Prokaryotic Argonautes for *In Vivo* Biotechnology and Molecular Diagnostics

Brett A Graver[1†] , Namrata Chakravarty[2†], Kevin V Solomon[2*]

[1] Department of Biological Sciences, University of Delaware, Newark, DE, 19716 USA

[2] Department of Chemical and Biomolecular Engineering, University of Delaware, Newark, DE, 19716 USA

[†]Equal contribution

*Correspondence: Kevin Solomon - kvs@udel.edu

Twitter Handles: Brett Graver - @graver_brett; Namrata Chakravarty - @NamrataChakr; Kevin Solomon - @drksolomon

ORCIDs: Brett Graver - 0009-0004-0673-6550; Namrata Chakravarty - 0000-0002-0720-110X ; Kevin Solomon - 0000-0003-2904-9118

Abstract

Prokaryotic Argonautes (pAgos) are an emerging class of programmable endonucleases that are believed to be more flexible than existing CRISPR-Cas systems with significant potential for biotechnology. Current applications of pAgos include a myriad of molecular diagnostics and *in vitro* DNA assembly tools. However, efforts have historically been centered on thermophilic pAgo variants. To enable *in vivo* biotechnological applications such as gene editing, focus has shifted to pAgos originating from mesophilic organisms. In this review, we discuss what is known of pAgos, how they are being developed for various applications and outline strategies to overcome current challenges to *in vivo* applications in prokaryotes and eukaryotes.

Keywords : Prokaryotic Argonaute, Gene Editing, Molecular Diagnostics, Mesophile

Endonucleases enable biotechnology

Biotechnology, at is core, is driven by a suite of DNA endonucleases that enable the manipulation of DNA sequences for diverse applications including large-scale production of recombinant therapeutic proteins, genetically modified crops, and sustainable fuel production [1–4]. The first DNA endonucleases to be developed were bacterial restriction enzymes that cleave DNA within or near specific DNA sequences (restriction sites) as a defense mechanism against invading DNA [5]. These enzymes have proved indispensable to genetic engineering and ushered in a new era of biotechnology by enabling recombinant DNA technology, genome mapping, and DNA sequencing among other applications [6]. Restriction enzymes recognize short sequence motifs (~ 6 nts) that are ubiquitous within genomes, restricting their use for precise *in vivo* genetic manipulations. Subsequent invention and discovery of **meganucleases** (see glossary) and **programmable endonucleases** with unique cleavage targets (~ 20 nts) introduced precise and targeted genetic modifications that facilitate the understanding of genome function, development of mutant or improved strains, and treatment of genetic diseases [7]. Among these are **homing endonucleases** such as I-SceI, synthetic zinc-finger and transcription-activator-like effector nucleases (**ZFN** and **TALENs**), and **Cas** nucleases, which exhibit diverse variability in target DNA recognition and their cleavage mechanisms [8]. However, these enzymes either have fixed sequence motifs that limit target selection or require nontrivial protein engineering to reprogram [9–12]. Even Cas nucleases, which can be easily reprogrammed via RNA guides, must bind adjacent to a protospacer adjacent motif sequence (**PAM**) that is typically 2-7 nts in length to recognize and cleave double stranded DNA [12]. Thus, there is a search for highly-specific

programmable DNA endonucleases that are more flexible with regards to choice of DNA target (e.g. PAM-less Cas endonucleases) [12,13].

Towards that end, there has been growing interest in prokaryotic Argonautes (**pAgos**) native to bacteria and archaea, which have the potential to recognize and cleave RNA and/or DNA targets [14,15]. Like Cas nucleases, these enzymes are easily programmed by short single stranded nucleic acid guides complementary to their target. However, they are believed to be more flexible than Cas nucleases as they are not known to have sequence restrictions such as PAMs, expanding the manipulation options for DNA regions with few or sparsely distributed PAMs. Moreover, pAgos offer several additional advantages that may overcome bottlenecks in delivery: 1) the majority of pAgos use ssDNA guides for targeting [14], which are inherently more stable than the RNA guides of Cas nucleases and may enhance the efficiency of nucleoprotein complex delivery; and 2) are more compact in size than Cas9, somewhat alleviating the capacity constraints of viral delivery vectors [16]. Finally, pAgos are relatively unencumbered with intellectual property claims and, in the near future, may be more economical to license and use for product development and manufacturing. Thus, pAgos are rapidly being characterized and developed for a number of biotechnologies. In this review, we discuss the properties of pAgo proteins, emerging techniques to elucidate their function, and strategies to develop them for synthetic biology applications (Figure 1, Key Figure). We focus on mesophilic pAgo candidates with potential for gene editing and provide a perspective on critical next steps for their development while outlining emerging high-throughput tools for the discovery and characterization of these systems.

Prokaryotic Argonautes: Programmable DNA Endonucleases

pAgos are nucleic acid-guided endonucleases that can be used as programmable molecular scissors when provided with short oligonucleotide guides, typically 16-21 nts in length [17,18]. Argonautes from different prokaryotic species may use either DNA or RNA guides or both, often phosphorylated or hydroxylated at the 5' end, to target DNA and/or RNA substrates [17,19]. The pairing of the bound guide to its complementary sequence activates the endonuclease activity of pAgo, cleaving the phosphodiester backbone of the substrate [20]. Because pAgos lack helicase function and only have a single active site unlike **CRISPR**-Cas9, they can only cleave one strand of DNA at a time. Double-stranded DNA cleavage necessitates loading pAgos with two complementary guides directed at the cutting site [21–23]. Soon after their discovery in prokaryotes (Box 1), these proteins were studied for their physiological role and cleavage mechanism. This endonuclease class has already found applications in biotechnology for targeted cleavage and detection of nucleic acids and can potentially be used for genome editing [15,24].

pAgos are divided into two major phylogenetic clades, short and long pAgos, which differ predominately in their architecture [14,17]. Long pAgos can be further subdivided into two classes depending on whether they are active or inactive [23]. Long pAgos are composed of 6 domains: N-terminal, L1, **PAZ**, L2, **MID**, and **PIWI** (Figure 2a). The PIWI domain is critical for nucleic acid cleaving activity and coordinates divalent metal cations via a catalytic tetrad of amino acids (DEDX, where X is D, H, or K) that cleaves substrate via nucleophilic attack [17,25–28]. Inactive long pAgos are frequently mutated in this tetrad, which abolishes cleavage activity. Four conserved amino acid residues in the MID domain (Y/R, K, Q, and K) form a charged binding pocket to anchor 5' phosphorylated

guides for target recognition. However, pAgos such as MpAgo and MhAgo, which accept 5’ hydroxylated RNA guides, have a hydrophobic pocket (I, V/I, P, L, Y) for guide binding [29,30]. These structural variations in the MID domain are hypothesized to be responsible for the RNA guide preferences in MpAgo [29]. The PAZ domain grips the 3’ end of the guide while the Argonaute searches for a complementary sequence on the target strand. Once the pAgo-guide complex encounters a complementary target sequence, the 3’ end of the guide is released for complete base pairing between the guide and target (Figure 2b). Consequently, this event activates the pAgo endonuclease activity in the PIWI domain. pAgo endonuclease activity cleaves the phosphodiester backbone generating a cleavage in the target nucleic acid. Subsequently, the target nucleic acid is released from the protein via the N domain [31,32]. L1 and L2 serve as linker sequences and are not known to participate in pAgo activity [20]. pAgos have no known restrictions for targeting DNA and, with its cleavage capabilities, can generate staggered "sticky" or “blunt” ends whereas most CRISPR/Cas systems only produce “blunt” fragments [33].

Short pAgos are comprised of a MID and an inactivated PIWI domain and can be fused with an analogous PAZ domain (APAZ) (Figure 2a) [14]. APAZ domains are homologous to the N domain of long pAgos, not the PAZ domains, and are usually associated with nucleases like **Sir2** and **TIR** [17,25]. As a bacterial defense system, short pAgos can form protein complexes with Sir2 causing the depletion of NAD(P)+ from plasmid or phage invaded bacterial cells, consequently leading to cell death [34]. This review focuses on long pAgos, as they have DNA cleaving activity, but a recent review by Koopal *et. al* (2021) covers short pAgos in more detail [21].

Long pAgos such as that of *Thermus thermophilus* (TtAgo) has been thoroughly investigated and its DNA interference mechanism has been identified [35]. TtAgo has been shown to derive its 5'-phosphorylated DNA guides (13-25 nt) from invading plasmids and neutralizes multi-copy foreign elements by cleaving the complementary strand as a mechanism to defend the host from the foreign DNA via DNA-guided DNA interference [35,36]. While lowering the transcription and copy number of the invading element, TtAgo also triggers CRISPR-Cas genes, representing a functional association between the immune systems in protecting the host cell [28]. Similar associations of pAgo and CRISPR-Cas genes have also been observed in MpAgo [25]. *In vitro,* TtAgo in its apo-form implements a 'chopping' mechanism that leads to random processing of dsDNA into short ssDNA fragments. These fragments are then loaded as functional guides and directs the subsequent targeted cleavage of invader DNA [36]. This mechanism autonomously generates guides and is conserved among pAgos [36–38]. However, cleavage in this apo-form is non-specific and less efficient than the guided-processing of the target. Argonautes can critically differentiate between self and non-self DNA via several determinants such as DNA chromatin state, the AT-biases of the invading DNA, gene copy number, DNA replication frequency, and the presence of **Chi-sites** [17,37,39,40].

Commercializing pAgos

pAgos have been developed as molecular diagnostics with attomolar sensitivities for single-nucleotide polymorphisms in a number of conditions (Figure 1) [41–45]. These

platforms have been extended for detection of pathogens like **SARS-CoV-2** infection and its variant D614 with significantly reduced time yet comparable performance to commercial **RT-qPCR** methods [41,42]. In this scheme, targeted cleavage of reverse transcribed SARS-CoV-2 DNA generated guide for subsequent cleavage of a quenched molecular beacon generating a fluorescent output. However, due to the multiple turnover of pAgos [46], unlike more popular SpCas9, signal amplification arises in part from more rapid enzymatic activity rather than slower PCR amplification in traditional qPCR assays. Similarly, multiplex platforms have been developed to independently and rapidly detect **COVID-19**, influenza and human papilloma viruses simultaneously via a one-pot assay [43,45,47,48]. Finally, thermophilic Argonautes have also been employed for early detection of cancer via discriminated cleavage of healthy alleles allowing for enrichment and subsequent detection of cancer-associated rare mutations [44,49]. In all these diagnostics, thermophilic pAgos have been primarily used and are easier to develop due to the high operating temperatures removing the need for helicases in target recognition and cleavage, which are not native to pAgos and must be supplemented (Box 2). pAgo-based diagnostics boast sensitivities on par with commercial systems, more rapid results as compared to the gold-standard of qPCR, and use more shelf-stable DNA guides for activity.

pAgos have also been developed as programmable DNA-guided artificial restriction enzymes (AREs) for DNA assembly and DNA fingerprinting, and are commercially available [50,51] (e.g. New England Biolabs, Ipswich, MA). These AREs can recognize and cleave DNA at any site programmed via a short DNA oligonucleotide. Thus, they can readily be implemented in DNA assembly to create truly unique overhangs

without the need to remove preexisting restriction sites. Moreover, by staggering the complementarity of provided guides, one can create varying lengths of sticky or blunt ends, making them more robust and efficient than natural restriction enzymes.

Although, a number of CRISPR-Cas and pAgo based *in vitro* detection methods have been developed, *in vivo* detection has yet to be fully explored. *In vitro* detection technologies are certainly rapid, low-cost and high-throughput due to ease of culture handling and control over the physical and chemical environment. However, development of *in vivo* pAgo tools can pave the way for personalized medicine, aid in development of disease models and drug discovery, and enhance crops for nutrition and resilience in the face of climate change (Figure 1).

The search for new pAgos

A handful of pAgos have been well-characterized and crystallized [14]. Notably, most of these enzymes are derived from thermophilic bacteria and archaea such as *Thermus thermophilus* (TtAgo), *Methanocaldococcus jannaschii* (MjAgo), and *Pyrococcus furiosus* (PfAgo) [26,52,53]. However, their high optimal activity temperature from 55 °C up to 99.9 °C prevents them from being used for *in vivo* gene editing in species of interest and restricts their applications in detection-based diagnostic devices [22,26,54,55]. Consequently, there has been a recent acceleration of the discovery of mesophilic candidates that can carry out the cleavage at moderate temperatures and thus can be translated to *in vivo* genome editing and other applications. Of note, lately, there have been reports of new biochemically characterized mesophilic pAgos like CbAgo (*Clostridium butyricum*), CpAgo (*Clostridium perfringen*), SeAgo (*Synechococcus elongatus*), and KmAgo (*Kurthia massiliensis*) (Table 1) that exhibit DNA-guided cleavage

activity. However, their application in gene editing/detection remains to be fully explored. New diagnostics utilizing a mesophilic pAgo demonstrated the detection of SARS-CoV-2 at a moderate temperature of 42 °C [56]. So far, numerous other studied mesophilic pAgos have broadened the understanding of their endonuclease activity that could advance the implementation of *in vivo* genome editing [57–59].

Recently, mesophilic pAgos have shown promise *in vivo* as a platform for gene editing [60–63]. The halophilic *Natronobacterium gregoryi* Argonaute (NgAgo) was first investigated as a mesophilic pAgo for DNA-guided genome editing, albeit sparking controversy in the scientific community due to issues of reproducibility in human cell lines [64–66]. However, NgAgo demonstrated reproducible guided DNA cleavage activity both *in vitro* and *in vivo* in *E. coli* [60] and enhanced homologous recombination-mediated gene insertions in *E. coli* and *Pasteurella multocida* or gene editing [60,61]. While it is unclear whether activity is due solely to NgAgo activity [61], there is evidence to suggest that gene editing is directed by exogenously supplied 5'P ssDNA guides and enhanced due to the specific cleavage activity of pAgo via an intact PIWI domain [60,61]. However, NgAgo is poorly soluble in mesophilic hosts and has an unusual architecture with an additional single stranded DNA binding protein domain making it unsuitable for further development [60]. Emerging reports with CbAgo confirm prokaryotic gene editing capabilities via pAgos [62,63]. More importantly, these reports suggest that guides may be endogenously derived from plasmids with specific features such as high copy number and/or the presence Chi-sites, rather than being supplied, presenting a potential avenue for *in vivo* guide generation. Guide generation and/or delivery may be a limiting factor for efficient gene editing and library generation making reliable methods for *in vivo* guide generation

desirable. However, to date, pAgo-based gene editing has been attempted without success in eukaryotic hosts [15,66].

Currently studied mesophilic pAgos appear to have limitations in guide nucleotide base preferences at certain positions, targetability based on GC content, and optimal catalytic temperature determined from *in vitro* experimentation [57–59,67,68]. The most promising mesophilic Argonaute to date appears to be CbAgo. CbAgo, utilizing a 5'P-DNA guide, has cleavage activity on DNA substrates across a moderate temperature range of 30°C to 50°C with near optimal activity at 37°C in *in vitro* experimentation [67,68]. CbAgo was also demonstrated to cleave DNA in *E. coli* at 37°C [37]. Multiple groups have independently investigated CbAgo and demonstrated that the supplementation of accessory proteins, such as *E.coli* nuclease deficient **helicase recBC**, assists CbAgo in cleavage of linear dsDNA *in vitro* (Box 2) [58,67,68]. However, CbAgo appears to prefer A/T rich DNA regions *in vitro* in the absence of accessory proteins [67,68].

More Optimal pAgos may Exist for *In Vivo* Application

While a handful of mesophilic pAgos have been experimentally validated, there are hundreds of potential pAgos yet to be studied (Figure 2c) [69]. Genome database mining and other bioinformatic efforts have identified dozens of active mesophilic long pAgos that may be suitable candidates for gene editing [58]. To further this search, the AGODB annotation database has been developed to provide the community with domain annotation of known eAgos and pAgos as well an analytical program to predict whether an input sequence is an Ago protein [69]. However, limiting this search to existing sequences may hinder the discovery of the most ideal pAgo. Different approaches such as bioprospecting environmental samples, as seen in industrial and other biotech fields,

may provide pAgos to study which are not readily available on genomic databases [70]. In fact, bioprospecting efforts are being undertaken for the identification of new CRISPR/Cas systems from ecological samples [71]. Environmental candidates, through this bioprospecting effort, can be assayed for mesophilic DNA cleavage by a robust positive selection screen, which does not exist as of today. Strategies like these could be expanded to other endonucleases such as pAgos.

While pAgo guide and substrate preferences need to be experimentally determined, gene neighborhoods and protein structure may provide insight to predict pAgo function. A subset of studied pAgos, MpAgo, TpAgo, and TsAgo, have preference for non-canonical 5'OH-RNA guides to target DNA [29,72,73]. MpAgo and TpAgo are found in an operon with primase, cas1, and cas2 genes, which all uses RNA substrates, immediately upstream of the pAgo gene. While MpAgo and TpAgo have been experimentally validated, a handful of pAgos have been predicted to use this 5'OH-RNA guide based on gene neighborhood assessment [14,29]. Gene neighborhoods surrounding other pAgos, such as CbAgo and NgAgo, are consistent with a DNA substrate preference and hint at potential accessory proteins (e.g. helicases, DNA methyltransferases, DNA recombinases, etc.) utilized by these Argonautes [14]. While the majority of pAgos are known to target DNA, there are few candidates that can precisely cleave both DNA and RNA such as TtAgo and KmAgo [19,52,74]. MbpAgo, from the psychrotolerant *Mucilaginibacter paludism*, preferentially cleaves RNA substrates at mesophilic temperatures *in vitro*, further expanding the applications towards RNA editing [75]. Bioinformatic tools for systematic prediction of functionally linked genes have been developed for bacterial and archaeal genomes, with pAgos being used as a published

dataset with this tool [76]. Structural analysis also provides insight to key residues and protein structure that may dictate a pAgo guide preference. For example, key residue substitutions to more hydrophobic amino acids in the MID domain and α-helix structuring of the PAZ domain steric hindrance in the 5’ phosphate pocket results in pAgo preference for 5’OH-guides [29].

Screening Methods for pAgo Candidates

To date, pAgos are most commonly studied *in vitro* with substrate cleavage assays [57,59,60,67,68]. This has been the primary method of investigation of pAgos uncovering key characteristics including nucleotide guide preferences, 5’ guide modification, and nucleotide substrate preferences [58–60,67,68,75,77]. Additionally, other characteristics such as optimal catalytic temperature, cation requirements and GC content of target can be determined [67,68]. However, the process of generating purified pAgo for *in vitro* cleavage assays can quickly become time consuming when investigating multiple proteins of interest.

A promising alternate strategy for *in vitro* experimentation is cell free expression for more rapid cleavage assessment of pAgos [58,60]. Cell-free protein expression, which forgoes protein purification, has been used to successfully investigate other programmable DNA endonucleases such as CRISPR/Cas systems [78–81]. This system can be used as an expedited preliminary screening method to quickly assess potential pAgo candidates for cleavage activity in a high throughput manner [58]. Cell-free expression pipelines have also been validated to investigate the characteristics of NgAgo function and functional domain mapping [60].

In vivo characterization strategies may allow for the quick screening of potential pAgo candidates for activity in a high throughput manner and allow for analysis of difficult to purify proteins. Moreover, *in vivo* approaches allow for natural interactions between pAgos and accessory protein homologs to ensure function [61,82]. Classically, **CcdB** and **barnase** have been used as counter selection toxins for endonuclease activity and directed evolution *in vivo* [83]. CcdB has been utilized for directed evolution of CRISPR/Cas9 as well as zinc finger nucleases [84–87]. However, CcdB is highly lethal and leaky expression leads to cell death, ultimately reducing sensitivity of the assay [88]. Recently, an *in vivo* I-sceI homing nuclease mediated assay has been demonstrated to assess the activity of **SpCas9** and its variants [89]. Positive selection was also demonstrated with this system with a greater sensitivity to endonuclease activity due to the less toxic I-sceI counterselection as compared to CcdB systems [89]. As of yet, no homing endonuclease counterselection *in vivo* assay has been demonstrated to investigate pAgos. However, solutions developed for Cas nucleases may be leveraged for higher throughput pAgo screening and access to directed evolution techniques.

Mesophilic pAgos can Fulfill the Needs of Biotechnology

As the field moves forward with engineering efforts with the aim of *in vivo* applications utilizing mesophilic pAgos, it will be important to cast as wide of a net as possible when considering candidates. However, this will require the development of high throughput methods to assess the candidates in a timely manner. The role of accessory proteins will also need to be considered. *In vitro* supplementation of helicases and Rec proteins enhance cleavage by pAgos raising the question of the role of these and other accessory proteins effect on pAgo function [58,90]. When assessing these accessory

proteins, effective delivery strategies for *in vivo* application will need to be considered. Design strategies may include the use of pAgo fusion proteins by engineering additional domains onto the endonuclease, similar to the native repA domain found on the N terminus of NgAgo and other halophilic Argonautes [14,60]. Alternatively, co-expression of a pAgo with necessary accessory proteins may be explored.

Another consideration would be nucleotide guide generation to target the pAgo to a correct location. To date, transformation has been successfully used to deliver guides for Argonaute-mediated cleavage in *E. coli*, however, new systems will need to be engineered to generate guides *in vivo* for eukaryotic application. RNA-guided active long pAgos would greatly simplify guide delivery; however, none have been discovered as of yet. Thus, efficient methods for ssDNA guide generation and delivery must be developed. **Apo-pAgos** generate their own guide from cleaving highly replicating plasmid DNA [37,62,63], which may also be exploited for gene editing.

Concluding Remarks

Mesophilic pAgos have significant potential for biotechnology and offer potential competitive advantages in flexibility of target selection, efficiency of cellular delivery, and fewer legal risks with regards to intellectual property (Figure 1). These advantages are already being exploited with thermophilic pAgos in a multitude of *in vitro* molecular diagnostics where specific nucleotide recognition is required [41,91]. However, to fully realize the potential of pAgos including *in vivo* applications, efficient mesophilic pAgos that function at moderate temperatures are needed and critical details of pAgo function must be elucidated (see Outstanding Questions). The emerging consensus is that pAgos, like their eukaryotic cousins, function *in vivo* in a multiprotein complex (Box 2) that

remains to be completely identified. Nonetheless, key components such as helicases that interact with long active pAgos have been recognized that are critical to function and significantly impact pAgo properties such as substrate topology preferences, DNA/RNA and sequence preferences, and optimal operating temperatures. Thus, a key challenge to identifying pAgos with optimal properties for development is elucidation of this complex to more accurately characterize pAgo properties and/or the creation of high throughput *in vivo* screens to more rapidly evaluate candidates where heterologous proteins naturally form these complexes with endogenous bacterial proteins [61,82]. The lack of reports of pAgo activity in eukaryotic systems may be due in part to incomplete complex formation, which may need to be reconstituted/engineered for proper function. Finally, an open question in the field is how best to efficiently deliver or generate ssDNA guides for pAgo targeting, which will be critical to scaling pAgo technologies to large genome scale libraries.

Having overcome these hurdles, pAgos would be ripe for development for a number of tools beyond simple DNA cleavage (Figure 3). Like CRISPR interference and activation, catalytic dead pAgos could be employed for gene expression regulation, epigenomic editing, and *in vivo* imaging [92–94]. However, unlike CRISPR-Cas, the lack of a PAM restriction would enable more precise binding of pAgos for more refined and potentially improved control of cellular phenotypes. Similarly, pAgo-based prime editors could be more streamlined as the ssDNA guide may remove the need for a bulky reverse transcriptase fusion. Ultimately, gene editing mediated by pAgo cleavage events would provide the field with a powerful alternative to CRISPR-Cas systems and have a significant scientific impact in several areas like cancer research, the generation of animal

models for human diseases, and crop development. Thus, pAgos are poised to revolutionize biotechnology with the significant potential for the creation of cross-cutting engineering tools and solving grand challenges in a number of fields.

Box 1. The native role of pAgos

The mechanism that led to co-suppression of chalcone synthase protein (CHS) in *Petunia hybrida* plants remained elusive until the discovery of RNA interference (RNAi) in *Caenorhabditis elegans* in 1998 [95–97]. This prompted studies revealing the role of Argonaute and associated RISC proteins in post-transcriptional gene regulation, anti-viral defense, and mobility of transposons in eukaryotes [98,99]. Two years later, homologous Argonaute proteins were identified in prokaryotes (pAgos) and determined to have a role in bacterial immune systems [25,100,101].

While pAgos are similar to eukaryotic Argonautes (eAgos), they have a greater range of catalytic and biological functions. pAgo proteins are known to have divergent nucleic acid interaction domains, which indicates their unusual specificity for recognizing and cleaving nucleic acids [14]. These proteins facilitate DNA- and/or RNA-guided DNA interference using single-stranded oligonucleotide guides or participate in RNA interference, unlike eAgos that strictly operate RNA-guided RNA interference. While eAgos assemble into a multiprotein complex with Dicer, various non-coding RNAs and other proteins to form RISC complex (small RNA-induced silencing complexes) for function [102], analogous complexes with long pAgos remain to be identified.

Box 2. The Role of Accessory Proteins

Gene editing in eukaryotic organisms is a heavily desired technology where the potential of mesophilic pAgos may be applied. However, attempts at gene editing with NgAgo in HEK293T cell lines have not seen reproducible success, raising the question if pAgos would be able to work effectively in eukaryotes [64,65,103]. Cleavage of dsDNA requires target recognition by an endonuclease via DNA duplex melting, which is achieved in other systems such as CRISPR-Cas and restriction enzymes by integrated helicase activity or by thermal energy released by enzyme-DNA interactions at nucleotide motifs (PAM site for Cas complexes) [14,17,25,39,104,105]. To date, pAgos have not been recognized to have a required nucleotide motif for recognition nor do they possess helicase activity. However, NgAgo and a small collection of halophilic pAgos have been found to have an additional N-terminal single stranded DNA binding (ssb) protein domain, repA [14,60]. This may hint at the need for additional domains or interacting *accessory* proteins, such as ssbs, helicases and DNA interacting proteins, which may be necessary for *in vivo* pAgo function.

Investigation of local gene neighborhoods surrounding pAgo genes in sequenced genomes supports this hypothesis as they also contain DNA interacting genes and potential pAgo complex partners, such as helicases, methylases, ssbs, and nucleases (Figure I) [14]. Moreover, the substrates of these neighboring genes may be indicative of the DNA/RNA substrate preference of the pAgo. pAgos have been demonstrated to copurify with accessory proteins in heterologous hosts [61,82]. *in vitro* pAgo activity assays supplemented with accessory proteins such as ssb and helicase enhances catalysis and removes target sequence preferences, suggesting a critical role for these

proteins [58,90]. The ability of heterologous bacterial hosts to copurify with endogenous homologs of these accessory proteins, may also help explain the validated *in vivo* activity of pAgos such as NgAgo in non-native prokaryotic systems, which remains undetected in eukaryotes or traditional *in vitro* assays [60,61,65,103]: eukaryotic homologs may be too divergent to associate with and support pAgo function and native activity without accessory proteins is too low to detect.

Acknowledgements

This work was supported by the National Science Foundation (NSF MCB-2143856). The authors declare no conflict of interest. Figures were generated with Biorender.com. We thank Rob Barlow and Topher Pirner for comments on the manuscript.

References

1. Dingermann, T. (2008) Recombinant therapeutic proteins: Production platforms and challenges. *Biotechnol. J.* 3, 90–97

2. Reichert, J.M. *et al.* (2005) Monoclonal antibody successes in the clinic. *Nat Biotechnol* 23, 1073–1078

3. Toenniessen, G.H. *et al.* (2003) Advances in plant biotechnology and its adoption in developing countries. *Current Opinion in Plant Biology* 6, 191–198

4. Lynd, L.R. *et al.* (2008) How biotech can transform biofuels. *Nat Biotechnol* 26, 169–172

5. Danna, K. and Nathans, D. (1971) Specific cleavage of simian virus 40 DNA by restriction endonuclease of Hemophilus influenzae *. *Proceedings of the National Academy of Sciences* 68, 2913–2917

6. Roberts, R.J. (2005) How restriction enzymes became the workhorses of molecular biology. *Proceedings of the National Academy of Sciences* 102, 5905–5908

7. Li, H. *et al.* (2020) Applications of genome editing technology in the targeted therapy of human diseases: mechanisms, advances and prospects. *Sig Transduct Target Ther* 5, 1–23

8. Guha, T.K. *et al.* (2017) Programmable Genome Editing Tools and their Regulation for Efficient Genome Engineering. *Computational and Structural Biotechnology Journal* 15, 146–160

9. Urnov, F.D. *et al.* (2010) Genome editing with engineered zinc finger nucleases. *Nat Rev Genet* 11, 636–646

10. Nemudryi, A.A. *et al.* (2014) TALEN and CRISPR/Cas Genome Editing Systems: Tools of Discovery. *Acta Naturae* 6, 19–40

11. Gimble, F.S. and Wang, J. (1996) Substrate recognition and induced DNA distortion by the PI-SceI endonuclease, an enzyme generated by protein splicing. *J Mol Biol* 263, 163–180

12. Collias, D. and Beisel, C.L. (2021) CRISPR technologies and the search for the PAM-free nuclease. *Nat Commun* 12, 555

13. Li, Y. *et al.* (2022) Comparison of CRISPR/Cas and Argonaute for nucleic acid tests. *Trends in Biotechnology* DOI: 10.1016/j.tibtech.2022.11.002

14. Ryazansky, S. *et al.* (2018) The Expanded Universe of Prokaryotic Argonaute Proteins. *mBio* 9, e01935-18

15. Hegge, J.W. *et al.* (2018) Prokaryotic Argonaute proteins: novel genome-editing tools? *Nat Rev Microbiol* 16, 5–11

16. Chew, W.L. *et al.* (2016) A multifunctional AAV-CRISPR-Cas9 and its host response. *Nat Methods* 13, 868–874

17. Swarts, D.C. *et al.* (2014) The evolutionary journey of Argonaute proteins. *Nat Struct Mol Biol* 21, 743–753

18. Kropocheva, E.V. *et al.* (2022) Prokaryotic Argonaute Proteins as a Tool for Biotechnology. *Mol Biol* 56, 854–873

19. Kropocheva, E. *et al.* (2021) A programmable pAgo nuclease with universal guide and target specificity from the mesophilic bacterium Kurthia massiliensis. *Nucleic Acids Res* 49, 4054–4065

20. Sheng, G. *et al.* (2014) Structure-based cleavage mechanism of *Thermus thermophilus* Argonaute DNA guide strand-mediated DNA target cleavage. *Proc. Natl. Acad. Sci. U.S.A.* 111, 652–657

21. Koopal, B. *et al.* (2022) A long look at short prokaryotic Argonautes. *Trends Cell Biol* DOI: 10.1016/j.tcb.2022.10.005

22. Swarts, D.C. *et al.* (2014) DNA-guided DNA interference by a prokaryotic Argonaute. *Nature* 507, 258–261

23. Lisitskaya, L. *et al.* (2018) DNA interference and beyond: structure and functions of prokaryotic Argonaute proteins. *Nat Commun* 9, 5165

24. Qin, Y. *et al.* (2022) Emerging Argonaute-based nucleic acid biosensors. *Trends in Biotechnology* 40, 910–914

25. Makarova, K.S. *et al.* (2009) Prokaryotic homologs of Argonaute proteins are predicted to function as key components of a novel system of defense against mobile genetic elements. *Biol Direct* 4, 29

26. Willkomm, S. *et al.* (2017) Structural and mechanistic insights into an archaeal DNA-guided Argonaute protein. *Nat Microbiol* 2, 1–7

27. Burroughs, A.M. *et al.* (2013) Two novel PIWI families: roles in inter-genomic conflicts in bacteria and Mediator-dependent modulation of transcription in eukaryotes. *Biology Direct* 8, 13

28. Swarts, D.C. *et al.* (2015) Effects of Argonaute on Gene Expression in Thermus thermophilus. *PLoS One* 10, e0124880

29. Kaya, E. *et al.* (2016) A bacterial Argonaute with noncanonical guide RNA specificity. *Proc Natl Acad Sci U S A* 113, 4057–4062

30. Wang, L. *et al.* (2022) A bacterial Argonaute with efficient DNA and RNA cleavage activity guided by small DNA and RNA. *Cell Reports* 41, 111533

31. Wang, Y. *et al.* (2008) Structure of an argonaute silencing complex with a seed-containing guide DNA and target RNA duplex. *Nature* 456, 921–926

32. Jung, S.-R. *et al.* (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

33. Liu, Z. *et al.* (2020) Application of different types of CRISPR/Cas-based systems in bacteria. *Microbial Cell Factories* 19, 172

34. Zaremba, M. *et al.* (2022) Short prokaryotic Argonautes provide defence against incoming mobile genetic elements through NAD+ depletion. *Nat Microbiol* 7, 1857–1869

35. Hur, J.K. *et al.* (2014) Prokaryotic Argonautes defend genomes against invasive DNA. *Trends in Biochemical Sciences* 39, 257–259

36. Swarts, D.C. *et al.* (2017) Autonomous Generation and Loading of DNA Guides by Bacterial Argonaute. *Mol Cell* 65, 985-998.e6

37. Kuzmenko, A. *et al.* (2020) DNA targeting and interference by a bacterial Argonaute nuclease. *Nature* 587, 632–637

38. Xing, J. *et al.* (2022) Prokaryotic Argonaute Protein from Natronobacterium gregoryi Requires RNAs To Activate for DNA Interference In Vivo. *mBio* 13, e0365621

39. Zander, A. *et al.* (2017) Guide-independent DNA cleavage by archaeal Argonaute from Methanocaldococcus jannaschii. *Nat Microbiol* 2, 17034

40. Rocha, E.P.C. and Danchin, A. (2002) Base composition bias might result from competition for metabolic resources. *Trends in Genetics* 18, 291–294

41. He, R. *et al.* (2019) Pyrococcus furiosus Argonaute-mediated nucleic acid detection. *Chemical Communications* 55, 13219–13222

42. Wang, F. *et al.* (2021) PfAgo-based detection of SARS-CoV-2. *Biosensors and Bioelectronics* 177, 112932

43. Ye, X. *et al.* (2022) Argonaute-integrated isothermal amplification for rapid, portable, multiplex detection of SARS-CoV-2 and influenza viruses. *Biosensors and Bioelectronics* 207, 114169

44. Liu, Q. *et al.* (2021) Argonaute integrated single-tube PCR system enables supersensitive detection of rare mutations. *Nucleic Acids Research* 49, e75

45. Xun, G. *et al.* (2021) A rapid, accurate, scalable, and portable testing system for COVID-19 diagnosis. *Nat Commun* 12, 2905

46. Swarts, D.C. *et al.* (2015) Argonaute of the archaeon Pyrococcus furiosus is a DNA-guided nuclease that targets cognate DNA. *Nucleic Acids Res* 43, 5120–5129

47. Lin, Q. *et al.* (2022) Programmable Analysis of MicroRNAs by Thermus thermophilus Argonaute-Assisted Exponential Isothermal Amplification for Multiplex Detection (TEAM). *Anal Chem* 94, 11290–11297

48. Xun, G. *et al.* (2021) Argonaute with stepwise endonuclease activity promotes specific and multiplex nucleic acid detection. *Bioresources and Bioprocessing* 8, 46

49. Song, J. *et al.* (2020) Highly specific enrichment of rare nucleic acid fractions using Thermus thermophilus argonaute with applications in cancer diagnostics. *Nucleic Acids Research* 48, e19–e19

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

51. Enghiad, B. *et al.* (2022) PlasmidMaker is a versatile, automated, and high throughput end-to-end platform for plasmid construction. *Nat Commun* 13, 2697

52. Wang, Y. *et al.* (2008) Structure of the guide-strand-containing argonaute silencing complex. *Nature* 456, 209–213

53. Song, J.-J. *et al.* (2004) Crystal structure of Argonaute and its implications for RISC slicer activity. *Science* 305, 1434–1437

54. Yuan, Y.-R. *et al.* (2005) Crystal Structure of A. aeolicus Argonaute, a Site-Specific DNA-Guided Endoribonuclease, Provides Insights into RISC-Mediated mRNA Cleavage. *Molecular cell* 19, 405

55. Guo, X. *et al.* (2021) A Hyperthermophilic Argonaute From Ferroglobus placidus With Specificity on Guide Binding Pattern. *Front Microbiol* 12, 654345

56. Li, X. *et al.* (2022) Mesophilic Argonaute-based isothermal detection of SARS-CoV-2. *Front Microbiol* 13, 957977

57. García-Quintans, N. *et al.* (2019) DNA interference by a mesophilic Argonaute protein, CbcAgo. *F1000Res* 8, 321

58. Vaiskunaite, R. *et al.* (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 Res* 50, 4616–4629

59. Liu, Y. *et al.* (2021) A programmable omnipotent Argonaute nuclease from mesophilic bacteria Kurthia massiliensis. *Nucleic Acids Res* 49, 1597–1608

60. Lee, K.Z. *et al.* (2021) NgAgo possesses guided DNA nicking activity. *Nucleic Acids Res* 49, 9926–9937

61. Fu, L. *et al.* (2019) The prokaryotic Argonaute proteins enhance homology sequence-directed recombination in bacteria. *Nucleic Acids Res* 47, 3568–3579

62. Huang, S. *et al.* (2023) Genome manipulation by guide-directed Argonaute cleavage. *Nucleic Acids Research* 51, 4078–4085

63. Esyunina, D. *et al.* (2023) Specific targeting of plasmids with Argonaute enables genome editing. *Nucleic Acids Research* 51, 4086–4099

64. Lee, S.H. *et al.* (2016) Failure to detect DNA-guided genome editing using Natronobacterium gregoryi Argonaute. *Nat Biotechnol* 35, 17–18

65. Javidi-Parsijani, P. *et al.* (2017) No evidence of genome editing activity from Natronobacterium gregoryi Argonaute (NgAgo) in human cells. *PLoS One* 12, e0177444

66. Martínez-Gálvez, G. *et al.* (2016) ssDNA and the Argonautes: The Quest for the Next Golden Editor. *Hum Gene Ther* 27, 419–422

67. Hegge, J.W. *et al.* (2019) DNA-guided DNA cleavage at moderate temperatures by Clostridium butyricum Argonaute. *Nucleic Acids Research* 47, 5809–5821

68. Kuzmenko, A. *et al.* (2019) Programmable DNA cleavage by Ago nucleases from mesophilic bacteria Clostridium butyricum and Limnothrix rosea. *Nucleic Acids Res* 47, 5822–5836

69. Li, B. *et al.* (2022) AGODB: a comprehensive domain annotation database of argonaute proteins. *Database* 2022, baac078

70. Tatta, E.R. *et al.* (2022) Bioprospecting of microbial enzymes: current trends in industry and healthcare. *Appl Microbiol Biotechnol* 106, 1813–1835

71. Burstein, D. *et al.* (2017) New CRISPR–Cas systems from uncultivated microbes. *Nature* 542, 237–241

72. Doxzen, K.W. and Doudna, J.A. (2017) DNA recognition by an RNA-guided bacterial Argonaute. *PLoS One* 12, e0177097

73. Xie, X. *et al.* (2023) A bacterial Argonaute from Tepiditoga spiralis with the ability of RNA guided plasmid cleavage. *Biochem Biophys Res Commun* 640, 157–163

74. Wang, Y. *et al.* (2009) Nucleation, propagation and cleavage of target RNAs in argonaute silencing complexes. *Nature* 461, 754–761

75. Li, W. *et al.* (2022) A programmable pAgo nuclease with RNA target preference from the psychrotolerant bacterium Mucilaginibacter paludis. *Nucleic Acids Res* 50, 5226–5238

76. Shmakov, S.A. *et al.* (2019) Systematic prediction of functionally linked genes in bacterial and archaeal genomes. *Nat Protoc* 14, 3013–3031

77. Lisitskaya, L. *et al.* (2022) Programmable RNA targeting by bacterial Argonaute nucleases with unconventional guide binding and cleavage specificity. *Nat Commun* 13, 4624

78. Marshall, R. *et al.* (2018) Rapid and Scalable Characterization of CRISPR Technologies Using an E. coli Cell-Free Transcription-Translation System. *Molecular Cell* 69, 146-157.e3

79. Wandera, K.G. and Beisel, C.L. (2022) Rapidly Characterizing CRISPR-Cas13 Nucleases Using Cell-Free Transcription-Translation Systems. *Methods Mol Biol* 2404, 135–153

80. Wandera, K.G. *et al.* (2020) An enhanced assay to characterize anti-CRISPR proteins using a cell-free transcription-translation system. *Methods* 172, 42–50

81. Wimmer, F. *et al.* (2022) Rapid cell-free characterization of multi-subunit CRISPR effectors and transposons. *Molecular Cell* 82, 1210-1224.e6

82. Jolly, S.M. *et al.* (2020) Thermus thermophilus Argonaute Functions in the Completion of DNA Replication. *Cell* 182, 1545-1559.e18

83. Gruen, M. *et al.* (2002) An in vivo selection system for homing endonuclease activity. *Nucleic Acids Research* 30, e29

84. Hand, T.H. *et al.* (2019) Directed evolution studies of a thermophilic Type II-C Cas9. *Methods Enzymol* 616, 265–288

85. Lee, J.K. *et al.* (2018) Directed evolution of CRISPR-Cas9 to increase its specificity. *Nat Commun* 9, 3048

86. Hand, T.H. *et al.* (2021) Catalytically Enhanced Cas9 Through Directed Protein Evolution. *CRISPR J* 4, 223–232

87. Paschon, D.E. *et al.* (2019) Diversifying the structure of zinc finger nucleases for high-precision genome editing. *Nat Commun* 10, 1133

88. Chen, Z. and Zhao, H. (2005) A highly sensitive selection method for directed evolution of homing endonucleases. *Nucleic Acids Research* 33, e154

89. Lee, K.Z. *et al.* (2022) Repurposing the Homing Endonuclease I-SceI for Positive Selection and Development of Gene-Editing Technologies. *ACS Synthetic Biology* DOI: 10.1021/acssynbio.1c00340

90. Hunt, E.A. *et al.* (2018) Single-stranded binding proteins and helicase enhance the activity of prokaryotic argonautes in vitro. *PLoS One* 13, e0203073

91. Wang, L. *et al.* (2021) Pyrococcus furiosus Argonaute coupled with modified ligase chain reaction for detection of SARS-CoV-2 and HPV. *Talanta* 227, 122154

92. Larson, M.H. *et al.* (2013) CRISPR interference (CRISPRi) for sequence-specific control of gene expression. *Nat Protoc* 8, 2180–2196

93. Schultenkämper, K. *et al.* (2020) Impact of CRISPR interference on strain development in biotechnology. *Biotechnology and Applied Biochemistry* 67, 7–21

94. Wu, Y. *et al.* (2023) CRISPR–dCas12a-mediated genetic circuit cascades for multiplexed pathway optimization. *Nat Chem Biol* 19, 367–377

95. van der Krol, A.R. *et al.* (1990) Flavonoid genes in petunia: addition of a limited number of gene copies may lead to a suppression of gene expression. *Plant Cell* 2, 291–299

96. Napoli, C. *et al.* (1990) Introduction of a Chimeric Chalcone Synthase Gene into Petunia Results in Reversible Co-Suppression of Homologous Genes in trans. *Plant Cell* 2, 279–289

97. Fire, A. *et al.* (1998) Potent and specific genetic interference by double-stranded RNA in Caenorhabditis elegans. *Nature* 391, 806–811

98. Bohmert, K. *et al.* (1998) AGO1 defines a novel locus of Arabidopsis controlling leaf development. *EMBO J* 17, 170–180

99. Meister, G. and Tuschl, T. (2004) Mechanisms of gene silencing by double-stranded RNA. *Nature* 431, 343–349

100. Cerutti, L. *et al.* (2000) Domains in gene silencing and cell differentiation proteins: the novel PAZ domain and redefinition of the Piwi domain. *Trends in Biochemical Sciences* 25, 481–482

101. Liu, M. *et al.* (2020) The Clustered Regularly Interspaced Short Palindromic Repeat System and Argonaute: An Emerging Bacterial Immunity System for Defense Against Natural Transformation? *Frontiers in Microbiology* 11

102. Fang, X. and Qi, Y. (2016) RNAi in Plants: An Argonaute-Centered View. *The Plant Cell* 28, 272–285

103. Khin, N.C. *et al.* (2017) No evidence for genome editing in mouse zygotes and HEK293T human cell line using the DNA-guided Natronobacterium gregoryi Argonaute (NgAgo). *PLoS One* 12, e0178768

104. Sternberg, S.H. *et al.* (2014) DNA interrogation by the CRISPR RNA-guided endonuclease Cas9. *Nature* 507, 62–67

105. McClelland, S.E. and Szczelkun, M.D. (2004) The Type I and III Restriction Endonucleases: Structural Elements in Molecular Motors that Process DNA. In *Restriction Endonucleases* (Pingoud, A. M., ed), pp. 111–135, Springer

106. Cao, Y. *et al.* (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

107. Olina, A. *et al.* (2020) Genome-wide DNA sampling by Ago nuclease from the cyanobacterium Synechococcus elongatus. *RNA Biol* 17, 677–688

108. Letunic, I. and Bork, P. (2016) Interactive tree of life (iTOL) v3: an online tool for the display and annotation of phylogenetic and other trees. *Nucleic Acids Res* 44, W242–W245

109. Mukherjee, S. *et al.* (2023) Twenty-five years of Genomes OnLine Database (GOLD): data updates and new features in v.9. *Nucleic Acids Research* 51, D957–D963

110. Chen, I.-M.A. *et al.* (2023) The IMG/M data management and analysis system v.7: content updates and new features. *Nucleic Acids Research* 51, D723–D732

Table 1 Experimentally Validated pAgos

Name	Organism	Protein ID	Organism Temperature Range[a]	Guide Nucleotide	Nucleotide Substrate	Ion	PAZ Type[b]	MID Type	References
AaAgo	*Aquifex aeolicus*	WP_010880937	Hyperthermophilic	5'P-DNA	RNA	Mg^{2+}, Mn^{2+}	normal	YK	[54]
CbAgo	*Clostridium butyricum*	WP_058142162.1	Mesophilic	5'P-DNA	DNA	Mn^{2+}, Mg^{2+}	normal	YK	[67,68]
CpAgo	*Clostridium perfringens*	WP_003477422.1	Mesophilic	5'P-DNA	DNA	Mn^{2+}, Mg^{2+}	normal	YK	[106]
IbAgo	*Intestinibacter bartlettii*	WP_007287731.1	Mesophilic	5'P-DNA	DNA	Mn^{2+}, Mg^{2+}	normal	YK	[106]
KmAgo	*Kurthia massiliensis*	WP_010289662.1	Mesophilic	5'P-DNA	DNA/RNA	Mn^{2+},Mg^{2+}	normal	YK	[59]
LrAgo	*Limnothrix rosea*	WP_075892274.1	Mesophilic	P-DNA	DNA	Mn^{2+},Mg^{2+}	normal	YK	[68]
MbpAgo	*Mucilaginibacter paludis*	WP_008504757.1	Psychrotolerant	5'P/OH-DNA	RNA	Mg^{2+}, Mn^{2+}	normal	HK	[75]
MhAgo	*Marinitoga hydrogenitolerans*	WP_072865986.1	Thermophilic	All guide variations	DNA/RNA	Mn^{2+}	small	5'OH	[30]
MjAgo	*Methanocaldococcus jannaschii*	WP_005580376.1	Hyperthermophilic	5'P-DNA	DNA	Mg^{2+}, Mn^{2+}	normal	YK	[26]

MpAgo	*Marinitoga piezophila*	WP_014295921.1	Thermophilic	5'OH-RNA	DNA	Mn^{2+},Mg^{2+}	small	5'OH	[29]
NgAgo	*Natronobacterium gregoryi*	WP_005580376.1	Mesophilic	5'P-DNA	DNA	Mg^{2+}, Mn^{2+}	normal	YK	[60]
PfAgo	*Pyrococcus furiosus*	WP_011011654.1	Hyperthermophilic	5'P-DNA	DNA	Mn^{2+}, Co^{2+}	normal	YK	[46]
PliAgo	*Pseudooceanicola lipolyticus*	WP_100161590.1	Mesopihilic	5'P/OH DNA	RNA	Mg^{2+}, Mn^{2+}	normal	RH	[77]
SeAgo	*Synechococcus elongatus*	WP_011244830.1	Mesophilic	5'P-DNA	DNA	Mn^{2+}	normal	KR	[107]
TsAgo	*Tepiditoga spiralis*	WP_190614489.1	Mesophilic	5'OH-RNA	DNA	Mn^{2+}, Mg^{2+}	small	5'OH	[73]
TtAgo	*Thermus thermophilus*	WP_011174533.1	Hyperthermophilic	5'P-DNA	DNA/RNA	Mn^{2+},Mg^{2+}	normal	RK	[82,90]

[a] Organism Temperature Range Designations: Hyperthermophilic: 80-110 ℃; Thermophilic: 50-80 ℃; Mesophilic: 20-45 ℃; Psychrotolerant: 0-30 ℃

[b] Small PAZ domain designation lacks the R3 subdomain

Figure 1. Current and potential applications of prokaryotic Argonautes. pAgos are currently used for *in vitro* molecular diagnostics and DNA assembly. There are significant opportunities for *in vivo* biotechnologies such as gene editing, therapeutics, imaging, and crop development with pAgos functional at mesophilic temperatures (20 – 45°C).

Figure 2. Diversity of long prokaryotic Argonautes. (A) Cartoon domain structure of a short pAgo with MID and PIWI* domains with associated proteins and a long pAgo with N-terminus, PAZ, MID, and PIWI domains. (B) Crystal structure representations of *Clostridium butyricum* (CbAgo, PDB:6QZK) bound to DNA target strand. Domain coloring matches cartoon diagram (N-terminus – forest green, linker 1 – pink, PAZ – wheat, linker 2 – grey, MID -blue, PIWI – red, DNA guide – purple, DNA target substrate - yellow). (C) Circular phylogenetic tree of long pAgos determined by AGODB [69]. Cyan branches indicate experimentally validated long pAgos. The outer circle denotes in a blue stripe long pAgos and in an orange stripe short pAgos which originate from a mesophilic host. Phylogenetic tree visualized by iTOL v6 [108].

Figure 3. Potential pAgo-derived tools. (A) Fusion of catalytically dead Argonautes (dpAgo, depicted in blue) with transcription repressors like KRAB and activators like VP64 fused with p65 and RTA may be developed to regulate the expression of a specific gene of interest (GOI). (B) Engineered dpAgos can be developed to visualize endogenous genomic elements in living cells by fusion of fluorescent proteins. (C) Epigenetic editors may be targeted to specific loci by fusing dpAgos that recognize programmed sequences to DNA modifying enzymes, such as DNA methylases and histone acetyltransferases that modify the genome. (D) DNA base-editing could theoretically be achieved by fusing dpAgo to deaminases to stimulate C>T or A>G base edits without DNA cleavage. (E)

Prime-editing by active pAgos (muti-colored) could potentially be developed via long ssDNA hairpin guides. Due to the ssDNA nature of the guide, reverse transcription would not be needed, and edits could potentially be integrated via single stranded DNA homologous recombination on either side of the ssDNA break in the target with the other half of the hairpin guide.

Figure I. pAgo gene neighborhood may predict substrate preference. Gene neighborhood of representative pAgos CbAgo, NgAgo, MpAgo, and MbpAgo. CbAgo and NgAgo, which both utilize 5'P DNA to target DNA substrates, are flanked by DNA modifying genes in the genome. MpAgo gene is found in a CRISPR locus downstream of a primase, Cas1, and Cas2 genes. Experimentally validated pAgos which use a 5'OH RNA guide appears in this same genomic neighborhood pattern. MbpAgo, which targets RNA, displays a different gene neighborhood landscape as compared to DNA modifying pAgos. Gene neighborhoods were modified from JGI's IMG/MER database [109,110].

Glossary

Apo-pAgos: Prokaryotic argonaute without a guide

Barnase: a bacterial protein with ribonuclease activity secreted by *Bacillus amyloliquefaciens*, it is lethal to the cell when expressed without its inhibitor barstar.

Cas: stands for CRISPR associated protein which is the part of bacterial adaptive immunity against viruses and plasmids, and is utilized in genetic engineering applications.

CcdB: toxin from the ccd system on the *E. coli* F plasmid, acts as a gyrase poison.

Chi-sites: a short stretch of DNA in the genome of a bacterium near which homologous recombination is more likely to occur. They serve as the stimulators of DNA double-strand break repair in bacteria.

COVID-19: Coronavirus disease, an infectious disease caused by the SARS-CoV-2 virus.

CRISPR: Clustered Regularly Interspaced Short Palindromic Repeats are repetitive DNA sequences found in the genomes of bacteria and archaea, which are derived from DNA fragments of infecting bacteriophages. These are used to detect and destroy DNA from similar bacteriophages during subsequent infections.

Helicase recBC: DNA helicase that functions in homologous recombination in *E. coli.*

Homing endonucleases: enzymes that catalyze a highly specific double-strand DNA break at few or even singular genomic locations due to a large (~20 nts) recognition sequence. Naturally found in introns and self-splicing inteins

Meganucleases: enzymes that catalyze a highly specific double-strand DNA break due to a large (~20 nts) recognition sequence. May be natural (e.g. homing endonuclease) or synthetic (e.g. ZFNs)

MID: Middle domain of prokaryotic argonautes

Molecular beacon: hairpin-shaped molecules with an internally quenched fluorophore whose fluorescence is restored when the hairpin is disrupted (e.g. through binding).

PAM: Protospacer adjacent motif - a short DNA sequence required for a Cas nuclease to cut that flanks the guide RNA.

PAZ: PIWI-Argonaute-Zwille domain of prokaryotic argonaute

PIWI: stands for the P-element Induced Wimpy Testis domain of prokaryotic Argonaute that cleaves RNA/DNA

Programmable endonucleases: nucleases that can be programmed with nucleic acids or protein engineering for site-specific cleavage

RT-qPCR: Quantitative reverse transcription polymerase chain reaction, used to detect and quantify RNA.

SARS-CoV-2: a member of a large family of viruses called coronaviruses, causative agent of COVID-19.

Sir2: Silent information regulator 2 (Sir2) enzymes that catalyze NAD+-dependent protein/histone deacetylation.

SpCas9: Cas9 protein of *Streptococcus pyogenes*

TALENs: Transcription activator-like effector nucleases (TALENs) comprise a nonspecific nuclease fused to a sequence-specific DNA-binding domain and used as a genome editing tool.

TIR: Toll/interleukin-1 receptor (TIR) homology domain is an intracellular signaling domain found in MyD88, interleukin 1 receptor and the Toll receptor. It contains three highly conserved regions, and mediates protein-protein interactions between the Toll-like receptors (TLRs) and signal-transduction components.

ZFN: Zinc finger nucleases (ZFNs) are a class of engineered DNA-binding proteins that facilitate targeted editing of the genome by creating double-strand breaks in DNA.

Figure 1

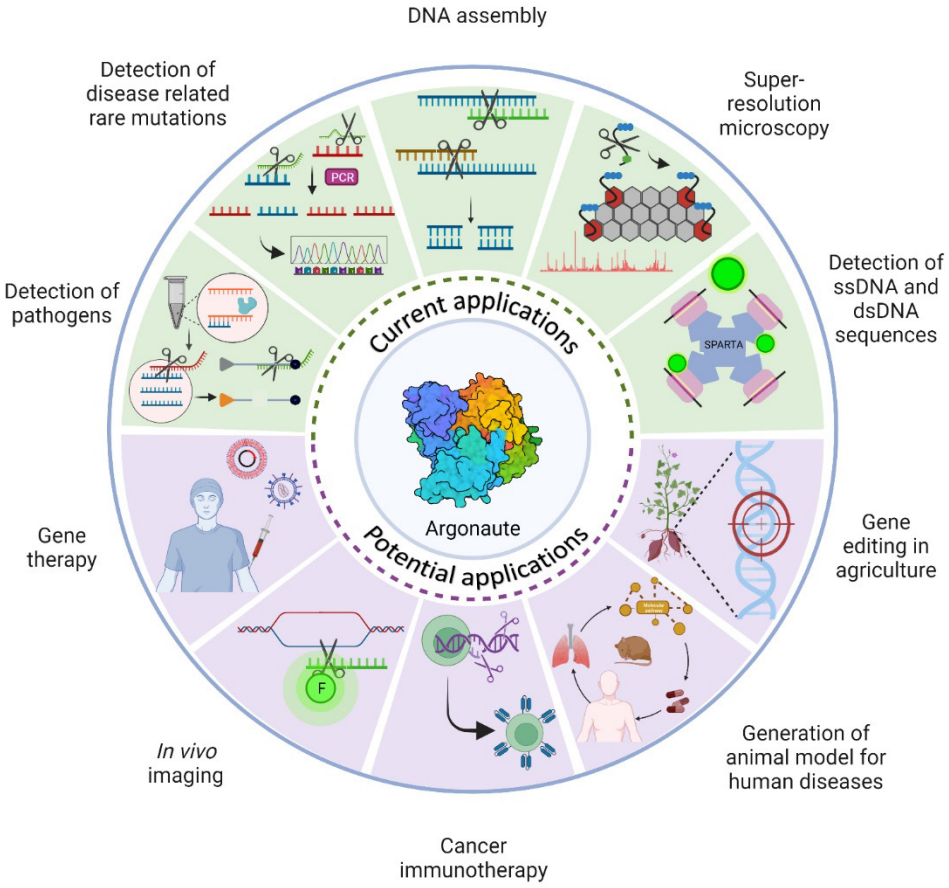

DNA assembly
Detection of disease related rare mutations
Super-resolution microscopy
Detection of pathogens
Detection of ssDNA and dsDNA sequences
Current applications
Argonaute
Potential applications
Gene therapy
Gene editing in agriculture
In vivo imaging
Generation of animal model for human diseases
Cancer immunotherapy
PCR
SPARTA

Figure 2

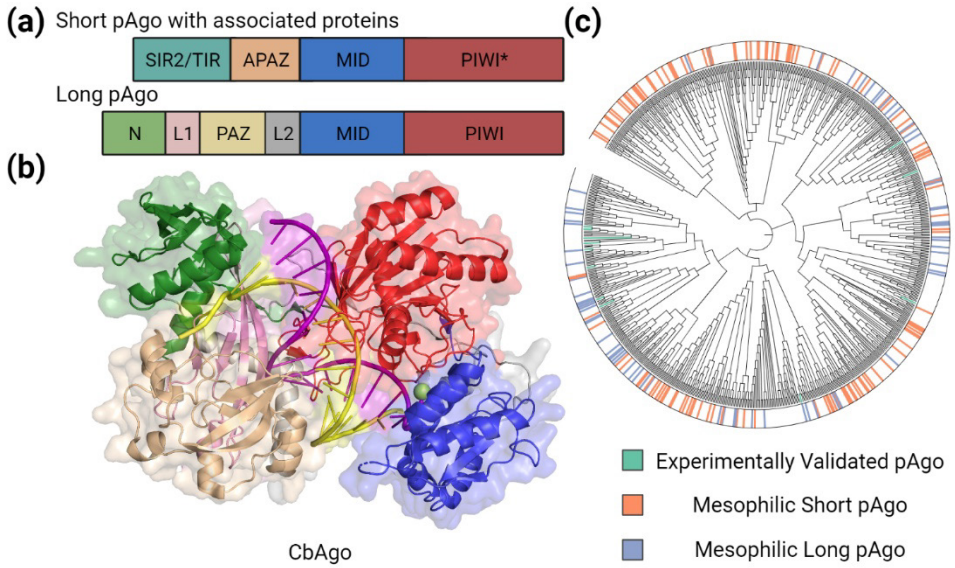

(a)
Short pAgo with associated proteins
SIR2/TIR
APAZ
MID
PIWI*
Long pAgo
N
L1
PAZ
L2
MID
PIWI
(b)
CbAgo
(c)
Experimentally Validated pAgo
Mesophilic Short pAgo
Mesophilic Long pAgo

Figure 3

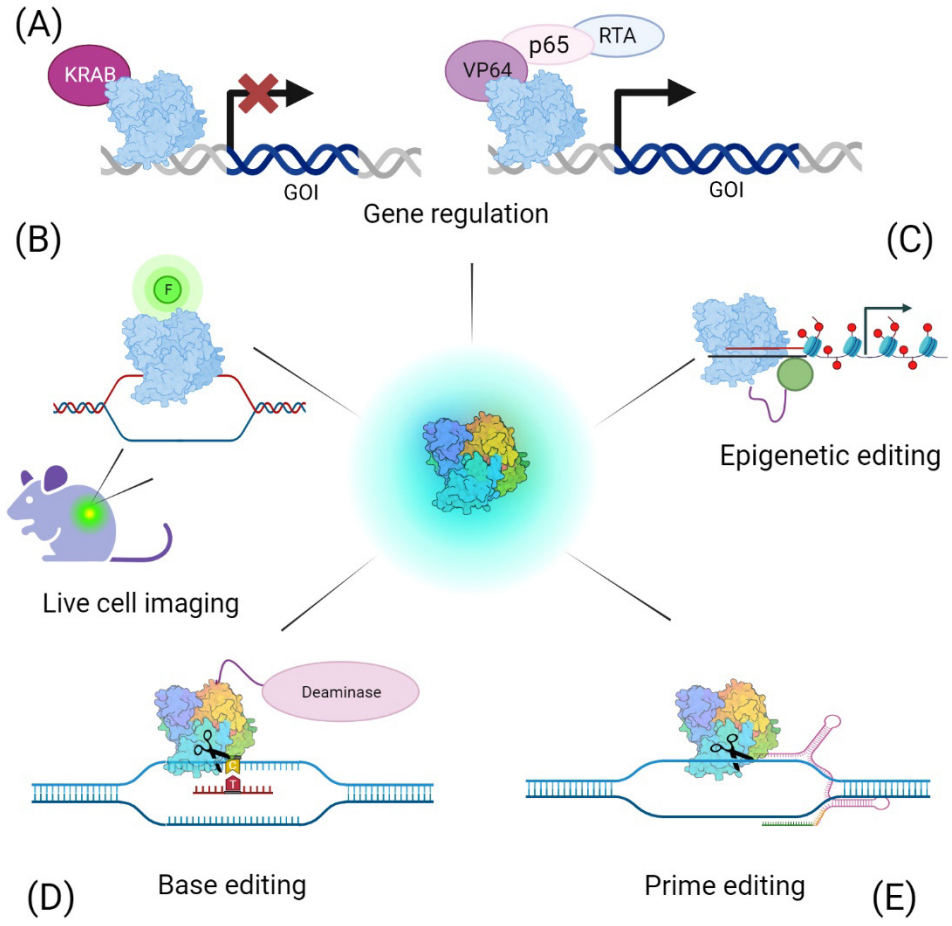
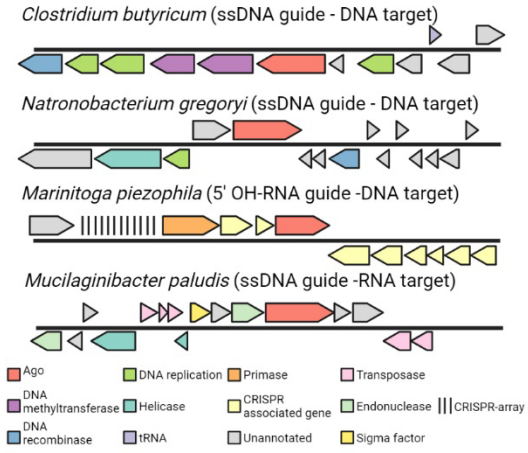

(A)
KRAB
GOI
VP64
p65
RTA
GOI
Gene regulation
(B)
F
Live cell imaging
(C)
Epigenetic editing
Deaminase
(D)
Base editing
Prime editing
(E)

Figure I


Clostridium butyricum (ssDNA guide - DNA target)
Natronobacterium gregoryi (ssDNA guide - DNA target)
Marinitoga piezophila (5' OH-RNA guide -DNA target)
Mucilaginibacter paludis (ssDNA guide -RNA target)
Ago
DNA replication
Primase
Transposase
DNA methyltransferase
Helicase
CRISPR associated gene
Endonuclease
CRISPR-array
DNA recombinase
tRNA
Unannotated
Sigma factor