STRUCTURE AND FUNCTION OF A PROKARYOTIC ARGONAUTE FROM *PSEUDOMONAS AERUGINOSA*

by

Reece Sheridan Erickson

A thesis submitted in partial fulfillment
of the requirements for the degree

of

Master of Science

in

Microbiology

MONTANA STATE UNIVERSITY
Bozeman, Montana

December 2020

ACKNOWLEDGEMENTS

I am grateful to friends and colleagues for thoughtful discussions and critical feedback. I would especially like to thank Dr. Heini Miettinen-Granger for generating the PACS2 mutants and Murat Büyükyörük for expertise in bioinformatics. Research in the Wiedenheft lab is supported by the National Institutes of Health (R35GM134867), a young investigator award from Amgen, the M. J. Murdock Charitable Trust, and the Montana State University Agricultural Experimental Station (USDA NIFA).

TABLE OF CONTENTS

LIST OF TABLES

Table Page

LIST OF FIGURES

Figure Page

ABSTRACT

Argonautes (Ago) are structurally and functionally diverse proteins present in all domains of life. A common feature of these ancient proteins is their ability to bind nucleic acid guides that target the protein to complementary sequences. Although eukaryotic argonautes (eAgo) have been well-studied, we still know very little about the function of prokaryotic argonautes (pAgo) in bacterial and archaeal species. To address this gap in our knowledge, my thesis focused on determining the biochemical properties as well as the cellular functions of a pAgo from the organism *Pseudomonas aeruginosa* PACS2 (PaAgo). Here, we show that PaAgo plays a role in regulating the expression of transposons within PACS2. I also present results indicating that deletion of the PaAgo gene and its neighboring genes causes toxicity to *P. aeruginosa*. Finally, I provide evidence that PaAgo and a neighboring protein are binding partners and form a multi-protein complex. Future work will focus on copurifying and sequencing PaAgo nucleic acid guides as well as clarifying the mechanisms guide acquisition and biological function.

CHAPTER ONE

THE STRUCTURE AND FUNCTION OF PROKARYOTIC ARGONAUTES

Abstract

The argonaute (Ago) family of proteins is present in all domains of life. In eukaryotes, Agos act as sequence-specific nucleases that use small RNA guides to target complimentary RNA targets. Many prokaryotic genomes also encode argonautes, and they are significantly more diverse than their eukaryotic counterparts. Despite the diversity of prokaryotic argonautes, few have been studied in detail and many critical aspects of their functions and mechanisms remain unknown. Here, we review available data on the structure and functions of prokaryotic argonautes as well as discuss potential cellular roles they may play.

Introduction

Argonautes (Ago) are an ancient and ubiquitous family of proteins with members in all domains of life [1]. In eukaryotes, argonautes are characterized by their ability to use small RNA guides to recognize, and sometimes cleave, complementary RNA targets [14]. This process, collectively known as RNA interference (RNAi), is responsible for a diverse array of functions including regulating gene expression and protecting against harmful genetic elements [21]. RNAi processes are present in the majority of eukaryotes [3] and likely predate the last eukaryotic common ancestor [2]. Bioinformatic analysis of prokaryotic genomes shows that Agos are also present in the genomes of both bacteria and archaea (~10% and ~30%, respectively) [4] and that prokaryotic argonautes (pAgo) are significantly more diverse than their eukaryotic counterparts (eAgo) [5]. Based on this increased level of diversity, it has been hypothesized that pAgos may have additional functions compared to eAgos. Despite this diversity and widespread presence of Agos in prokaryotes, few pAgos have been functionally or biochemically characterized to date. Understanding the structure, mechanisms, and cellular functions of pAgos will be important if they are to be adapted as research tools or therapeutics in a way similar to eAgo. Here, we review available data on the functions and structures of prokaryotic argonautes. We also discuss the seemingly diverse biological roles of these proteins and address outstanding questions in the field.

Structure of Argonaute proteins

All eAgos contain six conserved domains that are arranged in a conserved order (N-terminal, L1 (linker 1), PAZ (PIWI-Argonaute-Zwille), L2, MID (Middle), and PIWI (P-element Induced Wimpy Testis)) (**Figure 1.1**). In contrast, the structures of pAgos are significantly more diverse. pAgo structures can be divided roughly in half into two clades, known as "long" and "short" pAgos [4]. Long pAgos contain all the domains present in eAgos. Long pAgos can be further divided into long-A and long-B pAgos, where long-A's generally contain the conserved DEDX amino acid tetrad in their PIWI domain that is required for nuclease activity while long-Bs do not (**Figure 1.1**). Short pAgos are truncated and only contain the MID and PIWI domains. Additionally, similar to long-B pAgos, all short pAgos are missing a canonical DEDX tetrad on their PIWI domain, rendering the pAgo nuclease inactive. Many pAgos with inactive nucleases are fused to or located in the same operon as genes containing putative nucleases that are hypothesized to play roles in either the generation of nucleic acid guides or in the cleavage of complementary targets [5]. Short pAgos are also frequently encoded near proteins containing APAZ (analog of PAZ) domains [4]. While APAZ have no sequence similarity to PAZ or N domains, APAZ domains are found exclusively in or near short pAgos, and thus has been speculated to be a functional analog of either the PAZ or N domain [4,5,13].

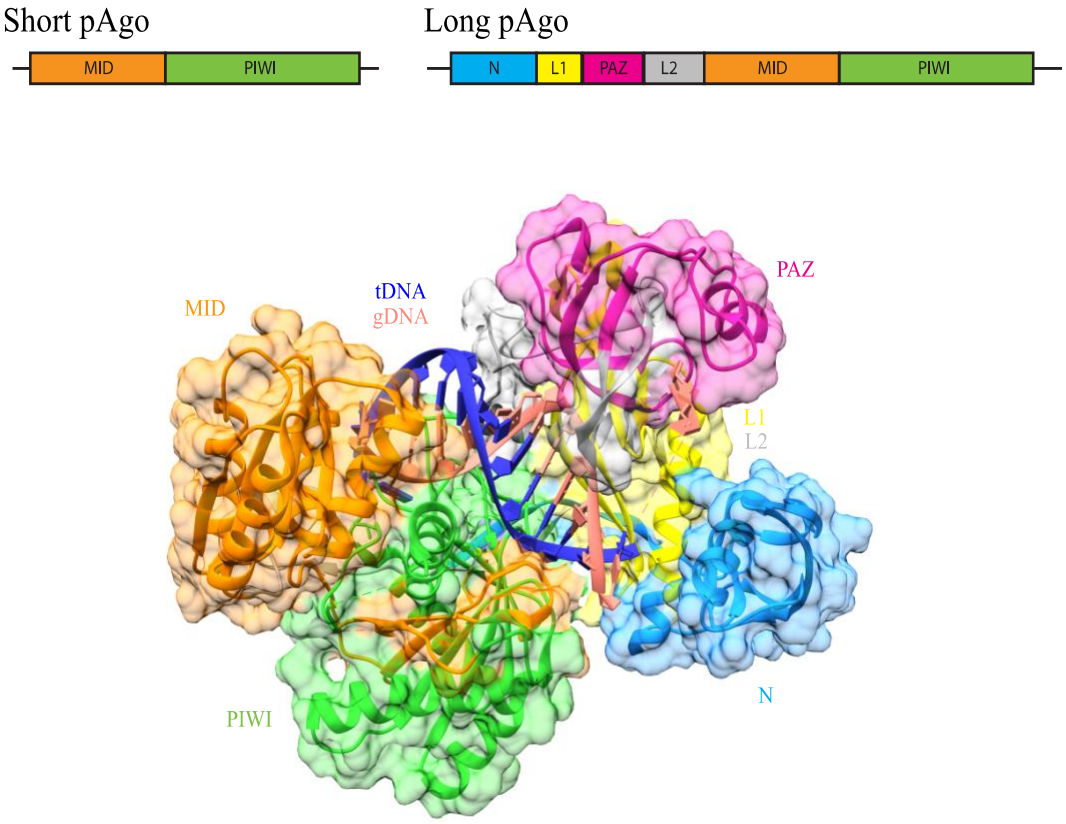

Figure 1.1. Structural organization of prokaryotic argonautes. Schematic of short pAgo (left) and long pAgo (right) domain architecture. Crystal structure of the long-A argonaute protein from *Thermus thermophilus* (TtAgo PDB: 4N41) bound to a ssDNA guide (orange) and ssDNA target (blue). The model was rendered using ChimeraX [45].

Functions of argonaute protein domains

The PIWI domain of argonaute proteins is responsible for the endonucleolytic ("slicing") action of Agos. PIWI domains include an RNase H-like active site that is defined by a conserved DEDX (X denotes D, H, or K) tetrad of acidic amino acids [1]. The glutamate of the DEDX motif is located on the "glutamic finger" [1], which undergoes a conformational change during hybridization of the guide to a complementary

target. In catalytically active pAgos, the glutamic finger switches from “unplugged” to “plugged in” conformation upon target binding. Shifting into the “plugged in” conformation completes the DEDX tetrad and allows for the binding of 2 divalent metal ions (Mg^{2+} or Mn^{2+}), which then triggers target cleavage between the 10^{th} and 11^{th} base pairs of the guide nucleic acid [1]. All long-B and short pAgos contain mutations that perturb the DEDX motif that are predicted to prevent conformational changes of the glutamic finger, or prevent metal ion binding, and thus long-B and short pAgos are predicted to be catalytically inactive [5].

The MID domain is responsible for binding the 5’ end of the nucleic acid guides by forming a basic nucleotide-binding pocket. The majority of Agos recognize and bind 5’-phosphorylated guides using a Mg^{2+} ion in the binding pocket [6]. pAgos display significant diversity in their preference for guides with specific nucleotides at the 5’ position. For example, TtAgo (*Thermus thermophilus*) preferentially binds guides with 5’-cysteine [8], MjAgo (*Methanocaldococcus jannaschii*) 5’-purines [9], RsAgo (*Rhodobacter sphaeroides*) 5’-uridines [7], while MpAgo (*Marinitoga piezophila*) shows no 5’ nucleotide preference [10]. Additionally, some pAgos do not bind 5’-phosphorylated guides, and instead, only associate with guides that contain a 5’- hydroxyl (MpAgo) [11]. This preference for 5’- hydroxylated guides is due to a MID domain that relies on a hydrophobic binding pocket rather than a metal ion to anchor the 5’ end of guides [46].

The PAZ domain is responsible for binding the 3’ end of guide nucleic acids [5]. Compared to PIWI and MID domains, the amino acid sequences of PAZ domains are

more divergent, but maintain structural similarity [5]. A full-length PAZ (such as that of TtAgo) possesses four nucleic acid binding regions that form a hydrophobic pocket that serves to anchor the 3’ end of guides [12, 29, 34, 47]. Some pAgos (e.g., RsAgo) with truncated PAZ domains can still bind guides with an orientation that differs from that of a complete PAZ [10], indicating that a complete PAZ domain is not required for guide binding. These variations in PAZ domains have been hypothesized to affect various aspects of guide binding such as the kinetics of pAgo-guide complex formation and stability, 3’ nucleotide preference, and the efficacy of target recognition [5]. Many pAgos, especially members of the short pAgo group, do not contain PAZ domains. However, previous studies have identified a proposed “analog of PAZ” (APAZ) domain that is present in many short pAgos as well as in neighboring genes to short pAgos [4]. Because it is exclusively found near short pAgos, the APAZ domain has been hypothesized to be a functional analog to either the PAZ or N domains that are missing from short pAgos [4, 13]. The function of APAZ domains has yet to be shown, however, the majority of short pAgos have an APAZ domain located within their operon [5], suggesting that APAZ-containing proteins play an integral role in short pAgo function.

The N domain of Ago proteins is involved in guide acquisition [16], target cleavage [14], and the dissociation of strands after cleavage [15]. In the process of guide acquisition, double stranded nucleic acids are loaded into the Ago protein and the strand with the least stable 5’ end will be retained as a guide [16]. During the removal of the non-guide (passenger) strand, the N domain acts as a wedge that blocks guide-passenger base pairing at the 3’ end (known as “active wedging”). N domains also block guide-

target base pairing downstream of the guide and help to dissociate guide target duplexes after target cleavage (known as “passive wedging”) [17]. The role of the N domain in target cleavage has only been demonstrated for human Ago proteins and a single pAgo (TtAgo) [14, 17]. The diversity of N domains in prokaryotes suggests there are likely other mechanisms of action that have yet to be investigated.

Biological roles of prokaryotic argonautes

Based on the various roles of eAgos, pAgos have been hypothesized to function as a defense against foreign genetic elements such as transposons, plasmids, and phage [4]. This proposition has gained some support through the study of numerous long-A and long-B pAgos [7,8,18] (**Table 1**). However, despite the work done on long pAgos, no short pAgos have been functionally or structurally characterized to date.

Of the pAgos that have been studied, the most well characterized is that of *Thermus thermophilus* (TtAgo). TtAgo is a long-A pAgo that relies on a DNA guide to bind and cleave single-stranded DNA [8]. This is in contrast to eAgos that exclusively use RNA guides to target other complementary RNAs. Diversity in guide-target preference is a common trend in pAgos [19]. *In vitro*, TtAgo uses its guides to bind and cleave single-stranded DNA but can also cleave double-stranded substrates if provided guides complementary to both strands of the duplex [8]. The deletion of TtAgo also leads to an increased efficiency of plasmid transformation *in vivo*, and when expressed in *E. coli*, TtAgo preferentially binds to guides derived from the expression plasmid [8]. These results suggest that TtAgo plays a role is the suppression of foreign plasmids. However,

TtAgo has also been shown to participate in decatenating chromosomes following cellular replication by acting as a type II topoisomerase [20], indicating it also plays a role in DNA replication.

In contrast to TtAgo, the long-B pAgo from *Rhodobacter sphaeroides* (RsAgo) uses RNA guides to target complementary DNA [7]. RsAgo contains substitutions in its PIWI domain that render it nuclease inactive. Despite no nuclease activity, it has still been observed that RsAgo lowers the expression level of plasmid-encoded genes while not reducing plasmid copy number [7]. Based on this, a model for RsAgo action has emerged where RsAgo will use RNA guides to bind coding regions of DNA and act as a roadblock for RNA polymerase, thus regulating transcription of targeted genes.

Table 1. Summary of prokaryotic argonautes.

Host	Argonaute	Guide	Target	Biological Function	References
Aquifex aeolicus	AaAgo	DNA	RNA	—	29-31
Archaeglobus fulgidus	AfAgo	DNA	DNA	—	6, 25, 26, 28
Clostridium butyricum	CbAgo	DNA	DNA	—	27
Limnothrix rosea	LrAgo	DNA	DNA	—	27
Marinitoga piezophile	MpAgo	RNA	DNA	—	43, 44
Methanocaldococcus jannaschii	MjAgo	DNA	DNA	Reduced plasmid content and transformation efficiency	9, 32, 33
Pyrococcus furiosus	PfAgo	DNA	DNA	Reduced transformation efficiency	34-36
Rhodobacter sphaeroides	RsAgo	RNA	DNA	Reduced transcription of genes	7, 42
Synechococcus elongatus	SeAgo	DNA	DNA	DNA processing, chromosomal replication	41
Thermus thermophilus	TtAgo	DNA	DNA	Reduced transformation efficiency, changes in gene expression, chromosomal replication	8, 12, 17, 20, 37-40

The biological roles of non-canonical pAgos

pAgos have been hypothesized to have a role in protection against foreign genetic elements such as plasmids, phage, and transposons [19]. This role has been investigated in numerous different pAgos and there have been examples of pAgos targeting foreign sequences [7] as well as reducing plasmid transformation efficiency [8]. However, there has been no published work demonstrating a pAgo providing phage defense. Further, the majority of evidence that suggests pAgo's role in defense comes from ectopic *E. coli*

expression systems, and pAgo behavior differs when expressed endogenously versus ectopically. For example, when heterologous TtAgo is expressed in *E. coli*, it is able to be loaded with guides and cleave complementary targets after guide-target hybridization [8], as well as independently obtain guides through a double-stranded DNA 'chopping' mechanism [40]. In contrast, when expressed in *T. thermophilus* TtAgo recognizes single-stranded DNA breaks in the chromosome and acquires guides from downstream sequences [20]. Additionally, endogenous TtAgo does not use its acquired guides to cleave complementary targets [20]. Instead, it appears that its guides are merely a byproduct of the decatenation process since no subsequent cleavage occurs after guide acquisition [20]. Similarly, another long-A pAgo from *Synechococcus elongatus* (SeAgo) is able to be loaded with guides and cleave targets *in vitro,* while *in vivo* it preferentially acquires guides to its own chromosome with no additional cleavage after initial guide loading [41]. Additionally, based on preferential guide acquisition from the terminus of replication, *in vivo* results from TtAgo and SeAgo do not suggest defense as their primary function, but instead point to a role in replication. Furthermore, both these pAgos use their nucleases for guide acquisition and not for target degradation, suggesting that their *in vivo* role may not be as a sequence-guided nuclease, which is in stark contrast to catalytically active eAgos. While it has been proposed that pAgos may play a role in defense based on their genomic neighbors [4], it appears that the major function of studied pAgos is related to regulation of host genes as well as genome replication.

Future Work

Short pAgos

Despite making up roughly half of all pAgos [4], there are currently no published experimental data on short pAgos. Short pAgos contain an inactive PIWI domain and are also lacking N and PAZ domains. Thus, the mechanisms responsible for guide generation, as well as how they anchor the 3' end of guides, remains unknown. Short pAgos are also often encoded in the same genomic neighborhood as nucleases as well as proteins containing APAZ domains [5], indicating they may use accessory proteins to fulfill their cellular functions.

Guide generation

The molecular basis for pAgo guide generation has yet to be shown. Moreover, the diversity of pAgo structure is almost certainly reflected by a similar diversity in guide biogenesis. For nucleolytically active pAgos such as TtAgo, it appears they have the ability to generate their own guides [20, 40]. However, for short and long-B pAgos without an active nuclease, it is unlikely that they are able to perform stand-alone guide generation. Instead, they would either have to draw upon a pool of premade guides or would require the action of an accessory nuclease for guide acquisition. Roughly 75% of pAgos contain mutations in their nuclease active site [4] and thus require additional proteins for guide generation. Therefore, identifying pAgo binding partners will be crucial in determining the mechanisms of guide biogenesis.

Target selection

In addition to the guide and target substrate diversity that pAgos possess, the preference of pAgos for certain targets is similarly diverse. For example, there is evidence that RsAgo samples the *R. sphaeroides* transcriptome and preferentially obtains RNA guides complementary to transposases and prophage genes [7]. This is in contrast to TtAgo, which appears to recognize single stranded breaks in chromosomal DNA near the terminus of replication and uses its endonuclease activity to acquire guides directly from the bacterial chromosome *in vivo* [20]. These results point to a likely trend in pAgo biology, that different substrate preferences lead to different cellular roles. Determining the recognition signal for guide acquisition as well as the mechanism for determining self from non-self is critical to deciphering pAgo biological function.

REFERENCES CITED

1. Swarts DC, Makarova K, Wang Y, et al. The evolutionary journey of Argonaute proteins. Nat Struct Mol Biol. 2014;21:743–753.

2. Makarova KS, et al. Ancestral paralogs and pseudoparalogs and their role in the emergence of the eukaryotic cell. Nucleic Acids Res. 2005;33:4626–4638.

3. Shabalina, Svetlana A, and Eugene V Koonin. "Origins and evolution of eukaryotic RNA interference." Trends in ecology & evolution vol. 23,10 (2008): 578-87. doi:10.1016/j.tree.2008.06.005

4. Makarova, K. S., Wolf, Y. I., van der Oost, J. & Koonin, E. V. 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 (2009).

5. Ryazansky S, Kulbachinskiy A, Aravin AA. 2018. The expanded universe of prokaryotic Argonaute proteins. mBio 9:e01935-18.

6. Ma JB, et al. Structural basis for 5′-end-specifc recognition of guide RNA by the *A. fulgidus* Piwi protein. Nature. 2005;434:666–670

7. Olovnikov, I., Chan, K., Sachidanandam, R., Newman, D. K. & Aravin, A. A. Bacterial argonaute samples the transcriptome to identify foreign DNA. Mol. Cell 51, 594–605 (2013).

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

9. Willkomm, S. et al. Structural and mechanistic insights into an archaeal DNA- guided Argonaute protein. Nat. Microbiol. 2, 17035 (2017).

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

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

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

13. Willkomm S, Makarova KS, Grohmann D. 2018. DNA silencing by prokaryotic Argonaute proteins adds a new layer of defense against invading nucleic acids. FEMS Microbiol Rev 42:376–387.

14. Hauptmann, J. et al. Turning catalytically inactive human Argonaute proteins into active slicer enzymes. Nat. Struct. Mol. Biol. 20, 814–817 (2013).

15. Faehnle, C.R., Elkayam, E., Haase, A.D., Hannon, G.J. & Joshua-Tor, L. The making of a slicer: activation of human Argonaute-1. Cell Reports 3, 1901–1909 (2013).

16. Schwarz, D.S. et al. Asymmetry in the assembly of the RNAi enzyme complex. Cell 115, 199–208 (2003).

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

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

19. Lisitskaya L, Aravin AA, Kulbachinskiy A. DNA interference and beyond: structure and functions of prokaryotic Argonaute proteins. Nat Commun. 2018;9(1):5165. Published 2018 Dec 4.

20. Jolly et al., *Thermus thermophilus* Argonaute Functions in the Completion of DNA Replication, Cell (2020).

21. Christine Ender, Gunter Meister. Journal of Cell Science 2010 123: 1819-1823

22. Swarts, D. C., Jore, M. M. and Oost, J. v. (2014). Expression and Purification of the *Thermus thermophilus* Argonaute Protein. *Bio-protocol* 4(19): e1253

23. Expression and purification of proteins using 6xHistidine-tag: a comprehensive manual. IBA Lifesciences. October 2012. Pages 19-20.

24. Livak KJ, Schmittgen TD. Methods. Vol. 25. San Diego, CA: 2001. Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method; pp. 402–408.

25. Parker, J. S., Roe, S. M. & Barford, D. Crystal structure of a PIWI protein suggests mechanisms for siRNA recognition and slicer activity. EMBO J. 23, 4727–4737 (2004).

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

27. Anton Kuzmenko, Denis Yudin, Sergei Ryazansky, Andrey Kulbachinskiy, Alexei A Aravin, Programmable DNA cleavage by Ago nucleases from mesophilic bacteria *Clostridium butyricum* and *Limnothrix rosea*, *Nucleic Acids Research*, Volume 47, Issue 11, 20 June 2019, Pages 5822–5836.

28. Parker, J. S., Parizotto, E. A., Wang, M., Roe, S. M. & Barford, D. Enhancement of the seed-target recognition step in RNA silencing by a PIWI/MID domain protein. Mol. Cell 33, 204–214 (2009)

29. Yuan, Y. R. et al. Crystal structure of *A. aeolicus* argonaute, a site-specific DNA-guided endoribonuclease, provides insights into RISC-mediated mRNA cleavage. Mol. Cell 19, 405–419 (2005).

30. Yuan, Y. R., Pei, Y., Chen, H. Y., Tuschl, T. & Patel, D. J. A potential protein-RNA recognition event along the RISC-loading pathway from the structure of *A. aeolicus* Argonaute with externally bound siRNA. Structure 14, 1557–1565 (2006).

31. Rashid, U. J. et al. Structure of *Aquifex aeolicus* argonaute highlights conformational flexibility of the PAZ domain as a potential regulator of RNA-induced silencing complex function. J. Biol. Chem. 282, 13824–13832 (2007).

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

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

34. Song, J. J., Smith, S. K., Hannon, G. J. & Joshua-Tor, L. Crystal structure of Argonaute and its implications for RISC slicer activity. Science 305, 1434–1437 (2004).
35. Rivas, F. V. et al. Purified Argonaute2 and an siRNA form recombinant human RISC. Nat. Struct. Mol. Biol. 12, 340–349 (2005).

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

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

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

39. Swarts, D. C., Koehorst, J. J., Westra, E. R., Schaap, P. J. & van der Oost, J. Effects of Argonaute on gene expression in *Thermus thermophilus*. PLoS One 10, e0124880 (2015).

40. Swarts, D. C. et al. Autonomous generation and loading of DNA guides by bacterial Argonaute. Mol. Cell 65, 985–998 (2017).

41. Anna Olina, Anton Kuzmenko, Maria Ninova, Alexei A. Aravin, Andrey Kulbachinskiy & Daria Esyunina. Genome-wide DNA sampling by Ago nuclease from the cyanobacterium *Synechococcus elongatus*, RNA Biology, 17:5, 677-688 (2020).

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

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

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

45. Pettersen EF, Goddard TD, Huang CC, Meng EC, Couch GS, Croll TI, Morris JH, Ferrin TE. UCSF ChimeraX: Structure visualization for researchers, educators, and developers. *Protein Sci.* (2020)

46. Emine Kaya, Kevin W. Doxzen, Kilian R. Knoll, Ross C. Wilson, Steven C.Strutt, Philip J. Kranzusch, Jennifer A. Doudna. A Bacterial Argonaute with noncanonical guide RNA specificity. Proceedings of the National Academy of Sciences Apr (2016).

47. Liu Y, Esyunina D, Olovnikov I, Teplova M, Kulbachinskiy A, Aravin AA, Patel DJ. Accommodation of helical imperfections in *Rhodobacter sphaeroides* Argonaute ternary complexes with guide RNA and target DNA. Cell Rep 24:453– 462. (2018).

CHAPTER TWO

THE FUNCTIONAL AND BIOCHEMICAL PROPERTIES OF A SHORT ARGONAUTE PROTEIN FROM *PSEUDOMONAS AERUGINOSA*

Abstract

Argonaute (Ago) proteins rely on small RNA or DNA guides that target these proteins to complementary RNA or DNA targets. Although eukaryotic argonautes (eAgo) have been well studied for their role in RNA interference as well as other small RNA pathways, we still know very little about the function of prokaryotic argonautes (pAgo) in bacteria and archaea. Here, I present results suggesting that the pAgo from the PACS2 strain of *Pseudomonas aeruginosa* (PaAgo) plays a role in regulating the expression of transposases and that deletion of the PaAgo and the genes immediately upstream and downstream results in a 'plaquing' phenotype in *P. aeruginosa*. I also show that the PaAgo protein physically interacts with a neighboring protein in its gene cassette, forming a multi-protein complex. Future work will focus on determining biological function and mechanisms of action.

Introduction

Argonaute proteins are well known for their role in the eukaryotic RNA interference (RNAi) pathway [11, 12]. However, argonautes are also present in numerous bacterial and archaeal species [2]. Prokaryotic argonautes (pAgos) are significantly more diverse than those of eukaryotes [3] and this increased diversity has led us to hypothesize that pAgos may have similarly diverse functions.

Phylogenetic analyses have delineated three distinct pAgo clades: long-A, long-B, and short pAgos [3]. Long-A pAgos contain six domains. The N-terminal domain (N-domain) involved in strand dissociation after target cleavage, two linkers (L1 and L2), an endonucleolytic PIWI domain, and a MID and PAZ domains that are responsible for anchoring the 5’ and 3’ ends of guides, respectively [2]. Long-B pAgos contain the same domains as long-A, but have mutations in their PIWI domains that are predicted to render the nuclease inactive. Short pAgos possess only the MID and PIWI domains, and similar to long-Bs, have no active nuclease. While long pAgos have been demonstrated to have numerous cellular functions, there are currently no publications on the role of short pAgos, leaving a critical gap in our understanding of prokaryotic argonautes.

In vivo and *in vitro* studies of long pAgos show that they bind and sometimes cleave nucleic acids in a guide-dependent or independent manner [4,5]. These studies also reveal that some long pAgos associate with additional proteins to perform their action [5]. While short pAgos have no active nuclease, they are often encoded near putative nucleases and DNA binding proteins [2,3], indicating that they likely have binding partners that assist in their biological roles.

Pseudomonas aeruginosa PACS2 contains a prokaryotic argonaute protein (PaAgo). PaAgo is a short pAgo with a length of 471 amino acids that contains complete MID and PIWI domains (**Figure 2.1**, PACS2_12720). The PaAgo gene is flanked by a downstream gene coding for 214 amino acids that is annotated as a domain of unknown function (DUF2026) (**Figure 2.1**, PACS2_12715) and an upstream gene 560 amino acids in length containing partial N and PAZ domains and an L1 linker (**Figure 2.1**, PACS2_12725). Here, we investigate the short PaAgo of *Pseudomonas aeruginosa* PACS2 and its neighboring genes. We show that deletion of PaAgo results in toxicity to PACS2 as well as an increased level of transposon expression. Additionally, we demonstrate the concurrence of PaAgo cassette genes is conserved in 63% of *Pseudomonas* species and that PaAgo and the gene immediately upstream copurify and form a complex *in vitro*.

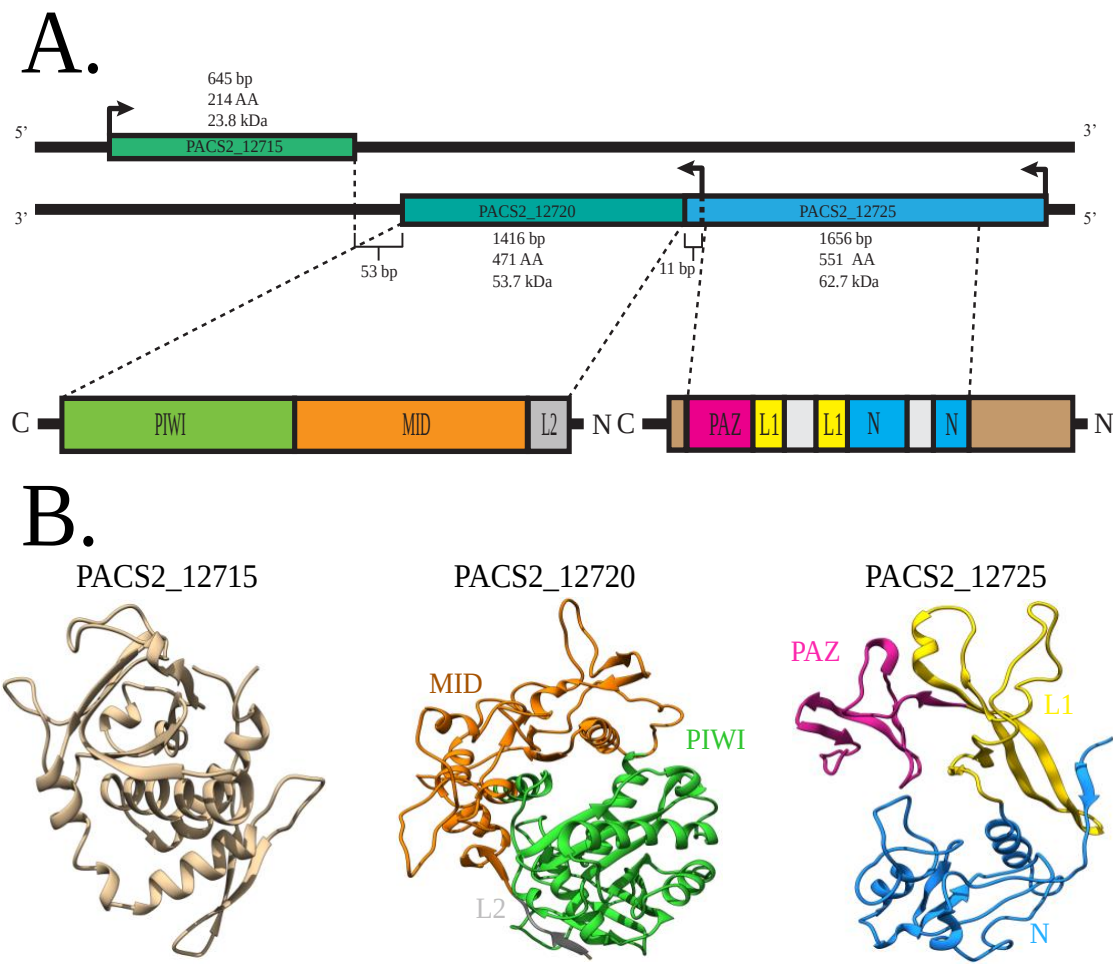

Figure 2.1. A: Schematic of *Pseudomonas aeruginosa* PACS2 argonaute cassette. Bottom two schematics depict the translated proteins of PACS2_12720 and PACS2_12725 with argonaute domains annotated. Unannotated portions represent either insertions into argonaute domains (grey) or regions without homology to argonaute proteins (brown). B: Homology based predictions of tertiary structure of the PACS2 argonaute cassette proteins with argonaute domains colored to correspond to schematic from part A. Models rendered using ChimeraX [7].

Materials and Methods

Strains and mutants

Pseudomonas aeruginosa strain PACS2 mutants were generated by Dr. Heini Miettinen-Granger. Genes of interest were first replaced by a gentamycin resistance gene (Gm^R) using the pEX-18T plasmid following the procedure described by Choi et al, 2005 [13]. Gm^R was then removed using the pFLP-2 plasmid containing the *Flp* recombinase following the procedure described in Choi et al, 2005 [13]. This procedure was used to generate PACS2 Δ12720::*gen* (Δ*ago*::*gen*), PACS2 Δ12720::16580 (Δ*ago*::*tran*), and PACS2 Δ12715-Δ12720-Δ12725 (Δall) mutants.

Recombinant expression of PACS2 Ago and flanking proteins was performed using *Escherichia coli* strain BL21 (DE3). First, PACS2 genomic DNA was used to amplify sequences for PACS2_12715, 12720, and 12725 via PCR using Q5 polymerase. Primers and PCR conditions described in **Table 2**. PCR fragments and expression plasmids were then digested with appropriate restriction enzymes (**Table 2**). A DNA Clean and Concentrator kit (Zymo Research; Cat. No D4014) was then used to purify digested DNA. Digested plasmid was subjected to Quick CIP (New England Biolabs; Cat. No M0525S) alkaline phosphotase. Inserts were then mixed with plasmid in a 3:1 insert:plasmid molar ratio and ligated using T4 DNA ligase. Ligations were then added to chemically competent BL21 (DE3) cells and transformed via heat-shock transformation.

Pseudomonas aeruginosa (PACS2) and *Escherishia coli* (BL21 (DE3)) strains were cultured in standard LB medium (Casein peptone 10 g/L, yeast extract 5 g/L, NaCl 5 g/L. Fischer Scientific; Cat no. BP9722-2) and on LB agar dishes (1.5% agar w/v).

Cells were always grown at 37º C and shaken at 225 RPM unless stated otherwise. pRSF-1b was maintained using 50 µg/mL kanamycin and pCDF-duet was maintained using 100 µg/mL streptomycin. Expression was induced for all plasmids using 0.1 mM Isopropyl β-D-1 Thiogalactopyranoside (IPTG). PACS2 Δ*ago::gen* was maintained using 50 µg/mL of gentamycin.

Table 2. Plasmids, primers, and PCR conditions used for PaAgo cassette cloning

Gene	Primer name and sequence	Plasmid	Restriction enzymes	Annealing temperature	Elongation Time
PACS2 12715	12715_AscI_F AAAAA**GGCGCGCC**AAATGAGCAAGAAGAAGCACCA	pCDF-Duet	AscI	65º C	30 seconds
	12715_SbfI_R AAAAA**CCTGCAGG**CTACCAACCCCCTTCGACAG		SbfI		
PACS2 12720	12720_HindIII_+5_R AAAAA**GGATCC**TCACATGTAAAAGCTATAGC	pRSF-1b	HindIII	55º C	75 seconds
	12720_BamHI_+5+F AAAAA**GGATCC**CTTGAAGATTAAAATCCTCAA		BamHI		
PACS2 12725	12725_S2_Fr_NcoI_F ATATA**CCATGG**AGTGGAGCCACCCGCAGTTCGAAAAGATGCCACTCCCAGCGGCCTC	pRSF-1b	NcoI	72º C	60 seconds
	12725_S2_Fr_Hind_R ATATA**AAGCTT**TTAATCTTCAAGGGCTAACTCCTCATCAAATTCAGCGCCAACC		HindIII		

Culturing and plaque assays

Cell were grown in liquid culture overnight. Overnight cultures (100 µL) were added to 4 mL of molten LB agarose (0.6% agarose w/v) and poured over 1.5% agar LB plates. Plates were then left to dry lid up until agarose became solid. 5 µL of phage serial dilutions were spotted onto plates and left to dry. Once dry, plates were incubated lid down at 37º C overnight. Plaquing efficiency was calculated by normalizing plaque counts from PACS2 mutants to PACS2 wild type.

Protein purification

Cells were grown at 37º C until they reached OD_{600} = 0.5, placed in 4º C for 30 minutes, Isopropyl β-D-1 Thiogalactopyranoside (IPTG) was added to a final concentration on 0.1 mM and the cultures were shaken (175 rpm) at 16º C overnight. Cells were then pelleted at 5000 rpm (Sorvall SLC-4000 rotor) for 15 minutes at 4º C and supernatant was discarded. The cell pellets were transferred into 50 mL tubes and frozen. Pellets were then thawed on ice and resuspended in 30 mL of appropriate binding buffer (Histidine-tag: 50 mM NaH_2PO_4, 300 mM NaCl, 10 mM imidazole, pH 8.0. Strep-tag: 20 mM Tris-HCl, 1 mM NaCl, 2 mM $MnCl_2 \cdot 4H_2O$, pH 8.0). Cell mixture was then lysed by sonication for 2 sessions of 3 minutes (2 second pulse, 3 second rest, 30% intensity, on ice). Lysate was then pelleted at 10000 rpm (Sorvall SS-34 rotor) for 1 hour at 4º C and remaining supernatant (cleared lysates) was saved.

Cleared cell lysates were used for all protein purifications. Histidine-tagged proteins were purified using Ni-NTA Superflow Cartridge (Qiagen; Cat No. 30760) following protocol from IBA Lifesciences [23]. Strep-tagged proteins were purified using

Strep-Tactin Superflow Plus Cartidge (Qiagen; Cat No. 30060) following protocol from Swarts et al [22].

Pull-down assay

Cells were grown at 37º C until they reached OD_{600} = 0.5, placed in 4º C for 30 minutes, Isopropyl β-D-1 Thiogalactopyranoside (IPTG) was added to a final concentration on 0.1 mM and the cultures were shaken (175 rpm) at 16º C overnight. Cells were then pelleted at 5000 rpm (Sorvall SLC-4000 rotor) for 15 minutes at 4º C and supernatant was discarded. The cell pellets were transferred into 50 mL tubes and frozen. Pellets were then thawed on ice and resuspended in 30 mL of appropriate binding buffer (Histidine-tag: 50 mM NaH_2PO_4, 300 mM NaCl, 10 mM imidazole, pH 8.0. Strep-tag: 20 mM Tris-HCl, 1 mM NaCl, 2 mM $MnCl_2 \cdot 4H_2O$, pH 8.0). Cell mixture was then lysed by sonication for 2 sessions of 3 minutes (2 second pulse, 3 second rest, 30% intensity, on ice). Lysate was then pelleted at 10000 rpm (Sorvall SS-34 rotor) for 1 hour at 4º C and remaining supernatant (cleared lysates) was saved.

Cleared lysates from induced BL21-DE3 strains containing pCDF_duet-his-12715 and/or pRSF_1b-his-12720 were mixed with cleared lysates from the BL21-DE3 strain containing pRSF_1b-strep-12725. Mixed lysates were then incubated at room temperature shaking gently for 1 hour and purified using a Strep-Tactin Superflow Plus Cartridge following the protocol from Swarts et al [22]. Affinity purified proteins were visualized on a 10% acrylamide SDS-PAGE denaturing gel.

Size exclusion chromatography

Affinity purified proteins were concentrated using 10K protein concentrator (ThemoFisher; Cat No. 88535) and loaded onto SuperDex 200 10/300 GL column (Sigma; Cat No. GE17-5175-01) equilibrated in a buffer containing 20 mM Tris-HCl, and 1 M NaCl (8.0 pH). The column was calibrated using using a standard curve generated using a Gel Filtration Standard Kit (BioRad; Cat No. 151-1901) in a buffer containing 50 mM HEPES, 300 mM KCl, 5% glycerol v/v, 1 mM tris(2-carboxyethyl)phosphine (TCEP), pH 7.5. SEC data was analyzed using Unicorn 5.31 software (Cytiva Lifesciences) and the fractions were run on a 10% SDS-PAGE denaturing gel. Gel were stained using Coomassie blue (BioRad; Cat No. 1610803). Stain was then removed using destain (20% MeOH v/v, 10% acetic acid v/v) and visualized using an Epson ET-3750 scanner. Ladder used was PageRuler Plus Prestained protein ladder (ThemoFisher; Cat. No 26620).

RT-qPCR

Cells were grown to mid-log phase (OD_{600} = 0.5) under appropriate antibiotic selection were applicable. 1 mL of culture was pelleted at 13,000 RPM (benchtop centrifuge) for 2 minutes and supernatant was discarded. Total RNA was extracted using NucleoSpin RNA prep kit (Macherey-Nagel; Cat No. 740955.10) and total RNA was measured using a NanoDrop One^{C} spectrophotometer (ThermoFisher Scientific). A cDNA library was then created using the High-Capacity cDNA Reverse Transcription Kit (Applied Biosystems; Cat No. 4368814). qPCR was then performed using the QuantiTect SYBR Green RT-PCR Kit (Qiagen; Cat No. 204243) and was results were recorded on a

LightCycler 96 (Roche Lifesciences). Results were then analyzed via the $2^{-\Delta\Delta CT}$ method [24] using the gene *RpoD* as a reference gene. Errors bars depict the standard error of two biological replicates with three technical replicates each. Significant difference was calculated using a two-tailed t-test [10].

Protein structure and domain prediction

Protein sequences from PACS2_12715, 12720, and 12725 were submitted to HHpred [25] and results were used to identify domains with homology to other argonaute proteins. Atomic coordinates were downloaded from the Protein Data Bank (PDB) and used as templates to model predicted structures in SWISS-MODEL [26-30]. Proteins were also submitted to Phyre^2 [31] and results were compared to SWISS-MODEL and HHpred results. The best models were selected based on MolProbity scores [32].

Results

PaAgo regulates transposase expression in PACS2

Allelic exchange was used to generate ΔPaAgo mutants in PACS2 following the procedure described by Choi et al, 2005. During the process, the gene of interest is replaced by a gene encoding gentamycin resistance (Gm^R). The gentamycin resistance gene is then removed using the flippase recombinase (*Flp*), resulting in a markerless deletion of the gene of interest [13]. Two independent attempts to remove the Gm^R maker from the PaAgo (PACS2_12720) deletion mutant resulted in the insertion of an IS4-like transposase (PACS2_16580) after introduction of *Flp* (**Figure 2.2, A**). Based on this observation, and established roles for both eukaryotic and prokaryotic Agos targeting transposons [6, 14, 15], we hypothesized that the PaAgo cassette is involved in regulating the activity of transposases in PACS2. To test this hypothesis, we performed RT-qPCR to measure transcription of all transposases present in the genomes of PACS2 wild type (WT), a strain in which the gentamycin gene replaces PaAgo (Δ12720::*gen* or Δ*ago*::*gen*), a strain in which the PaAgo is replaced by a transposase (Δ12720::16580 or Δ*ago*::*tran*), and a clean deletion that removes all three genes in the cassette (Δ12715-Δ12720-Δ12725 or Δall) (**Figure 2.2, B**). We find that three transposases (PACS2_16580, 17525, and 17530) in the Δ*ago*::*tran* mutant show significantly increased (p-value < 0.05) expression compared to wild type. Similarly, we also show that the Δ*ago*::*gen* mutant displays significantly increased expression for five out of six transposases (PACS2_16580, 17530, 17535, 23295, and 23300; p-value < 0.05). Interestingly, the Δall mutant showed expression levels similar to wild type PACS2 with

only two transposases (PACS2_23295 and 23300) exhibiting significantly increased expression (p-value < 0.05). Based on the fact that at least two transposases are upregulated whenever any proteins in the cassette are deleted, these results suggest that the PACS2 PaAgo cassette plays a role in regulating the transcription of transposases within PACS2. However, the differential expression of transposases in each of the three mutants indicates that the precise mechanisms of the regulation remain elusive.

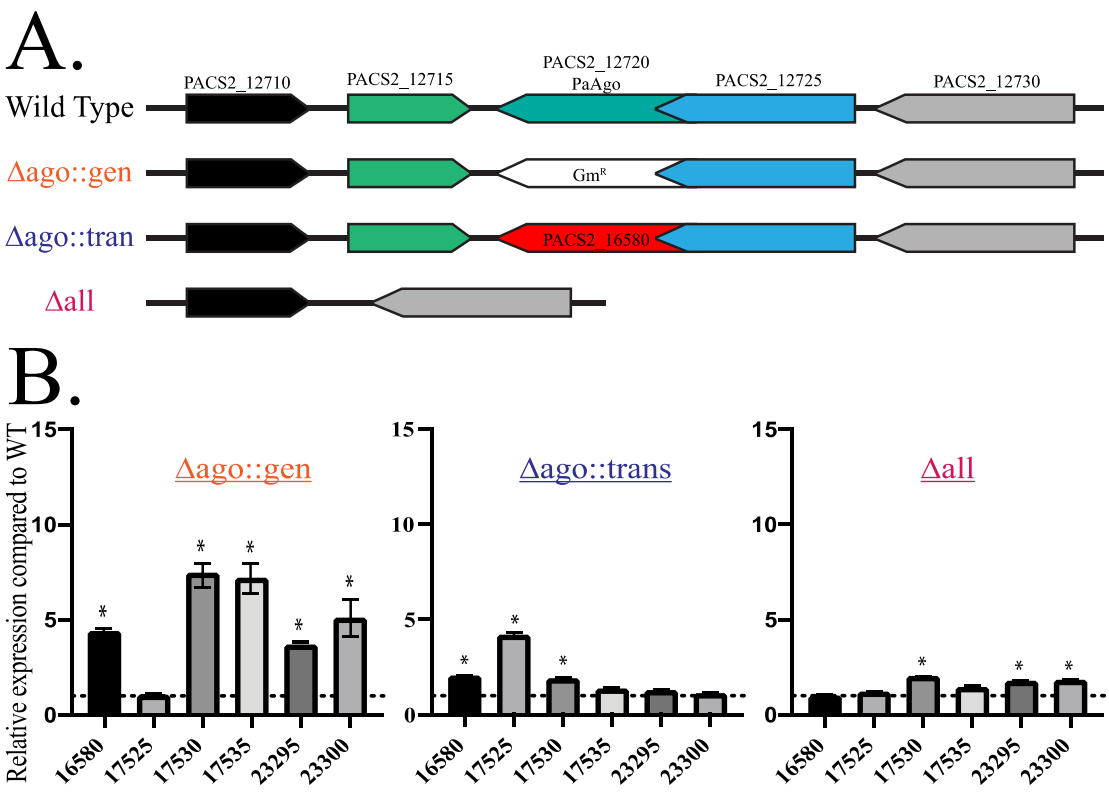

Figure 2.2. RT-qPCR results measuring PACS2 transposase expression levels in PACS2 mutants relative to PACS2 wild type. A: *P. aeruginosa* strains: Wild type PACS2; Δ*ago*::*gen*, PaAgo deletion containing gentamycin resistance (Gm^R); Δ*ago*::*tran*, PaAgo deletion with transposase (PACS2_16580) insertion; Δall, complete deletion of PaAgo cassette. B: Results from RT-qPCR. Graphs show level of transcription of PACS2 transposases for each mutant relative to wild type.

PaAgo deletion leads to toxicity in PACS2

In order to determine if the deletion of PaAgo and its adjacent proteins led to decreased protection from bacteriophage we performed plaque assays using three serially diluted phage in order to measure differences in plaquing for the ΔPaAgo mutants compared to wild type PACS2. After exposure to two lysogenic (JBD-25 and JBD-18) and one lytic (F8) phage, we saw no difference in plaquing for any of the ΔPaAgo mutants compared to wild type (data not shown). However, when a plaque assay is performed without any phage, the Δall (Δ12715-Δ12720-Δ12725) mutant displays a 'plaquing' phenotype while all other strains show no plaquing (**Figure 2.3**). The same results are observed when performing plaque assays using the phage DMS3*vir* (data not shown). These results indicate that deletion of PaAgo and its neighboring genes is detrimental to cellular health and leads to some level of toxicity for PACS2. The results also show that the plaquing phenotype is not due to the presence of DMS3*vir.*

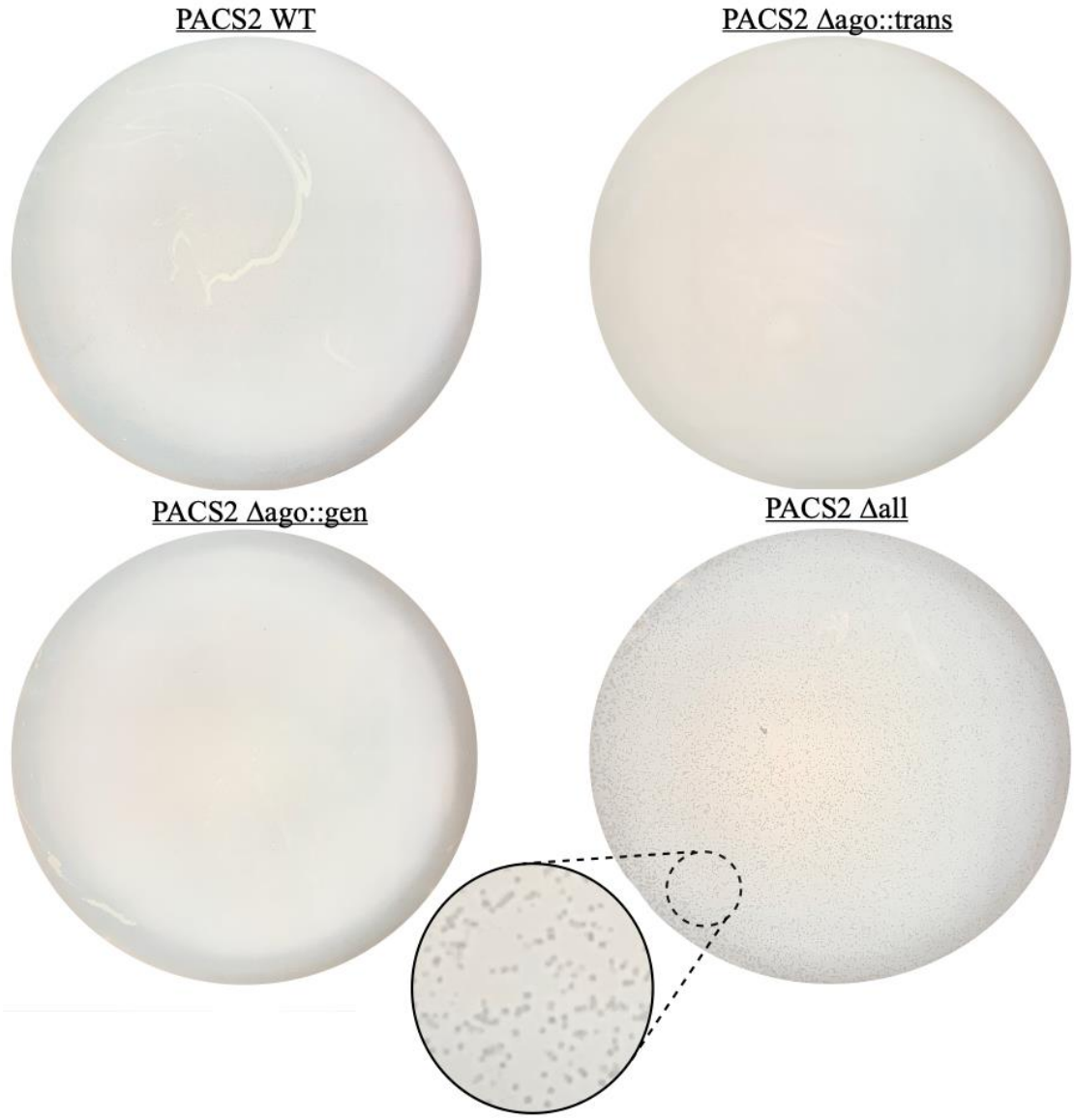

Figure 2.3. Plates of PACS2 strains depicting ‘plaquing’ phenotype in PACS2 Δall mutant after no exposure to phage.

Structure prediction

We performed structural analysis to investigate whether PaAgo and its flanking genes possessed regions of homology to other argonaute proteins. We first submitted the amino acid sequences of the three proteins to HHpred [25] in order to determine if any contained domains with homology to other argonaute proteins. We find that PaAgo shares the most homology with the PIWI and MID domains of the *Rhodobacter sphaeroides* argonaute protein (100% confidence, 17% identity, 1.3e-45 Evalue) and that PACS2_12725 shares the most homology with the N and PAZ domains of an argonaute from *Marinitoga piezophila* (99% confidence, 12% identity, 1.4e-9 Evalue). PACS2_12715 had no similarity to any documented argonautes, however it did share homology with a putative cysteine protease from *Agrobacterium tumefaciens* (100% confidence, 34% identity, 7.1e-52 Evalue). Based on these results, we generated homology-based structural predictions for all three proteins in SWISS-MODEL (**Figure 2.1**). The models were evaluated using 'clashscores', Ramachandran values, and percentage of bad side-chain rotamers using MoProbity [8] (**Figure 2.1**). The results of PaAgo having a PIWI and MID domain while PACS2_12725 possesses probable N and PAZ domains (**Figure 2.1**), coupled with previous results showing that short pAgos are often encoded near accessory nucleases and DNA binding proteins [5], led us to hypothesize that the proteins in the PaAgo cassette may assemble into a complex.

The PaAgo cassette proteins form a complex *in vitro*

Structural predictions of the proteins in the PaAgo cassette show that PACS2_12720 contains a PIWI and MID domain while PACS2_12725 contain PAZ and N domains. Together, these four regions constitute a potential split long pAgo. Based on this, we hypothesized that the proteins in the PaAgo cassette form a complex to perform their functions. To test this hypothesis, we performed pull-down assays where all three proteins (PACS2_12715, 12720, and 12725) are overexpressed in *E. coli* expression systems and are tagged with different affinity tags (PACS2_12715 and 12720 are His-tag, and PACS2_12725 is strep-tag; **Figure 2.4, A**). We mixed the lysates from the three cell lines in various combinations and pulled-down on the strep-tagged PACS2_12725. Neither PACS2_12720 nor PACS2_12725 copurify with PACS2_12715 (**Figure 2.4, B**). In contrast, pull-downs performed using the affinity tag on PACS2_12725 (PaAgo) resulted in copurification of PACS2_12720 (**Figure 2.4, B**), indicating that PACS2_12720 and 12725 physically interact and form a complex *in vitro*.

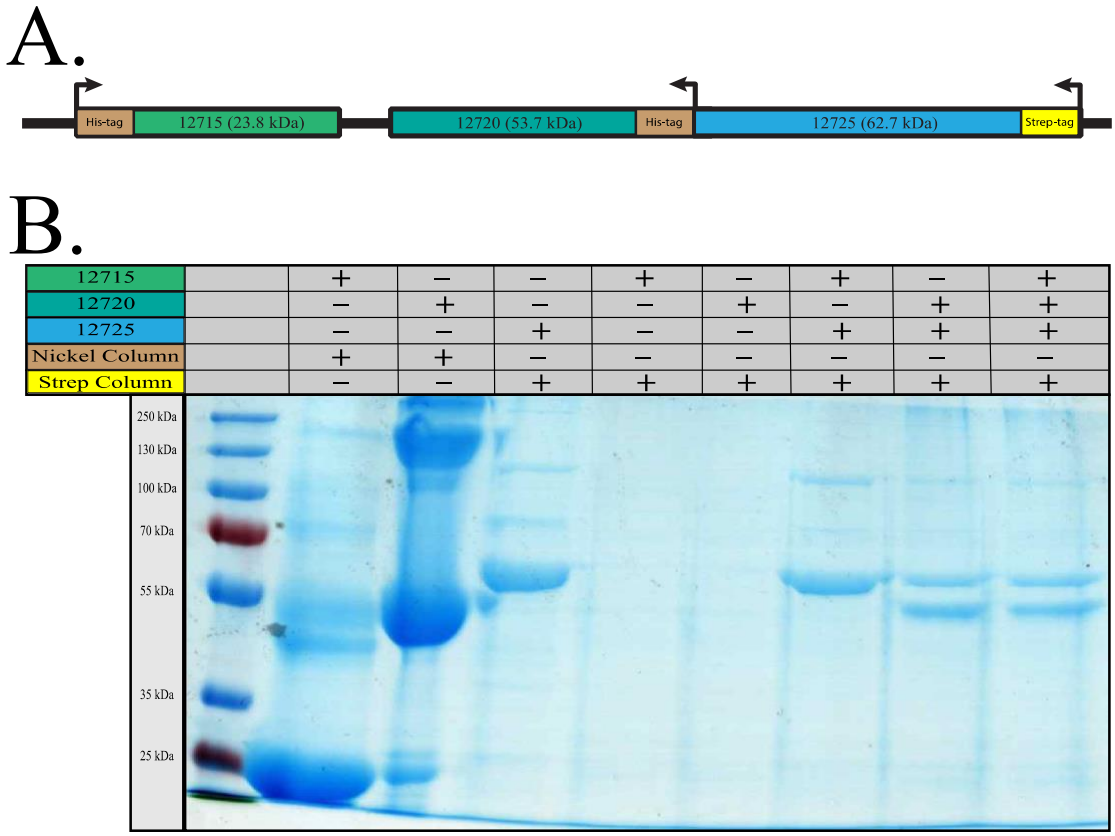

Figure 2.4. PaAgo specifically associates with one of the proteins encoded by an adjacent gene. A: Schematic shows PaAgo cassette with affinity tags annotated as well as molecular weight of each protein. B: SDS-PAGE gell showing results from pull-down assay. Gel legend is color coded to match cassette schematic.

To confirm this result, we performed size exclusion chromatography (SEC) on the three individually purified proteins and the mixture of all three (**Figure 2.5**). We found that PACS2_12715 elutes as a single peak corresponding to 31 kDa (**Figure 2.5, A)** and 12720 results in two peaks corresponding to 54 kDa and 31 kDa (**Figure 2.5, B**). 12725 displays two distinct peaks estimated at 86 kDa, and 192 kDa (**Figure 2.5, C**). The mixed sample with all three proteins results in three peaks with estimated molecular weights of 90 kDa, 180 kDa, and 354 kDa **(Figure 2.5, D)**. These results suggest that PACS2_12715

and 12720 purify as monomers (**Figure 2.5, A-B**). Additionally, it appears that PACS2_12720 undergoes some proteolysis upon purification, which can be seen by the peak corresponding to 31 kDa (**Figure 2.5, B**) Based on PACS2_12725 chromatogram presenting two peaks while showing a single band on an SDS-PAGE gel (**Figure 2.5, C**), it appears that 12725 purifies as a mixture of multimers made up of only 12725. The mixed sample of PACS2_12715, 12720, and 12725 similarly shows multiple peaks on its chromatogram while the SDS-PAGE of the SEC peaks show two bands corresponding to PACS2_12720 and 12725 (**Figure 2.5, D**). This confirms that PACS2_12720 and 12725 form a complex and suggests that the 12720/12725 complex forms multimers composed of both proteins.

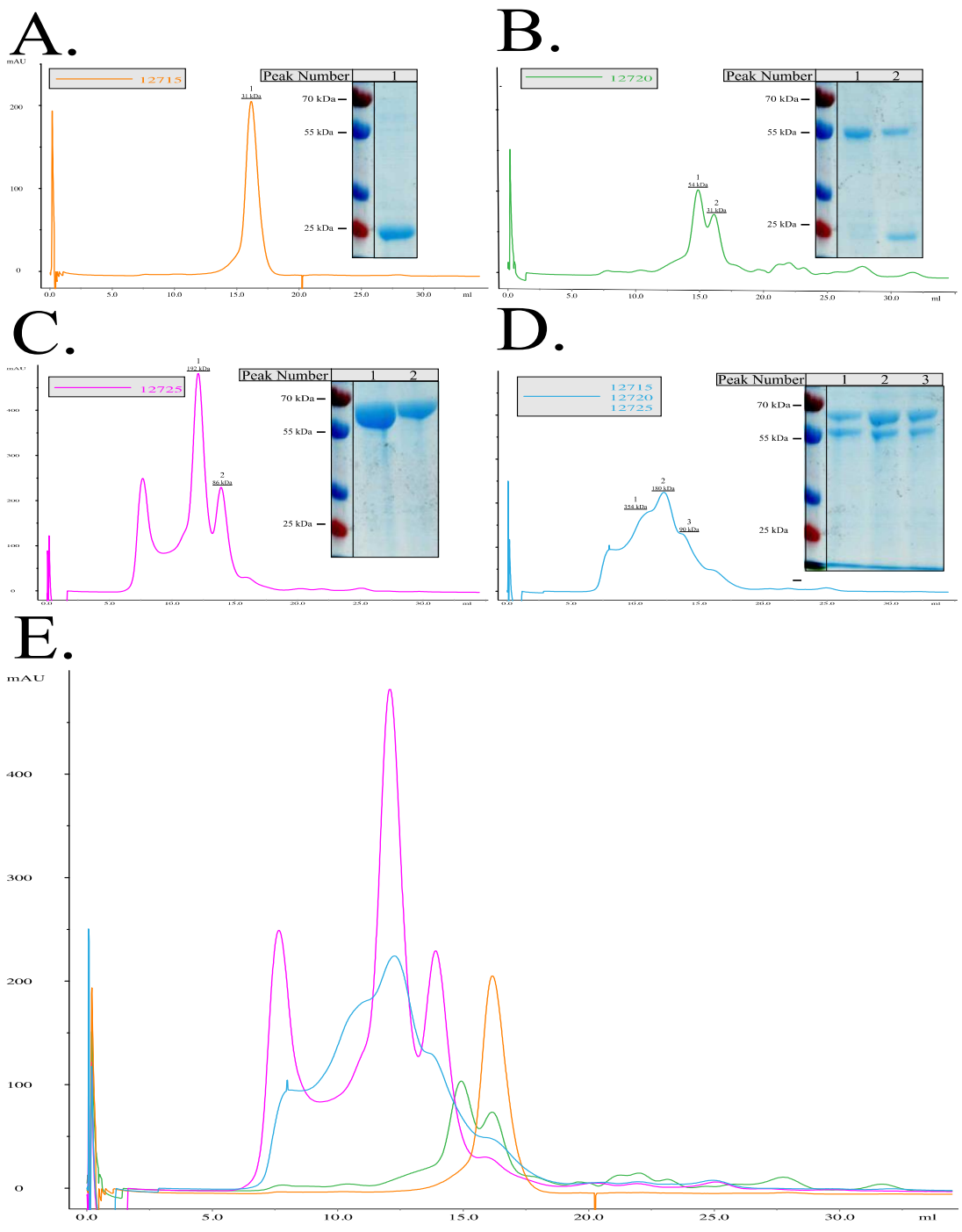

Figure 2.5. Chromatograms showing abs$_{280}$ of pull-down assay results from Figure 2.4 and denaturing SDS-PAGE gels of chromatogram peak fractions. Lane numbers for A-D correspond to peaks from chromatograms. A: Results from 12715 purified on nickel column. B: Results from 12720 purified on nickel column. C: Results from 12725 purified on strep-tactin column. D: Results of 12715, 12720, and 12725 combined and purified on strep-tactin column. E: Overlay of chromatograms from A-D.

Phylogenetic distribution of PaAgo cassette

PaAgo (PACS2_12720) contains both MID and PIWI domains while its neighboring protein, PACS2_12725, contains an N and PAZ domain. Our pull-down assays demonstrate that this short pAgo (PACS2_12720) combines with an adjacent argonaute domain-containing protein (PACS2_12725) to form a complex (**Figure 2.4 and 2.5**). To investigate whether the concurrence of PACS2_12720 and 12725 is conserved in other species, we searched for the presence of PaAgo cassette homologs in 17,660 sequenced genomes and mapped their presence on a phylogenetic tree (**Figure 2.6**). We find that 216 (1.2%) genomes encode homologs to PACS2_12715, while 444 (2.5%) and 546 (3.1%) have homologs for PACS2_12720 and 12725, respectively. Of the 444 genomes that encode a PaAgo homolog, 18% also encode a homolog to PACS2_12725, indicating there is not widespread co-occurrence of the two proteins. However, when looking exclusively at the genus *Pseudomonas*, 12720 and 12725 have a concurrence frequency of 63%. This suggests that these two proteins are encoded together more often than not and their ability to form a complex may be a common trait in *Pseudomonas* species.

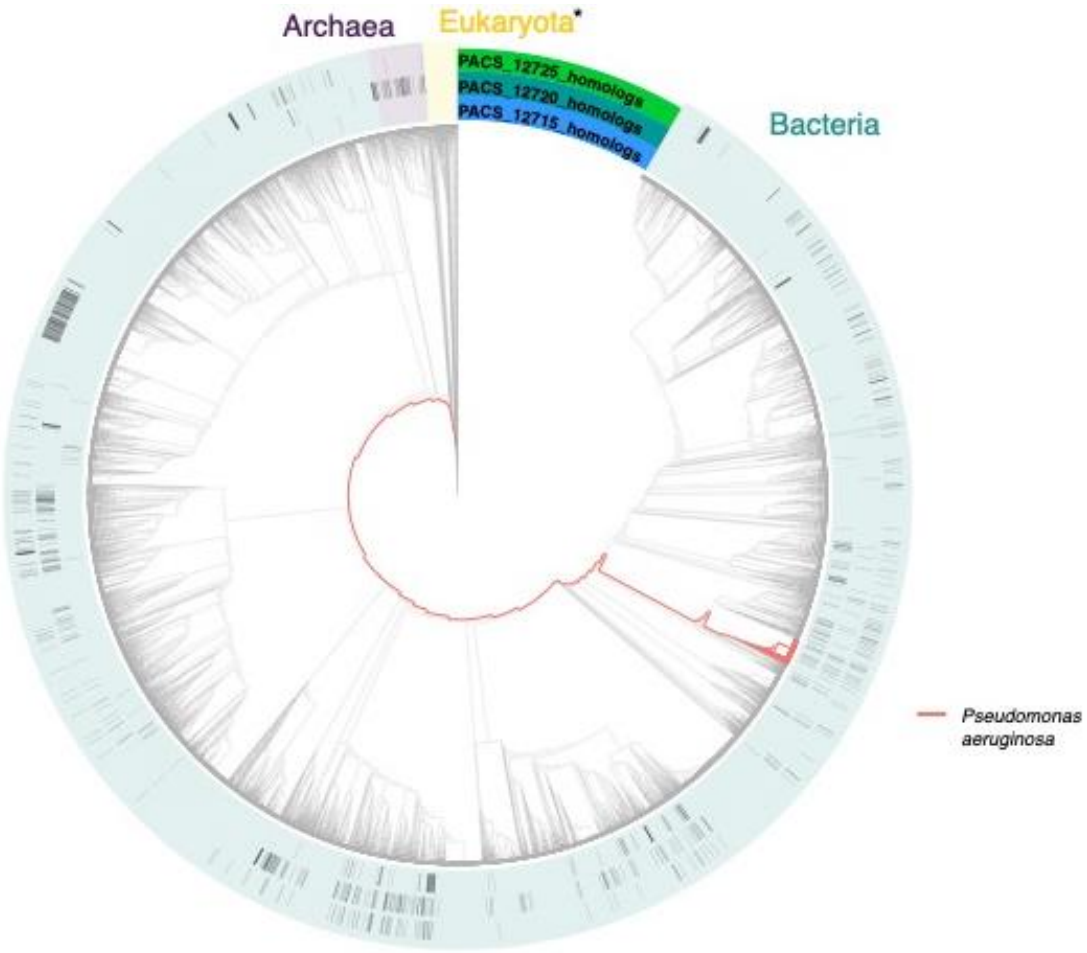

Figure 2.6. Phylogenetic tree showing the presence of PaAgo cassette homologs in bacterial, archaeal, and eukaryotic species. Red branches indicate *Pseudomonas aeruginosa* strains. Figure courtesy of Murat Büyükyörük.

Discussion

Short pAgos make up roughly 50% of all pAgos [3], yet functional studies of short pAgos are currently missing. In this study, we show that the short pAgo from *Pseudomonas aeruginosa* PACS2 is involved in regulation of transposases and its absence is detrimental to cell health. We also demonstrate that the PaAgo protein forms a complex with a neighboring protein (PACS2_12725). We speculate that PACS2_12725 is an APAZ-like protein and that the complex is necessary for biological activity of the the short PaAgo.

The ability for Agos to regulate transposases is a well-documented role in eukaryotes [2], and the ability to preferentially obtain guides complementary to transposons has been shown in pAgos as well [6]. These examples of transposon regulation have been limited to eukaryotic and long pAgos while our work suggests that this trend may also extend into short pAgos. Based on the increased amount of transposase transcripts in two ΔPaAgo mutants (Δ*ago::tran* and Δ*ago::gen*, **Figure 2.2**) it is likely that PaAgo (PACS2_12720) plays a role in transposase regulation. It appears that deletion of PaAgo (PACS2_12720) causes an increase in transposase transcripts while a deletion of all three genes (PACS2_12715-12725) returns transposase levels to roughly wild type (**Figure 2.2**). These results suggest that either PACS2_12715 or 12725 may raise transposase transcription while PACS2_12720 acts to lower transcript levels. Measuring transposase transcript levels from Δ12715 and Δ12725 single mutants will be critical to determine their role in transposase regulation. Whether PaAgo lowers the level of transcripts by targeting the PACS2 genome and preventing transcription, or if it

actively degrades mRNA transcripts remains unknown. Determining the guide preference (DNA or RNA) of PaAgo will also be an important step in determining its mechanism of regulation.

The toxicity that is observed in the absence of the PaAgo cassette is not unprecedented, as it has been shown that a Δ*ago* mutant of *T. thermophilus* is outcompeted by a wild type strain [5]. The 'plaquing' phenotype that PACS2 Δall (Δ12715-Δ12720-Δ12725) displays (**Figure 2.3**) may be due to the modest increase in transposase expression (**Figure 2.2**). The fact that plaquing is only seen in the Δall mutant, which has the smallest increase in transposase transcripts, and not in either the Δ*ago::gen* or Δ*ago::tran* suggest that increased transposase expression is not primarily responsible for toxicity. Instead, it is possible that the PaAgo cassette is regulating additional processes whose deregulation would lead to cell death. Sequencing PaAgo guides will be critical in order to determine targets and identify any other regulation roles that PaAgo plays in PACS2.

PACS2 also contains at least one putative prophage with strong homology to filamentous prophage Pf1[9]. It's possible that the plaquing phenotype may be due to increased excision of this or another endogenous prophage. The PaAgo cassette may act to suppress prophage activity and, if true, its deletion would lead to increased prophage induction and the uniform plaquing pattern that we observe (**Figure 2.3**). The next step will be to perform experiments aimed is isolating phages form the plates.

Binding partners appear to be an important factor in the action of pAgos. Even long-A pAgos, which have all active domains and should be able to act independently,

have been shown to interact with a suite of proteins [5], and the binding partners that pAgos use influence the biological functions they perform. Because short pAgos are missing the N and PAZ domains it has been speculated that they may use accessory proteins to perform their actions. Our data support this hypothesis by showing that PaAgo and its neighboring protein form a complex *in vitro* (**Figure 2.4**). Based on the structure of existing argonautes and the fact that PACS2_12720 and 12725 combine to possess all the domains of a full Ago, we hypothesize that these two proteins (PACS2_12720 and 12725) assemble around a nucleic acid guide and bind to one another at the MID-PAZ interface. Determining the structure of this complex will be necessary to confirm this hypothesis. It will also be important to determine the necessity of guide presence for complex formation.

12720-12725 complex formation also points to a possible strategy for short pAgos. PACS2_12720 contains only a MID and PIWI domain (**Figure 2.1**), classifying it as a short pAgo. The protein immediately upstream, 12725, contains both a PAZ and N domain and its association with 12720 indicates that the two proteins combine to form a "split" long-B pAgo. Based on the frequency of concurrence of these proteins in *Pseudomonas* species (~60%), it appears that 12720/12725 complex formation is important for PaAgo function in *Pseudomonas* sp. While there is very little concurrence between 12720 and 12725 outside of *Pseudomonas* sp. (18%), our analyses only probed for the presence of these specific proteins. Based on previous work showing that short pAgos are frequently encoded near N, PAZ, or APAZ domain-containing proteins [3], PACS2_12720 homologs present in non-Pseudomonads may be encoded near other

argonaute domain proteins that aren't related to PACS2_12725. Experiments desinged to identify N or PAZ domain proteins in genomes containing homologs to PaAgo will help to determine is "split" long pAgo formation is conserved. This is the first example of a short pAgo using an N or PAZ domain containing protein as a binding partner. It will be necessary to identify the binding partners of other short pAgos in order to determine if this is a common strategy for the clade.

REFERENCES CITED

1. Hauptmann, J. et al. Turning catalytically inactive human Argonaute proteins into active slicer enzymes. Nat. Struct. Mol. Biol. 20, 814–817 (2013).

2. Makarova, K. S., Wolf, Y. I., van der Oost, J. & Koonin, E. V. 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 (2009).

3. Ryazansky S, Kulbachinskiy A, Aravin AA. 2018. The expanded universe of prokaryotic Argonaute proteins. mBio 9:e01935-18.

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

5. Jolly et al., Thermus thermophilus Argonaute Functions in the Completion of DNA Replication, Cell (2020).

6. Olovnikov, I., Chan, K., Sachidanandam, R., Newman, D. K. & Aravin, A. A. Bacterial argonaute samples the transcriptome to identify foreign DNA. Mol. Cell 51, 594–605 (2013).

7. Pettersen EF, Goddard TD, Huang CC, Meng EC, Couch GS, Croll TI, Morris JH, Ferrin TE. UCSF ChimeraX: Structure visualization for researchers, educators, and developers. Protein Sci. (2020).

8. Chen, Vincent B et al. “MolProbity: all-atom structure validation for macromolecular crystallography.” Acta crystallographica. Section D, Biological crystallography vol. 66,Pt 1 (2010).

9. Mathee, Kalai et al. “Dynamics of Pseudomonas aeruginosa genome evolution.” Proceedings of the National Academy of Sciences of the United States of America vol. 105,8 (2008).

10. William Sealy Gosset . The probable error of a mean. Biometrika. **6** (1): 1–25. (1908).

11. Ender, C. & Meister, G. Argonaute proteins at a glance. J. Cell Sci. 123, 1819–1823 (2010).

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

13. Choi, KH., Schweizer, H.P. An improved method for rapid generation of unmarked Pseudomonas aeruginosa deletion mutants. BMC Microbiol 5, 30 (2005).

14. Underwood, Charles J, and Robert A Martienssen. “Argonautes team up to silence transposable elements in Arabidopsis.” *The EMBO journal* vol. 34,5 (2015).

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

REFERENCES CITED

Anna Olina, Anton Kuzmenko, Maria Ninova, Alexei A. Aravin, Andrey Kulbachinskiy & Daria Esyunina. Genome-wide DNA sampling by Ago nuclease from the cyanobacterium *Synechococcus elongatus*, RNA Biology, 17:5, 677-688 (2020).

Anton Kuzmenko, Denis Yudin, Sergei Ryazansky, Andrey Kulbachinskiy, Alexei A Aravin, Programmable DNA cleavage by Ago nucleases from mesophilic bacteria *Clostridium butyricum* and *Limnothrix rosea*, *Nucleic Acids Research*, Volume 47, Issue 11, 20 June 2019, Pages 5822–5836.

Chen, Vincent B et al. "MolProbity: all-atom structure validation for macromolecular crystallography." Acta crystallographica. Section D, Biological crystallography vol. 66,Pt 1 (2010).

Choi, KH., Schweizer, H.P. An improved method for rapid generation of unmarked Pseudomonas aeruginosa deletion mutants. BMC Microbiol 5, 30 (2005).

Christine Ender, Gunter Meister. Journal of Cell Science 2010 123: 1819-1823

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

Ender, C. & Meister, G. Argonaute proteins at a glance. J. Cell Sci. 123, 1819–1823 (2010).

Expression and purification of proteins using 6xHistidine-tag: a comprehensive manual. IBA Lifesciences. October 2012. Pages 19-20.

Faehnle, C.R., Elkayam, E., Haase, A.D., Hannon, G.J. & Joshua-Tor, L. The making of a slicer: activation of human Argonaute-1. Cell Reports 3, 1901–1909 (2013).

Hauptmann, J. et al. Turning catalytically inactive human Argonaute proteins into active slicer enzymes. Nat. Struct. Mol. Biol. 20, 814–817 (2013).

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

Jolly et al., *Thermus thermophilus* Argonaute Functions in the Completion of DNA Replication, Cell (2020).

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

Lisitskaya L, Aravin AA, Kulbachinskiy A. DNA interference and beyond: structure and functions of prokaryotic Argonaute proteins. Nat Commun. 2018;9(1):5165. Published 2018 Dec 4.

Liu Y, Esyunina D, Olovnikov I, Teplova M, Kulbachinskiy A, Aravin AA, Patel DJ. Accommodation of helical imperfections in *Rhodobacter sphaeroides* Argonaute ternary complexes with guide RNA and target DNA. Cell Rep 24:453– 462. (2018).

Livak KJ, Schmittgen TD. Methods. Vol. 25. San Diego, CA: 2001. Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method; pp. 402–408.

Ma JB, et al. Structural basis for 5′-end-specifc recognition of guide RNA by the *A. fulgidus* Piwi protein. Nature. 2005;434:666–670

Makarova KS, et al. Ancestral paralogs and pseudoparalogs and their role in the emergence of the eukaryotic cell. Nucleic Acids Res. 2005;33:4626–4638.

Makarova, K. S., Wolf, Y. I., van der Oost, J. & Koonin, E. V. 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 (2009).

Mathee, Kalai et al. “Dynamics of Pseudomonas aeruginosa genome evolution.” Proceedings of the National Academy of Sciences of the United States of America vol. 105,8 (2008).

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

Olovnikov, I., Chan, K., Sachidanandam, R., Newman, D. K. & Aravin, A. A. Bacterial argonaute samples the transcriptome to identify foreign DNA. Mol. Cell 51, 594–605 (2013).

Parker, J. S., Roe, S. M. & Barford, D. Crystal structure of a PIWI protein suggests mechanisms for siRNA recognition and slicer activity. EMBO J. 23, 4727–4737 (2004).

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

Parker, J. S., Parizotto, E. A., Wang, M., Roe, S. M. & Barford, D. Enhancement of the seed-target recognition step in RNA silencing by a PIWI/MID domain protein. Mol. Cell 33, 204–214 (2009)

Pettersen EF, Goddard TD, Huang CC, Meng EC, Couch GS, Croll TI, Morris JH, Ferrin TE. UCSF ChimeraX: Structure visualization for researchers, educators, and developers. *Protein Sci.* (2020)

Rashid, U. J. et al. Structure of *Aquifex aeolicus* argonaute highlights conformational flexibility of the PAZ domain as a potential regulator of RNA-induced silencing complex function. J. Biol. Chem. 282, 13824–13832 (2007).

Rivas, F. V. et al. Purified Argonaute2 and an siRNA form recombinant human RISC. Nat. Struct. Mol. Biol. 12, 340–349 (2005).

Ryazansky S, Kulbachinskiy A, Aravin AA. 2018. The expanded universe of prokaryotic Argonaute proteins. mBio 9:e01935-18.

Shabalina, Svetlana A, and Eugene V Koonin. "Origins and evolution of eukaryotic RNA interference." Trends in ecology & evolution vol. 23,10 (2008): 578-87.

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

Song, J. J., Smith, S. K., Hannon, G. J. & Joshua-Tor, L. Crystal structure of Argonaute and its implications for RISC slicer activity. Science 305, 1434–1437 (2004).

Swarts DC, Makarova K, Wang Y, et al. The evolutionary journey of Argonaute proteins. Nat Struct Mol Biol. 2014;21:743–753.

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

Swarts, D. C., Jore, M. M. and Oost, J. v. (2014). Expression and Purification of the *Thermus thermophilus* Argonaute Protein. *Bio-protocol* 4(19): e1253

Schwarz, D.S. et al. Asymmetry in the assembly of the RNAi enzyme complex. Cell 115, 199–208 (2003).

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

Swarts, D. C., Koehorst, J. J., Westra, E. R., Schaap, P. J. & van der Oost, J. Effects of Argonaute on gene expression in *Thermus thermophilus*. PLoS One 10, e0124880 (2015).

Swarts, D. C. et al. Autonomous generation and loading of DNA guides by bacterial Argonaute. Mol. Cell 65, 985–998 (2017).

Underwood, Charles J, and Robert A Martienssen. “Argonautes team up to silence transposable elements in Arabidopsis.” *The EMBO journal* vol. 34,5 (2015).

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

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

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

William Sealy Gosset . The probable error of a mean. Biometrika. **6** (1): 1–25. (1908).

Willkomm, S. et al. Structural and mechanistic insights into an archaeal DNA- guided Argonaute protein. Nat. Microbiol. 2, 17035 (2017).

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

Yuan, Y. R. et al. Crystal structure of *A. aeolicus* argonaute, a site-specific DNA-guided endoribonuclease, provides insights into RISC-mediated mRNA cleavage. Mol. Cell 19, 405–419 (2005).

Yuan, Y. R., Pei, Y., Chen, H. Y., Tuschl, T. & Patel, D. J. A potential protein-RNA recognition event along the RISC-loading pathway from the structure of *A. aeolicus* Argonaute with externally bound siRNA. Structure 14, 1557–1565 (2006).

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

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