AMERICAN SOCIETY FOR MICROBIOLOGY | Journal of Bacteriology®

COMMENTARY

Short, Rich, and Powerful: a New Family of Arginine-Rich Small Proteins Have Outsized Impact in *Agrobacterium tumefaciens*

Melene A. Alakavuklar,[a] Clay Fuqua[a]

[a]Department of Biology, Indiana University, Bloomington, Indiana, USA

ABSTRACT Due to minute size and limited sequence complexity, small proteins can be challenging to identify but are emerging as important regulators of diverse processes in bacteria. In this issue of the *Journal of Bacteriology*, Kraus and coworkers (A. Kraus, M. Weskamp, J. Zierles, M. Balzer, et al., J Bacteriol 202:e00309-20, 2020, https://doi.org/10.1128/JB.00309-20) report a comprehensive analysis of a fascinating subfamily of arginine-rich small proteins in *Agrobacterium tumefaciens*, conserved among *Alphaproteobacteria*. Their findings reveal that these small proteins are under complex regulation and have a disproportionately large impact on metabolism and behavior.

KEYWORDS *Agrobacterium tumefaciens*, *Alphaproteobacteria*, domain of unknown function, metabolic control, regulatory RNA, small proteins, biofilms

The issue of scale has a major impact on our ability to recognize important constituents in the world around us. Objects of small scale may have a tremendous impact but can be easily obscured by the larger context in which they reside. An example of this phenomenon, among many, is the relatively meager recognition and understanding of small proteins compared to that of their larger, more readily identifiable counterparts. Small proteins are encoded by short coding sequences that specify 50 amino acids or fewer and often lack clearly annotated domains or motifs indicative of their structure or function. This makes identification and characterization of genes for small proteins in genomic sequences more challenging than for larger proteins. Recent advances have led to the prediction and initial characterization of a growing number of small proteins in *Escherichia coli* and other bacteria (1, 2). Emerging evidence suggests that, often, small proteins interact with cellular membranes or protein complexes and can play influential roles in regulating diverse processes, such as sugar transport, protection from phage invasion, and regulation of sensor kinase function (2–4).

A related but even broader challenge is the prediction of protein function from sequence information. Even for larger proteins possessing predicted secondary structure motifs and recognized domains, a challenge remains in characterizing genes with domains of unknown function (DUFs). More than 20% of protein families in the Pfam database are annotated as DUFs (5). Many DUFs are highly conserved and may serve important or even essential functions (6), but more studies into the structure and function of these domains are needed to understand their cellular roles.

Recent work has identified a family of conserved arginine-rich proteins, classified as DUF1127, that are typically quite short (7). Almost all DUF1127 domain-containing proteins are composed of single domains and are fewer than 100 amino acids (aa) in length, with approximately 15% of these under 51 aa, that can be classified as true small proteins. DUF1127 is found primarily in members of the *Alphaproteobacteria* and

Citation Alakavuklar MA, Fuqua C. 2020. Short, rich, and powerful: a new family of arginine-rich small proteins have outsized impact in *Agrobacterium tumefaciens*. J Bacteriol 202:e00450-20. https://doi.org/10.1128/JB.00450-20.

Editor Anke Becker, Philipps University Marburg

Address correspondence to Clay Fuqua, cfuqua@indiana.edu.

For the article discussed see https://doi.org/10.1128/JB.00309-20.

The views expressed in this article do not necessarily reflect the views of the journal or of ASM.

Accepted manuscript posted online 24 August 2020
Published 22 October 2020

Gammaproteobacteria (7, 8). The *Gammaproteobacteria* typically harbor a single DUF1127 protein, annotated as YjiS, the function of which has not been reported (9).

In this issue of the *Journal of Bacteriology*, Kraus et al. report compelling initial analysis of multiple DUF1127-containing proteins in the plant pathogen *Agrobacterium tumefaciens*, providing new insights into their functions and distribution among the members of the *Alphaproteobacteria* (8). There have been a few prior studies on the *Alphaproteobacteria* DUF1127 proteins. The DUF1127 protein RSP_6037 in *Rhodobacter sphaeroides* is in an operon with and affects expression of four small RNAs (sRNAs), with the whimsical name of the cuckoo RNAs (CcsR1 to CcsR4 [CcsR1-R4]) for their conserved sequence (CCUCCUCCC) motif (10, 11). The CcsR RNAs and RSP_6037 regulate C_1 carbon metabolism (10). Expression of RSP_6037 and another DUF1127-containing gene, RSP_0557, are induced in the transition from exponential to stationary phase in the presence of oxygen (12). In *Brucella abortus*, three DUF1127 proteins are activated by VtlR (13), a LysR-type regulatory protein that is required for virulence (14).

Rather than relying on genome annotations, which often overlook small proteins (1), Kraus et al. (8) took an unbiased approach to identify *Agrobacterium tumefaciens* proteins containing the DUF1127 domain. *A. tumefaciens* is a facultative plant pathogen and the causative agent of crown gall disease (15, 16). To find potential *A. tumefaciens* DUF1127 proteins, Kraus et al. (8) translated all possible reading frames from the *A. tumefaciens* C58 genome sequence and used BLAST to search the *in silico* amino acid sequence against known DUF1127 proteins (17). This targeted approach led to the identification of seven proteins in *A. tumefaciens* containing the arginine-rich sequence of DUF1127. Of these seven, four were previously annotated to contain DUF1127, two were listed as hypothetical proteins of unknown function, and one was encoded by a newly annotated open reading frame (ORF) that had formerly been described as an sRNA. Three of the *A. tumefaciens* DUF1127 proteins are true small proteins (<50 aa), and are designated short DUF1127 proteins (SDPs), and the other four are considered long DUF1127 proteins (LDPs), ranging from 72 to 101 aa. Kraus et al. (8) observed that the short and long DUF1127 proteins in *A. tumefaciens* appeared to fall into distinct protein groups and asked whether other DUF1127 proteins in the InterPro database fall into similar subdivisions (7). Indeed, they found three distinct subclasses of DUF1127 proteins: those similar to the *A. tumefaciens* SDPs that were found mostly in *Alphaproteobacteria*, those similar to the *E. coli* YjiS protein and largely restricted to the *Gammaproteobacteria*, and a third group, including the *A. tumefaciens* LDPs distinct from the first two categories. Kraus et al. (8) also observed that in the *Alphaproteobacteria*, the DUF1127 SDP proteins are almost always associated with a gene encoding an LsrB-like regulatory protein and cuckoo sRNAs. LsrB is a LysR-type protein and is homologous to VltR from *B. abortus*. Prior studies on the *A. tumefaciens* LsrB protein found that it stimulates biofilm formation and attachment to host plants to drive infection, and the Δ*lsrB* mutant is deficient in plant transformation (18). Kraus et al. (8) tested whether deletion of *lsrB* affects DUF1127 protein expression and observed that the absence of LsrB decreases expression of the SDPs and reciprocally increases LDP expression. During growth in culture, each SDP is preferentially expressed at different stages of the growth curve, but they all appear to be induced at elevated temperature, suggesting some common regulation as well as some specific control of each SDP.

Kraus et al. (8) investigated the role for the SDPs in *A. tumefaciens* physiology. Although individual deletion of the SDP genes had little obvious phenotypic effect, simultaneous deletion of all three SDP genes resulted in a profound growth defect during the transition to stationary phase. The differential in growth between the triple SDP mutant and the wild type was dependent on levels of both sucrose and NaCl in the medium, reflecting potential connections to metabolism and perhaps osmotic stress. Kraus et al. (8) looked for changes in protein abundance between wild-type and the triple SDP mutant cell extracts by one-dimensional SDS-PAGE analysis. They observed an increase in one obvious band in protein extracts from the triple SDP mutant in late exponential phase, which they identified through mass spectrometry to contain the phosphate ABC transporter component PtsS and NADP-dependent alcohol dehydro-

genase Adh. Strong induction of the corresponding genes in the triple SDP mutant was confirmed by Northern blot analysis. Transcriptome sequencing (RNA-seq) analysis comparing the transcripts of the wild-type strain to that of the triple SDP mutant revealed that, strikingly, thousands of genes were differentially expressed in the triple SDP mutant at various time points in the growth cycle. Among the genes upregulated in the triple SDP mutant at all growth phases were the *soxBDAG* and *glyA* genes involved in glycine/serine homeostasis and C_1 metabolism and the *norDBC* genes involved in denitrification. Regulation of carbon metabolism has been linked to SDPs in other *Alphaproteobacteria*, as deletion of a *B. abortus* SDP gene affects fucose utilization, and in *R. sphaeroides*, overexpression of the cuckoo sRNAs with an SDP modulates C_1 metabolism (10, 13). Genes associated with the phosphorus starvation response, including *ptsS* and *phoB*, were upregulated in the triple SDP mutant in late exponential phase. Two LDP genes were upregulated in the triple SDP mutant, suggesting a genetic interaction between the two classes of DUF1127 proteins.

Deletion of all three SDPs led to an increase in cell-cell aggregation and biofilm formation. Colonies of this mutant are elevated for exopolysaccharides that bind the dye Congo red (predominantly cellulose and the unipolar polysaccharide) and decreased for binding of calcofluor white, indicative of decreases in succinoglycan (19, 20). This pattern of polysaccharide production is also consistent with elevated biofilm formation in *A. tumefaciens*. The RNA-seq results suggested that the triple SDP mutant mimics cells undergoing phosphorus limitation, a condition that has been shown to trigger elevated attachment and biofilm formation in *A. tumefaciens*, and may be at the root of this observed phenotype (21, 22). *A. tumefaciens* is well known as a plant pathogen that genetically transforms plants by DNA transfer (16). Kraus et al. (8) report that when assayed using a qualitative *Arabidopsis* seedling infection test, the triple SDP mutant retained *A. tumefaciens* virulence; in contrast, the transcriptional activator of the SDPs, LsrB, is required for efficient transformation (18). Therefore, LsrB likely regulates additional cellular components that influence plant transformation and virulence. This result is reminiscent of data from *B. abortus* suggesting that, while the regulator protein VtlR is required for virulence, deletion of the SDPs does not significantly influence *B. abortus* virulence (13). There may be yet-undiscovered conserved LsrB (VtlR)-activated genes or pathways that participate more directly in these virulence programs. Deletion of all the three SDP genes led to increased expression of the cuckoo sRNA L5, immediately downstream of the gene for SDP3, suggesting a genetic interaction between the small proteins and the sRNA. The roles of this and the other cuckoo sRNAs in *A. tumefaciens* remain unclear.

Kraus et al. (8) provide new insight into the different classes of DUF1127 proteins in bacteria and, particularly, the role of the SDP-type DUF1127 proteins in *A. tumefaciens*. This study lays the groundwork for investigations into the mechanism by which these tiny proteins impact large-scale phenotypes such as growth, metabolism, and biofilm formation. While the single SDP mutants did not have strong phenotypes, suggesting redundancy, the loss of all three SDP proteins led to dramatic changes in growth, metabolism, and biofilm formation. It is interesting that the four LDPs of *A. tumefaciens* are oppositely impacted in the *lsrB* mutant relative to the SDPs, but the function of these proteins has not been characterized. The strong positive charge imparted through the arginine-rich sequence of DUF1127 proteins, particularly in the SDPs (8), may facilitate interactions with RNA, consistent with their proximity to and regulatory impact on the cuckoo sRNAs. Interestingly, the three SDPs of *B. abortus* localized to the membrane in cell fractionation experiments, and this may be a conserved attribute of the SDPs (13). The positive charge of the SDPs could also mediate interactions with negatively charged membrane phospholipids. These two potential interaction targets are not mutually exclusive. Future work should build from the current studies and reveal how these fascinating small proteins are integrated into the control of cellular physiology and how they impart their far-reaching effects in *A. tumefaciens* and other members of the *Alphaproteobacteria*.

ACKNOWLEDGMENT

Research on *Agrobacterium* physiology, attachment, and biofilm formation in the Fuqua lab is supported by National Institutes of Health grant GM120337.

REFERENCES

1. Hemm MR, Weaver J, Storz G. 2020. *Escherichia coli* small proteome. EcoSal Plus 9:ESP.0031.2019. https://doi.org/10.1128/ecosalplus.ESP-0031-2019.
2. Storz G, Wolf YI, Ramamurthi KS. 2014. Small proteins can no longer be ignored. Annu Rev Biochem 83:753–777. https://doi.org/10.1146/annurev-biochem-070611-102400.
3. Ragunathan PT, Vanderpool CK. 2019. Cryptic-prophage-encoded small protein DicB protects *Escherichia coli* from phage infection by inhibiting inner membrane receptor proteins. J Bacteriol 201:e00475-19. https://doi.org/10.1128/JB.00475-19.
4. Yadavalli SS, Goh T, Carey JN, Malengo G, Vellappan S, Nickels BE, Sourjik V, Goulian M, Yuan J. 2020. Functional determinants of a small protein controlling a broadly conserved bacterial sensor kinase. J Bacteriol 202:e00305-20. https://doi.org/10.1128/JB.00305-20.
5. Mudgal R, Sandhya S, Chandra N, Srinivasan N. 2015. De-DUFing the DUFs: deciphering distant evolutionary relationships of domains of unknown function using sensitive homology detection methods. Biol Direct 10:38. https://doi.org/10.1186/s13062-015-0069-2.
6. Goodacre NF, Gerloff DL, Uetz P. 2014. Protein domains of unknown function are essential in bacteria. mBio 5:e00744-13. https://doi.org/10.1128/mBio.00744-13.
7. Mitchell AL, Attwood TK, Babbitt PC, Blum M, Bork P, Bridge A, Brown SD, Chang H-Y, El-Gebali S, Fraser MI, Gough J, Haft DR, Huang H, Letunic I, Lopez R, Luciani A, Madeira F, Marchler-Bauer A, Mi H, Natale DA, Necci M, Nuka G, Orengo C, Pandurangan AP, Paysan-Lafosse T, Pesseat S, Potter SC, Qureshi MA, Rawlings ND, Redaschi N, Richardson LJ, Rivoire C, Salazar GA, Sangrador-Vegas A, Sigrist CJA, Sillitoe I, Sutton GG, Thanki N, Thomas PD, Tosatto SCE, Yong S-Y, Finn RD. 2019. InterPro in 2019: improving coverage classification and access to protein sequence annotations. Nucleic Acids Res 47:D351–D360. https://doi.org/10.1093/nar/gky1100.
8. Kraus A, Weskamp M, Zierles J, Balzer M, Busch R, Eisfeld J, Lambertz J, Nowaczyk MM, Narberhaus F. 2020. Arginine-rich small proteins with a domain of unknown function, DUF1127, play a role in phosphate and carbon metabolism of *Agrobacterium tumefaciens*. J Bacteriol 202:e00309-20. https://doi.org/10.1128/JB.00309-20.
9. Sibley MH, Raleigh EA. 2004. Cassette-like variation of restriction enzyme genes in *Escherichia coli* C and relatives. Nucleic Acids Res 32:522–534. https://doi.org/10.1093/nar/gkh194.
10. Billenkamp F, Peng T, Berghoff BA, Klug G. 2015. A cluster of four homologous small RNAs modulates C_1 metabolism and the pyruvate dehydrogenase complex in *Rhodobacter sphaeroides* under various stress conditions. J Bacteriol 197:1839–1852. https://doi.org/10.1128/JB.02475-14.
11. Reinkensmeier J, Giegerich R. 2015. Thermodynamic matchers for the construction of the cuckoo RNA family. RNA Biol 12:197–207. https://doi.org/10.1080/15476286.2015.1017206.
12. McIntosh M, Eisenhardt K, Remes B, Konzer A, Klug G. 2019. Adaptation of the alphaproteobacterium *Rhodobacter sphaeroides* to stationary phase. Environ Microbiol 21:4425–4445. https://doi.org/10.1111/1462-2920.14809.
13. Budnick JA, Sheehan LM, Kang L, Michalak P, Caswell CC. 2018. Characterization of three small proteins in *Brucella abortus* linked to fucose utilization. J Bacteriol 200:e00127-18. https://doi.org/10.1128/JB.00127-18.
14. Sheehan LM, Budnick JA, Blanchard C, Dunman PM, Caswell CC. 2015. A LysR-family transcriptional regulator required for virulence in *Brucella abortus* is highly conserved among the α-proteobacteria. Mol Microbiol 98:318–328. https://doi.org/10.1111/mmi.13123.
15. Escobar MA, Dandekar AM. 2003. *Agrobacterium tumefaciens* as an agent of disease. Trends Plant Sci 8:380–386. https://doi.org/10.1016/S1360-1385(03)00162-6.
16. Nester EW. 2015. Agrobacterium: nature's genetic engineer. Front Plant Sci 5:730. https://doi.org/10.3389/fpls.2014.00730.
17. Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, Madden TL. 2009. BLAST+: architecture and applications. BMC Bioinformatics 10:421–429. https://doi.org/10.1186/1471-2105-10-421.
18. Tang G, Li Q, Xing S, Li N, Tang Z, Yu L, Yan J, Li X, Luo L. 2018. The LsrB protein is required for *Agrobacterium tumefaciens* interaction with host plants. Mol Plant Microbe Interact 31:951–961. https://doi.org/10.1094/MPMI-02-18-0041-R.
19. Xu J, Kim J, Koestler BJ, Choi JH, Waters CM, Fuqua C. 2013. Genetic analysis of *Agrobacterium tumefaciens* unipolar polysaccharide production reveals complex integrated control of the motile-to-sessile switch. Mol Microbiol 89:929–948. https://doi.org/10.1111/mmi.12321.
20. Tomlinson AD, Ramey-Hartung B, Day TW, Merritt PM, Fuqua C. 2010. *Agrobacterium tumefaciens* ExoR represses succinoglycan biosynthesis and is required for biofilm formation and motility. Microbiology 156:2670–2681. https://doi.org/10.1099/mic.0.039032-0.
21. Danhorn T, Hentzer M, Givskov M, Parsek MR, Fuqua C. 2004. Phosphorus limitation enhances biofilm formation of the plant pathogen *Agrobacterium tumefaciens* through the PhoR-PhoB regulatory system. J Bacteriol 186:4492–4501. https://doi.org/10.1128/JB.186.14.4492-4501.2004.
22. Xu J, Kim J, Danhorn T, Merritt PM, Fuqua C. 2012. Phosphorus limitation increases attachment in *Agrobacterium tumefaciens* and reveals a conditional functional redundancy in adhesin biosynthesis. Res Microbiol 163:674–684. https://doi.org/10.1016/j.resmic.2012.10.013.