DNA-guided genome
manipulation in *Escherichia coli*

Thesis by
Shan Huang

In Partial Fulfillment of the Requirements
for the degree of
Doctor of Philosophy

Caltech

CALIFORNIA INSTITUTE OF TECHNOLOGY
Pasadena, California

2023
(Defended July 19, 2022)

© 2023

Shan Huang
ORCID: 0000-0002-4436-3327

ACKNOWLEDGEMENTS

iv

ABSTRACT

Argonaute proteins (Agos) were initially discovered in eukaryotes as key players in RNA interference (RNAi) pathways and later found in prokaryotes. Some prokaryotic argonautes (pAgos) have been shown to mediate nucleic acid-guided cleavage of DNA targets, reminiscent of the nucleic acid-guided DNase activity of the CRISPR/Cas9 system. It has been postulated that pAgo variants might be used as a novel genome-editing tool. However, genome manipulation induced by pAgo-mediated DNA cleavage has never been established. To demonstrate that pAgo-mediated DNA cleavage can introduce genomic mutations in *Escherichia coli*, we first created a recombination system and showed that CbAgo, a pAgo from *Clostridium butyricum*, can be directed by plasmid-encoded guide sequences to cleave the genome target site and induce chromosome recombination between downstream direct repeat sequences. Results from testing different pAgo variants suggest that the recombination rate correlates well with pAgo DNA cleavage activity, and the mechanistic study suggests the recombination involves DSB generation and RecBCD processing. In RecA-deficient *E. coli* strain, guide-directed CbAgo cleavage on chromosomes severely impairs cell growth, which can be utilized as counter-selection to assist Lambda-Red recombineering. These findings demonstrate the guide-directed cleavage of pAgo on the host genome is mutagenic and can lead to different outcomes according to the function of the host DNA repair machinery. Furthermore, we created a dCbAgo-based deaminase and showed that it can not only act as a random mutagen in vivo but also has the potential to be directed by plasmid-encoded guide sequences. We anticipate the novel DNA-guided interference by pAgo only or by its fusion protein to be useful in broader genetic manipulation. My work of engineering fluorescent protein-based nicotine biosensors via computational design and experimental evolution is also described in the thesis.

PUBLISHED CONTENT AND CONTRIBUTIONS

S. Huang, K.Wang, S.L. Mayo (2021). "Guide-directed DNA cleavage by a prokaryotic Argonaute protein induces chromosome recombination in *Escherichia coli*". *bioRxiv*. DOI: 10.1101/2021.11.17.469050.

Contribution: S. Huang participated in the project's conceptualization and designed the study. S. Huang performed all experiments, analyzed and interpreted all experimental data, and prepared the draft of the manuscript, which was reformatted into the majority of Chapters 2 and 3 of this thesis.

TABLE OF CONTENTS

LIST OF ILLUSTRATIONS AND/OR TABLES

C h a p t e r   1

INTRODUCTION

Argonaute proteins (Agos) have first been demonstrated to be the key player in eukaryotic RNA interference (RNAi) pathways in 2001 (Hammond et al., 2001). The characterization of RNAi reveals that typical eukaryotic Agos (eAgos) function as nucleases that use small RNAs (smRNAs) as guides to target and cleave complementary RNA molecules (Meister, 2013). The currently identified eAgo family consists of three phylogenetically distant clades: (I) the AGO clade, (II) the PIWI clade, and (III) the *Caenorhabditis elegans*-specific WAGO clade (Peters and Meister, 2007; Hutvagner and Simard, 2008; Siomi et al., 2011; Meister, 2013). The AGO clade proteins are the actual effector of the RNA-induced silencing complex (RISC) that mediates cytoplasmic RNAi processes, including mRNA degradation and translational inhibition (Treiber et al., 2019). Before triggering RNAi, RISC has to first load double-stranded RNA (dsRNAs), which are generated from the DICER processing of small interfering RNAs (siRNAs) or microRNAs (miRNAs) in the cytoplasm. The PIWI clade proteins mainly function in germline cells, where they mediate transcriptional silencing of transposons assisted by PIWI-interacting RNAs (piRNAs) (Malone and Hannon, 2009; Siomi et al., 2011). Interestingly, piRNAs are generated from single-stranded RNAs (ssRNAs), which are expressed from genomic piRNA clusters (Aravin et al., 2006). The WAGO clade proteins are found in *Caenorhabditis elegans*, which also participate in RNAi but lack mRNA cleavage activity. They acquire their smRNAs from "primary" eAgos or through small RNA amplification processes (Yigit et al., 2006).

The initial Ago research has focused on eAgos, but recent years have witnessed the expanding research and knowledge about prokaryotic Agos (pAgos), as described below.

1.1 Structure and function of pAgos

Analysis of prokaryotic genomes shows that ~30% of archaea genomes and ~10% of bacteria genomes contain pAgo genes (Makarova et al., 2009; Swarts et al., 2014b) and they appear to be much more diverse than their eAgo counterparts. The pAgo family can be very roughly divided into two groups: long pAgos and short pAgos. Long pAgos contain all four conserved domains present in eAgos including the N-terminal (N), PIWI-Argonaute-Zwille (PAZ), middle (MID), and P-element-induced wimpy testis (PIWI) domains. The N domain facilitates the target binding and cleavage. The MID and PAZ domains are responsible for guide and target binding. The PIWI domain contains a ribonuclease H (RNase H) site that mediates the target cleavage (Parker et al., 2005; Ryazansky et al., 2018). It is worth noting that both the guide and target molecules of pAgos are single-stranded DNAs (ssDNAs) or RNAs (ssRNAs), similar to their eukaryotic counterparts.

Structural analysis of pAgo reveals its bilobed structure, where the N and PAZ domains form one lobe and the MID and PIWI domains form another, and the two lobes accommodate both the guide and the target molecules. The 5'- and 3'-ends of the guide are anchored in the binding pockets of the MID and PAZ domains respectively (Wang et al., 2008). The extensive interactions between the guide and the PAZ domain appear to be important for pAgo target specificity (Jung et al., 2013; Zander et al., 2014) and may also increase the stability of the guide molecule by blocking cellular nuclease attack (Hur et al., 2013). The position 2–6 of guide nucleotides are pre-oriented in a helical conformation and exposed to the solution to facilitate the initial target binding, via the reduction of the entropic penalty for the duplex formation (Wang et al., 2008b; Elkayam et al., 2012; Kaya et al., 2016; Willkomm et al., 2017). Target binding of this region of the guide molecule results in the exposure of its downstream nucleotides through conformational changes and facilitates their pairing with the target (Wang et al., 2009; Sheng et al., 2014; Schirle et al., 2014). Extensive guide-target pairing results in the conformational changes in the PAZ and PIWI domains, accompanied by the release of the 3'-end of the guide previously anchored in the PAZ pocket. Eventually, the cleavage-favored configuration is formed in the active

site and the subsequent cleavage is enabled (Wang et al., 2008a; Wang et al., 2009; Sheng et al., 2014).

The nucleic acid cleavage site is highly conserved in active pAgos. It consists of four amino acid residues, DEDX (X stands for D, H, or K), which binds two divalent metal ions such as Mg^{2+} or Mn^{2+} and cleaves specific phosphodiester bonds in the target molecules (Makarova et al., 2009; Swarts et al., 2014b). When there is no target bound to the pAgo, one glutamate residue stays outside the active site, and therefore the metal ion chelation in the active site is substantially impaired (Wang et al., 2008b; Kaya et al., 2016; Willkomm et al., 2017). On the contrary, the guide-target pairing in the pAgo results in dramatic conformational changes in the PIWI and PAZ domains which direct that glutamate residue into the active site and enable its metal ion binding and subsequent target cleavage (Wang et al., 2009; Sheng et al., 2014). For most active pAgos and eAgos, target cleavage happens between the two nucleotides that pair with the 10th and 11th nucleotides of the guide (Yuan et al, 2005; Wang et al., 2008a; Wang et al., 2009; Zander et al., 2014; Kaya et al., 2016; Willkomm et al., 2017).

The last step of the catalytic cycle is the product release, which appears to be the rate-limiting step for many pAgos. For example, kinetic studies show that the gDNA-bound CbAgo remains bound to the cleaved target for a prolonged time before it dissociates and cleaves new DNA targets (Hegge et al., 2019; Kuzmenko et al., 2019). Such persisting complementary guide-target interactions are also observed in eAgo kinetic analysis (Haley and Zamore, 2004; Rivas et al., 2005; Ameres et al., 2007; Salomon et al., 2015). However, the target release is not rate-limiting in TtAgo-mediated DNA cleavage, because a more rapid target release is allowed at the high temperature used in the assays (Wang et al., 2009).

All short pAgos and some long pAgos such as AfAgo (*Archaeoglobus fulgidus*) (Ma et al., 2005) and RsAgo (*Rhodobacter sphaeroides*) (Miyoshi et al., 2016), are unable to perform target cleavage because their target cleavage sites are either compromised or completely absent. However, that catalytically inactive pAgos can still load guides and bind to complementary targets. Their cellular functions remain largely unknown, but their

encoding genes are often found in the same clusters with genes encoding putative nucleases, implicating these pAgos and nucleases may collaborate in the generation and loading of guide molecules and the subsequent target binding (Swarts et al., 2014b).

1.2 Characterized pAgo nucleases

Many active pAgos from diverse species have been identified and characterized in the past decade, among which most exhibit DNA-guided DNA cleavage activity, including TtAgo (*Thermus thermophilus*) (Swarts et al., 2014a), PfAgo (*Pyrococcus furiosus*) (Swarts et al., 2015), MjAgo (*Methanocaldococcus jannaschii*) (Zander et al., 2017; Willkomm et al., 2017), CbAgo (*Clostridium butyricum*) (Hegge et al., 2019; Kuzmenko et al., 2019; García-Quintans et al., 2020), LrAgo (*Limnothrix rosea*) (Kuzmenko et al., 2019), CpAgo (*Clostridium perfringens*), IbAgo (*Intestinibacter bartlettii*) (Cao et al., 2019), SeAgo (*Synechococcus elongatus*) (Olina et al., 2020), KmAgo (*Kurthia massiliensis*) (Liu et al., 2021; Kropocheva et al., 2021), and FpAgo (*Ferroglobus placidus*) (Guo et al., 2021). 5'-phosphorylated gDNA (5'P-gDNA) can be readily utilized by these pAgos to cleave their targets, but only CbAgo, LrAgo, CpAgo, IbAgo, SeAgo, and KmAgo can utilize 5'-hydroxylated gDNA (5'OH-gDNA) as guides in the cleavage. Some pAgos can use small guide RNAs (gRNAs) to cleave ssDNA substrates, including AaAgo (*Aquifex aeolicus*) (Yuan et al., 2005), MpAgo (*Marinitoga piezophila*), TpAgo (*Thermotoga profunda*) (Kaya et al., 2016), and KmAgo (Liu et al., 2021; Kropocheva et al., 2021). Additionally, some pAgos such as AaAgo, TtAgo, MpAgo, CpAgo, and KmAgo, can cleave ssRNA substrates as well.

In fact, 5'OH-gDNA, 5'P-gDNA, 5'OH-gRNA, and 5'P-gRNA have all been shown to support guide-directed pAgo cleavage on ssDNA and ssRNA targets. The most omnipotent pAgo that has ever been reported is KmAgo, as it can use 5'OH-gDNA, 5'P-gDNA, and 5'P-gRNA as guides for target cleavage (the 5'P-gRNA-directed cleavage is relatively inefficient), except 5'OH-gRNA (Liu et al., 2021; Kropocheva et al., 2021). Interestingly, MpAgo and TpAgo preferentially utilize 5'OH-gRNA to target nucleic acids, which distinguish them from other pAgos (Kaya et al., 2016).

Recently, Vaiskunaite et al. directly compared the ssDNA cleavage activity of different putative pAgo nucleases using 5'P-gDNAs as guides and found that CbAgo is the most active ssDNA nuclease among the tested 45 pAgo candidates (Vaiskunaite et al. 2022). They further demonstrated the *E. coli* endogenous helicase RecBCD can efficiently assist CbAgo in cleaving linear dsDNA at 37°C, while previous reports showed that CbAgo alone can only cleave linear dsDNA at elevated temperature (Kuzmenko et al., 2019).

1.3 Recent advances in pAgo research

Recent research demonstrated that TtAgo participates in the host DNA replication, during which TtAgo binds to gDNAs that are derived from the termination region on chromosomes and helps the host disentangle the newly generated, catenated circular chromosomes with the help of other protein factors (Jolly et al., 2020). Besides, NgAgo has been shown to stimulate homologous recombination (Fu et al., 2019; Lee et al., 2021). However, the involvement of the pAgo-mediated cleavages on chromosomes and the resulting chromosomal single-stranded or double-stranded breaks (DSBs) in the above processes have not been confirmed. By contrast, observations in these reports are against the assumption that DSBs are required for those processes.

In the area of short pAgo, a novel in vivo pathway of short pAgo has been recently reported (Koopal et al., 2022), where researchers found that short pAgo can associate with a fusion protein of a PAZ analog (APAZ) domain and a Toll/interleukin-1 receptor (TIR)-like domain, and form a short pAgo TIR-APAZ (SPARTA) complex. When a high-copy plasmid is present in the cell, the pAgo abstracts gRNAs from the plasmid transcripts and mediate the binding of the SPARTA complex to the plasmid DNA. Target binding of the SPARTA complex results in the activation of the TIR domain, which depletes cellular $NAD(P)^+$ and reduces cell viability, which slows down the invasion of the plasmids in the bacterial population.

For the application of pAgos, their potential in nucleic acid detection has been extensively explored (He et al., 2019; Song et al., 2020; Wang et al., 2021a,b; Liu et al., 2021). Besides, PfAgo has been used as an artificial restriction enzyme to develop a molecular cloning

platform termed PlasmidMaker (Enghiad and Zhao, 2017; Enghiad et al., 2022). This platform is a plasmid construction streamlined with a robotic system, out of which more than 100 plasmids have been created by combining different DNA fragments.

1.4 In vivo DNA interference by pAgo

One of the major functions of pAgo is to help their hosts defend against invading DNAs, which has been observed for certain long pAgos (Swarts et al., 2014b, 2015; Zander et al., 2017).

Recently, Kuzmenko et al. showed a detailed defense mechanism against invading DNAs mediated by CbAgo in *E. coli* (Kuzmenko et al., 2020), which is essentially an in vivo DNA interference pathway. First, small DNAs generated from spontaneous degradation of plasmids or other multicopy genetic elements are loaded by CbAgo as guides (Figure 1.1). Next, gDNA-bound CbAgo introduces DSBs at the corresponding sites, including chromosomal regions, and triggers *E. coli* exonuclease RecBCD to cause DNA degradation. Invading DNAs such as plasmids and phages can thus be targeted and efficiently eliminated through this mechanism, which is confirmed by experiments in CbAgo-expressing *E. coli*. This research is very interesting from the genome editing point of view because it demonstrates that CbAgo can be directed to the specific genomic regions, generate DSBs there, and trigger the *E. coli* endogenous DNA repair mechanism, which may be utilized as a source to introduce genomic mutations. However, whether such a CbAgo-induced DSB is mutagenic has not been explored in the report.

1.5 Thesis outline

The chapters that follow detail the work I have done toward the goal of implementing DNA-guided genome manipulation in *E. coli* using pAgo. Chapter 2 details the creation of a recombination system in *E. coli*. Chapter 3 describes the characterization of the aforementioned recombination system as well as the introduction of CbAgo-assisted recombineering. Chapter 4 describes the development of the dCbAgo-based deaminase as a random or DNA-guided mutator. Chapter 5 describes the engineering of nicotine

biosensors. Chapter 6 summarizes the major contributions of this work, and discusses future work that will be required to develop a DNA-guided genome editing system by pAgo.
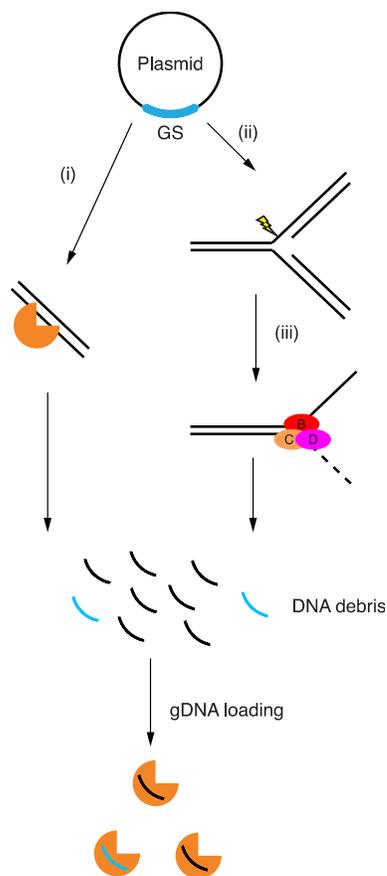
Figure 1.1: Proposed model for gDNA biogenesis of CbAgo. As commonly found in pAgos studies, CbAgo generates and binds small gDNA from plasmids (Hegge et al., 2019; Kuzmenko et al., 2019; Kuzmenko et al., 2020). Small DNA fragments can be generated from guide-free CbAgo mediated plasmid degradation ("chopping" activity), likely at partially unwound dsDNA in AT-rich regions (step i). Alternatively, RecBCD-dependent plasmid degradation during plasmid replication (steps ii and iii) can also generate small DNA debris, a process similar to the spacer acquisition in CRISPR adaptation (Levy et al., 2015; Ivančić-Baće et al., 2015; Modell et al., 2017). The resulting small DNA fragments can serve as substrates for gDNA loading by CbAgo. The RecBCD-dependent plasmid degradation may also be responsible for dCbAgo gDNA loading.

BIBLIOGRAPHY

Ameres, S.L., Martinez, J., Schroeder, R., 2007. Molecular Basis for Target RNA Recognition and Cleavage by Human RISC. Cell 130, 101–112.

Aravin, A., Gaidatzis, D., Pfeffer, S., Lagos-Quintana, M., Landgraf, P., Iovino, N., Morris, P., Brownstein, M.J., Kuramochi-Miyagawa, S., Nakano, T., Chien, M., Russo, J.J., Ju, J., Sheridan, R., Sander, C., Zavolan, M., Tuschl, T., 2006. A novel class of small RNAs bind to MILI protein in mouse testes. Nature 442, 203–207.

Cao, Y., Sun, W., Wang, J., Sheng, G., Xiang, G., Zhang, T., Shi, W., Li, C., Wang, Y., Zhao, F., Wang, H., 2019. Argonaute proteins from human gastrointestinal bacteria catalyze DNA-guided cleavage of single- and double-stranded DNA at 37 °C. Cell Discov 5, 1–4.

Elkayam, E., Kuhn, C.-D., Tocilj, A., Haase, A.D., Greene, E.M., Hannon, G.J., Joshua-Tor, L., 2012. The Structure of Human Argonaute-2 in Complex with miR-20a. Cell 150, 100–110.

Enghiad, B., Xue, P., Singh, N., Boob, A.G., Shi, C., Petrov, V.A., Liu, R., Peri, S.S., Lane, S.T., Gaither, E.D., Zhao, H., 2022. PlasmidMaker is a versatile, automated, and high throughput end-to-end platform for plasmid construction. Nat Commun 13, 2697.

Enghiad, B., Zhao, H., 2017. Programmable DNA-Guided Artificial Restriction Enzymes. ACS Synth. Biol. 6, 752–757.

Fu, L., Xie, C., Jin, Z., Tu, Z., Han, L., Jin, M., Xiang, Y., Zhang, A., 2019. The prokaryotic Argonaute proteins enhance homology sequence-directed recombination in bacteria. Nucleic Acids Research 47, 3568–3579.

García-Quintans, N., Bowden, L., Berenguer, J., Mencía, M., 2020. DNA interference by a mesophilic Argonaute protein, CbcAgo. F1000Res 8, 321.

Guo, X., Sun, Y., Chen, L., Huang, F., Liu, Q., Feng, Y., 2021. A Hyperthermophilic Argonaute From Ferroglobus placidus With Specificity on Guide Binding Pattern. Frontiers in Microbiology 12.

Haley, B., Zamore, P.D., 2004. Kinetic analysis of the RNAi enzyme complex. Nat Struct Mol Biol 11, 599–606.

Hammond, S.M., Boettcher, S., Caudy, A.A., Kobayashi, R., Hannon, G.J., 2001. Argonaute2, a Link Between Genetic and Biochemical Analyses of RNAi. Science 293, 1146–1150.

He, R., Wang, L., Wang, F., Li, W., Liu, Y., Li, A., Wang, Y., Mao, W., Zhai, C., Ma, L., 2019. Pyrococcus furiosus Argonaute-mediated nucleic acid detection. Chemical Communications 55, 13219–13222.

Hegge, J.W., Swarts, D.C., Chandradoss, S.D., Cui, T.J., Kneppers, J., Jinek, M., Joo, C., van der Oost, J., 2019. DNA-guided DNA cleavage at moderate temperatures by Clostridium butyricum Argonaute. Nucleic Acids Research 47, 5809–5821.

Hur, J.K., Zinchenko, M.K., Djuranovic, S., Green, R., 2013. Regulation of Argonaute Slicer Activity by Guide RNA 3′ End Interactions with the N-terminal Lobe *. Journal of Biological Chemistry 288, 7829–7840.

Hutvagner, G., Simard, M.J., 2008. Argonaute proteins: key players in RNA silencing. Nat Rev Mol Cell Biol 9, 22–32.

Ivančić-Baće, I., Cass, S.D., Wearne, S.J., Bolt, E.L., 2015. Different genome stability proteins underpin primed and naïve adaptation in E. coli CRISPR-Cas immunity. Nucleic Acids Research 43, 10821–10830.

Jolly, S.M., Gainetdinov, I., Jouravleva, K., Zhang, H., Strittmatter, L., Bailey, S.M., Hendricks, G.M., Dhabaria, A., Ueberheide, B., Zamore, P.D., 2020. Thermus thermophilus Argonaute Functions in the Completion of DNA Replication. Cell 182, 1545-1559.e18.

Jung, S.-R., Kim, E., Hwang, W., Shin, S., Song, J.-J., Hohng, S., 2013. Dynamic Anchoring of the 3′-End of the Guide Strand Controls the Target Dissociation of Argonaute–Guide Complex. J. Am. Chem. Soc. 135, 16865–16871.

Kaya, E., Doxzen, K.W., Knoll, K.R., Wilson, R.C., Strutt, S.C., Kranzusch, P.J., Doudna, J.A., 2016. A bacterial Argonaute with noncanonical guide RNA specificity. Proceedings of the National Academy of Sciences 113, 4057–4062.

Koopal, B., Potocnik, A., Mutte, S.K., Aparicio-Maldonado, C., Lindhoud, S., Vervoort, J.J.M., Brouns, S.J.J., Swarts, D.C., 2022. Short prokaryotic Argonaute systems trigger cell death upon detection of invading DNA. Cell 185, 1471-1486.e19.

Kropocheva, E., Kuzmenko, A., Aravin, A.A., Esyunina, D., Kulbachinskiy, A., 2021. A programmable pAgo nuclease with universal guide and target specificity from the mesophilic bacterium Kurthia massiliensis. Nucleic Acids Research 49, 4054–4065.

Kuzmenko, A., Oguienko, A., Esyunina, D., Yudin, D., Petrova, M., Kudinova, A., Maslova, O., Ninova, M., Ryazansky, S., Leach, D., Aravin, A.A., Kulbachinskiy, A., 2020. DNA targeting and interference by a bacterial Argonaute nuclease. Nature 587, 632–637.

Kuzmenko, A., Yudin, D., Ryazansky, S., Kulbachinskiy, A., Aravin, A.A., 2019. Programmable DNA cleavage by Ago nucleases from mesophilic bacteria Clostridium butyricum and Limnothrix rosea. Nucleic Acids Research 47, 5822–5836.

Lee, K.Z., Mechikoff, M.A., Kikla, A., Liu, A., Pandolfi, P., Fitzgerald, K., Gimble, F.S., Solomon, K.V., 2021. NgAgo possesses guided DNA nicking activity. Nucleic Acids Research 49, 9926–9937.

Levy, A., Goren, M.G., Yosef, I., Auster, O., Manor, M., Amitai, G., Edgar, R., Qimron, U., Sorek, R., 2015. CRISPR adaptation biases explain preference for acquisition of foreign DNA. Nature 520, 505–510.

Liu, Q., Guo, X., Xun, G., Li, Z., Chong, Y., Yang, L., Wang, H., Zhang, F., Luo, S., Cui, L., Zhao, P., Ye, X., Xu, H., Lu, H., Li, X., Deng, Z., Li, K., Feng, Y., 2021. Argonaute integrated single-tube PCR system enables supersensitive detection of rare mutations. Nucleic Acids Research 49, e75.

Liu, Y., Li, W., Jiang, X., Wang, Y., Zhang, Z., Liu, Q., He, R., Chen, Q., Yang, J., Wang, L., Wang, F., Ma, L., 2021. A programmable omnipotent Argonaute nuclease from mesophilic bacteria Kurthia massiliensis. Nucleic Acids Research 49, 1597–1608.

Ma, J.-B., Yuan, Y.-R., Meister, G., Pei, Y., Tuschl, T., Patel, D.J., 2005. Structural basis for 5′-end-specific recognition of guide RNA by the A. fulgidus Piwi protein. Nature 434, 666–670.

Makarova, K.S., Wolf, Y.I., van der Oost, J., Koonin, E.V., 2009. Prokaryotic homologs of Argonaute proteins are predicted to function as key components of a novel system of defense against mobile genetic elements. Biology Direct 4, 29.

Malone, C.D., Hannon, G.J., 2009. Small RNAs as Guardians of the Genome. Cell 136, 656–668.

Meister, G., 2013. Argonaute proteins: functional insights and emerging roles. Nat Rev Genet 14, 447–459.

Miyoshi, T., Ito, K., Murakami, R., Uchiumi, T., 2016. Structural basis for the recognition of guide RNA and target DNA heteroduplex by Argonaute. Nat Commun 7, 11846.

Modell, J.W., Jiang, W., Marraffini, L.A., 2017. CRISPR–Cas systems exploit viral DNA injection to establish and maintain adaptive immunity. Nature 544, 101–104.

Nazer, E., Gómez Acuña, L., Kornblihtt, A.R., 2022. Seeking the truth behind the myth: Argonaute tales from "nuclearland." Molecular Cell 82, 503–513.

Olina, A., Kuzmenko, A., Ninova, M., Aravin, A.A., Kulbachinskiy, A., Esyunina, D., 2020. Genome-wide DNA sampling by Ago nuclease from the cyanobacterium Synechococcus elongatus. RNA Biology 17, 677–688.

Parker, J.S., Roe, S.M., Barford, D., 2005. Structural insights into mRNA recognition from a PIWI domain–siRNA guide complex. Nature 434, 663–666.

Peters, L., Meister, G., 2007. Argonaute Proteins: Mediators of RNA Silencing. Molecular Cell 26, 611–623.

Rivas, F.V., Tolia, N.H., Song, J.-J., Aragon, J.P., Liu, J., Hannon, G.J., Joshua-Tor, L., 2005. Purified Argonaute2 and an siRNA form recombinant human RISC. Nat Struct Mol Biol 12, 340–349.

Ryazansky, S., Kulbachinskiy, A., Aravin, A.A., 2018. The Expanded Universe of Prokaryotic Argonaute Proteins. mBio 9, e01935-18.

Salomon, W.E., Jolly, S.M., Moore, M.J., Zamore, P.D., Serebrov, V., 2015. Single-Molecule Imaging Reveals that Argonaute Reshapes the Binding Properties of Its Nucleic Acid Guides. Cell 162, 84–95.

Schirle, N.T., Sheu-Gruttadauria, J., MacRae, I.J., 2014. Structural basis for microRNA targeting. Science 346, 608–613.

Sheng, G., Zhao, H., Wang, J., Rao, Y., Tian, W., Swarts, D.C., van der Oost, J., Patel, D.J., Wang, Y., 2014. 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.

Siomi, M.C., Sato, K., Pezic, D., Aravin, A.A., 2011. PIWI-interacting small RNAs: the vanguard of genome defence. Nat Rev Mol Cell Biol 12, 246–258.

Song, J., Hegge, J.W., Mauk, M.G., Chen, J., Till, J.E., Bhagwat, N., Azink, L.T., Peng, J., Sen, M., Mays, J., Carpenter, E.L., van der Oost, J., Bau, H.H., 2020. Highly specific enrichment of rare nucleic acid fractions using Thermus thermophilus argonaute with applications in cancer diagnostics. Nucleic Acids Research 48, e19.

Swarts, D.C., Hegge, J.W., Hinojo, I., Shiimori, M., Ellis, M.A., Dumrongkulraksa, J., Terns, R.M., Terns, M.P., van der Oost, J., 2015. Argonaute of the archaeon Pyrococcus furiosus is a DNA-guided nuclease that targets cognate DNA. Nucleic Acids Research 43, 5120–5129.

Swarts, D.C., Jore, M.M., Westra, E.R., Zhu, Y., Janssen, J.H., Snijders, A.P., Wang, Y., Patel, D.J., Berenguer, J., Brouns, S.J.J., van der Oost, J., 2014a. DNA-guided DNA interference by a prokaryotic Argonaute. Nature 507, 258–261.

Swarts, D.C., Makarova, K., Wang, Y., Nakanishi, K., Ketting, R.F., Koonin, E.V., Patel, D.J., van der Oost, J., 2014b. The evolutionary journey of Argonaute proteins. Nat Struct Mol Biol 21, 743–753.

Treiber, T., Treiber, N., Meister, G., 2019. Regulation of microRNA biogenesis and its crosstalk with other cellular pathways. Nat Rev Mol Cell Biol 20, 5–20.

Vaiskunaite, R., Vainauskas, J., Morris, J.J.L., Potapov, V., Bitinaite, J., 2022. Programmable cleavage of linear double-stranded DNA by combined action of Argonaute CbAgo from Clostridium butyricum and nuclease deficient RecBC helicase from E. coli. Nucleic Acids Research 50, 4616–4629.

Wang, F., Yang, J., He, R., Yu, X., Chen, S., Liu, Y., Wang, L., Li, A., Liu, L., Zhai, C., Ma, L., 2021. PfAgo-based detection of SARS-CoV-2. Biosensors and Bioelectronics 177, 112932.

Wang, L., He, R., Lv, B., Yu, X., Liu, Y., Yang, J., Li, W., Wang, Y., Zhang, H., Yan, G., Mao, W., Liu, L., Wang, F., Ma, L., 2021. Pyrococcus furiosus Argonaute coupled with modified ligase chain reaction for detection of SARS-CoV-2 and HPV. Talanta 227, 122154.

Wang, Y., Juranek, S., Li, H., Sheng, G., Tuschl, T., Patel, D.J., 2008a. Structure of an argonaute silencing complex with a seed-containing guide DNA and target RNA duplex. Nature 456, 921–926.

Wang, Y., Juranek, S., Li, H., Sheng, G., Wardle, G.S., Tuschl, T., Patel, D.J., 2009. Nucleation, propagation and cleavage of target RNAs in Ago silencing complexes. Nature 461, 754–761.

Wang, Y., Sheng, G., Juranek, S., Tuschl, T., Patel, D.J., 2008b. Structure of the guide-strand-containing argonaute silencing complex. Nature 456, 209–213.

Willkomm, S., Makarova, K.S., Grohmann, D., 2018. DNA silencing by prokaryotic Argonaute proteins adds a new layer of defense against invading nucleic acids. FEMS Microbiology Reviews 42, 376–387.

Willkomm, S., Oellig, C.A., Zander, A., Restle, T., Keegan, R., Grohmann, D., Schneider, S., 2017. Structural and mechanistic insights into an archaeal DNA-guided Argonaute protein. Nat Microbiol 2, 1–7.

Yigit, E., Batista, P.J., Bei, Y., Pang, K.M., Chen, C.-C.G., Tolia, N.H., Joshua-Tor, L., Mitani, S., Simard, M.J., Mello, C.C., 2006. Analysis of the C. elegans Argonaute Family Reveals that Distinct Argonautes Act Sequentially during RNAi. Cell 127, 747–757.

Yuan, Y.-R., Pei, Y., Ma, J.-B., Kuryavyi, V., Zhadina, M., Meister, G., Chen, H.-Y., Dauter, Z., Tuschl, T., Patel, D.J., 2005. Crystal Structure of A. aeolicus Argonaute, a Site-Specific DNA-Guided Endoribonuclease, Provides Insights into RISC-Mediated mRNA Cleavage. Molecular Cell 19, 405–419.

Zander, A., Holzmeister, P., Klose, D., Tinnefeld, P., Grohmann, D., 2014. Single-molecule FRET supports the two-state model of Argonaute action. RNA Biology 11, 45–56.

Zander, A., Willkomm, S., Ofer, S., van Wolferen, M., Egert, L., Buchmeier, S., Stöckl, S., Tinnefeld, P., Schneider, S., Klingl, A., Albers, S.-V., Werner, F., Grohmann, D., 2017. Guide-independent DNA cleavage by archaeal Argonaute from Methanocaldococcus jannaschii. Nat Microbiol 2, 1–10.

C h a p t e r   2

DEVELOPMENT OF RECOMBINATION SYSTEM IN *ESCHERICHIA COLI*

It has been postulated that pAgo may have potential applications in genome manipulation ever since the discovery of its DNA nuclease activity (Swarts et al., 2014; Hegge et al., 2018). However, to the best of our knowledge, mutations induced by guide-directed cleavage of pAgos in the host genome have never been firmly established. In 2016, NgAgo from *Natronobacterium gregoryi* was reported to utilize gDNAs and actively cleave target DNA both in vitro and in human cell lines, which appears to be irreproducible and the original paper got eventually retracted (Cyranoski, 2017). Later, NgAgo has been shown to stimulate homologous recombination in *E. coli* (Fu et al., 2019; Lee et al., 2021), but the observations in these reports suggest that the DNA cleavage activity of NgAgo is not strictly required in the stimulation of homologous recombination and therefore the DSB generation is unlikely the prerequisite for the NgAgo-induced chromosome recombination.

2.1 Creation of the recombination system

Previous observation that guide-directed CbAgo cleavage at the *E. coli* chromosome efficiently triggers RecBCD activity (Kuzmenko et al., 2020) inspired us to hypothesize that RecBCD-dependent chromosome recombination should be triggered by guide-directed CbAgo cleavage as well. It has been shown that DSBs introduced by SbcCD cleavage at a 246-bp chromosomal palindrome (pal246) stimulate RecBCD-dependent recombination between two downstream direct repeat sequences (Eykelenboom et al., 2008). To determine whether similar recombination can be induced by guide-directed CbAgo cleavage (Figure 2.1a), we integrated a recombination cassette at cynX loci on *E. coli* strain DL1777, which is ~6 kb away from the target lacZ locus (Figure 2.1b). This recombination cassette contains an EM7 promoter and a kanamycin resistance gene whose function is abolished by the insertion of a stop codon array and a 270-bp direct repeat sequence. A recombination event between the two direct repeat sequences removes the insertion,

restores the gene function, and confers kanamycin resistance to the host. Considering the introduction of the Chi site, an 8-base 5'-GCTGGTGG-3' motif recognized by RecBCD (Dillingham et al., 2008; Sinha et al., 2020) has been shown to stimulate RecBCD recombination (Eykelenboom et al., 2008), and we also incorporated a varied number of Chi sites into the genome, with their 5' ends oriented toward the recombination cassette (Figure 2.1b).

To target the lacZ locus, we created a targeting plasmid pTet_CbAgo/GS encoding a CbAgo expression cassette under the control of a tetracycline-inducible promoter (pTet), and a 1000-bp guide sequence (GS) homologous to the lacZ gene (Figure 2.1c). Importantly, GS is the only sequence on the plasmid (except for an 80-bp rrnB T1 terminator sequence) that is homologous to the genome, assuring only the lacZ locus will be effectively targeted. For controls, plasmids with no GS, no CbAgo gene, or neither, were also created. To determine the dependence of CbAgo cleavage activity, we created plasmids encoding a CbAgo mutant (dCbAgo: CbAgo D541A D611A) that contains mutations of two catalytic residues in its active site which were previously shown to abolish its endonuclease activity in vitro (Hegge et al., 2019; Kuzmenko et al., 2019) and DSB generation activity in vivo (Kuzmenko et al., 2020).

2.2 Effects of the introduction of Chi sites

We then combined the obtained plasmids and strains, induced CbAgo expression, and recovered induced cells to measure recombination frequencies, which was calculated as the fraction of ampicillin-resistant cells that became resistant to kanamycin (kanamycin-resistant and ampicillin-resistant colony-forming units (cfu)/ampicillin-resistant cfu), because only the recombinants have a restored functional kanamycin resistance gene. When there are three or six Chi sites adjacent to the recombination cassette (corresponding strain 3×ChikanS and 6×ChikanS), recombination frequencies by CbAgo/GS were significantly higher than the rest control groups (Figure 2.2a), suggesting a recombination pathway that is mediated by guide-directed CbAgo cleavage.

Interestingly, we also observed remarkable recombination frequencies in dCbAgo/GS groups in some conditions (Figure 2.2a). When there is no Chi site adjacent to the recombination cassette (corresponding strain nonChikanS), the cell bearing dCbAgo/GS had a recombination frequency being ~10-fold higher than the one bearing CbAgo/GS. These results suggest there is a recombination pathway that is mediated by the non-cleavage function of CbAgo and is outperformed by the cleavage-dependent pathway in the presence of the CbAgo active site. Although the exact mechanism remains unknown, this dCbAgo-mediated recombination pathway should be independent of DSB and RecBCD because dCbAgo/GS was previously shown unable to generate DSB or trigger RecBCD activity in vivo (Kuzmenko et al., 2020).

2.2 Effects of guide sequences with different lengths

We also examined the effects of GS length on the recombination frequency (Figure 2.2b) and found that in the range of 50 to 500 bp, recombination frequency increases as the length of GS increases. We decided to use 1000 bp as the length of GS and 3×ChikanS as the model strain to perform fluctuation analysis to estimate recombination rates (Figure 2.3). The cell bearing CbAgo/GS had a recombination rate that is fivefold higher than the ones with dCbAgo, with and without GS, 14-fold higher than the one with CbAgo-only, and 30-fold higher over the rest control groups. The actual contribution of dCbAgo-mediated recombination to the total recombination events in 3×ChikanS/CbAgo/GS should be much smaller than one-fifth because it should be largely outperformed by the cleavage-dependent pathway as previous observation suggests. Together, these findings reveal novel chromosome recombination that is induced by guide-directed CbAgo cleavage. Its Chi site dependence implies DSB formation and RecBCD processing during recombination.

2.3 Experimental methods

Culture conditions. *E. coli*, cultured in Luria-Bertani (LB) medium and agar, was incubated at 37 °C or 30 °C. When appropriate, antibiotics were added to the medium at the following final concentrations: ampicillin, 100 μg/ml; chloramphenicol, 20 μg/ml; kanamycin, 35

µg/ml. Bacterial cell growth was monitored periodically by measuring the optical density of culture aliquots at 600 nm.

Determination of recombination frequency. Cells were transformed with appropriate plasmids and plated on LB plates supplemented with ampicillin. The next day, 5 ml of LB medium supplemented with ampicillin was inoculated with single colony and aerated at 37 °C until OD600 = 0.3–0.4. The temperature was then adjusted to 18 °C and after 30 min protein expression was induced by adding anhydrotetracycline to 200 ng/ml for 16 h. Cultures were then cooled down on ice for 10 min, washed with ice-cold PBS (pH 7.2), resuspended in 5 ml of LB medium supplemented with ampicillin, and recovered at 37 °C for 5 h. Serial dilutions of cells were plated on the LB plates supplemented with appropriate antibiotics to determine cfu.

Fluctuation analysis. Cells were transformed with appropriate plasmids and plated on LB plates supplemented with ampicillin. The next day, 1 ml of LB medium supplemented with ampicillin and 200 ng/ml anhydrotetracycline was inoculated by single colony and aerated at 37 °C for 12 h before making serial dilutions of cultures and plating on the LB plates supplemented with appropriate antibiotics to determine cfu. The Ma-Sandri-Sarkar Maximum Likelihood Estimator (MSS-MLE) Method or the Lea-Coulson Method of the Median in the Fluctuation AnaLysis CalculatOR (FALCOR)[25] were used to calculate recombination rates and 95% confidence intervals. The online FALCOR tool is available at https://lianglab.brocku.ca/FALCOR/.

Plasmid construction. pTet_wtCas9 was a gift from Dr. Stanley Qi (Addgene plasmid #44250). pET28b_CbAgo and pET28b_dCbAgo were gifts from Dr. Alexei Aravin. pDL1999 was a gift from Dr. David Leach. Plasmids pnonChikanS, pChikanS, p3×ChikanS, and p6×ChikanS were created to serve as templates for PCR amplification to generate dsDNA donors for Lambda-Red recombineering. Plasmid pnonChikanS was constructed as follows. Plasmid backbone consisting of two 400-bp homology arms to the cynX gene and an EM7 promoter was amplified from pDL1999 using primers kan-bb.F/kan-bb.R. Two copies of kan gene were amplified from plasmid pET28b using

primers kan1.F/kan1.R and kan2.F/kan2.R, respectively. Each PCR product was gel-purified and all three fragments were fused by Gibson Assembly. This way, the recombination cassette containing two non-functional copies of the kanamycin resistance gene separated by a stop codon array was created and inserted adjacent to the EM7 promoter, between the two cynX homology arms. Plasmid pChikanS was obtained through amplification of pnonChikanS with primers Chi.F/Chi.R. PCR product was gel-purified and self-ligated by Gibson assembly. Plasmid p3×ChikanS was obtained through amplification of pChikanS with primers 3×Chi.F/3×Chi.R. PCR product was gel-purified and self-ligated by Gibson assembly. Plasmid p6×ChikanS was obtained through amplification of p3×ChikanS with primers 6×Chi.F/6×Chi.R. PCR product was gel-purified and self-ligated by Gibson assembly. To construct plasmids pTet_CbAgo and pTet_dCbAgo, plasmid backbone was amplified from pTet_wtCas9 using primers pTet-bb.F/pTet-bb.R. CbAgo and dCbAgo sequences were amplified from pET28b_CbAgo and pET28b_dCbAgo respectively using primers Cb.F/Cb.R, gel-purified, and ligated with the plasmid backbone individually via Gibson Assembly. To construct plasmids pTet_CbAgo/GS and pTet_dCbAgo/GS, a 1000-bp sequence homologous to lacZ gene was amplified from DL1777 genomic DNA via colony PCR using primers GS-lacZ.F/GS-lacZ.R. Plasmid backbones of pTet_CbAgo and pTet_dCbAgo were also PCR amplified with primers GS-bb.F/GS-bb.R, gel-purified, and ligated with the 1000-bp lacZ homologous sequence individually via Gibson Assembly. To construct plasmids pTet_CpAgo/GS, pTet_CaAgo/GS, pTet_CdAgo/GS, pTet_IbAgo/GS, plasmid backbones were amplified from pTet_CbAgo/GS using primers pTet-bb.F/pTet-bb.R. To construct plasmids pEmpty and pGS, pTet_CbAgo and pTet_CbAgo/GS were amplified using primers empty.F/empty.R respectively to remove the CbAgo gene from the plasmid. The resulting PCR products were gel-purified and self-ligated via Gibson Assembly. Plasmid pCas9-Red was purchased from Sigma-Aldrich (Catalog Number CAS9BAC1P), which contains the gene for Cas9 from *Streptococcus pyogenes* (spCas9) expressed from its native promoter, and the genes for Lambda-Red recombinases Exo, Beta, and Gam under control of the arabinose-inducible ParaB promoter. To construct plasmid pgRNA_cynX, plasmid pgRNA_lacZ was used as the template, which encodes a CRISPR-

gRNA targeting lacZ under control of the J23119 promoter and a sacB gene from *Bacillus subtilis* for counter-selection. Plasmid backbone was amplified from pgRNA_lacZ using primers gRNA-cynX.F/gRNA-cynX.R, gel-purified, and self-ligated by Gibson Assembly, resulting in the replacement of the original spacer sequence with the synthetic spacer sequence targeting cynX.

Strain construction. Strain DL1777 were gifts from Dr. David Leach. To generate strain nonChikanS, ChikanS, 3×ChikanS, 6×ChikanS, plasmid pnonChikanS, pChikanS, p3×ChikanS, and p6×ChikanS were amplified with primers cynX-arm.F/cynX-arm.R respectively, and PCR products were gel-purified as dsDNA donors. DL1777 was transformed with pCas9-Red, plated on LB plates supplemented with kanamycin, and incubated overnight at 30 °C. One of the transformants was inoculated in 5 ml LB media supplemented with kanamycin and grown at 30 °C until OD600 = 0.3–0.4. The Lambda-Red genes were then induced with 15 mM L-arabinose for 45 min. The culture was used to prepare electrocompetent cells by washing twice with 10% glycerol and resuspending in 50 μl 10% glycerol. 5 μl mixture of ~250 ng dsDNA and 100-ng plasmid pgRNA_lacZ were added to the cells, which were then subject to electroporation and allowed to recover in 1 ml LB for 2 h at 30 °C. The recovered cells were plated on LB plates supplemented with ampicillin and kanamycin and incubated overnight at 30 °C. Colonies with correct genomic integration were verified by colony PCR and sequencing. Upon obtaining positive hits, plasmids were cured by growing the cells overnight in LB media without antibiotics at 37 °C and plating on LB plates supplemented with 5% sucrose. Successful plasmid curing was verified by colony PCR.

Statistical analyses. GraphPad Prism 9 was used to evaluate statistical significance. Student's t-test (two-tailed) was used for the statistical analysis of the experiments. *P* values < 0.05 were considered significant.
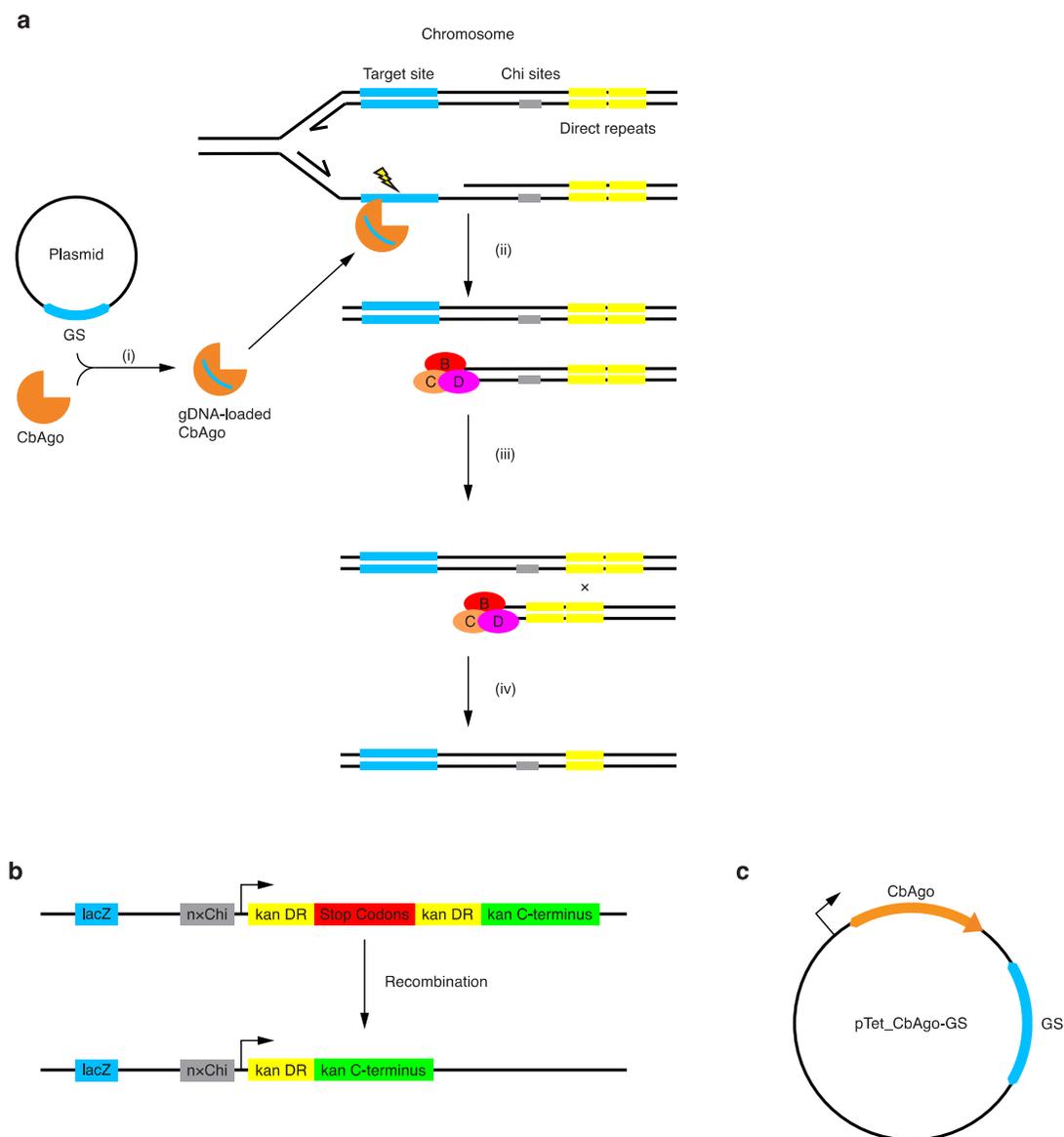
Figure 2.1: Guide-directed CbAgo cleavage stimulates chromosome recombination. a, Proposed mechanism for guide-directed CbAgo cleavage and RecBCD-dependent recombination. (i) CbAgo acquires gDNA from plasmid-encoded GS. See Extended Data Fig. 1 for the proposed model of gDNA biogenesis. (ii) Guide-directed CbAgo cleavage of lagging strand during chromosome replication generates DSB and (iii) triggers RecBCD binding and processing of chromosome DNA. (iv) Recognition of Chi sites by RecBCD attenuates its DNA degradation activity and facilitates recombination between sister chromosomes via RecA. **b**, Structure of the recombination cassette. kan, kanamycin resistance gene. DR, direct repeat. **c**, Structure of the CbAgo targeting plasmid.
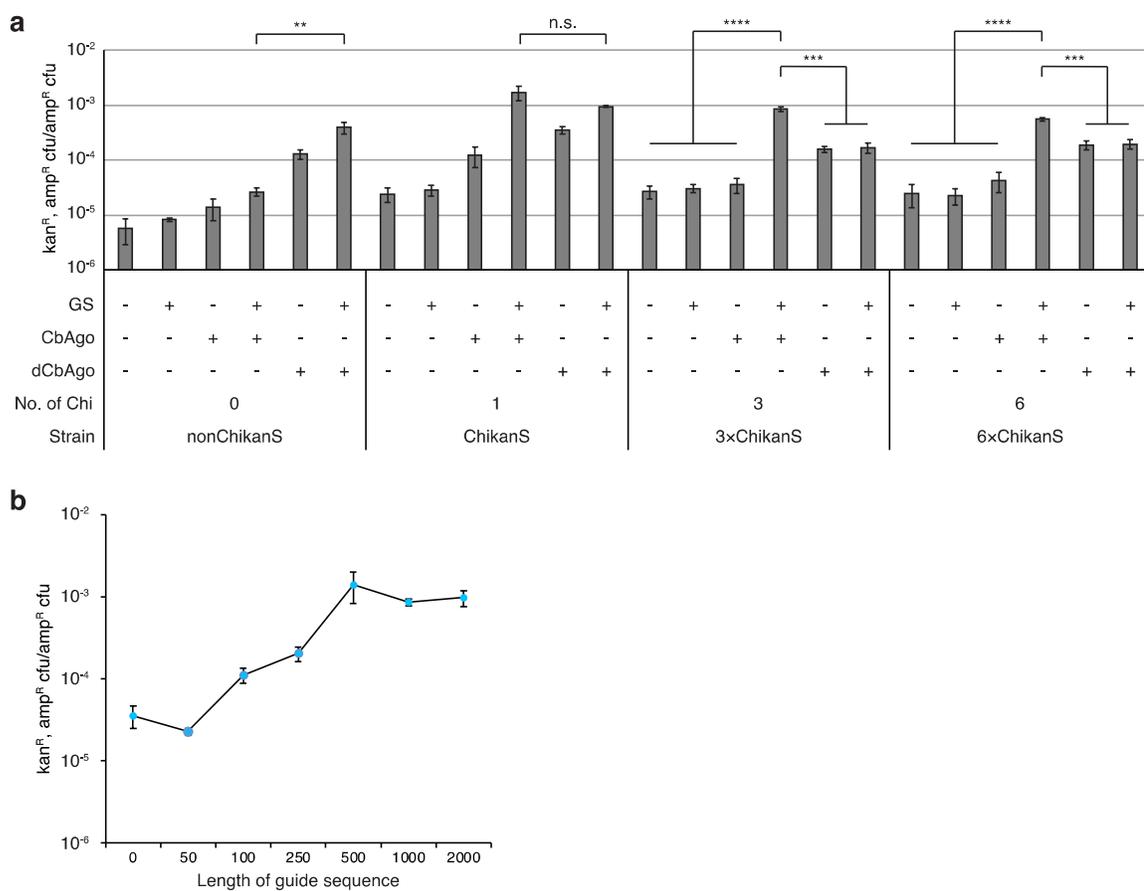
Figure 2.2: Effects of the number of Chi sites and GS length on the recombination frequency. a, Recombination frequencies in different genetic contexts. Error bars, mean ± s.d. from three independent cultures. **b**, Recombination frequencies of strain 3×ChikanS expressing CbAgo in the presence of GS with varying lengths. Error bars, mean ± s.d. from three independent cultures. *P* values were calculated by two-tailed unpaired Student's t-test; n.s. *P* > 0.05, *P < 0.05, **P < 0.01, ***P < 0.001, ****P < 0.0001.
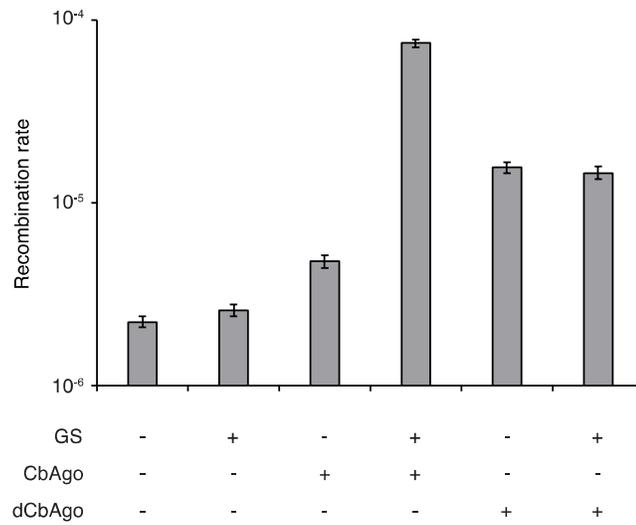
Figure 2.3: Guide-directed CbAgo cleavage stimulates chromosome recombination. Recombination rates of strain 3×ChikanS in different genetic contexts were determined by fluctuation analysis from eight independent cultures. Error bars represent 95% confidence intervals.

BIBLIOGRAPHY

Bierne, H., Ehrlich, S.D., Michel, B., 1997. Deletions at stalled replication forks occur by two different pathways. The EMBO Journal 16, 3332–3340.

Capaldo, F.N., Barbour, S.D., 1975. DNA content, synthesis and integrity in dividing and non-dividing cells of rec− strains of Escherichia coli K12. Journal of Molecular Biology 91, 53–66.

Citorik, R.J., Mimee, M., Lu, T.K., 2014. Sequence-specific antimicrobials using efficiently delivered RNA-guided nucleases. Nat Biotechnol 32, 1141–1145.

Cyranoski, D., 2017. Authors retract controversial NgAgo gene-editing study. Nature.

Dillingham, M.S., Kowalczykowski, S.C., 2008. RecBCD Enzyme and the Repair of Double-Stranded DNA Breaks. Microbiology and Molecular Biology Reviews 72, 642–671.

Eykelenboom, J.K., Blackwood, J.K., Okely, E., Leach, D.R.F., 2008. SbcCD Causes a Double-Strand Break at a DNA Palindrome in the Escherichia coli Chromosome. Molecular Cell 29, 644–651.

Fu, L., Xie, C., Jin, Z., Tu, Z., Han, L., Jin, M., Xiang, Y., Zhang, A., 2019. The prokaryotic Argonaute proteins enhance homology sequence-directed recombination in bacteria. Nucleic Acids Research 47, 3568–3579.

Hall, B.M., Ma, C.-X., Liang, P., Singh, K.K., 2009. Fluctuation AnaLysis CalculatOR: a web tool for the determination of mutation rate using Luria–Delbrück fluctuation analysis. Bioinformatics 25, 1564–1565.

Hegge, J.W., Swarts, D.C., van der Oost, J., 2018. Prokaryotic Argonaute proteins: novel genome-editing tools? Nat Rev Microbiol 16, 5–11.

Kuzmenko, A., Oguienko, A., Esyunina, D., Yudin, D., Petrova, M., Kudinova, A., Maslova, O., Ninova, M., Ryazansky, S., Leach, D., Aravin, A.A., Kulbachinskiy, A., 2020. DNA targeting and interference by a bacterial Argonaute nuclease. Nature 587, 632–637.

Kuzmenko, A., Yudin, D., Ryazansky, S., Kulbachinskiy, A., Aravin, A.A., 2019. Programmable DNA cleavage by Ago nucleases from mesophilic bacteria Clostridium butyricum and Limnothrix rosea. Nucleic Acids Research 47, 5822–5836.

Kuzminov, A., Stahl, F.W., 1997. Stability of linear DNA in recA mutant Escherichia coli cells reflects ongoing chromosomal DNA degradation. Journal of Bacteriology 179, 880–888.

Lee, K.Z., Mechikoff, M.A., Kikla, A., Liu, A., Pandolfi, P., Fitzgerald, K., Gimble, F.S., Solomon, K.V., 2021. NgAgo possesses guided DNA nicking activity. Nucleic Acids Research 49, 9926–9937.

Michel, B., 2000. Replication fork arrest and DNA recombination. Trends in Biochemical Sciences 25, 173–178.

Pennington, J.M., Rosenberg, S.M., 2007. Spontaneous DNA breakage in single living Escherichia coli cells. Nat Genet 39, 797–802.

Sinha, A.K., Possoz, C., Leach, D.R.F., 2020. The Roles of Bacterial DNA Double-Strand Break Repair Proteins in Chromosomal DNA Replication. FEMS Microbiology Reviews 44, 351–368.

Skarstad, K., Boye, E., 1993. Degradation of individual chromosomes in recA mutants of Escherichia coli. Journal of Bacteriology 175, 5505–5509.

Smith, G.R., 2012. How RecBCD Enzyme and Chi Promote DNA Break Repair and Recombination: a Molecular Biologist's View. Microbiology and Molecular Biology Reviews 76, 217–228.

Swarts, D.C., Jore, M.M., Westra, E.R., Zhu, Y., Janssen, J.H., Snijders, A.P., Wang, Y., Patel, D.J., Berenguer, J., Brouns, S.J.J., van der Oost, J., 2014. DNA-guided DNA interference by a prokaryotic Argonaute. Nature 507, 258–261.

Taylor, A.F., Smith, G.R., 1985. Substrate specificity of the DNA unwinding activity of the RecBC enzyme of Escherichia coli. Journal of Molecular Biology 185, 431–443.

Vaiskunaite, R., Vainauskas, J., Morris, J.J.L., Potapov, V., Bitinaite, J., 2022. Programmable cleavage of linear double-stranded DNA by combined action of

Argonaute CbAgo from Clostridium butyricum and nuclease deficient RecBC helicase from E. coli. Nucleic Acids Research 50, 4616–4629.

Wigley, D.B., 2013. Bacterial DNA repair: recent insights into the mechanism of RecBCD, AddAB and AdnAB. Nat Rev Microbiol 11, 9–13.

Chapter 3

CHARACTERIZATION OF AGO-MEDIATED RECOMBINATION

3.1 Reliability of recombination system

To demonstrate the reliability of our recombination system, we tested four additional pAgos in strain 3×ChikanS, including CaAgo, CdAgo, CpAgo, and IbAgo. In the presence of GS, the rates of recombination induced by different pAgos correlate with the rank order of their reported in vitro ssDNA cleavage activity (Figure 3.1, see Vaiskunaite et al., 2022). We also tested an engineered DSB in our system by integrating a pal246 into the lacZ loci on the genome of strain 3×ChikanS and its *sbcCD* knockout mutant and measuring their recombination rates (Figure 3.2). The sbcCD$^+$, lacZ::pal246 strain yielded a ~100-fold increase in recombination rate compared to the sbcCD$^+$, lacZ$^+$ strain, and the *sbcCD*, lacZ::pal246 strain. The bigger fold change stimulated by SbcCD-pal246 over CbAgo/GS is consistent with the previous observation that SbcCD-pal246 is more efficient in DSB generation than CbAgo/GS in vivo (Kuzmenko et al., 2020). These findings indicate a strong correlation between pAgo DNA cleavage activity, DSB generation efficiency, and recombination rate in our system.

3.2 Involvement of DSB

To gain more insight into our recombination system, we first sought to provide direct evidence that CbAgo can be directed to attack *E. coli* chromosomes and cause DNA damage (Figure 2.1, steps i and ii). Since DNA damage induces the cellular SOS response, we used an *E. coli* strain carrying a chromosomally located gfp gene controlled by an SOS-inducible *sulA* promoter (Pennington et al., 2007) and performed flow cytometry to quantify single-cell fluorescence level (Figure 3.3a). We observed a sixfold increase of fluorescence in cells expressing CbAgo and a 15-fold increase in cells containing CbAgo/GS, while the cells containing dCbAgo/GS exhibited no difference in cellular fluorescence compared to the cells bearing empty plasmids (Figure 3.3b). The fluorescence increase in cells expressing CbAgo without GS can be explained by previous observations

that CbAgo actively degrades plasmids (Kuzmenko et al., 2020), and plasmid degradation triggers SOS-response (Citorik et al., 2014). Importantly, the significant increase of fluorescence in cells expressing CbAgo in the presence of GS confirms that CbAgo can be guided to attack chromosomes, while dCbAgo/GS cannot.

3.3 Dependence of RecBCD and RecA

Then we sought to determine the involvement of *E. coli* endogenous DNA repair machinery RecA and RecBCD in the recombination by creating and testing *recA* and *recBCD* knockout mutants of strain 3×ChikanS. Since the viabilities of different knockout strains greatly varied after induction (Figure 3.4), we determined the fluctuation analysis is no longer suitable and decided to directly analyze recombination frequencies. In *recBCD* strain, recombination frequency by CbAgo/GS had no difference with the CbAgo-only control, representing a profound change from the result using recBCD$^+$ strain (Figure 3.5), indicating CbAgo/GS-induced recombination depends on RecBCD (Figure 2.1, step iii). Since RecBCD works closely with DSB (Taylor et al., 1985), this observation also suggests DSB generation in CbAgo/GS-induced recombination. An interesting discovery was that the *recBCD* strain bearing dCbAgo/GS showed ~1000-fold decreased ampicillin-resistant cfu and only ~20-fold decreased kanamycin-resistant and ampicillin-resistant cfu compared to its RecBCD$^+$ counterpart (Figure 3.4). These changes resulted in increased recombination frequency (Figure 3.5), suggesting dCbAgo-mediated recombination is RecBCD-independent. Moreover, this growth inhibition was reduced by ~60-fold in the presence of the CbAgo active site, suggesting it is again outperformed or inhibited when the CbAgo active site is present.

The CbAgo/GS-induced DSBs can further explain the extremely low viability of *recA* strain bearing CbAgo/GS (Figure 3.4): without the protection of RecA, continuously introduced DSBs trigger extensive DNA degradation by RecBCD, causing enormous loss of chromosomal DNA and subsequent cell death (Capaldo et al., 1975; Skarstad et al., 1993; Kuzminov et al., 1997). On the other hand, these cells had a high recombination frequency close to 0.1 (Figure 3.5), indicating RecA is not essential in the recombination induced by

guide-directed CbAgo cleavage even though it actively repairs DSBs generated during the process. These findings prompt us to modify our model that the final step of recombination (Figure 2.1, step iv) should be independent of RecA, reminiscent of the RecA-independent, replication arrest-induced deletion (Bierne et al., 1997; Michel, 2000).

3.4 CbAgo-assisted lambda Red recombineering

The observation that cfu of *recA* strain was reduced by three orders of magnitude when its genome is targeted by CbAgo (Figure 3.4) is very intriguing, as it supports a strategy to leverage CbAgo/GS targeting as counter-selection to facilitate recombineering (Figure 3.6a). For comparison, a self-targeting CRISPR-Cas9 system was reported to reduce cfu by three orders of magnitude in *E. coli* (Jiang et al., 2013). Co-expressing the CRISPR-Cas9 system to eliminate unedited cells, Lambda-Red recombineering achieved an increase of efficiency by ~10^4 fold and a 65% overall mutation rate. We sought to combine the CbAgo targeting system with Lambda-Red recombineering by introducing CbAgo expression plasmids into strain SIJ488_ΔrecA, which is RecA-deficient and has arabinose inducible Lambda-Red recombineering genes integrated into its genome. We first performed the standard recombineering procedure with dsDNA donor encoding kanamycin resistance cassette flanked by 1000-bp homology arms to replace the genomic lacZ gene. The recombineering efficiency was 1.2×10^{-4}, calculated from the fraction of cells that became kanamycin resistant. Then we performed recombineering in CbAgo-plasmid-containing cells, induced CbAgo expression, and recovered the cells to characterize the proposed counter-selection effect. The cell transformed with pTet_CbAgo/GS had a mutation efficiency of 2.3×10^{-2}, representing a ~100-fold increase in efficiency from standard recombineering (Figure 3.6b). Other control groups did not yield improvement, therefore the increased proportion of the edited cell population depends on guide-directed CbAgo cleavage.

3.5 Experimental methods

Culture conditions. *E. coli*, cultured in Luria-Bertani (LB) medium and agar, was incubated at 37°C or 30°C. When appropriate, antibiotics were added to the medium at the following

final concentrations: ampicillin, 100 µg/ml; chloramphenicol, 20 µg/ml; kanamycin, 35 µg/ml. Bacterial cell growth was monitored periodically by measuring the optical density of culture aliquots at 600 nm.

Determination of cfu and recombination frequency. Cells were transformed with appropriate plasmids and plated on LB plates supplemented with ampicillin. The next day, 5 ml of LB medium supplemented with ampicillin was inoculated with single colony and aerated at 37 °C until OD600 = 0.3–0.4. The temperature was then adjusted to 18 °C and after 30 min protein expression was induced by adding anhydrotetracycline to 200 ng/ml for 16 h. Cultures were then cooled down on ice for 10 min, washed with ice-cold PBS (pH 7.2), resuspended in 5 ml of LB medium supplemented with ampicillin, and recovered at 37 °C for 5 h. Serial dilutions of cells were plated on the LB plates supplemented with appropriate antibiotics to determine cfu.

Fluctuation analysis. Cells were transformed with appropriate plasmids and plated on LB plates supplemented with ampicillin. The next day, 1 ml of LB medium supplemented with ampicillin and 200 ng/ml anhydrotetracycline was inoculated by single colony and aerated at 37 °C for 12 h before making serial dilutions of cultures and plating on the LB plates supplemented with appropriate antibiotics to determine cfu. For plasmid-free strain 3×ChikanS, 3×ChikanS_ pal246_ΔsbcCD, and 3×ChikanS_pal246, 1 ml of LB medium was inoculated by single colony and aerated at 37 °C for 12 h before making serial dilutions of cultures and plating on the LB plates without antibiotic or supplemented with kanamycin to determine cfu. The Ma-Sandri-Sarkar Maximum Likelihood Estimator (MSS-MLE) Method or the Lea-Coulson Method of the Median in the Fluctuation AnaLysis CalculatOR (FALCOR) were used to calculate recombination rates and 95% confidence intervals. The online FALCOR tool is available at https://lianglab.brocku.ca/FALCOR/.

Flow cytometry analysis. Strain SMR6669 cells were transformed with appropriate plasmids and plated on LB plates supplemented with ampicillin. The next day, 1 ml of LB

medium supplemented with ampicillin and 200 ng/ml anhydrotetracycline was inoculated with single colony and aerated at 37 °C for 12 h. Cultures were then washed with ice-cold PBS (pH 7.2), diluted 1:500 into ice-cold PBS (pH 7.2), passed through 40-μm cell strainers. Then 1 ug/ml propidium iodide was added to cell cultures to determine cell viability for 30 minutes and the cell cultures were analyzed on a Beckman Coulter Cytoflex S Flow Cytometer. For each experiment, 10^5 cells per culture and 3 cultures per strain were analyzed. Flow cytometry data were analyzed using FlowJo software version 10.8.1. To comparatively quantify green cells, a green "gate" was set arbitrarily as the window in which ~0.9% of the control strain, SMR6669/pEmpty fall, according to the spontaneous SOS induction level (Pennington et al., 2007).

Lambda-Red recombineering. To calculate mutation efficiency for Lambda-Red recombineering (referred to as standard recombineering procedure), kanamycin-resistance cassette flanked by two 1000-bp homology arms to the lacZ gene was amplified from genomic DNA of SIJ488_ΔlacZ via colony PCR with primers lacZ-arm.1kb.F/lacZ-arm.1kb.R. Resulting PCR product was gel-purified as dsDNA donor. 5 ml of LB medium was inoculated with a single colony of strain SIJ488_ΔrecA and aerated at 37 °C until OD600 = 0.3–0.4. The Lambda-Red genes were then induced with 15 mM L-arabinose for 45 min. The culture was used to prepare electrocompetent cells by washing twice with 10% glycerol and resuspending in 50 μl 10% glycerol. 2 μl mixture of ~300 ng dsDNA was added to the cells, which were then subject to electroporation and allowed to recover in 1 ml LB for 2 h at 37 °C. Serial dilutions of cells were plated on the LB plates with no antibiotic or supplemented with kanamycin to determine cfu. To calculate mutation efficiency for CbAgo-assisted Lambda-Red recombineering, the kanamycin-resistance cassette flanked by two 1000-bp homology arms to the lacZ gene was amplified from the genomic DNA of SIJ488_ΔlacZ via colony PCR with primers lacZ-arm.1kb.F/lacZ-arm.1kb.R. Resulting PCR product was gel-purified as dsDNA donor. Cells of strain SIJ488_ΔrecA were transformed with appropriate plasmids and plated on LB plates supplemented with ampicillin. The next day, 5 ml of LB medium supplemented with ampicillin was inoculated with single colony and aerated at 37 °C until OD600 = 0.3–0.4.

The Lambda-Red genes were then induced with 15 mM L-arabinose for 45 min. The culture was used to prepare electrocompetent cells by washing twice with 10% glycerol and resuspending in 50 µl 10% glycerol. 2 µl mixture of ~300 ng dsDNA was added to the cells, which were then subject to electroporation and allowed to recover in 1 ml LB supplemented with ampicillin for 2 h at 37 °C. The recovered cells were diluted into 5 ml LB supplemented with ampicillin and 0.2% glucose and continued growing for 2 hours at 37 °C. The temperature was then adjusted to 18 °C and after 30 min, protein expression was induced by adding anhydrotetracycline to 200 ng/ml for 16 h. Upon induction, cultures were cooled down on ice for 10 min, washed with ice-cold PBS (pH 7.2), resuspended in 5 ml of LB medium supplemented with ampicillin, and recovered at 37 °C for 5 h. Serial dilutions of cells were plated on the LB plates supplemented with appropriate antibiotics to determine cfu.

Plasmid construction. To construct plasmids pTet_CpAgo/GS, pTet_CaAgo/GS, pTet_CdAgo/GS, pTet_IbAgo/GS, plasmid backbones were amplified from pTet_CbAgo/GS using primers pTet-bb.F/pTet-bb.R. CpAgo, CaAgo, CdAgo, IbAgo gene fragments were ordered from IDT (Integrated DNA Technologies) or Twist Bioscience, amplified using primers pTet-Ago.F/pTet-Ago.R individually, gel-purified and ligated with the plasmid backbones individually via Gibson Assembly.

Strain construction. Strain DL2859 and DL2874 were gifts from Dr. David Leach. SMR6669 was a gift from Dr. Susan Rosenberg. SIJ488 was a gift from Dr. Alex Nielsen (Addgene bacterial strain #68246). Strain 3×ChikanR was generated by previously described CbAgo/GS mediated recombination using strain 3×ChikanS. Kanamycin-resistant colonies with correct recombination were verified by colony PCR and sequencing. Plasmids were then cured by growing the cells overnight in LB media without antibiotics at 37 °C and plating on LB plates. Successful plasmid curing was confirmed by the cell sensitivity to ampicillin. To generate strain 3×ChikanS_pal246 and 3×ChikanS_pal246_ΔsbcCD, genomic DNA of 3×ChikanS was amplified via colony PCR with primers cynX-arm.F/gmR-arm.R and PCR product was gel-purified as dsDNA donor.

DL2859 and DL2874 were individually transformed with pCas9-Red, plated on LB plates supplemented with kanamycin, and incubated overnight at 30 °C. One of the transformants was inoculated in 5 ml LB media supplemented with kanamycin and grown at 30 °C until OD600 = 0.3–0.4. The Lambda-Red genes were then induced with 15 mM L-arabinose for 45 min. The culture was used to prepare electrocompetent cells by washing twice with 10% glycerol and resuspending in 50 µl 10% glycerol. 5 µl mixture of ~250 ng dsDNA and 100 ng plasmid pgRNA_lacZ was added to the cells, which were then subject to electroporation and allowed to recover in 1 ml LB for 2 h at 30 °C. The recovered cells were plated on LB plates supplemented with ampicillin and kanamycin and incubated overnight at 30 °C. Colonies with correct genomic integration were verified by colony PCR and sequencing. Upon obtaining positive hits, plasmids were cured by growing the cells overnight in LB media without antibiotics at 37 °C and plating on LB plates supplemented with 5% sucrose. Successful plasmid curing was verified by colony PCR. To generate strain 3×ChikanS_ΔrecBCD and 3×ChikanS_ΔrecA, chloramphenicol-resistance cassettes were amplified from plasmid pDL1999 using primers recA.cmR.F/recA.cmR.R and recBCD.cmR.F/recBCD.cmR.R respectively. PCR products were gel-purified and used as dsDNA donors. 3×ChikanS was transformed with pKD46 plasmid, plated on LB plates supplemented with ampicillin, and incubated overnight at 30 °C. A 5 ml culture inoculated from a single colony was grown at 30°C until OD600 = 0.3–0.4. The Lambda-Red genes were then induced with 15 mM L-arabinose for 45 min. The culture was used to prepare electrocompetent cells by washing twice with 10% glycerol and resuspending in 50 µl 10% glycerol. These cells were transformed with ~250 ng dsDNA by electroporation and allowed to recover in 1 ml LB for 2 h at 30 °C. The recovered cells were plated on LB plates supplemented with chloramphenicol and incubated overnight at 37 °C. The correct knockouts were verified by colony PCR and sequencing. Upon obtaining positive hits, the pKD46 plasmid was cured by growing the cell in LB media at 37°C overnight and plating on LB plates. Successful plasmid curing was confirmed by the cell sensitivity to ampicillin. To generate strain SIJ488_ΔrecA, the chloramphenicol-resistance cassette was amplified from plasmid pDL1999 using primers recA.cmR.F/recA.cmR.R. PCR products were gel-purified and used as dsDNA donor. A 5 ml culture inoculated from single colony of SIJ488

was grown at 37°C until OD600 = 0.3–0.4. The Lambda-Red genes were then induced with 15 mM L-arabinose for 45 min. The culture was used to prepare electrocompetent cells by washing twice with 10% glycerol and resuspending in 50 µl 10% glycerol. These cells were transformed with ~250 ng dsDNA by electroporation and allowed to recover in 1 ml LB for 2 h at 37 °C. The recovered cells were plated on LB plates supplemented with chloramphenicol and incubated overnight at 37 °C. Correct knockouts were verified by colony PCR and sequencing. To generate strain SIJ488_ΔlacZ, the kanamycin-resistance cassette was amplified from genomic DNA of 3×ChikanR via colony PCR with primers lacZ.kanR.F/lacZ.kanR.R, and the PCR product was gel-purified as dsDNA donor. A 5 ml culture inoculated from single colony of SIJ488 was grown at 37°C until OD600 = 0.3–0.4. The Lambda-Red genes were then induced with 15 mM L-arabinose for 45 min. The culture was used to prepare electrocompetent cells by washing twice with 10% glycerol and resuspending in 50 µl 10% glycerol. These cells were transformed with ~250 ng dsDNA by electroporation and allowed to recover in 1 ml LB for 2 h at 37 °C. The recovered cells were plated on LB plates supplemented with kanamycin and incubated overnight at 37 °C. Correct knockouts were verified by colony PCR and sequencing.

Statistical analyses. GraphPad Prism 9 was used to evaluate statistical significance. Student's t-test (two-tailed) was used for the statistical analysis of the experiments. *P* values < 0.05 were considered significant.
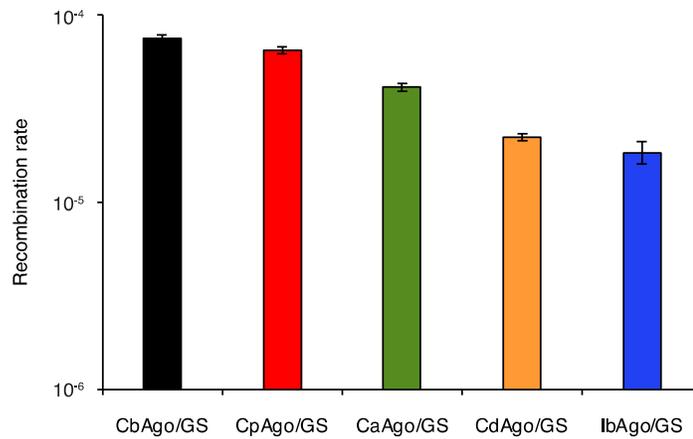
Figure 3.1: Recombination rate by pAgo/GS correlates with pAgo in vitro **DNA cleavage activity.** Recombination rates using different pAgos in the presence of GS in strain 3×ChikanS. See Ref. (Vaiskunaite et al., 2022) for comparisons of in vitro DNA cleavage activity among different pAgos.
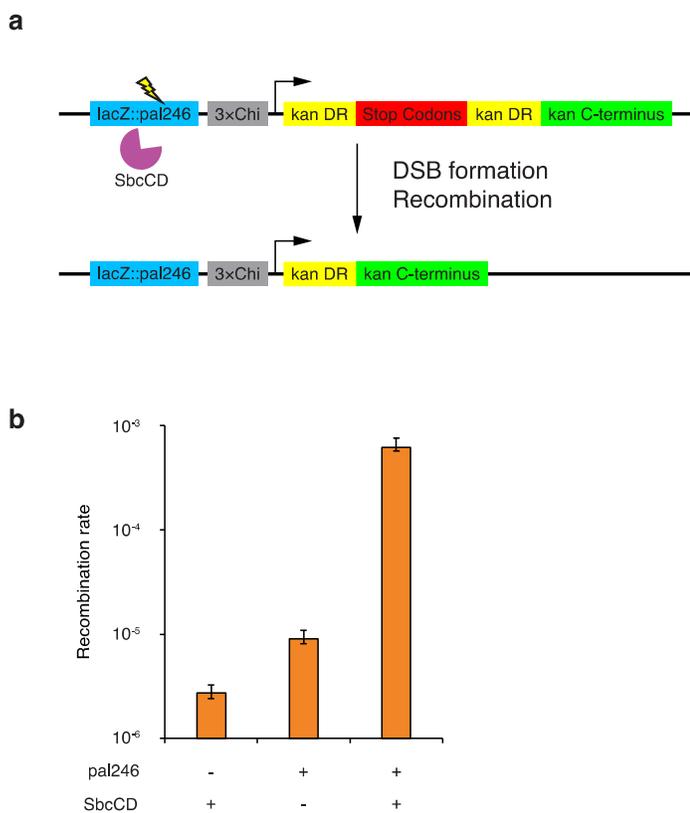
Figure 3.2: Recombination rate by pAgo/GS correlates with in vivo **DSB generation efficiency. a**, Genetic structure of engineered DSB. A pal246 was inserted into lacZ loci, whose cleavage by *E. coli* endogenous nuclease SbcD forms DSB and stimulates recombination. **b**, Recombination rates using sbcCD$^+$/ΔsbcCD, lacZ$^+$/lacZ::pal246 strains. Strains used here were 3×ChikanS, 3×ChikanS_pal246_ΔsbcCD and 3×ChikanS_pal246. See Ref. (Kuzmenko et al., 2020) for the comparison of in vivo DSB generation efficiency between SbcCD/pal246 and CbAgo/GS. Recombination rates were determined by fluctuation analysis from eight independent cultures. Error bars represent 95% confidence intervals.
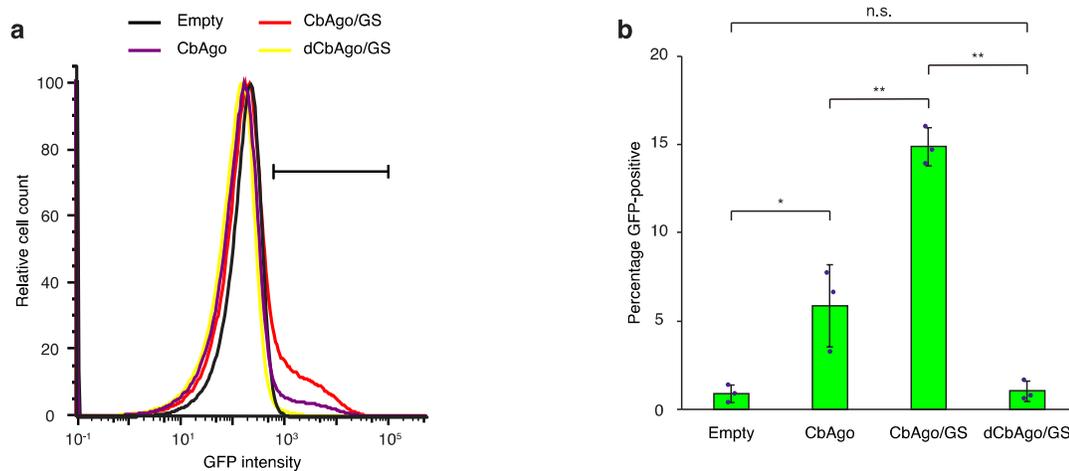
Figure 3.3: CbAgo/GS generates DSB generation on chromosomes. a, CbAgo/GS induces cellular SOS response. Flow cytometry histograms normalized to the mode of the population, combining data from three independent cultures. Horizontal bar represents the GFP+ gate. **b**, Quantification of GFP-positive cells within the GFP+ gate. Error bars, mean ± s.d. from three independent cultures. *P* values were calculated by two-tailed unpaired Student's t-test; n.s. *P* > 0.05, *P < 0.05, **P < 0.01.
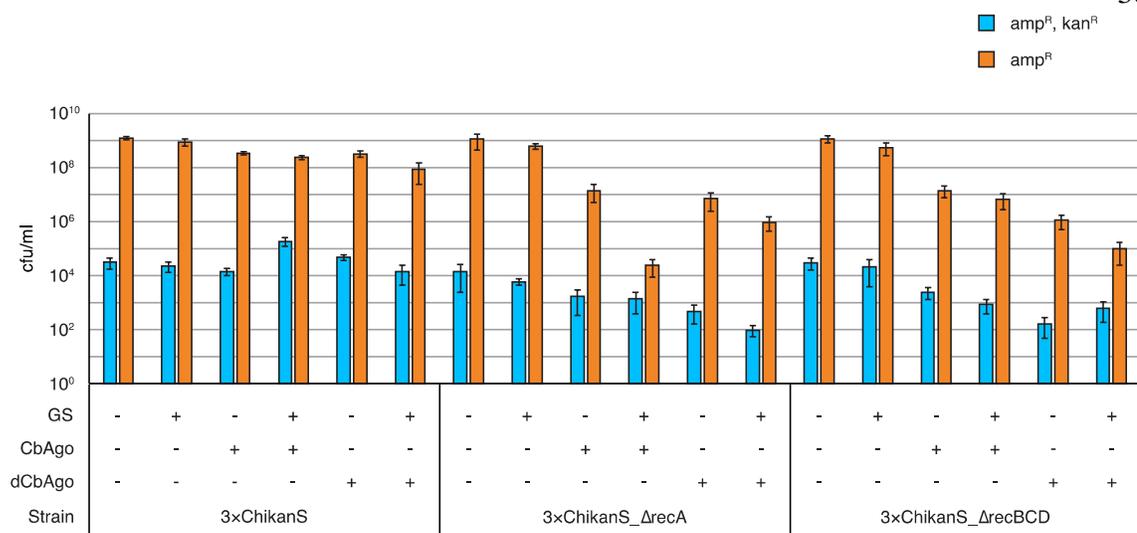
Figure 3.4: Viabilities of cells in different genetic contexts. Error bars, mean ± s.d. from eight independent cultures.
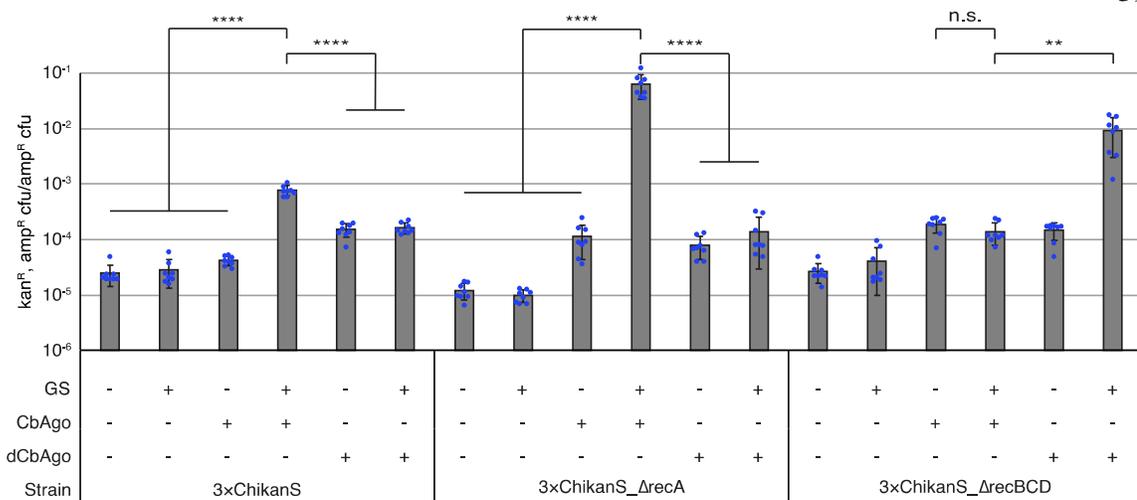
Figure 3.5: Recombination by CbAgo/GS depends on RecBCD processing but is independent of RecA. Recombination frequencies in different genetic contexts. Error bars, mean ± s.d. from eight independent cultures. *P* values were calculated by two-tailed unpaired Student's t-test; n.s. *P* > 0.05, *P < 0.05, **P < 0.01, ***P < 0.001, ****P < 0.0001.
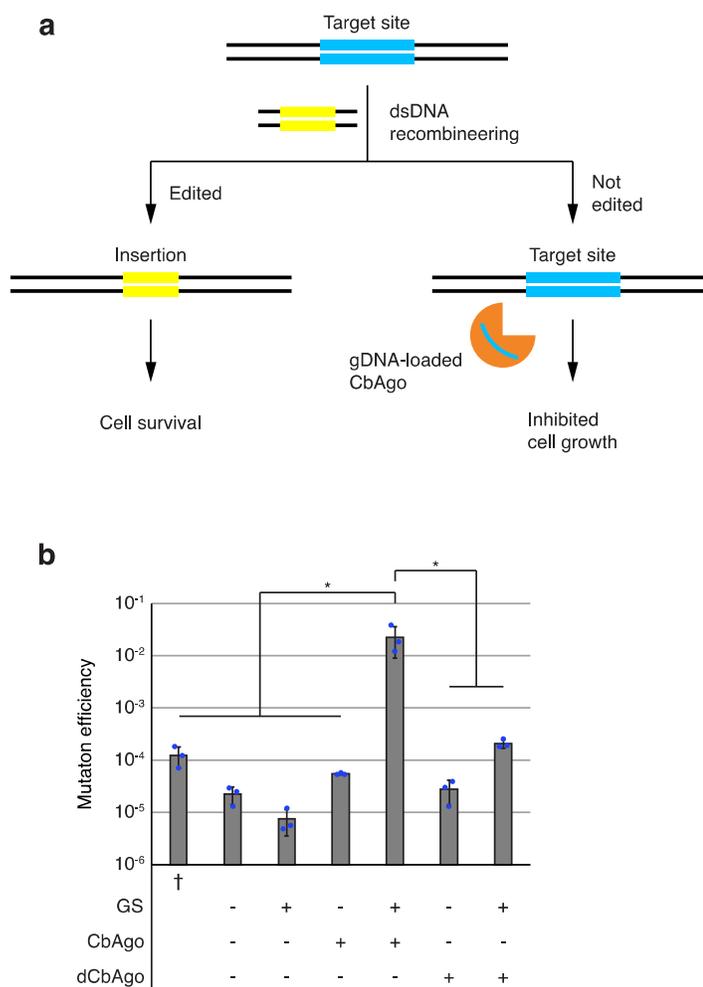
Figure 3.6: Targeted CbAgo cleavage assists Lambda Red-mediated recombineering in RecA-deficient *E. coli* strain. a, Scheme of CbAgo-assisted recombineering. After recombineering, the growth of unedited cells will be suppressed by CbAgo cleavage and subsequent DNA degradation, while successfully edited cells will be resistant to CbAgo cleavage and exhibit kanamycin resistance. Yellow, kanamycin resistance cassette. Blue, lacZ gene. **b**, Mutation rates using strain SIJ488_ΔrecA in different genetic contexts. †Standard recombineering procedure was applied using plasmid-free cells. Error bars, mean ± s.d. from three independent experiments. *P* values were calculated by two-tailed unpaired Student's t-test; **P* < 0.05.

BIBLIOGRAPHY

Bierne, H., Ehrlich, S.D., Michel, B., 1997. Deletions at stalled replication forks occur by two different pathways. The EMBO Journal 16, 3332–3340.

Capaldo, F.N., Barbour, S.D., 1975. DNA content, synthesis and integrity in dividing and non-dividing cells of rec− strains of Escherichia coli K12. Journal of Molecular Biology 91, 53–66.

Citorik, R.J., Mimee, M., Lu, T.K., 2014. Sequence-specific antimicrobials using efficiently delivered RNA-guided nucleases. Nat Biotechnol 32, 1141–1145.

Jiang, W., Bikard, D., Cox, D., Zhang, F., Marraffini, L.A., 2013. RNA-guided editing of bacterial genomes using CRISPR-Cas systems. Nat Biotechnol 31, 233–239.

Kuzmenko, A., Oguienko, A., Esyunina, D., Yudin, D., Petrova, M., Kudinova, A., Maslova, O., Ninova, M., Ryazansky, S., Leach, D., Aravin, A.A., Kulbachinskiy, A., 2020. DNA targeting and interference by a bacterial Argonaute nuclease. Nature 587, 632–637.

Kuzminov, A., Stahl, F.W., 1997. Stability of linear DNA in recA mutant Escherichia coli cells reflects ongoing chromosomal DNA degradation. Journal of Bacteriology 179, 880–888.

Makarova, K.S., Wolf, Y.I., van der Oost, J., Koonin, E.V., 2009. Prokaryotic homologs of Argonaute proteins are predicted to function as key components of a novel system of defense against mobile genetic elements. Biology Direct 4, 29.

Michel, B., 2000. Replication fork arrest and DNA recombination. Trends in Biochemical Sciences 25, 173–178.

Pennington, J.M., Rosenberg, S.M., 2007. Spontaneous DNA breakage in single living Escherichia coli cells. Nat Genet 39, 797–802.

Skarstad, K., Boye, E., 1993. Degradation of individual chromosomes in recA mutants of Escherichia coli. Journal of Bacteriology 175, 5505–5509.

Swarts, D.C., Makarova, K., Wang, Y., Nakanishi, K., Ketting, R.F., Koonin, E.V., Patel, D.J., van der Oost, J., 2014. The evolutionary journey of Argonaute proteins. Nat Struct Mol Biol 21, 743–753.

Taylor, A.F., Smith, G.R., 1985. Substrate specificity of the DNA unwinding activity of the RecBC enzyme of Escherichia coli. Journal of Molecular Biology 185, 431–443.

Vaiskunaite, R., Vainauskas, J., Morris, J.J.L., Potapov, V., Bitinaite, J., 2022. Programmable cleavage of linear double-stranded DNA by combined action of Argonaute CbAgo from Clostridium butyricum and nuclease deficient RecBC helicase from E. coli. Nucleic Acids Research 50, 4616–4629.

Chapter 4

DNA-GUIDED AGO-MEDIATED GENOME MUTATOR

In previous chapters, we demonstrated that CbAgo can be directed by plasmid-originated gDNAs to target the genomic sites, generate DSBs on chromosomes, and lead to different genomic mutations depending on host DNA repair mechanisms. While the exact mechanism remains elusive, dCbAgo appears to follow a similar path and interact with the target site on chromosomes. Since the target molecules for gDNA-bound (d)CbAgo must be ssDNA, we hypothesize that gDNA-bound (d)CbAgo should be able to direct ssDNA-acting enzymes and introduce other types of mutations, reminiscent of the dCas9-based epigenome editing (Gilbert et al., 2013; Maeder et al., 2013; Perez-Pinera et al., 2013; Lawhorn et al., 2014; Chavez et al., 2015; Brocken et al., 2018).

Here we explored the possibility of attaching cytidine deaminases to (d)CbAgo to create novel DNA-guided deaminases. Since many cytidine deaminases such as rApo1 function on ssDNAs, we rationalize that a fusion protein of rApo1 and (d)CbAgo might be able to target the specific genomic region through the guide-directed (d)CbAgo binding and introduce C to T (or G to A) mutations near the binding site if the (d)CbAgo moiety retains its ability of gDNA acquisition and target binding, and the rApo1 deaminase activity is not impaired by the fusion. The obtained deaminase has the potential to serve as a novel DNA-guided mutator, which can be directed to and mutate the target genome region by the plasmid-encoded GS (Figure 4.1).

4.1 Introduction of in vivo mutagenesis

Experimental evolution is a critical technique in protein engineering and evolutionary biology, whose success relies on efficient target gene diversification. In vitro methods to diversify target genes always require intensive and repetitive labor and are often hindered by limited sequence diversity because of inefficient cellular DNA uptake. On the other hand, in vivo global mutagenesis methods suffer from toxicity, false positive artifact, and rapid silencing of mutagens. Therefore, targeted in vivo mutagenesis strategies have been

developed to overcome these deficiencies toward the ideal in vivo mutagenesis method that comes with high gene specificity with a low off-target rate and introduces evenly distributed mutations within a wide target space.

However, current in vivo mutagenesis methods still have their drawbacks. For example, Cas9-based base editor (Rees and Liu, 2018) can only mutate <10 nucleotides per sgRNA, while other Cas9-based methods like CRISPR-X (Hess et al, 2016) and EvolvR (Halperin et al., 2018) can target several hundreds of nucleotides from PAM but have no distinguishing gene boundary. Furthermore, the requirement of PAM sequences adjacent to the target sequence limits the broader application of Cas9-based mutagenesis. Methods using orthogonal error-prone DNA polymerases mutate the whole or part of a plasmid, which requires the target genes to be placed on that plasmid (Camps et al., 2003; Ravikumar et al., 2018). Similarly, the phage-assisted continuous evolution (PACE), which links the desired protein activity to phage production, requires the target genes to be placed on a specific plasmid encoding a phage protein pIII (Esvelt et al., 2011; Morrison et al., 2020).

The gene-specific in vivo mutagenesis using a fusion protein of deaminase and T7 RNA polymerase has been demonstrated successful in *E. coli* (Moore et al., 2018; Park and Kim, 2021), yeast (Cravens et al., 2021), and human cells (Chen et al., 2020). This method allows the introduction of C to T and G to A mutations to the target genes in vivo with minimal off-target rates. However, it also requires the target genes to be placed between a T7 promoter and terminator, which greatly limits its applicability.

4.2 Creation and validation of the mutator

To construct the fusion protein of deaminase and (d)CbAgo, we created a mutator plasmid encoding rApo1-XTEN-dCbAgo (XTEN is a linker used in Cas9-based base editor) under the control of the pTet promoter. An *E. coli* strain Δung, whose genomic uracil N-glycosylase (UNG) gene has been knocked out, was also created. The UNG knockout has been shown to inhibit the repair of uridine to cytidine after deamination and increase the mutagenesis rates in previous reports (Hess et al., 2017).

To verify the activity of the fusion protein, we transformed the strain Δung with the obtained mutator plasmid, grew the transformed cells in the presence of aTc, isolated plasmids from the cultures, and sequenced those plasmids. All the plasmids isolated from cells expressing rApo1-XTEN-dCbAgo had C and T mixed nucleotides or G and A mixed nucleotides in varied positions, while the plasmids isolated from the control strain containing empty plasmids did not bear virtually any mutations (Figure 4.2). This observation confirms that the fusion protein is an active deaminase in vivo.

In addition, we examined the ability of rApo1-XTEN-dCbAgo to randomly introduce mutations in *E. coli* genomes by testing its assistance in developing rifampicin (Rif) resistance. Rifampicin is an antibiotic with broad-spectrum antibacterial activity owing to its RNAP inhibition effect. Rifampicin can bind to RNAP at a site near its active center and block the formation of phosphodiester bonds during RNA synthesis (McClure and Cech, 1978). In *E. coli*, resistance to rifampicin arises from mutations in the rpoB gene, which encodes the β subunit of RNAP, because these mutations decrease the affinity of RNAP against rifampicin (Xu et al., 2005). Previous reports have demonstrated that single-point mutation in different positions of the rpoB gene is able to generate rifampicin resistance. The survival rate under rifampicin selection correlates with random mutation rates in the genome (Garibyan et al., 2003; Badran and Liu, 2015).

The cells expressing rApo1-XTEN-dCbAgo, rApo1-XTEN-CbAgo, rApo1, and dCbAgo together as separate proteins (rApo1/dCbAgo), or no mutators were subject to rifampicin and ampicillin selection. The rifampicin resistance frequency was calculated as the fraction of ampicillin-resistant cells that became resistant to rifampicin (rifampicin-resistant and ampicillin-resistant cfu/ampicillin-resistant cfu). Selection on rifampicin plates revealed that rApo1-XTEN-dCbAgo–expressing cells displayed rifampicin resistance frequency $\sim10^5$-fold higher than background, indicating rApo1-XTEN-dCbAgo can serve as a strong global mutagen (Figure 4.3). The rifampicin resistance frequency generated by rApo1-XTEN-dCbAgo was also significantly higher than the one by rApo1/dCbAgo, suggesting tethering dCbAgo to rApo1 efficiently enhances its deamination activity on genomic DNAs,

likely via dCbAgo-mediated DNA binding. On the other hand, the presence of CbAgo DNA cleavage activity in the mutator dramatically decreases the rifampicin resistance frequency compared to the one generated by rApo1-XTEN-dCbAgo, which can be explained by one (or both) of the following models: (i) Target release after CbAgo cleavage decreases the interacting time between the tethered cytidine deaminase and the genomic DNA; (ii) CbAgo cleavage on genomic DNAs generates DSBs, which triggers RecBCD-mediated DNA degradation, which has a good chance to remove the uridine introduced by deaminase.

4.3 Evaluation of in vivo mutagenesis using dCbAgo-based mutator

After confirmation of the deaminase activity of rApo1-XTEN-dCbAgo, we sought to examine if it can acquire its guides from plasmid-encoded guide sequence and introduce mutations on the target site in vivo. The mutator plasmid was used to transform the Δung strain together with a temperature-sensitive guide plasmid encoding a 1000-bp guide sequence, which is homologous to the lacZ gene in the genome (Figure 4.4c). The transformed cells were allowed to grow in the inducing media containing aTc and the guide plasmid was eliminated from the induced cells. Colonies of cells that were free of guide plasmid were subject to colony PCR to amplify the genomic region around the target locus and Sanger sequencing to evaluate mutations.

Only the cells that had expressed rApo1-XTEN-dCbAgo in the presence of the guide plasmid yielded mutations from the sequenced region, which were all G to A mutations. This indicates the deamination by rApo1-XTEN-dCbAgo on the genome is strongly strand-dependent, which has also been observed in the case of mutaT7 (Moore et al., 2018; Park and Kim, 2021). Surprisingly, most of the observed mutations (5 out of 6) located in the regions that are outside of the target region, while there was only 1 out of 6 mutations that is within the target region (Figure 4.4b). This is quite against our expectation and our model (Figure 4.1), where the dCbAgo moiety of the mutator should be directed to the target region by small gDNAs generated from the guide plasmid. If dCbAgo moiety were to use small gDNAs and bind to ssDNA of chromosomes (during replication), we would expect

many more mutations within or slightly beyond the target region, because the linker length between the rApo1 moiety and dCbAgo moiety is only 16 amino acids long, and therefore the position where the deamination happens and where the mutation is introduced should be within close range to the dCbAgo binding site. For comparison, the rApo1-XTEN-dCas9 deaminase can only efficiently generate C to T mutation to the position that is 19 nucleotides away from PAM, which is the Cas9 binding site (Komor et al., 2016).

On the other hand, the cells expressing rApo1 and dCbAgo as separate proteins with GS, and the cells expressing rApo1-XTEN-dCbAgo without GS did not yield detectable mutations within the sequenced region (Figure 4.4a). This observation suggests that rApo1-XTEN-dCbAgo can be "guided" to the target region and introduce mutations around the target region. The exact mechanism of this guide-directed deamination remains unknown, but it is unlikely via small gDNAs that are generated from the guide plasmid.

4.4 Experimental methods

Culture conditions. *E. coli*, cultured in Luria-Bertani (LB) medium and agar, was incubated at 30°C. When appropriate, antibiotics were added to the medium at the following final concentrations: ampicillin, 100 μg/ml; chloramphenicol, 20 μg/ml; kanamycin, 35 μg/ml; rifampicin, 100 μg/ml. Bacterial cell growth was monitored periodically by measuring the optical density of culture aliquots at 600 nm.

Determination of cfu and resistance frequency. Cells were transformed with appropriate plasmids and plated on LB plates supplemented with ampicillin. The next day, 1 ml of LB medium supplemented with ampicillin and 200 ng/ml aTc was inoculated by single colony and aerated at 30 °C for 24 h before making serial dilutions of cultures and plating on the LB plates supplemented with appropriate antibiotics to determine cfu.

Genome sequencing. The guide plasmids were first eliminated by streaking single colonies on plates and incubating the streaked plates at elevated temperatures. Ampicillin sensitive colonies from the streaked plates were subject to colony PCR using primers seq-lacZ.F/ seq-lacZ.R and PCR products were subject to Sanger sequencing.

Plasmid construction. pET42b-BE3 was a gift from Dr. David Liu (Addgene plasmid # 87437). To construct plasmids pTet_rApo1-XTEN-dCbAgo, pTet_rApo1-XTEN-CbAgo, plasmid backbones were amplified from pTet_CbAgo and pTet_dCbAgo using primers pTet-Cb.F/pTet-Cb.R. The rApo1 gene fragment with the XTEN linker sequence was amplified from plasmid pET42b-BE3 using primers rApo1.F/rApo1.R. The resulting PCR product was gel-purified and ligated with the plasmid backbones individually via Gibson Assembly. Plasmid pTet_rApo1/dCbAgo was obtained through amplification of pTet_rApo1-XTEN-dCbAgo with primers stop.F/stop.R. PCR product was gel-purified and self-ligated by Gibson assembly.

Strain construction. To generate strain Δung, the kanamycin-resistance cassette was amplified from genomic DNA of 3×ChikanR via colony PCR with primers ung.kanR.F/ung.kanR.R, and the PCR product was gel-purified as dsDNA donor. Strain DL1777 was transformed with pKD46 plasmid, plated on LB plates supplemented with ampicillin, and incubated overnight at 30 °C. A 5 ml culture inoculated from single colony was grown at 30°C until OD600 = 0.3–0.4. The Lambda-Red genes were then induced with 15 mM L-arabinose for 45 min. The culture was used to prepare electrocompetent cells by washing twice with 10% glycerol and resuspending in 50 µl 10% glycerol. These cells were transformed with ~250 ng dsDNA by electroporation and allowed to recover in 1 ml LB for 2 h at 30 °C. The recovered cells were plated on LB plates supplemented with kanamycin and incubated overnight at 37 °C. The correct knockouts were verified by colony PCR and sequencing. Upon obtaining positive hits, the pKD46 plasmid was cured by growing the cell in LB media at 37°C overnight and plating on LB plates. Successful plasmid curing was confirmed by the cell sensitivity to ampicillin.

Statistical analyses. GraphPad Prism 9 was used to evaluate statistical significance. Student's t-test (two-tailed) was used for the statistical analysis of the experiments. *P* values < 0.05 were considered significant.
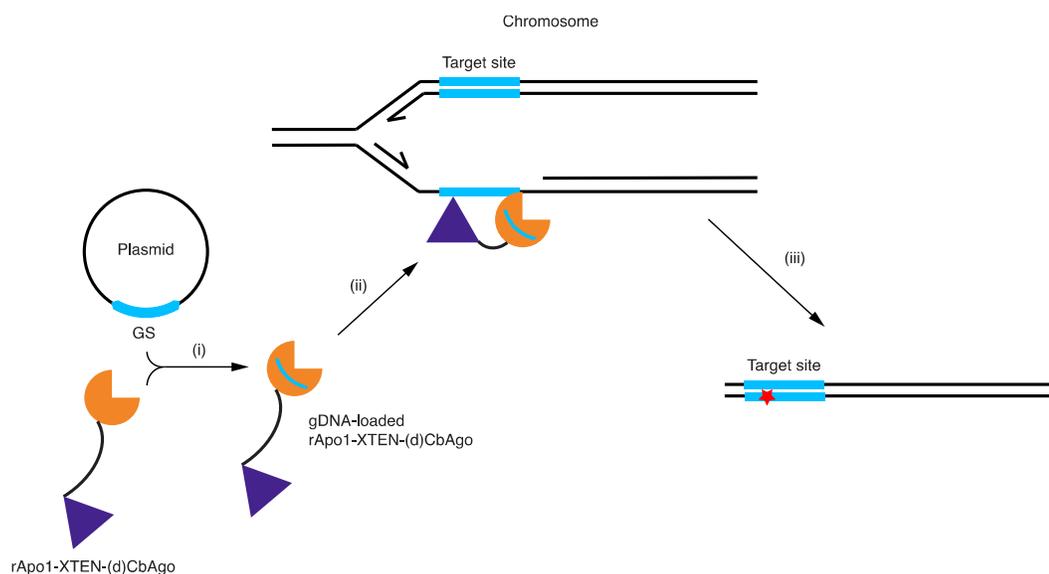
Figure 4.1: Proposed mechanism for the DNA-directed mutator. XTEN is a linker with 16 amino acids. (i) The (d)CbAgo moiety of the mutator acquires gDNAs from the plasmid-encoded GS. (ii) Guide-directed binding of the (d)CbAgo moiety to the chromosome (binding to the lagging strand may be favored) during chromosome replication. (iii) The rApo1 moiety deaminates chromosomal DNAs and eventually introduces C to T or G to A mutation to the genome via DNA replication or repair.
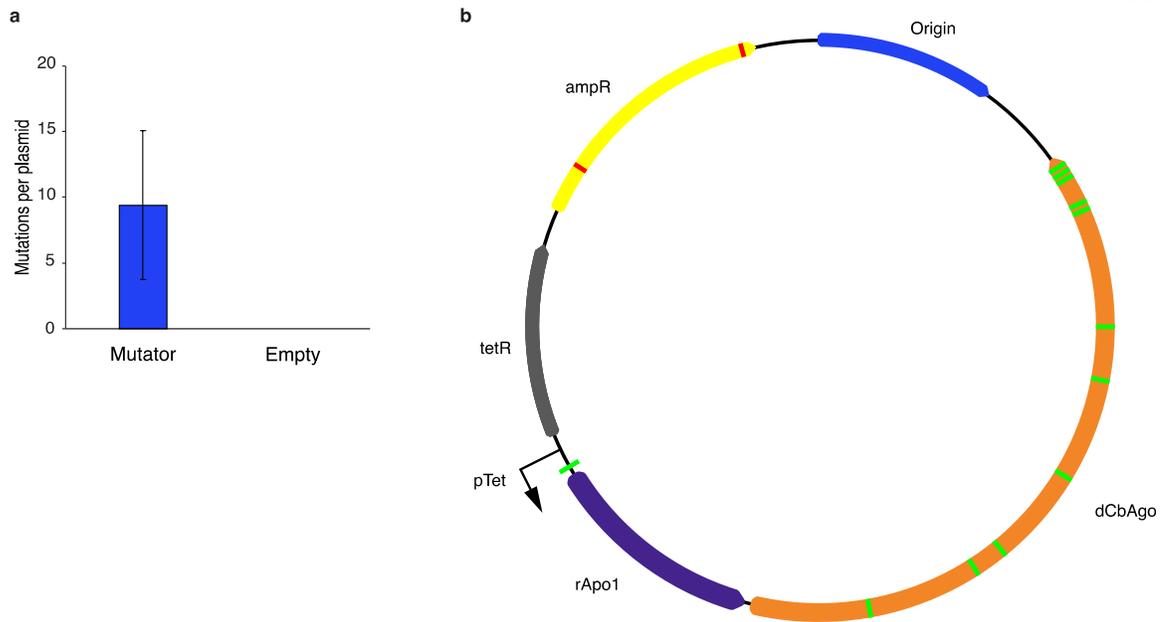
Figure 4.2: rApo1-XTEN-dCbAgo randomly mutates plasmids in vivo. **a**, The number of mutations identified from sequenced plasmids isolated from cells expressing rApo1-XTEN-dCbAgo or containing the empty plasmid. One mutation is defined as one position from the sequence data showing a mixture of nucleotides. Error bars, mean ± s.d. from ten independent experiments. **b**, Distribution of mutations on a plasmid isolated from cells expressing rApo1-XTEN-dCbAgo. Green: mixture of C and T. Red: mixture of G and A.
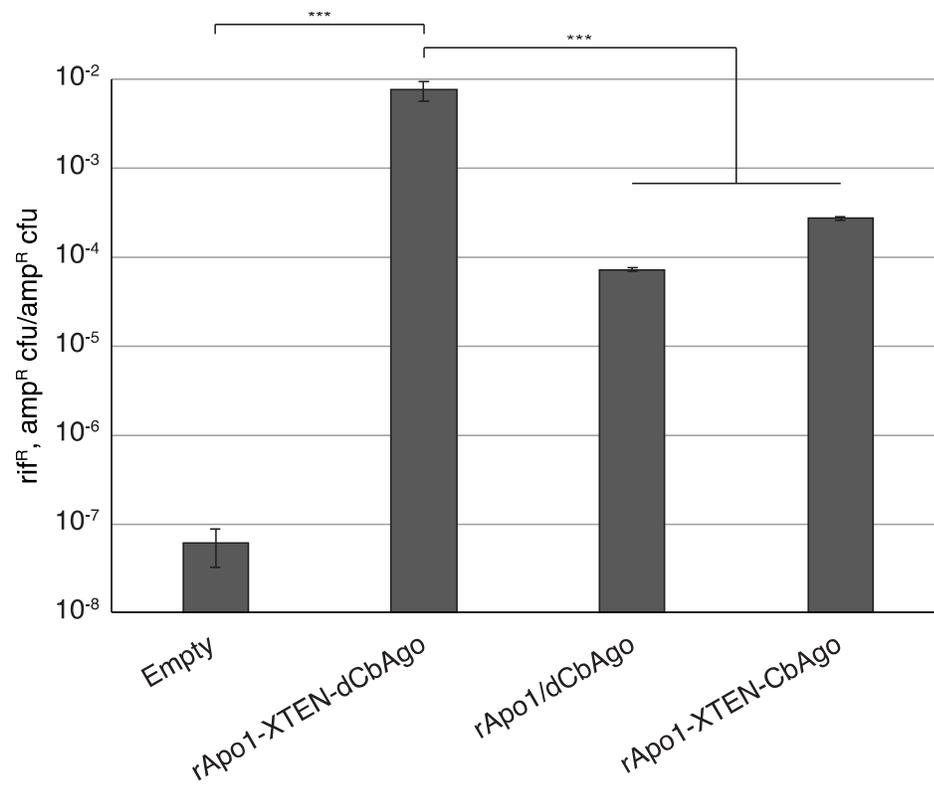
Figure 4.3: rApo1-XTEN-dCbAgo randomly mutates genomes in vivo**.** Rifampicin resistance assay using strain Δung expressing different mutator constructs. Error bars, mean ± s.d. from three independent cultures. *P* values were calculated by two-tailed unpaired Student's t-test; ***$P < 0.01$.
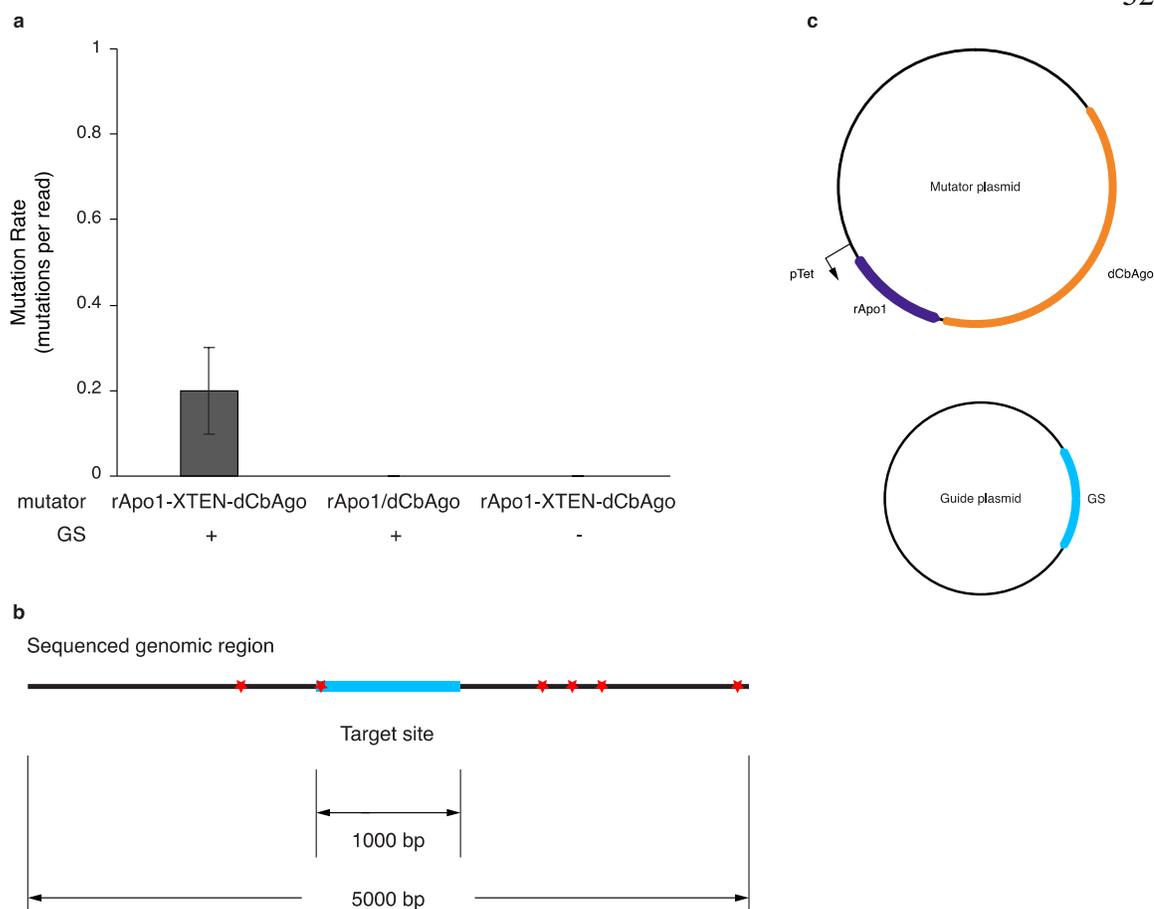
Figure 4.4: rApo1-XTEN-dCbAgo can be guided to introduce genomic mutations in vivo**. a**, The mutation rates of the cells expressing rApo1-XTEN-dCbAgo with GS, the cells expressing rApo1 and dCbAgo separately with GS, and the cells expressing rApo1-XTEN-dCbAgo without GS. Error bars, mean ± s.d. from three biological replicates. For each experiment, ten single colonies were sequenced to determine the mutation rate. **b**, Distribution of all identified mutations from the cells co-transformed with pTet_rApo1-XTEN-dCbAgo and pTs_GS. All identified mutations were G to A mutations from 30 sequenced colonies. **c**, Structures of the mutator plasmid and the guide plasmid.

BIBLIOGRAPHY

Allen, J.M., Simcha, D.M., Ericson, N.G., Alexander, D.L., Marquette, J.T., Van Biber, B.P., Troll, C.J., Karchin, R., Bielas, J.H., Loeb, L.A., Camps, M., 2011. Roles of DNA polymerase I in leading and lagging-strand replication defined by a high-resolution mutation footprint of ColE1 plasmid replication. Nucleic Acids Research 39, 7020–7033.

Badran, A.H., Liu, D.R., 2015. Development of potent in vivo mutagenesis plasmids with broad mutational spectra. Nat Commun 6, 8425.

Brocken, D.J.W., Tark-Dame, M., Dame, R.T., 2018. dCas9: A Versatile Tool for Epigenome Editing. Current Issues in Molecular Biology 26, 15–32.

Camps, M., Loeb, L.A., 2003. Use of Pol I-Deficient E. coli for Functional Complementation of DNA Polymerase, in: Arnold, F.H., Georgiou, G. (Eds.), Directed Enzyme Evolution: Screening and Selection Methods, Methods in Molecular Biology™. Humana Press, Totowa, NJ, pp. 11–18.

Camps, M., Naukkarinen, J., Johnson, B.P., Loeb, L.A., 2003. Targeted gene evolution in Escherichia coli using a highly error-prone DNA polymerase I. Proceedings of the National Academy of Sciences 100, 9727–9732.

Chavez, A., Scheiman, J., Vora, S., Pruitt, B.W., Tuttle, M., P R Iyer, E., Lin, S., Kiani, S., Guzman, C.D., Wiegand, D.J., Ter-Ovanesyan, D., Braff, J.L., Davidsohn, N., Housden, B.E., Perrimon, N., Weiss, R., Aach, J., Collins, J.J., Church, G.M., 2015. Highly efficient Cas9-mediated transcriptional programming. Nat Methods 12, 326–328.

Chen, H., Liu, S., Padula, S., Lesman, D., Griswold, K., Lin, A., Zhao, T., Marshall, J.L., Chen, F., 2020. Efficient, continuous mutagenesis in human cells using a pseudo-random DNA editor. Nat Biotechnol 38, 165–168.

Cravens, A., Jamil, O.K., Kong, D., Sockolosky, J.T., Smolke, C.D., 2021. Polymerase-guided base editing enables in vivo mutagenesis and rapid protein engineering. Nat Commun 12, 1579.

Esvelt, K.M., Carlson, J.C., Liu, D.R., 2011. A system for the continuous directed evolution of biomolecules. Nature 472, 499–503.

Garibyan, L., Huang, T., Kim, M., Wolff, E., Nguyen, A., Nguyen, T., Diep, A., Hu, K., Iverson, A., Yang, H., Miller, J.H., 2003. Use of the rpoB gene to determine the specificity of base substitution mutations on the Escherichia coli chromosome. DNA Repair 2, 593–608.

Gilbert, L.A., Larson, M.H., Morsut, L., Liu, Z., Brar, G.A., Torres, S.E., Stern-Ginossar, N., Brandman, O., Whitehead, E.H., Doudna, J.A., Lim, W.A., Weissman, J.S., Qi, L.S., 2013. CRISPR-Mediated Modular RNA-Guided Regulation of Transcription in Eukaryotes. Cell 154, 442–451.

Halperin, S.O., Tou, C.J., Wong, E.B., Modavi, C., Schaffer, D.V., Dueber, J.E., 2018. CRISPR-guided DNA polymerases enable diversification of all nucleotides in a tunable window. Nature 560, 248–252.

Hess, G.T., Tycko, J., Yao, D., Bassik, M.C., 2017. Methods and Applications of CRISPR-Mediated Base Editing in Eukaryotic Genomes. Molecular Cell 68, 26–43.

Jin, D.J., Gross, C.A., 1988. Mapping and sequencing of mutations in the Escherichia colirpoB gene that lead to rifampicin resistance. Journal of Molecular Biology 202, 45–58.

Komor, A.C., Kim, Y.B., Packer, M.S., Zuris, J.A., Liu, D.R., 2016. Programmable editing of a target base in genomic DNA without double-stranded DNA cleavage. Nature 533, 420–424.

Lawhorn, I.E.B., Ferreira, J.P., Wang, C.L., 2014. Evaluation of sgRNA Target Sites for CRISPR-Mediated Repression of TP53. PLOS ONE 9, e113232.

Maeder, M.L., Linder, S.J., Cascio, V.M., Fu, Y., Ho, Q.H., Joung, J.K., 2013. CRISPR RNA–guided activation of endogenous human genes. Nat Methods 10, 977–979.

McClure, W.R., Cech, C.L., 1978. On the mechanism of rifampicin inhibition of RNA synthesis. Journal of Biological Chemistry 253, 8949–8956.

Moore, C.L., Papa, L.J., Shoulders, M.D., 2018. A Processive Protein Chimera Introduces Mutations across Defined DNA Regions In vivo. J. Am. Chem. Soc. 140, 11560–11564.

Morrison, M.S., Podracky, C.J., Liu, D.R., 2020. The developing toolkit of continuous directed evolution. Nat Chem Biol 16, 610–619.

Park, H., Kim, S., 2021. Gene-specific mutagenesis enables rapid continuous evolution of enzymes in vivo. Nucleic Acids Research 49, e32.

Perez-Pinera, P., Kocak, D.D., Vockley, C.M., Adler, A.F., Kabadi, A.M., Polstein, L.R., Thakore, P.I., Glass, K.A., Ousterout, D.G., Leong, K.W., Guilak, F., Crawford, G.E., Reddy, T.E., Gersbach, C.A., 2013. RNA-guided gene activation by CRISPR-Cas9–based transcription factors. Nat Methods 10, 973–976.

Ravikumar, A., Arzumanyan, G.A., Obadi, M.K.A., Javanpour, A.A., Liu, C.C., 2018. Scalable, Continuous Evolution of Genes at Mutation Rates above Genomic Error Thresholds. Cell 175, 1946-1957.e13.

Rees, H.A., Liu, D.R., 2018. Base editing: precision chemistry on the genome and transcriptome of living cells. Nat Rev Genet 19, 770–788.

Xu, M., Zhou, Y.N., Goldstein, B.P., Jin, D.J., 2005. Cross-Resistance of Escherichia coli RNA Polymerases Conferring Rifampin Resistance to Different Antibiotics. Journal of Bacteriology 187, 2783–2792.

Chapter 5

ENGINEERING NICOTINE BIOSENSOR FOR IMPROVED SENSITIVITY

Nicotine intake has been shown to evoke nicotinic acetylcholine (ACh) receptors (nAChRs) in neurons via two pathways. The first pathway is termed the "outside-in" pathway, where exogenous nicotine enters the cell and directly binds to the nAChRs on the plasma membrane (PM). The nicotine binding of nAChRs results in their channel opening and Na^+ and Ca^{2+} ions entry, leading to the increase of the frequency of neuronal action potentials, similar to the activation process by the endogenous ACh (Miwa et al., 2011; Nees, 2015). The second pathway is termed the "inside-out" pathway. In this scenario, exogenous nicotine enters the cell, permeates into the endoplasmic reticulum (ER), and binds to nascent nAChRs as chaperones. The nicotine binding of nAChRs stabilizes the α4- and β2-subunit-containing (α4β2*) nAChRs and leads to their exit from the ER (Kuryatov et al., 2005; Sallette et al., 2005; Lester et al., 2009). As a result, nAChR on the PM is up-regulated post-translationally (Henderson and Lester, 2015).

To systematically study the outside-in and inside-out pathways as well as other nicotine-mediated processes during smoking/vaping, nicotine biosensors that can probe the pharmacokinetics of nicotine, which is the concentration and dynamics of nicotine in human cells and systems during nicotine intake, are required.

Fluorescent protein (FP)-based biosensors are fusion proteins of FPs and ligand binding proteins which have been developed for the measurement of a variety of neurotransmitters and neuromodulators including glutamate (Marvin et al., 2013), dopamine (Patriarchi et al., 2018; Sun et al., 2018), GABA (Marvin et al., 2019), ACh (Jing et al., 2020), and serotonin (Unger et al., 2020). Current developed FP-based biosensors use either bacterial periplasmic binding proteins (PBPs) or G-protein-coupled-receptors (GPCRs) as their ligand-binding moieties. The PBP-based biosensors are soluble proteins, which means they

can potentially be localized and image different subcellular locations in vivo. From the engineering point of view, those soluble biosensors can be readily purified and characterized in vitro, easy to evolve via bacterial expression and high-throughput screening, and amenable to building transformative devices.

5.1 Introduction to intensity-based nicotine sensing fluorescent reporters (iNicSnFRs)

To develop a tool for nicotine pharmacokinetic measurement, a series of FP-based nicotine biosensors (iNicSnFRs) has been developed (Shivange et al., 2019). The construct of the iNicSnFR biosensor is a fusion protein of a PBP and a circularly permuted GFP (cpGFP) with optimized linkers (and appropriate tags). Its development was inspired by the creation of the glutamate sensor iGluSnFR, which is constructed by inserting the cpGFP into the *E. coli* glutamate-binding PBP, GltI (Marvin et al., 2013).

PBPs are specialized receptors that are ubiquitously produced by bacteria and archaea to capture important small molecules such as carbohydrates, amino acids, vitamins, and ions, and transport them into their cytoplasm. PBPs are bilobed proteins and adopt two major forms: the open form and the closed form. The open form predominates in the absence of a ligand (Carrithers and Lerner, 1996; Shilton et al., 1996). When the ligand is present, PBP actively grabs the ligand between its two Venus flytrap–like lobes, accompanied by large conformational changes in the protein structure (Marvin et al., 2011, 2013, 2018). Notably, the resulting closed form of the ligand-bound PBP generates a protein-binding surface that is not present in the open form, so that the ligand-bound PBPs can be recognized by membrane-bound receptors and transport the captured ligand (Hollenstein et al., 2007).

The specific PBPs that bind quaternary amine molecules such as choline, glycine betaine, and proline betaine were selected for the construction of iNicSnFRs assuming their affinity for choline and/or betaine could be potentially tuned toward nicotine, a tertiary amine, via protein engineering. In support of this assumption, structural studies of the quaternary amine binding PBPs demonstrate that its binding of the quaternary amine involves cation-π interactions mediated by four aromatic amino acid residues (Schiefner et al., 2004). Such

cation-π interactions have also been observed in the protein complex of nAChRs (Zhong et al., 1998; Brejc et al., 2001; Morales-Perez et al., 2016) or GPCRs (Van Arnam et al., 2013) with their primary and secondary amine ligands.

The insertion of cpGFP into a quaternary amine-binding PBP (OpuBC) from *Thermoanaerobacter sp X513* yielded a weakly responding iNicSnFR, whose fluorescence increases upon nicotine binding (quantified by the measurement of $\Delta F/F_0$, using excitation at 485 nm and measuring emission at 510 nm), confirming the binding event between the *T. sp X513* OpuBC and nicotine, and the successful allosteric transduction of conformational change from PBP into the cpGFP fluorophore pocket. Iterative directed evolution using site saturation mutagenesis (SSM) and site-directed mutagenesis (SDM) improved its sensitivity against nicotine by a factor of $\sim10^5$ and yielded iNicSnFR3a and iNicSnFR3b (Shivange et al., 2019).

Furthermore, the obtained iNicSnFR3 has been successfully directed to the PM and ER in different mammalian cell lines. Both the iNicSnFR3a and iNicSnFR3b can detect the intracellular nicotine when the applied extracellular nicotine concentration is above 1 μM. All the in vivo measurements show that nicotine equilibrates in the ER or completely depletes from the ER within a 10 s dynamic range, responding to the different extracellular nicotine concentrations. The data also show that the ER and PM have similar nicotine levels when the same extracellular nicotine concentration is applied (Shivange et al., 2019).

Later, crystal structures of ligand-bound iNicSnFR3a were obtained (Nichols et al., 2022). In the resolved structure obtained from co-crystalizing iNicSnFR3a and nicotine (PDB 7S7U), the ligand-bound PBP moiety of iNicSnFR3a adopts a closed form, but the ligand (most likely, nicotine) in the binding pocket was poorly resolved, possibly due to the micromolar affinity of iNicSnFR3a against nicotine. On the other hand, the crystal structure of varenicline-bound iNicSnFR3a was resolved with much less ambiguity (PDB 7S7T). A comparison of these two structures shows that nicotine and varenicline share the same interacting residues in the binding pocket and varenicline forms cation-π interactions with Tyr65 and Tyr357, indicating the nicotine and varenicline binding in iNicSnFR3a

follows the common binding mode observed in other OpuBC PBPs (Schiefner et al., 2004). Notably, in the unbound state of iNicSnFR3a, the Glu78 in the linker region between PBP and cpGFP stays at a short distance to the oxygen atom of the fluorophore, while the conformational change occurred in the ligand binding directs the Glu78 away from the fluorophore, which may stabilize the fluorophore and increase the fluorescence.

However, when it comes to nicotine measurement in other human systems, for example, blood, the sensitivity of iNicSnFR3 is not enough. This is because the blood levels of nicotine during and after nicotine intake like smoking, oral snuff, chewing tobacco, or nicotine gum, typically fall in the range between 50 nM to 150 nM during the 5–10 min time frame (Benowitz et al., 1988; Hukkanen et al., 2005). We aimed to further increase the sensitivity of iNicSnFR3 via a combination of computational design and directed evolution. To better characterize the sensitivity in the submicromolar range, the sensitivity in terms of S-slope is defined as $(\Delta F/F_0)/(\Delta[\text{nicotine}])$. Because the iNicSnFRs respond with a Hill coefficient of unity, their dose-response curves have linearity in low ligand concentrations (below 5% of EC50, EC50 equals the dissociation constant K_d for ligand binding and has the usual pharmacological meaning) and thus their binding affinity can be reflected by S-slope in this scenario. Preliminary data suggests the S-slope for iNicSnFR3a is ~0.25 μM^{-1} at pH 7.40 (Figure 5.1), and we aimed to increase it by at least 10-fold as a remarkable improvement for a better nicotine sensor.

5.2 Computational prediction

During the development of iNicSnFR3, the ligand binding pocket between the two lobes (Venus flytrap domains) of the PBP has been optimized by extensive SSM on "first-shell" residues that lie within 7 Å of the bound ligand as well as on several "second-shell" residues that have interactions with the first-shell residues (Shivange et al., 2019). To predict the mutation hotspots outside the "first shell," including those amino acid residues that are far away from the ligand and are not directly involved during ligand binding, we applied computational design for this purpose.

We first used AutoDock Vina to dock nicotine into iNicSnFR3a (Trott and Olson, 2010). The structure of iNicSnFR3a was extracted from the crystal structure of varenicline-bound iNicSnFR3a (PDB ID: 7S7T) and the varenicline molecule was deleted from the structure to provide an empty binding pocket for nicotine docking. The structure was prepared in AutoDockTools by removing waters, adding polar hydrogens, and assigning Gasteiger charges. To verify the docking procedure, we first docked varenicline into the prepared structure and the highest scoring pose showed only a ~0.5 Å RMSD from the original varenicline molecule in 7S7T, confirming the reliability of the docking procedure. Then, we docked nicotine into the structure and the conformation with the highest score was selected for computational design.

The nicotine-docked iNicSnFR3a structure was first standardized by our in-house computational protein design suite TRIAD. Then we used the single mutation stability module from TRIAD to rank every possible amino acid substitution at every residue position in the PBP moiety of iNicSnFR3a, one at a time, by predicting the change in free energy of folding upon mutation ($\Delta\Delta G$). The top 100 variants were selected (Table 1), from which we selected 34 sites for subsequent sequence design: S9, N20, E24, E27, R36, N39, N46, K51, R52, E64, E78, S325, K336, R341, K342, K362, L384, T413, K414, M418, E429, Q431, D434, D439, D452, D453, K454, R467, E476, N497, K499, D501, E513, E517, excluding all proline sites and the sites that have been optimized in previous evolution (Shivange et al., 2019).

Finally, we applied the sequence design module from TRIAD to these selected positions individually. All twenty amino acid residues were used to substitute each position and the resulting free energy changes were evaluated. This is essentially the same prediction strategy as the one used in the TRIAD single mutation stability module, but the overall free energy was calculated more comprehensively and reliably using the Rosetta algorithm. Sequence design on each selected position performed detailed free energy analysis and generated 20 entries of energy scores, from which we identified the positions that could generate the biggest energy change when mutated from their wild-type sequences (as

termed "free energy gain", Table 2). The top 10 positions: N20, E27, N46, K51, R52, K342, Q431, K454, R467, and E476, which gave the biggest energy changes from wild-type sequence upon mutation, plus N11 which was mutated in iNicSnFR3b, were selected for SSM screening.

5.3 Directed evolution

The full evolution path from iNicSnFR3a to iNicSnFR12 is shown in Figure 5.2. We performed the first round of SSM on positions N20, K342, Q431, and R467 on iNicSnFR3a individually and obtained mutations with the most improved sensitivity (N20R, K342Y, Q431A, and R467S). Testing different combinations of these mutations via SDM yielded a double mutant (iNicSnFR3a Q431A R467S) with further improved sensitivity.

Starting from iNicSnFR3a Q431A R467S, we performed the second round of SSM on N20 and K342 again, as well as on other positions from the list including N11, E27, N46, K51, R52, K454, E476, and identified N11E as a beneficial mutation and yielded iNicSnFR3a N11E Q431A R467S, which we termed iNicSnFR11.

However, the third round of SSM on iNicSnFR11 did not yield any beneficial mutations and we reasoned that iNicSnFR11 might be trapped at local fitness maxima (Packer and Liu, 2015). We decided to revisit the previously mutated positions by combinatorically screening multiple previously obtained beneficial mutations at the same time. We made the first combinatorial library of iNicSnFR11 incorporating K342, Y342, Q431, A431, G431, D431, E431, R467, G467, S467, H467, screened the library, and obtained iNicSnFR11 A431D with improved nicotine sensitivity. Then we made the second combinatorial library of iNicSnFR11 A431D incorporating K51, L51, Q51, R51, R52, V52, T52, E476, K476, M476, V476 and obtained iNicSnFR11 A431D E476K with further improved nicotine sensitivity, which we termed iNicSnFR12. The iNicSnFR12 is iNicSnFR3a with mutations: N11E, Q431D, R467S, and E476K.

5.4 Characterization of iNicSnFR11 and iNicSnFR12

We used excitation at 485 nm, performed emission measurements at 510 nm, and generated dose-response relations for iNicSnFR11 and iNicSnFR12 against four nicotinic agonists: nicotine, ACh, choline, and varenicline. For iNicSnFR11, the data suggests it has tighter binding against varenicline and Ach compared to nicotine (Figure 5.3), which is also true for its iNicSnFR3a parent. The estimated EC50 for choline is more than 17-fold greater than for nicotine (the actual EC50 is likely even bigger because the maximum fluorescence has not been reached when 1 mM nicotine is added), which is much higher than the usual value for choline in the brain ($\sim$10 μM; Klein et al.,1992). Linear dose-response relation of iNicSnFR11 in the nanomolar range gives the S-slope of 1.5 μM^{-1}, representing a 6-fold improvement over iNicSnFR3a (Figure 5.4).

For iNicSnFR12, similar to iNicSnFR11, the data suggests it has tighter binding against varenicline and Ach compared to nicotine (Figure 5.5). The estimated EC50 for choline is more than 12-fold greater than for nicotine (the actual EC50 is likely even bigger because the maximum fluorescence has not been reached when 1 mM nicotine is added), which is again much higher than the usual value for choline in the brain ($\sim$10 μM; Klein et al.,1992). Linear dose-response relation of iNicSnFR12 in the nanomolar range gives the S-slope of 2.6 μM^{-1}, representing a 10-fold improvement over iNicSnFR3a (Figure 5.6).

However, in healthy adults, the concentration of choline in plasma is around 10 μM (Bligh, 1952). This suggests even though iNicSnFR12 represents the most sensitive nicotine biosensor reported so far and has a much stronger response to nicotine than to choline, its usage in nicotine measurement for single cell or plasma might still be interfered with by choline, because the lowest nicotine concentration to be detected in plasma is 10 nM (which is enough to activate the "inside-out" pathway), which gives only 0.026 as fluorescence increase according to the S-slope, while 10 μM choline gives ~1 as fluorescence increase, which may dramatically decrease the signal-to-noise ratio. Therefore, desensitizing the iNicSnFRs against choline, while continuing to improve its sensitivity against nicotine, is still required for the application of iNicSnFRs.

5.5 Experimental methods

Mutagenesis and library screening. SSM libraries were generated employing the "22c-trick" method (Kille et al., 2013). The combinatorial mutagenesis (CM) libraries were generated in a way similar to the "22c-trick," with the modification that mixtures of primers, with each primer incorporated with one desired mutation, were used for PCR. The PCR products were gel purified, digested with DpnI, ligated using the Gibson assembly Mix, and used to directly transform *E. coli* BL21-Gold (DE3) chemically competent cells (Agilent Technologies, Santa Clara, CA). The expression and screening of iNicSnFR variants were performed in 96-well plate format, and 93 variants per SSM library were screened. Individual colonies from iNicSnFR libraries were cultivated in 1 ml of ZYM5052 autoinduction media (Studier, 2005) supplemented with 100 μg/ml ampicillin for 30 h at 30 °C, 250 rpm. Then the cells were harvested (3000 × g, 10 min, 4 ℃), washed with PBS, pH 7.40, and resuspended in 3×PBS, pH 7.40. Resuspended cells were lysed by freezing and thawing using liquid nitrogen and a room-temperature water bath. Intact cells and cell debris were removed by centrifugation at 3500 × g for 30 min at 4°C. The supernatants of lysates were tested with excitation at 485 nm and emission at 535 nm. A Tecan Spark M10 96-well fluorescence plate reader (Tecan, Männedorf, Switzerland) was used to measure baseline fluorescence (F_0) and nicotine-induced fluorescence (ΔF). Promising clones were amplified and sequenced, and the beneficial mutations were confirmed by the measurement of dose-response relation against nicotine (described below). The optimally responding construct in each round of mutagenesis was used as a template for the next round of SSM, SDM, or CM.

Purification and measurement of iNicSnFRs. All proteins were overexpressed in *E. coli* BL21-Gold (DE3) cells using 50 ml ZYM5052 autoinduction media (Studier, 2005). Cells were collected by centrifugation and stored at –80°C until use. For purification, frozen cell pellets were resuspended in PBS, pH 7.40, and lysed by sonication. Intact cells and cell debris were removed by centrifugation at 15,000 × g for 30 min at 4°C. The supernatant was collected and loaded onto a prewashed Ni NTA column with wash buffer at 4°C. Ni NTA wash buffer contained 20 mM imidazole in PBS, pH 7.4. Elution was achieved using an imidazole gradient (20–200 mM). Proteins were concentrated by centrifugation through

Amicon Ultra 15 filter units (Millipore, Burlington, MA) with a 30-kD cutoff and then dialyzed against 3×PBS, pH 7.40. The dialyzed protein was then subject to dose-response studies to characterize its responses to various drugs.

Fluorescence dose-response relation measurement. Purified biosensor and drug solutions were mixed by a liquid handling robot (epMotion) to yield the mixture of 100 nM biosensor and the drug with desired concentrations. Samples were prepared in triplicate. All solutions were in 3×PBS, pH 7.4. A Tecan Spark 10M was used to read the plate with 485 nm excitation and 535 nm emission wavelengths to measure GFP fluorescence (F_0 and ΔF). The resulting data were fit with the Origin 9.2 software (OriginLabs) to the Hill equation. As the Hill coefficient is near 1.0, we, therefore, calculated S-slope = $(\Delta F_{max}/F_0)/EC50$ in units of μM^{-1}.

Figure 5.1: Dose-response relations for purified iNicSnFR3a. Purified iNicSnFR3a was used in the concentration of 100 nM in 3×PBS, pH 7.40. Data are mean ± SEM (n = 3).

| Starting point | SSM | SDM | SSM | CM | CM |

iNicSnFR3a
wild-type
EC50 84 µM
ΔF/F0 = 6

N20R

K342Y

Q431A
EC50 60 µM
ΔF/F0 = 11

R467S

iNicSnFR11

R467S
EC50 21 µM
ΔF/F0 = 8

N11E
EC50 9.8 µM
ΔF/F0 = 7

Q431D
EC50 7.0 µM
ΔF/F0 = 7

iNicSnFR12
E476K
EC50 6.2 µM
ΔF/F0 = 8

Figure 5.2: Directed evolution of the iNicSnFRs. The path of protein engineering of iNicSnFR constructs. EC50 and ΔF/F0 were measured using bacterial lysates. SSM: site saturation mutagenesis; SDM: site-directed mutagenesis; CM: combinatorial mutagenesis.

Figure 5.3: Dose-response relations for purified iNicSnFR11 against different drugs. Purified iNicSnFR11 was used in the concentration of 100 nM in 3×PBS, pH 7.40. Data are mean ± SEM (n = 3). The S-slope here is defined as $(\Delta F_{max}/F_0)/EC50$ because iNicSnFRs respond with a Hill coefficient of unity.
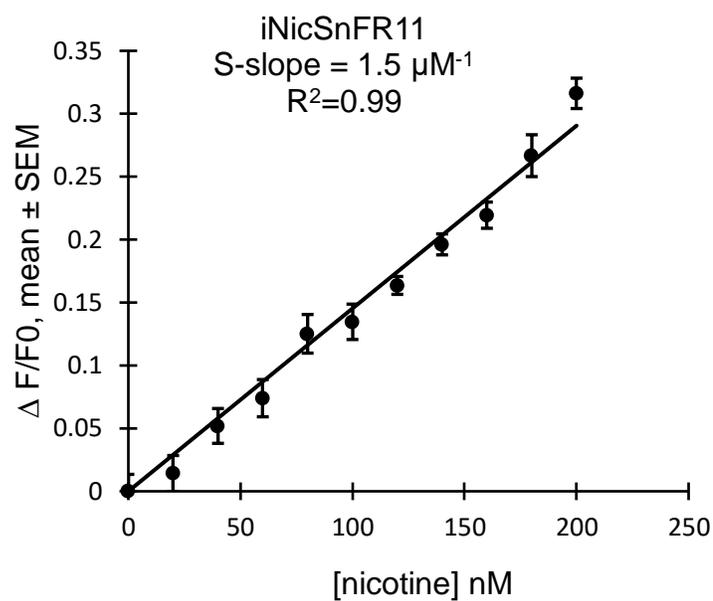
Figure 5.4: Linear dose-response relation for purified iNicSnFR11 against nicotine.
Purified iNicSnFR11 was used in the concentration of 100 nM in 3×PBS, pH 7.40. Data
are mean ± SEM (n = 16).
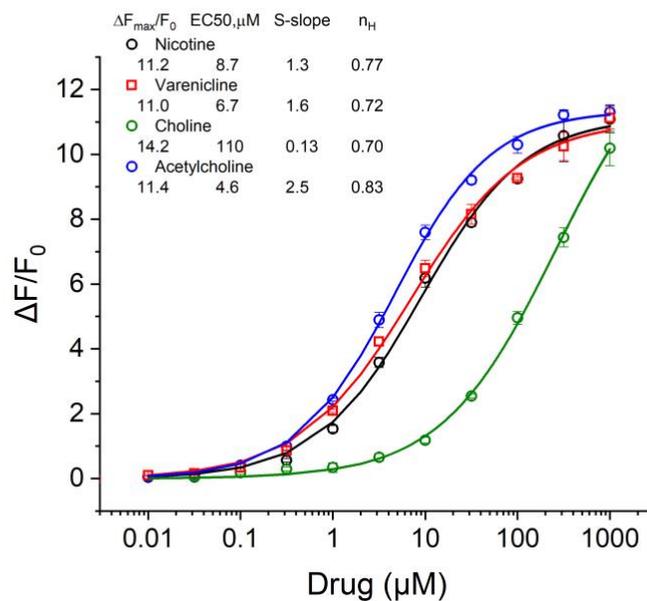
Figure 5.5: Dose-response relations for purified iNicSnFR12 against different drugs.
Purified iNicSnFR12 was used in the concentration of 100 nM in 3×PBS, pH 7.40. Data are mean ± SEM (n = 3). The S-slope here is defined as $(\Delta F_{max}/F_0)/EC50$ because iNicSnFRs respond with a Hill coefficient of unity.
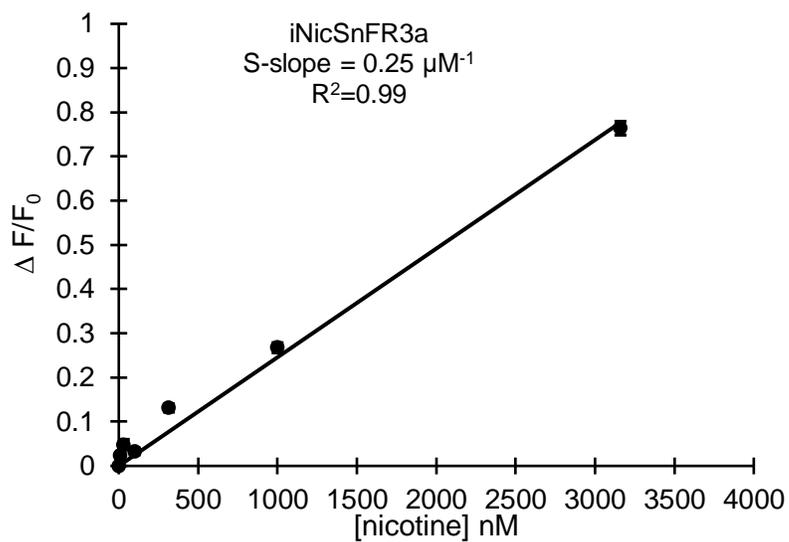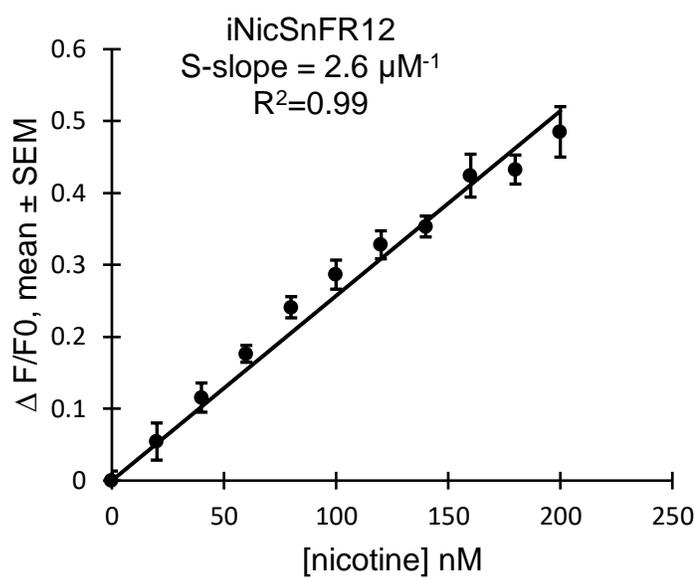
Figure 5.6: Linear dose-response relation for purified iNicSnFR12 against nicotine.
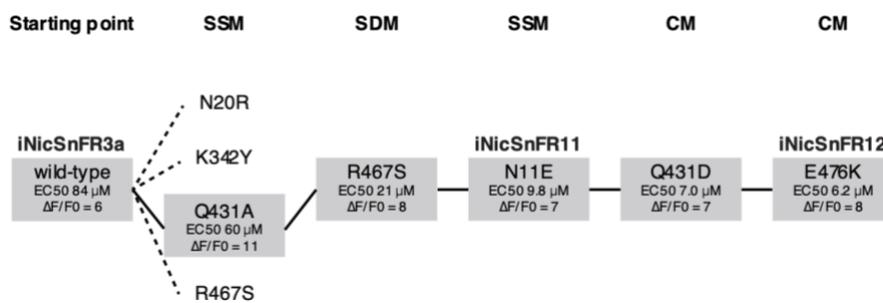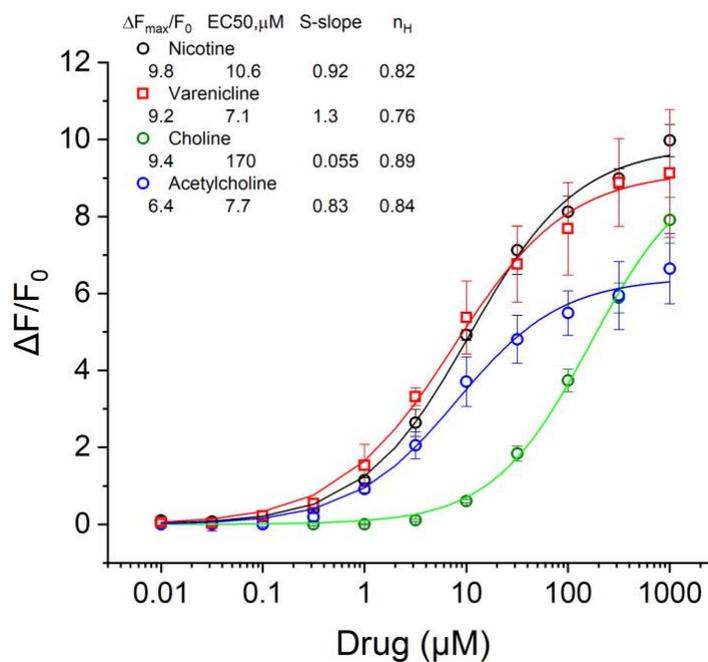Purified iNicSnFR12 was used in the concentration of 100 nM in 3×PBS, pH 7.40. Data are mean ± SEM (n = 8).

Ranking	Mutation	Energy Score
1	R467W	-7.1078127
2	D72Y	-6.7707508
3	D72F	-6.2820601
4	N39W	-5.9692582
5	P475W	-5.9475177
6	P79W	-5.7863615
7	R36F	-5.505194
8	E24W	-5.2833046
9	P328W	-5.2793392
10	R36Y	-5.1795235
11	P79H	-5.1449738
12	N20W	-4.8246239
13	P79M	-4.7779608
14	N46W	-4.6925254
15	Q431Y	-4.6010919
16	K499W	-4.535685
17	P79F	-4.4745986
18	N39F	-4.4237193
19	R36W	-4.42128
20	K342W	-4.411696
21	D72H	-4.3542099
22	E24Y	-4.3187315
23	P400W	-4.1939649
24	E24F	-4.1337034
25	P328Y	-4.1313307
26	N39Y	-4.1266322
27	E513W	-4.1100203
28	E24L	-4.1067959
29	S325W	-4.0312502
30	P79L	-3.9890606
31	P475F	-3.9846658
32	P507A	-3.9595498
33	P475Y	-3.9553714
34	D453W	-3.8811972
35	N20Y	-3.8492186
36	K454Y	-3.8384634

37	R36I	-3.8167268
38	R52W	-3.7538574
39	P400Y	-3.7526092
40	E64L	-3.7279936
41	P328F	-3.7270317
42	R467Y	-3.7269856
43	S325F	-3.7228232
44	E476W	-3.653812
45	Q431F	-3.6227739
46	N497Y	-3.6135927
47	N497W	-3.5791251
48	R467L	-3.5493039
49	N497F	-3.5249293
50	T413W	-3.5164357
51	R36V	-3.4819131
52	P79Y	-3.4751594
53	E78W	-3.4723928
54	S325Y	-3.4395971
55	R36L	-3.4302272
56	S9Y	-3.4162288
57	K336W	-3.4139543
58	P77W	-3.3913841
59	K454W	-3.3772821
60	P323W	-3.3653378
61	T413Y	-3.3546179
62	D439M	-3.3411551
63	E27L	-3.3357544
64	M418Y	-3.3332307
65	P79I	-3.3173147
66	E78Y	-3.3082286
67	P323F	-3.3074791
68	K454F	-3.2698812
69	P475L	-3.267489
70	P464I	-3.258727
71	L384W	-3.2441745
72	E24H	-3.2351242
73	P400F	-3.2295172

74	E517W	-3.2149863
75	P475H	-3.2145723
76	D72W	-3.1951895
77	Q431W	-3.1933558
78	H68W	-3.1910837
79	N46L	-3.1854905
80	P79T	-3.1647957
81	P350I	-3.1506409
82	K342F	-3.1496806
83	K51W	-3.1408016
84	D434V	-3.133205
85	P464V	-3.1317006
86	E78F	-3.1227254
87	K362W	-3.1148366
88	D501I	-3.1110394
89	R341W	-3.0742049
90	P79Q	-3.0611639
91	D452H	-3.0518135
92	E429W	-3.044675
93	P350W	-3.0379412
94	R467F	-3.0377459
95	P507T	-3.023746
96	N497V	-3.0210514
97	P400L	-3.0143553
98	N20F	-2.9747545
99	D453H	-2.9694544
100	K414W	-2.9650854

Table 5.1: Top 100 mutation outputs from Single Mutant Stability (ΔΔG) analysis on iNicSnFR3a (ranked by energy score).

Ranking	Position	Mutation	Free Energy Gain
1	K342	F	-4.510815498
2	R52	K	-3.981761265
3	Q431	Y	-3.961053323
4	K51	I	-3.853883659
5	K454	R	-3.83684393
6	N20	D	-3.560516383
7	R467	K	-3.56029382
8	E476	S	-2.877748373
9	N46	L	-2.742869843
10	E27	L	-2.043275748
11	D452	P	-2.04233217
12	T413	I	-1.954859494
13	D453	S	-1.849582601
14	E64	D	-1.729077026
15	N497	A	-1.728372151
16	K414	I	-1.612479529
17	M418	D	-1.588025107
18	S325	F	-1.429726312
19	R36	I	-1.416204
20	S9	T	-1.110264931
21	N39	H	-0.672257512
22	E517	R	-0.578694504
23	E78	Y	-0.244509768
24	E24	-	0
25	K499	-	0
26	E513	-	0
27	K336	-	0
28	D439	-	0
29	L384	-	0
30	D434	-	0
31	K362	-	0
32	D501	-	0
33	R341	-	0
34	E429	-	0

Table 5.2: Mutation outputs from sequence design on iNicSnFR3a on 34 previously-identified positions (ranked by free energy gain).

BIBLIOGRAPHY

Arnam, E.B.V., Blythe, E.E., Lester, H.A., Dougherty, D.A., 2013. An Unusual Pattern of Ligand-Receptor Interactions for the α7 Nicotinic Acetylcholine Receptor, with Implications for the Binding of Varenicline. Mol Pharmacol 84, 201–207.

Benowitz, N.L., Porchet, H., Sheiner, L., Jacob III, P., 1988. Nicotine absorption and cardiovascular effects with smokeless tobacco use: Comparison with cigarettes and nicotine gum. Clinical Pharmacology & Therapeutics 44, 23–28.

Bligh, J., 1952. The level of free choline in plasma. J Physiol 117, 234–240.

Brejc, K., van Dijk, W.J., Klaassen, R.V., Schuurmans, M., van der Oost, J., Smit, A.B., Sixma, T.K., 2001. Crystal structure of an ACh-binding protein reveals the ligand-binding domain of nicotinic receptors. Nature 411, 269–276.

Carrithers, M.D., Lerner, M.R., 1996. Synthesis and characterization of bivalent peptide ligands targeted to G-protein-coupled receptors. Chemistry & Biology 3, 537–542.

Chen, X., Schauder, S., Potier, N., Van Dorsselaer, A., Pelczer, I., Bassler, B.L., Hughson, F.M., 2002. Structural identification of a bacterial quorum-sensing signal containing boron. Nature 415, 545–549.

Feng, J., Zhang, C., Lischinsky, J.E., Jing, M., Zhou, J., Wang, H., Zhang, Y., Dong, A., Wu, Z., Wu, H., Chen, W., Zhang, P., Zou, J., Hires, S.A., Zhu, J.J., Cui, G., Lin, D., Du, J., Li, Y., 2019. A Genetically Encoded Fluorescent Sensor for Rapid and Specific In vivo Detection of Norepinephrine. Neuron 102, 745-761.e8.

Henderson, B.J., Lester, H.A., 2015. Inside-out neuropharmacology of nicotinic drugs. Neuropharmacology, The Nicotinic Acetylcholine Receptor: From Molecular Biology to Cognition 96, 178–193.

Hollenstein, K., Frei, D.C., Locher, K.P., 2007. Structure of an ABC transporter in complex with its binding protein. Nature 446, 213–216.

Hukkanen, J., Jacob, P., Benowitz, N.L., 2005. Metabolism and Disposition Kinetics of Nicotine. Pharmacol Rev 57, 79–115.

Jing, M., Li, Yuexuan, Zeng, J., Huang, P., Skirzewski, M., Kljakic, O., Peng, W., Qian, T., Tan, K., Zou, J., Trinh, S., Wu, R., Zhang, S., Pan, S., Hires, S.A., Xu, M., Li, H., Saksida, L.M., Prado, V.F., Bussey, T.J., Prado, M.A.M., Chen, L., Cheng, H., Li, Yulong, 2020. An optimized acetylcholine sensor for monitoring in vivo cholinergic activity. Nat Methods 17, 1139–1146.

Kille, S., Acevedo-Rocha, C.G., Parra, L.P., Zhang, Z.-G., Opperman, D.J., Reetz, M.T., Acevedo, J.P., 2013. Reducing Codon Redundancy and Screening Effort of Combinatorial Protein Libraries Created by Saturation Mutagenesis. ACS Synth. Biol. 2, 83–92.

Klein, J., Köppen, A., Löffelholz, K., Schmitthenner, J., 1992. Uptake and Metabolism of Choline by Rat Brain After Acute Choline Administration. Journal of Neurochemistry 58, 870–876.

Kuryatov, A., Luo, J., Cooper, J., Lindstrom, J., 2005. Nicotine Acts as a Pharmacological Chaperone to Up-Regulate Human α4β2 Acetylcholine Receptors. Mol Pharmacol 68, 1839–1851.

Lester, H.A., Xiao, C., Srinivasan, R., Son, C.D., Miwa, J., Pantoja, R., Banghart, M.R., Dougherty, D.A., Goate, A.M., Wang, J.C., 2009. Nicotine is a Selective Pharmacological Chaperone of Acetylcholine Receptor Number and Stoichiometry. Implications for Drug Discovery. AAPS J 11, 167–177.

Lobas, M.A., Tao, R., Nagai, J., Kronschläger, M.T., Borden, P.M., Marvin, J.S., Looger, L.L., Khakh, B.S., 2019. A genetically encoded single-wavelength sensor for imaging cytosolic and cell surface ATP. Nat Commun 10, 711.

Marvin, J.S., Borghuis, B.G., Tian, L., Cichon, J., Harnett, M.T., Akerboom, J., Gordus, A., Renninger, S.L., Chen, T.-W., Bargmann, C.I., Orger, M.B., Schreiter, E.R., Demb, J.B., Gan, W.-B., Hires, S.A., Looger, L.L., 2013. An optimized fluorescent probe for visualizing glutamate neurotransmission. Nat Methods 10, 162–170.

Marvin, J.S., Scholl, B., Wilson, D.E., Podgorski, K., Kazemipour, A., Müller, J.A., Schoch, S., Quiroz, F.J.U., Rebola, N., Bao, H., Little, J.P., Tkachuk, A.N., Cai, E., Hantman, A.W., Wang, S.S.-H., DePiero, V.J., Borghuis, B.G., Chapman, E.R., Dietrich, D., DiGregorio, D.A., Fitzpatrick, D., Looger, L.L., 2018. Stability,

affinity, and chromatic variants of the glutamate sensor iGluSnFR. Nat Methods 15, 936–939.

Marvin, J.S., Schreiter, E.R., Echevarría, I.M., Looger, L.L., 2011. A genetically encoded, high-signal-to-noise maltose sensor. Proteins: Structure, Function, and Bioinformatics 79, 3025–3036.

Marvin, J.S., Shimoda, Y., Magloire, V., Leite, M., Kawashima, T., Jensen, T.P., Kolb, I., Knott, E.L., Novak, O., Podgorski, K., Leidenheimer, N.J., Rusakov, D.A., Ahrens, M.B., Kullmann, D.M., Looger, L.L., 2019. A genetically encoded fluorescent sensor for in vivo imaging of GABA. Nat Methods 16, 763–770.

Miwa, J.M., Freedman, R., Lester, H.A., 2011. Neural Systems Governed by Nicotinic Acetylcholine Receptors: Emerging Hypotheses. Neuron 70, 20–33.

Morales-Perez, C.L., Noviello, C.M., Hibbs, R.E., 2016. X-ray structure of the human α4β2 nicotinic receptor. Nature 538, 411–415.

Nees, F., 2015. The nicotinic cholinergic system function in the human brain. Neuropharmacology, The Nicotinic Acetylcholine Receptor: From Molecular Biology to Cognition 96, 289–301.

Nichols, A.L., Blumenfeld, Z., Fan, C., Luebbert, L., Blom, A.E., Cohen, B.N., Marvin, J.S., Borden, P.M., Kim, C.H., Muthusamy, A.K., Shivange, A.V., Knox, H.J., Campello, H.R., Wang, J.H., Dougherty, D.A., Looger, L.L., Gallagher, T., Rees, D.C., Lester, H.A., 2022. Fluorescence activation mechanism and imaging of drug permeation with new sensors for smoking-cessation ligands. eLife 11, e74648.

Packer, M.S., Liu, D.R., 2015. Methods for the directed evolution of proteins. Nat Rev Genet 16, 379–394.

Patriarchi, T., Cho, J.R., Merten, K., Howe, M.W., Marley, A., Xiong, W.-H., Folk, R.W., Broussard, G.J., Liang, R., Jang, M.J., Zhong, H., Dombeck, D., von Zastrow, M., Nimmerjahn, A., Gradinaru, V., Williams, J.T., Tian, L., 2018. Ultrafast neuronal imaging of dopamine dynamics with designed genetically encoded sensors. Science 360, eaat4422.

Quiocho, F.A., Ledvina, P.S., 1996. Atomic structure and specificity of bacterial periplasmic receptors for active transport and chemotaxis: variation of common themes. Molecular Microbiology 20, 17–25.

Sallette, J., Pons, S., Devillers-Thiery, A., Soudant, M., Prado de Carvalho, L., Changeux, J.-P., Corringer, P.J., 2005. Nicotine Upregulates Its Own Receptors through Enhanced Intracellular Maturation. Neuron 46, 595–607.

Schiefner, A., Breed, J., Bösser, L., Kneip, S., Gade, J., Holtmann, G., Diederichs, K., Welte, W., Bremer, E., 2004. Cation-π Interactions as Determinants for Binding of the Compatible Solutes Glycine Betaine and Proline Betaine by the Periplasmic Ligand-binding Protein ProX from Escherichia coli*. Journal of Biological Chemistry 279, 5588–5596.

Shilton, B.H., Flocco, M.M., Nilsson, M., Mowbray, S.L., 1996. Conformational Changes of Three Periplasmic Receptors for Bacterial Chemotaxis and Transport: The Maltose-, Glucose/Galactose- and Ribose-binding Proteins. Journal of Molecular Biology 264, 350–363.

Shivange, A.V., Borden, P.M., Muthusamy, A.K., Nichols, A.L., Bera, K., Bao, H., Bishara, I., Jeon, J., Mulcahy, M.J., Cohen, B., O'Riordan, S.L., Kim, C., Dougherty, D.A., Chapman, E.R., Marvin, J.S., Looger, L.L., Lester, H.A., 2019. Determining the pharmacokinetics of nicotinic drugs in the endoplasmic reticulum using biosensors. Journal of General Physiology 151, 738–757.

Studier, F.W., 2005. Protein production by auto-induction in high-density shaking cultures. Protein Expression and Purification 41, 207–234.

Sun, F., Zeng, J., Jing, M., Zhou, J., Feng, J., Owen, S.F., Luo, Y., Li, F., Wang, H., Yamaguchi, T., Yong, Z., Gao, Y., Peng, W., Wang, L., Zhang, S., Du, J., Lin, D., Xu, M., Kreitzer, A.C., Cui, G., Li, Y., 2018. A Genetically Encoded Fluorescent Sensor Enables Rapid and Specific Detection of Dopamine in Flies, Fish, and Mice. Cell 174, 481-496.e19.

Surin, B.P., Rosenberg, H., Cox, G.B., 1985. Phosphate-specific transport system of Escherichia coli: nucleotide sequence and gene-polypeptide relationships. Journal of Bacteriology 161, 189–198.

Tam, R., Saier, M.H., 1993. Structural, functional, and evolutionary relationships among extracellular solute-binding receptors of bacteria. Microbiological Reviews 57, 320–346.

Trott, O., Olson, A.J., 2010. AutoDock Vina: Improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. Journal of Computational Chemistry 31, 455–461.

Unger, E.K., Keller, J.P., Altermatt, M., Liang, R., Matsui, A., Dong, C., Hon, O.J., Yao, Z., Sun, J., Banala, S., Flanigan, M.E., Jaffe, D.A., Hartanto, S., Carlen, J., Mizuno, G.O., Borden, P.M., Shivange, A.V., Cameron, L.P., Sinning, S., Underhill, S.M., Olson, D.E., Amara, S.G., Temple Lang, D., Rudnick, G., Marvin, J.S., Lavis, L.D., Lester, H.A., Alvarez, V.A., Fisher, A.J., Prescher, J.A., Kash, T.L., Yarov-Yarovoy, V., Gradinaru, V., Looger, L.L., Tian, L., 2020. Directed Evolution of a Selective and Sensitive Serotonin Sensor via Machine Learning. Cell 183, 1986-2002.e26.

Zhong, W., Gallivan, J.P., Zhang, Y., Li, L., Lester, H.A., Dougherty, D.A., 1998. From ab initio quantum mechanics to molecular neurobiology: A cation–π binding site in the nicotinic receptor. Proceedings of the National Academy of Sciences 95, 12088–12093.

Chapter 6

CONCLUSIONS AND FUTURE DIRECTIONS

The present work represents only a first step toward our long-term goal of developing a DNA-guided genome editing system that functions robustly in prokaryotes and eukaryotes. Here, I would like to summarize the progress we have made and discuss future directions. We have demonstrated that CbAgo, the pAgo from *Clostridium butyricum*, can cleave *E. coli* chromosomes and introduce genomic mutations, by either directly leading to chromosome recombination or serving as a counter-selection to assist Lambda-Red recombineering. To our knowledge, until now there are no existing examples of genetic mutation induced by guide-directed cleavage via pAgos. Our reports demonstrate that the double-stranded break (DSB) introduced by CbAgo is mutagenic and suggests the potential of establishing DNA-directed genome editing systems using pAgos. Additionally, we have created a novel dCbAgo-based deaminase, verified its activity in vivo, and showed that it mediates interesting DNA-directed deamination on genomes.

6.1 DNA-guided genome editing in bacteria

CbAgo alone has been shown to introduce DSBs and subsequent mutations in *E. coli* genomes, directed by plasmid-encoding guide sequence. As described in Chapters 2 and 3, we have observed the CbAgo-induced recombination between direct repeat sequences on chromosomes, and we anticipate that similar homologous recombination could happen between chromosomes and plasmids, which may provide an efficient mutagenesis method to replace genomic genes with mutated ones if the mutated genes are on the plasmid flanked by homologous sequences. It is worth noting that *E. coli* is not a highly recombinogenic species, therefore the plasmid-based homologous recombination may not have satisfying efficiency. However, in highly recombinogenic species such as *Streptococcus pneumoniae*, the aforementioned guide-directed CbAgo cleavage/plasmid-based homologous recombination system is likely to achieve much higher recombination rates.

Besides, CbAgo targeting has been shown able to assist recombineering in RecA-deficient strains and achieved mutation efficiency of ~1%, which may be especially useful in the genome editing for species that are not highly recombinogenic, such as *E. coli*. To further improve the mutation efficiency, engineering pAgo to improve its DNA cleavage activity should be necessary. Another factor that negatively affects the recombineering efficiency is the knockout of RecA, as confirmed by previous reports (Poteete and Fenton, 1993), which can be overcome by temporarily expressing RecA before recombineering. To extend the pAgo-assisted recombineering to RecA-active strains, RecA activity may be temporarily inhibited in the pAgo-eliminating step, via, for example, expressing RecA inhibitors (Moreb et al., 2017). This CbAgo-assisted recombineering can be useful in relatively large (>1000 bp) genomic replacement or deletion. Notably, a recent study also reported NgAgo-assisted recombineering (NgAgo, pAgo from *Natronobacterium gregoryi*), but the fold change was smaller than 2 and the enhancement of editing was not dependent on NgAgo endonuclease activity (Lee et al., 2021).

6.2 DNA-guided genome editing in eukaryotes

The broader impact will be achieved if pAgo-induced genome editing can be successfully applied in eukaryotes and especially in mammalian cells. However, efficiently unwinding chromatin DNAs should be the bottleneck of this technique because pAgo does not have a helicase domain and chromatinized DNA is resistant to pAgo targeting or cleavage (Zander et al., 2017). Interestingly, a patent from Dr. George Church's group has shown that TtAgo, in the presence of appropriate gDNAs, can induce target gene activation or suppression at AT-rich regions in human cells, and these effects can be enhanced when dCas9 is co-expressed and bound to the neighboring region (Church et al., 2017). This observation suggests that local unwinding of mammalian chromosomes (in AT-regions or adjacent to dCas9 binding sites) is sufficient for pAgo targeting and functioning. Therefore, the development and use of more potent DNA unwinding/strand displacing enzymes or reagents, such as peptide nucleic acids (PNAs), may eventually make the pAgo-mediated mammalian genome editing come true.

Besides, we have demonstrated the fusion of cytidine deaminase and dCbAgo has a stronger deamination activity in vivo compared to the cytidine deaminase itself, indicating the DNA binding by dCbAgo moiety facilitates the deamination by cytidine deaminase. We anticipate that such a dCbAgo-based deaminase could have interesting applications in mammalian cell genome editing if the aforementioned DNA unwinding issue could be readily solved by, for example, the usage of PNAs (Figure 6.1). This will make a novel base editing strategy that does not have any sequence requirement and can theoretically target any genomic region. Current CRISPR-based base editors have been demonstrated successful for many genome positions in a variety of cell types (Rees and Liu, 2018). However, their requirement of properly positioned PAM sequences limits its flexibility, especially when the editing window is smaller than 5 nucleotides in some base editor variants.

6.3 Exploration of dCbAgo-mediated mechanism

The mechanism of guide-directed recombination by dCbAgo in our system remains unknown, although this pathway appears to be independent of RecBCD and DSB. The initial model is that the dCbAgo works similar to CbAgo in the sense that it acquires small gDNAs from plasmids in vivo, binds to the chromosome (preferentially the lagging strand during replication), and induces recombination by recruiting unknown *E. coli* nucleases or recombinases. This speculation is supported by findings in other pAgo research (Fu et al., 2019; Jolly et al., 2020) and the fact that many pAgo genes have been found associated with a variety of genes including nuclease and helicase (Makarova et al., 2009; Swarts et al., 2014).

However, this model cannot explain our observation in Chapter 4 very well. When the fusion protein rApo1-XTEN-dCbAgo was expressed in *E. coli* in the presence of plasmid-encoded guide sequences, which is a 1000-bp DNA sequence homologous to the lacZ gene, the major mutations were observed within the 2000-bp regions that are direct downstream or upstream of the targeted region while the number of mutations inside the targeted 1000-bp region was minimal. This is quite against our expectation because if the dCbAgo moiety

were to use small gDNAs and bind to ssDNA of chromosomes, mutations should be distributed across the target region with a slightly extended boundary (less than 100 bp). This is because the XTEN linker between the rApo1 moiety and dCbAgo moiety is only 16 amino acids long, therefore the position where the deamination happens should be in close distance to the dCbAgo binding site. For comparison, the rApo1-XTEN-dCas9 deaminase can only efficiently generate C to T mutation to the positions within 19 nucleotides from PAM, which is the dCas9 binding site (Komor et al., 2016).

One explanation for the mutations introduced by rApo1-XTEN-dCbAgo that are far away (in a distance >500 bp) from the target region can be the following. The assumption is there is some extensive but unknown DNA interaction between the guide sequence encoded plasmid and the chromosome. This interaction could be the well-characterized Holliday junction generated during homologous recombination. Since the guide sequence encoded plasmid has 1000-bp homology to the chromosome (which is the guide sequence itself), the homologous recombination between it and the chromosome should be frequent (Shen and Huang, 1986). In the case of the Holliday junction, ssDNAs are generated outside the homology region like bridges connecting chromosome DNA and plasmid DNA, which can be good targets for dCbAgo binding and deamination.

If such DNA interaction exists, our results described in Chapter 3 will suggest the following model. The dCbAgo can interact with the chromosome-plasmid complex and induce homologous recombination ~6 kb away. On the other hand, RecBCD, the *E. coli* endogenous exonuclease and helicase, can resolve this complex. The prolonged effect of the dCbAgo interaction with the complex can be detrimental to the host, as supported by the observation that the *E. coli* RecBCD knockout strain suffered from severely inhibited cell growth in the presence of dCbAgo and guide sequence.

As most of the above arguments are merely speculation, more research about dCbAgo activity in vivo is required to clarify its role. More data about the dCbAgo mediated mechanism may help us understand the functions of those naturally occurring pAgos that are unable to cleave nucleic acids. Combined with the modern DNA sequencing technique,

the rApo1-XTEN-dCbAgo described here could serve as a probe to reveal the binding of dCbAgo in vivo.

6.4 In vitro application of pAgo

The nature of pAgo nuclease that it uses small gDNAs and cleaves target ssDNA substrates without any sequence requirement suggests that pAgo may serve as a highly versatile restriction enzyme in vitro. The pAgo-based artificial restriction enzyme (ARE) may come with the capability of targeting theoretically any DNA sequence using small oligonucleotides as guides. As briefly discussed in Chapter 1, PfAgo has been used as an ARE to develop an automatic molecular cloning platform termed PlasmidMaker and create more than 100 plasmids from different DNA fragments (Enghiad and Zhao, 2017; Enghiad et al., 2022). However, even the most active pAgo nuclease identified so far, which is CbAgo (Vaiskunaite et al., 2022), has a DNA cleavage activity that is ~10-fold lower than normal restriction endonucleases (Kuzmenko et al., 2019). Therefore, engineering existing pAgos for improved DNA cleavage activity may not only be critical for pAgo in vivo application in genome editing but also be very important in realizing its in vitro potential as a powerful and flexible ARE.

6.5 Engineering pAgo DNA cleavage activity

Since we observed a strong correlation between in vitro pAgo cleavage activity and in vivo recombination rate, as described in Chapter 2, our recombination system presented here can potentially serve as a reporter and selection platform to engineer pAgo with improved DNA cleavage activity. Our recombination system links pAgo activity to the development of antibiotic resistance so that cells expressing pAgo mutants with enhanced DNA cleavage activity are likely to exhibit higher survival rates and be selected under antibiotic selection.

In addition, phage-assisted evolution of pAgo may be another interesting and feasible direction to engineer pAgo cleavage activity. This is because the previous report has demonstrated that CbAgo can efficiently help its *E. coli* expression host defend against phage attack and significantly increases its growth rate during phage infection (Kuzmenko et al., 2020). This defense is likely due to the spontaneous gDNA loading of CbAgo from

phage DNAs and the subsequent guide-directed cleavage. Therefore, cells expressing pAgo mutants with enhanced DNA cleavage activity are likely to exhibit stronger resistance against phage attack and be selected.

Moreover, as we have demonstrated in previous chapters, CbAgo can efficiently acquire gDNAs from plasmid-encoded guide sequences and use them as guides to cleave DNAs. We hypothesize that when the plasmid-encoded guide sequence is homologous to phage DNAs, this CbAgo-mediated defense could be further strengthened and fine-tuned by changing the length of the guide sequence. This feature may facilitate the development of the phage selection assay proposed above.
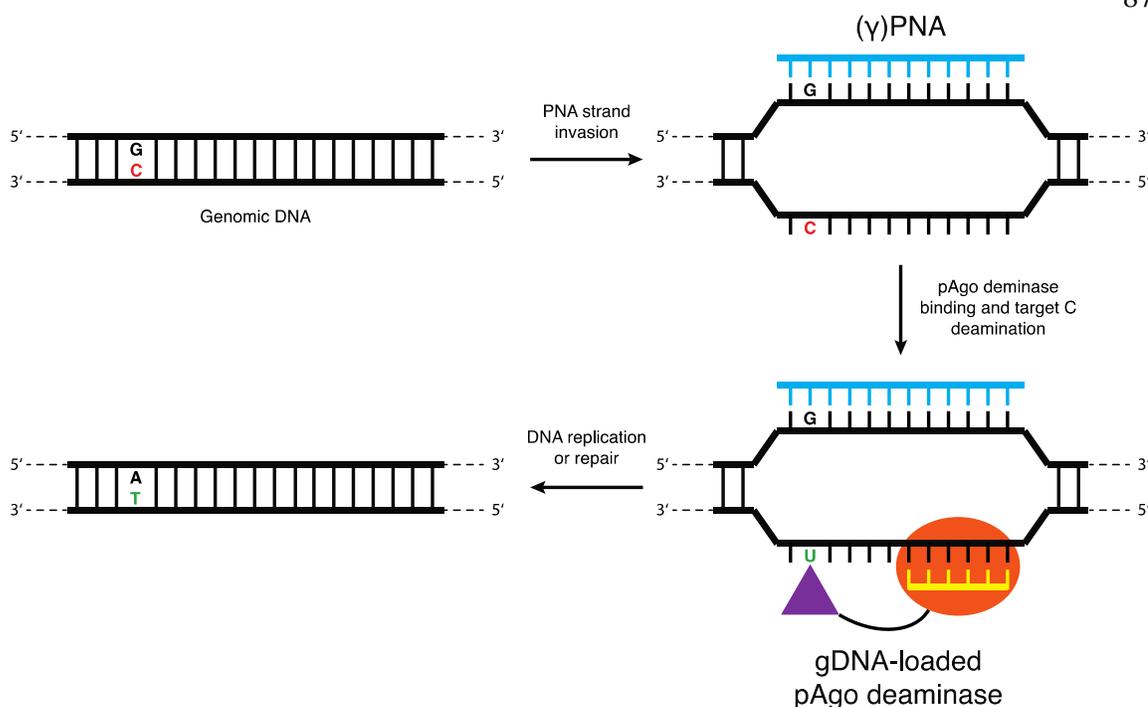
Figure 6.1: Proposed mechanism of PNA-assisted deamination. The bottom strand with a target C (red) is the target strand and is displaced by a specific PNA or γ-substituted PNA (blue), which is complementary to the non-target strand, via strand invasion. The gDNA-loaded pAgo moiety (orange) binds to the displaced target strand and direct the tethered CDA (purple) to the target C. Cytidine deamination converts the single-stranded target C to U and the resulting G:U pair can be converted to an A:T base pair during DNA replication or DNA repair.

BIBLIOGRAPHY

Church, G.M., Yang, L., MONROE, M.R., Huang, P.-Y., 2017. DNA-guided gene editing and regulation. WO2017139264A1. https://patents.google.com/patent/WO2017139264A1/en

Enghiad, B., Xue, P., Singh, N., Boob, A.G., Shi, C., Petrov, V.A., Liu, R., Peri, S.S., Lane, S.T., Gaither, E.D., Zhao, H., 2022. PlasmidMaker is a versatile, automated, and high throughput end-to-end platform for plasmid construction. Nat Commun 13, 2697.

Enghiad, B., Zhao, H., 2017a. Programmable DNA-Guided Artificial Restriction Enzymes. ACS Synth. Biol. 6, 752–757.

Enghiad, B., Zhao, H., 2017b. Programmable DNA-Guided Artificial Restriction Enzymes. ACS Synth. Biol. 6, 752–757.

Fu, L., Xie, C., Jin, Z., Tu, Z., Han, L., Jin, M., Xiang, Y., Zhang, A., 2019. The prokaryotic Argonaute proteins enhance homology sequence-directed recombination in bacteria. Nucleic Acids Research 47, 3568–3579.

Jolly, S.M., Gainetdinov, I., Jouravleva, K., Zhang, H., Strittmatter, L., Bailey, S.M., Hendricks, G.M., Dhabaria, A., Ueberheide, B., Zamore, P.D., 2020. Thermus thermophilus Argonaute Functions in the Completion of DNA Replication. Cell 182, 1545-1559.e18.

Komor, A.C., Kim, Y.B., Packer, M.S., Zuris, J.A., Liu, D.R., 2016. Programmable editing of a target base in genomic DNA without double-stranded DNA cleavage. Nature 533, 420–424.

Kuzmenko, A., Oguienko, A., Esyunina, D., Yudin, D., Petrova, M., Kudinova, A., Maslova, O., Ninova, M., Ryazansky, S., Leach, D., Aravin, A.A., Kulbachinskiy, A., 2020. DNA targeting and interference by a bacterial Argonaute nuclease. Nature 587, 632–637.

Kuzmenko, A., Yudin, D., Ryazansky, S., Kulbachinskiy, A., Aravin, A.A., 2019. Programmable DNA cleavage by Ago nucleases from mesophilic bacteria

Clostridium butyricum and Limnothrix rosea. Nucleic Acids Research 47, 5822–5836.

Lee, K.Z., Mechikoff, M.A., Kikla, A., Liu, A., Pandolfi, P., Fitzgerald, K., Gimble, F.S., Solomon, K.V., 2021. NgAgo possesses guided DNA nicking activity. Nucleic Acids Research 49, 9926–9937.

Moreb, E.A., Hoover, B., Yaseen, A., Valyasevi, N., Roecker, Z., Menacho-Melgar, R., Lynch, M.D., 2017. Managing the SOS Response for Enhanced CRISPR-Cas-Based Recombineering in E. coli through Transient Inhibition of Host RecA Activity. ACS Synth. Biol. 6, 2209–2218.

Poteete, A.R., Fenton, A.C., 1993. Efficient double-strand break-stimulated recombination promoted by the general recombination systems of phages lambda and P22. Genetics 134, 1013–1021.

Rees, H.A., Liu, D.R., 2018. Base editing: precision chemistry on the genome and transcriptome of living cells. Nat Rev Genet 19, 770–788.

Shen, P., Huang, H.V., 1986. HOMOLOGOUS RECOMBINATION IN ESCHERICHIA COLI: DEPENDENCE ON SUBSTRATE LENGTH AND HOMOLOGY. Genetics 112, 441–457.

Zander, A., Willkomm, S., Ofer, S., van Wolferen, M., Egert, L., Buchmeier, S., Stöckl, S., Tinnefeld, P., Schneider, S., Klingl, A., Albers, S.-V., Werner, F., Grohmann, D., 2017. Guide-independent DNA cleavage by archaeal Argonaute from Methanocaldococcus jannaschii. Nat Microbiol 2, 1–10.