enzymes. *ACS Synth. Biol.* **6**, 752–757 (2017).
  29. Swarts, D. C. *et al.* Argonaute of the archaeon *Pyrococcus furiosus* is a DNA-guided nuclease that targets cognate DNA. *Nucleic Acids Res.* **43**, 5120–5129 (2015).
  30. Lee, K. Z. *et al.* NgAgo-enhanced homologous recombination in *E. Coli* is mediated by



- DNA endonuclease activity. *bioRxiv* (2019). doi:10.1101/597237
31. Fu, L. *et al.* The prokaryotic Argonaute proteins enhance homology sequence-directed recombination in bacteria. *Nucleic Acids Res.* **47**, 3568–3579 (2019).
  32. Gao, F., Shen, X. Z., Jiang, F., Wu, Y. & Han, C. DNA-guided genome editing using the *Natronobacterium gregoryi* Argonaute. *Nat. Biotechnol.* **34**, (2017).
  33. Kaya, E. *et al.* A bacterial Argonaute with noncanonical guide RNA specificity. *Proc. Natl. Acad. Sci. U. S. A.* **113**, 3–8 (2016).
  34. Gatti-Lafranconi, P. *et al.* Evolution of Stability in a Cold-Active Enzyme Elicits Specificity Relaxation and Highlights Substrate-Related Effects on Temperature Adaptation. *J. Mol. Biol.* **395**, 155–166 (2010).
  35. Hawkins, R. E., Russell, S. J. & Winter, G. Selection of phage antibodies by binding affinity. Mimicking affinity maturation. *J. Mol. Biol.* **226**, 889–896 (1992).
  36. Shaikh, F. A. & Withers, S. G. Teaching old enzymes new tricks: Engineering and evolution of glycosidases and glycosyl transferases for improved glycoside synthesis. *Biochem. Cell Biol.* **86**, 169–177 (2008).
  37. Packer, M. S. & Liu, D. R. Methods for the directed evolution of proteins. *Nat. Rev. Genet.* **16**, 379–394 (2015).
  38. Shafee, T. Evolvability of a Viral Protease: Experimental Evolution of Catalysis , Robustness and Specificity. (2013).
  39. Costa, J. R. *et al.* Genome Editing Using Engineered Nucleases and Their Use in Genomic Screening. *Assay Guid. Man.* 1–24 (2004).
  40. Reyrat, J. M., Pelicic, V., Gicquel, B. & Rappuoli, R. Counterselectable markers: Untapped tools for bacterial genetics and pathogenesis. *Infect. Immun.* **66**, 4011–4017 (1998).
  41. Bochner, B. R., Huang, H. C., Schieven, G. L. & Ames, B. N. Positive selection for loss of tetracycline resistance. *J. Bacteriol.* **143**, 926–933 (1980).
  42. Maloy, S. R. & Nunn, W. D. Selection for loss of tetracycline resistance by *Escherichia coli*. *J. Bacteriol.* **145**, 1110–1112 (1981).
  43. Ryu, Y. S., Chandran, S.-P., Kim, K. & Lee, S. K. Oligo- and dsDNA-mediated genome editing using a tetA dual selection system in *Escherichia coli*. *PLOS* **12**, (2017).
  44. Stavropoulos, T. A. & Strathdee, C. A. Expression of the tetA ( C ) tetracycline efflux pump in *Escherichia coli* confers osmotic sensitivity. *FEMS Microbiol. Lett.* **190**, 147–150 (2000).

45. Gruen, M., Chang, K., Serbanescu, I. & Liu, D. R. An in vivo selection system for homing endonuclease activity. *Nucleic Acids Res.* **30**, e29–e29 (2002).
46. Chen, Z. & Zhao, H. A highly sensitive selection method for directed evolution of homing endonucleases. *Nucleic Acids Res.* **33**, e154–e154 (2005).
47. Guo, J., Gaj, T. & Barbas III, C. F. Directed Evolution of an Enhanced and Highly Efficient FokI Cleavage Domain for Zinc Finger Nucleases. *J. Mol. Biol.* **400**, 96–107 (2010).
48. Hu, J. H. *et al.* Evolved Cas9 variants with broad PAM compatibility and high DNA specificity. *Nature* **556**, 57–63 (2018).
49. Slaymaker, I. M. *et al.* Rationally engineered Cas9 nucleases with improved specificity. *Science* (80-. ). **351**, 84–88 (2016).
50. Doyon, J. B., Pattanayak, V., Meyer, C. B. & Liu, D. R. Directed evolution and substrate specificity profile of homing endonuclease I-SceI. *J. Am. Chem. Soc.* **128**, 2477–2484 (2006).
51. Chayot, R., Montagne, B., Mazel, D. & Ricchetti, M. An end-joining repair mechanism in *Escherichia coli*. *Proc. Natl. Acad. Sci. U. S. A.* **107**, 2141–2146 (2010).
52. SOC Medium. *Cold Spring Harb. Protoc.* **2018**, pdb.rec098863 (2018).
53. Kuhlman, T. E. & Cox, E. C. Site-specific chromosomal integration of large synthetic constructs. *Nucleic Acids Res.* **38**, (2010).
54. Lauritsen, I., Porse, A., Sommer, M. O. A. & Nørholm, M. H. H. A versatile one-step CRISPR-Cas9 based approach to plasmid-curing. *Microb. Cell Fact.* **16**, 1–10 (2017).
55. Chen, J. S. *et al.* Enhanced proofreading governs CRISPR-Cas9 targeting accuracy. *Nature* **550**, 407–410 (2017).
56. Greenfield, L., Boone, T. & Wilcox, G. DNA sequence of the araBAD promoter in *Escherichia coli* B/r. *Proc. Natl. Acad. Sci. U. S. A.* **75**, 4724–4728 (1978).
57. Egan, S. M. & Schleif, R. F. A regulatory cascade in the induction of rhaBAD. *Journal of Molecular Biology* **234**, 87–98 (1993).
58. Kuzminov, A. & Stahl, F. W. Stability of Linear DNA in recA Mutant *Escherichia coli* Cells Reflects Ongoing Chromosomal DNA Degradation. *J. Bacteriol.* **179**, 880–888 (1997).
59. Morlon, J., Sherratt, D. & Lazdunski, C. Identification of functional regions of the colicinogenic plasmid ColA. *MGG Mol. Gen. Genet.* **211**, 223–230 (1988).
60. Guo, M. *et al.* Structural insights into a high fidelity variant of SpCas9. *Cell Res.* **29**, 183–

- 192 (2019).
61. Nishimasu, H. *et al.* Engineered CRISPR-Cas9 nuclease with expanding targeting space. *Science* (80-. ). **361**, 1259–1262 (2018).
  62. Tadeo, X. *et al.* Structural Basis for the Aminoacid Composition of Proteins from Halophilic Archea. *PLoS Biol.* **7**, 1–9 (2009).
  63. Elcock, A. H. & McCammon, J. A. Electrostatic contributions to the stability of halophilic proteins. *J. Mol. Biol.* **280**, 731–748 (1998).

# PUBLICATION

bioRxiv preprint first posted online Apr. 4, 2019; doi: <http://dx.doi.org/10.1101/597237>. The copyright holder for this preprint (which was not peer-reviewed) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

## 1 **NgAgo-enhanced homologous recombination in *E. coli* is mediated by DNA** 2 **endonuclease activity**

3 Kok Zhi Lee<sup>1</sup>, Michael A. Mechikoff<sup>1</sup>, Archana Kikla<sup>2</sup>, Arren Liu<sup>2</sup>, Paula Pandolfi<sup>2</sup>, Frederick S. Gimble<sup>3,4</sup>,  
4 and Kevin V. Solomon<sup>1,2\*</sup>

5 <sup>1</sup>Department of Agricultural and Biological Engineering, Purdue University, West Lafayette, IN 47906,  
6 USA.

7 <sup>2</sup>Department of Biological Sciences, Purdue University, West Lafayette, IN 47906, USA.

8 <sup>3</sup>Purdue University Interdisciplinary Life Science Program (PULSe), Purdue University, West Lafayette,  
9 IN, 47906, USA.

10 <sup>4</sup>Department of Biochemistry, Purdue University, West Lafayette, IN, 47906, USA.

11

12 \* To whom correspondence should be addressed. Tel: +1 765-494-1134; Fax: +1 765-496-1115; Email:  
13 [kvs@purdue.edu](mailto:kvs@purdue.edu)

14

## 15 **ABSTRACT**

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

29

## 30 INTRODUCTION

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

38 Long pAgos are predicted to serve as a form of adaptive defense against invading nucleic acids such as  
39 phage/viral DNA and RNA<sup>6,7</sup>. With a single-stranded DNA and/or RNA as a guide, long pAgos cleave  
40 complementary target DNA, RNA, or both via the conserved catalytic tetrad, DEDX<sup>1</sup>. To create a double-  
41 stranded DNA break, long pAgos require two guides. Target recognition and cleavage is enabled by four  
42 canonical domains<sup>3</sup>: N (N-terminal), PAZ (PIWI-Argonaute-Zwille), MID (middle), and PIWI (P element-  
43 induced wimpy testis). The N-terminal domain is essential in target cleavage<sup>8,9</sup> and dissociation of cleaved  
44 strands<sup>9,10</sup>, though the detailed mechanism remains poorly understood. The MID domain interacts with the  
45 5'-end of the guide<sup>11</sup> and promotes binding of the guide to its target nucleic acids<sup>12</sup>. The PAZ domain  
46 interacts with the 3' end of a guide<sup>13-16</sup>, protecting it from degradation<sup>17</sup>. The PIWI domain plays a pivotal  
47 role in nucleic acid cleavage via the conserved catalytic tetrad, DEDX (D: aspartate, E: glutamate, X:  
48 histidine, aspartate or asparagine)<sup>5</sup>. Despite the presence of these canonical domains in all long pAgos,  
49 currently characterized pAgos including TtAgo<sup>2</sup>, MpAgo<sup>5</sup>, PfAgo<sup>18</sup> and MjAgo<sup>3,19</sup> work at very high  
50 temperatures (>55 °C)<sup>2,3,5,18</sup>, making them infeasible for gene editing in common mesophilic organisms.

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

## 64 MATERIAL AND METHODS

## 65 **Strains and plasmids**

66 *E. coli* strains and plasmids used in this study are listed in Table 1. Cloning was carried out according to  
67 standard practices<sup>28</sup> with primers, template, and purpose listed in Supplementary Table 1. Plasmids were  
68 maintained in *E. coli* DH5 $\alpha$ . NgAgo variants (wildtype, D663A/D738A, N-del, and repA with GST or His tag)  
69 that were used for *in vitro* activity assays were cloned into a IPTG-inducible T7 plasmid via the pET32a-  
70 GST-ELP64 (provided by Professor Xin Ge, University of California, Riverside).

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

## 74 **NgAgo expression and purification**

75 All GST-NgAgo or His-NgAgo variants were transformed into BL21 (DE3) electrocompetent cells and were  
76 plated on agar plates containing ampicillin (100  $\mu$ g/ml). A single colony was inoculated in LB with ampicillin  
77 for 16 hours and then cultured in 100 ml of LB containing ampicillin. Expression was induced with IPTG at  
78 0.1 mM final concentration when the culture OD<sub>600</sub> reached 0.5. After 4 hours incubation at 37 °C or 22  
79 °C overnight, cells were collected by centrifuge 7500 rpm at 4 °C for 5 minutes. The cell pellet was  
80 resuspended in TN buffer (10 mM Tris and 100mM NaCl, pH 7.5) and lysed via sonication at a medium  
81 power setting (~50 W) in 10 s intervals, with intervening 10 s incubations on ice to reduce heat denaturation.  
82 Cell lysates were then clarified at 12000 rpm at 4 °C for 30 minutes. The supernatant was collected as a  
83 soluble protein fraction. Both soluble and insoluble (cell pellet) fractions were purified via His-IDA nickel  
84 column (Clontech Laboratories, Mountain View, CA. Cat. No: 635657) according to the manufacturer  
85 instructions. Insoluble NgAgo protein was refolded on the column after denaturation with guanidium chloride  
86 according to manufacturer instructions. GST-tagged NgAgo variants were purified by glutathione agarose  
87 (Thermo Fisher Scientific, Waltham, MA. Cat. No: 16100) according to the manufacturer protocol.

## 88 ***In vitro* activity assay**

89 For the reloading protocol, five micrograms of purified NgAgo were mixed with one microgram total of  
90 phosphorylated single-stranded DNA (P-ssDNA) targeting mNeonGreen (Supplementary Table 2) and  
91 incubated at 55 °C for an hour. 200-300 ng of substrate plasmid DNA (pNCS-mNeonGreen) was then  
92 added to the sample. The final volume of the reaction was 50  $\mu$ l (working concentration: 20 mM Tris-Cl,  
93 300 mM KCl, 500  $\mu$ M MgCl<sub>2</sub>, and 2 mM DTT). The sample was then incubated at 37 °C for three hours. 0.8  
94 units of Proteinase K (NEB, Ipswich, MA. Cat. No: P8107S) were added to the sample to digest the protein  
95 for 5 minutes at 37 °C. The nucleic acids were then cleaned up by the DNA Clean & Concentrator™-5 kit  
96 (Zymo Research, Irvine, CA. Cat. No: D4003T) according to manufacturer instructions and mixed with 6X  
97 loading dye containing SDS (Thermo Fisher S, Waltham, MA. Cat. No: R1151) before gel electrophoresis.

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

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

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

#### 110 **Electrophoretic mobility shift assay (EMSA)**

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

#### 118 **Gene-editing assay**

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

#### 129 **Phyre 2 and HHpred analysis**

NgAgo protein (IMG/M Gene ID: 2510572918) was analyzed via Phyre 2<sup>34</sup> 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<sup>34</sup>. NgAgo was analyzed via HHpred<sup>35</sup> (<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<sup>34</sup> and HHpred<sup>35</sup> 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.



163 Structural analysis also identified an uncharacterized oligonucleotide/oligosaccharide-binding (OB) fold  
164 domain between residues 13-102 of NgAgo that is known to bind single-stranded DNA in eukaryotes and  
165 prokaryotes<sup>37</sup> (Fig. 1b). This OB domain has recently been identified as a new feature of pAgo<sup>38</sup>. As repA  
166 proteins were the most common matches on both Phyre 2 and HHpred, we will refer to this OB domain as  
167 repA (Supplementary Tables 5 and 6). While the repA domain is absent in all characterized pAgos, at least  
168 12 Ago homologs from various species deposited on IMG/M (<https://img.jgi.doe.gov/>) share this domain.  
169 Phylogenetic analysis showed that all the repA-containing pAgos were from halophilic Archaea forming a  
170 clade that is distinct from that of the current well-characterized pAgos (Fig. 1c). This monophyletic group of  
171 repA-containing pAgos may represent a new class of pAgos that is currently unrecognized in the literature<sup>39</sup>.  
172 Moreover, its unique presence within halophiles suggests that the repA domain may be required for function  
173 in high salt environments, potentially replacing the role of the canonical N-terminal domain, which was then  
174 truncated through evolution.

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

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

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

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

## 208 **RepA and PIWI domains are responsible for NgAgo DNA cleavage**

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

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

## 228 **repA and PIWI domains are essential for programmable DNA editing**

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

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

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

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

## 258 DISCUSSION

### 259 NgAgo may represent a new class of mesophilic pAgos

260 To our knowledge, NgAgo is the first studied pAgo with an uncharacterized repA domain, which indeed  
261 binds to single-stranded DNA (Fig. 2f). Surprisingly, we found that repA alone contributes to DNA cleavage  
262 activity (Fig. 2e). Moreover, repA aids the PIWI domain in NgAgo targeted DNA cleavage as homologous  
263 recombination is reduced in N-del mutants relative to wildtype (Fig. 3). Interestingly, all repA domain-  
264 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<sup>44</sup>; 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<sup>6</sup>, 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*<sup>41,44</sup>, 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

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

#### 301 **FUNDING**

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

#### 304 **Acknowledgment**

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

#### 308 **CONFLICT OF INTEREST**

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

#### 310 **Author contributions**

311 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  
312 the experiments. K.V.S., F.G., and K.Z.L. supervised research and experimental design. K.V.S., K.Z.L.,  
313 M.A.M., and F.G. wrote the manuscript.

314

## 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. Faehle, 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).

- 351 18. Swarts, D. C. *et al.* Argonaute of the archaeon *Pyrococcus furiosus* is a DNA-guided nuclease that  
352 targets cognate DNA. *Nucleic acids research* **43**, 5120-5129 (2015).
- 353 19. Zander, A. *et al.* Guide-independent DNA cleavage by archaeal Argonaute from *Methanocaldococcus*  
354 *jannaschii*. *Nature Microbiology* **2**, 17034 (2017).
- 355 20. Cyranoski, D. Authors retract controversial NgAgo gene-editing study. *Nature News*,  
356 doi:10.1038/nature.2017.22412 (2017).
- 357 21. Javidi-Parsijani, P. *et al.* No evidence of genome editing activity from *Natronobacterium gregoryi*  
358 Argonaute (NgAgo) in human cells. *Plos One* **12**, 14, doi:10.1371/journal.pone.0177444 (2017).
- 359 22. Wu, Z. *et al.* NgAgo-gDNA system efficiently suppresses hepatitis B virus replication through  
360 accelerating decay of pregenomic RNA. *Antiviral Research* (2017).
- 361 23. Burgess, S. *et al.* Questions about NgAgo. *Protein & Cell* **7**, 913-915, doi:10.1007/s13238-016-0343-9  
362 (2016).
- 363 24. Khin, N. C., Lowe, J. L., Jensen, L. M. & Burgio, G. No evidence for genome editing in mouse zygotes  
364 and HEK293T human cell line using the DNA-guided *Natronobacterium gregoryi* Argonaute (NgAgo).  
365 *PloS one* **12**, e0178768 (2017).
- 366 25. Qin, Y. Y., Wang, Y. M. & Liu, D. NgAgo-based fabp11a gene knockdown causes eye developmental  
367 defects in zebrafish. *Cell Research* **26**, 1349-1352, doi:10.1038/cr.2016.134 (2016).
- 368 26. Elcock, A. H. & McCammon, J. A. Electrostatic contributions to the stability of halophilic proteins.  
369 *Journal of molecular biology* **280**, 731-748 (1998).
- 370 27. Tadeo, X. *et al.* Structural basis for the amino acid composition of proteins from halophilic archaea. *PLoS*  
371 *biology* **7**, e1000257 (2009).
- 372 28. Sambrook, J., Fritsch, E. F. & Maniatis, T. *Molecular cloning: a laboratory manual*. (Cold spring harbor  
373 laboratory press, 1989).
- 374 29. Wood, W. B. Host specificity of DNA produced by *Escherichia coli*: bacterial mutations affecting the  
375 restriction and modification of DNA. *Journal of molecular biology* **16**, 118-113 (1966).
- 376 30. Tseng, H.-C., Martin, C. H., Nielsen, D. R. & Prather, K. L. J. Metabolic engineering of *Escherichia coli*  
377 for enhanced production of (R)- and (S)-3-hydroxybutyrate. *Applied and environmental microbiology* **75**,  
378 3137-3145 (2009).
- 379 31. Niu, Y., Tenney, K., Li, H. & Gimble, F. S. Engineering variants of the I-SceI homing endonuclease with  
380 strand-specific and site-specific DNA-nicking activity. *Journal of molecular biology* **382**, 188-202 (2008).
- 381 32. Rhodius, V. A. *et al.* Design of orthogonal genetic switches based on a crosstalk map of  $\sigma$ s, anti- $\sigma$ s,  
382 and promoters. *Molecular systems biology* **9**, 702 (2013).
- 383 33. Reisch, C. R. & Prather, K. L. J. The no-SCAR (Scarless Cas9 Assisted Recombineering) system for  
384 genome editing in *Escherichia coli*. *Scientific reports* **5**, 15096 (2015).
- 385 34. Kelley, L. A., Mezulis, S., Yates, C. M., Wass, M. N. & Sternberg, M. J. E. The Phyre2 web portal for  
386 protein modeling, prediction and analysis. *Nature protocols* **10**, 845-858 (2015).

387 35. Zimmermann, L. *et al.* A completely Reimplemented MPI bioinformatics toolkit with a new HHpred  
388 server at its Core. *Journal of molecular biology* **430**, 2237-2243 (2018).

389 36. Söding, J., Biegert, A. & Lupas, A. N. The HHpred interactive server for protein homology detection  
390 and structure prediction. *Nucleic acids research* **33**, W244-W248 (2005).

391 37. Flynn, R. L. & Zou, L. Oligonucleotide/oligosaccharide-binding fold proteins: a growing family of  
392 genome guardians. *Critical reviews in biochemistry and molecular biology* **45**, 266-275 (2010).

393 38. Ryazansky, S., Kulbachinskiy, A. & Aravin, A. The expanded universe of prokaryotic Argonaute  
394 proteins. *bioRxiv*, 366930 (2018).

395 39. Ryazansky, S., Kulbachinskiy, A. & Aravin, A. A. The expanded universe of prokaryotic Argonaute  
396 proteins. *mBio* **9**, e01935-01918 (2018).


397 40. Müller-Santos, M. *et al.* First evidence for the salt-dependent folding and activity of an esterase from  
398 the halophilic archaea Haloarcula marismortui. *Biochimica et Biophysica Acta (BBA)-Molecular and*  
399 *Cell Biology of Lipids* **1791**, 719-729 (2009).

400 41. Sunghyeok, Y. *et al.* DNA-dependent RNA cleavage by the Natronobacterium gregoryi Argonaute.  
401 *BioRxiv*, 101923 (2017).

402 42. Swarts, D. C. *et al.* Autonomous Generation and Loading of DNA Guides by Bacterial Argonaute.  
403 *Molecular Cell* **65**, 985-998 (2017).

404 43. Fu, L. *et al.* The prokaryotic Argonaute proteins enhance homology sequence-directed recombination  
405 in bacteria. *Nucleic acids research* (2019).

406 44. Hunt, E. A., Evans Jr, T. C. & Tanner, N. A. Single-stranded binding proteins and helicase enhance the  
407 activity of prokaryotic argonautes in vitro. *PloS one* **13**, e0203073 (2018).

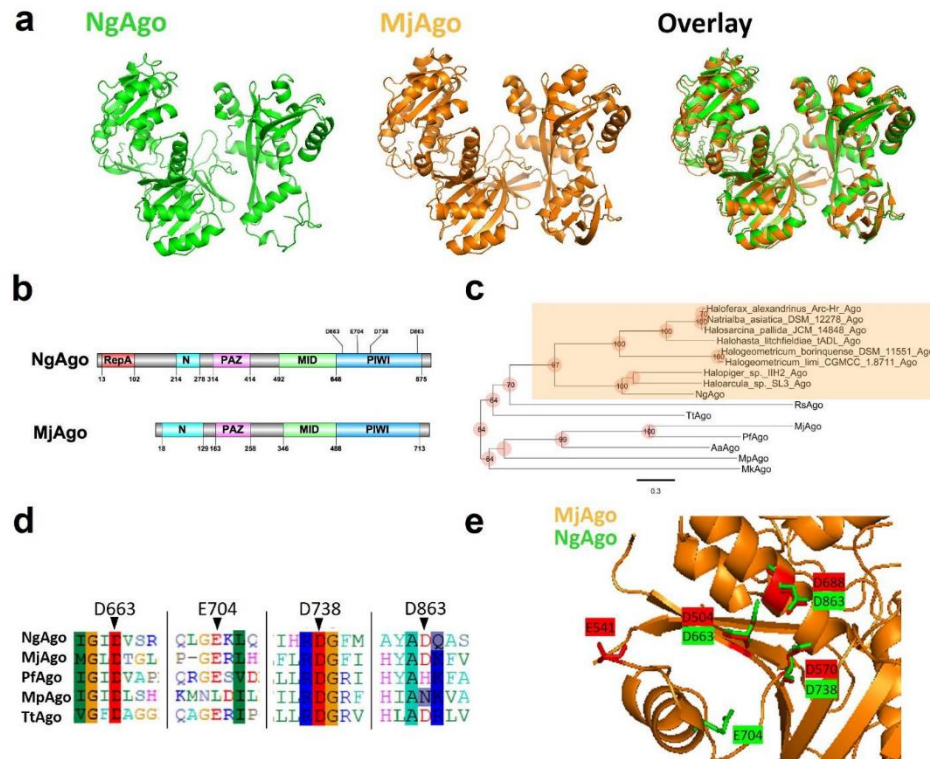
408 

409

410



# 411 FIGURES

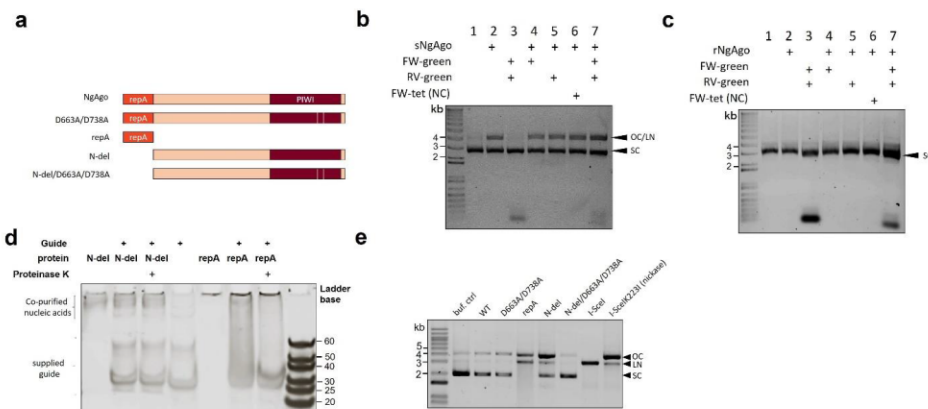


412

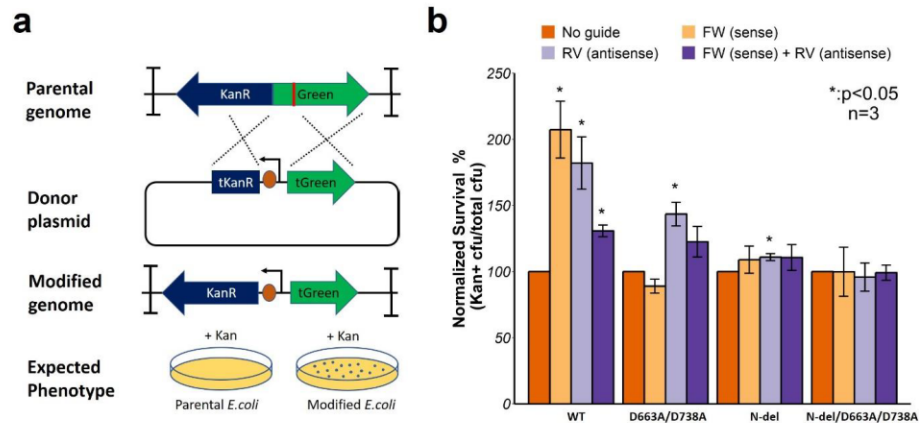
413

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 copurify 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  $\lambda$ -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
<b>Strains</b>				
BL21 (DE3)	F- ompT gal dcm lon hsdSB(rB- mB-) λ (DE3) [lacI lacUV5-T7p07 ind1 sam7 nin5] [malB+]K-12(λS)			<sup>29</sup>
MG1655 (DE3)	K-12 F- λ- ilvG- rfb-50 rph-1 (DE3)			<sup>30</sup>
MG1655 (DE3) <i>atpI::KanR</i> -mNeonGreen	K-12 F- λ- ilvG- rfb-50 rph-1 (DE3) <i>atpI::KanR</i> -mNeonGreen			This study
<b>Plasmids</b>				
pBSI-Scel(E/H)	<i>bla</i>		ColE1 derivative	<sup>31</sup>
pET32a-GST-ELP64	<i>bla</i> , <i>lacI</i> , P <sub>T7</sub> -GST-ELP64			Professor Xin Ge (University of California, Riverside)
pTKDP-hph	<i>bla</i> , <i>hph</i> , <i>sacB</i>		pMB1	<sup>32</sup>
pCas9-CR4	<i>cat</i> , P <sub>Tet</sub> -Cas9		p15A	<sup>33</sup>
pET-GST-Ago-His	<i>bla</i> , <i>lacI</i> , P <sub>T7</sub> -GST-NgAgo-His	pET32a-GST-ELP64	pBR322	This study
pET32a-His-Ago	<i>bla</i> , <i>lacI</i> , P <sub>T7</sub> -GST-NgAgo-His	pET32a-GST-ELP64	pBR322	This study
pET32a-His-repA	<i>bla</i> , <i>lacI</i> , P <sub>T7</sub> -His-repA	pET32a-GST-ELP64	pBR322	This study
pET-GST-N-del-His	<i>bla</i> , <i>lacI</i> , P <sub>T7</sub> -GST-N-del-His	pET32a-GST-ELP64	pBR322	This study
pET-GST-N-del/D663A/D738A-His	<i>bla</i> , <i>lacI</i> , P <sub>T7</sub> -GST-N-del/D663A/D738A-His	pET32a-GST-ELP64	pBR322	This study
pTKDP-KanR-mNeonGreen-hph	<i>bla</i> , <i>hph</i> , KanR-mNeonGreen	pTKDP-hph	pMB1	This study
p15-KanR-PtetRed	<i>cat</i> , KanR-mNeonGreen, P <sub>Tet</sub> -gam-beta-exo	pCas9-CR4	p15A	This study
pET32-BFP	<i>Amp</i> , <i>lacI</i> , P <sub>T7</sub> -BFP	pET32a-GST-ELP64 and pBAD-mTagBFP2	pBR322	This study

449