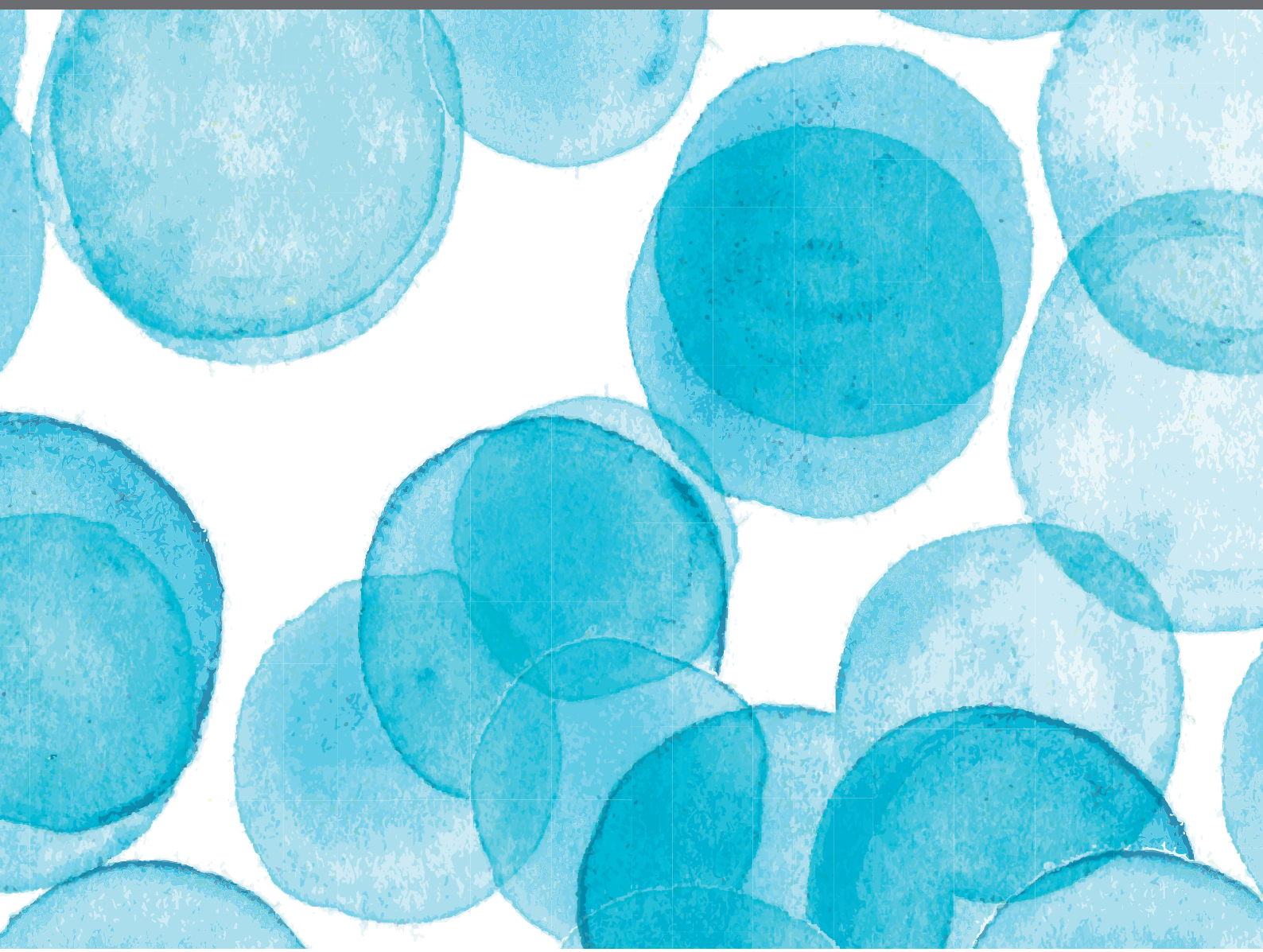
INSIGHTS IN MICROBIAL PHYSIOLOGY AND METABOLISM: 2021

EDITED BY: Biswarup Mukhopadhyay, Ulrike Kappler, Katarzyna Potrykus, Xinqing Zhao and Cornelia Welte

PUBLISHED IN: Frontiers in Microbiology



frontiers Research Topics

frontiers

ISSN 1664-8714
ISBN 978-2-83250-180-1
DOI 10.3389/978-2-83250-180-1

About Frontiers

Frontiers is more than just an open-access publisher of scholarly articles: it is a pioneering approach to the world of academia, radically improving the way scholarly research is managed. The grand vision of Frontiers is a world where all people have an equal opportunity to seek, share and generate knowledge. Frontiers provides immediate and permanent online open access to all its publications, but this alone is not enough to realize our grand goals.

Frontiers Journal Series

The Frontiers Journal Series is a multi-tier and interdisciplinary set of open-access, online journals, promising a paradigm shift from the current review, selection and dissemination processes in academic publishing. All Frontiers journals are driven by researchers for researchers; therefore, they constitute a service to the scholarly community. At the same time, the Frontiers Journal Series operates on a revolutionary invention, the tiered publishing system, initially addressing specific communities of scholars, and gradually climbing up to broader public understanding, thus serving the interests of the lay society, too.

Dedication to Quality

Each Frontiers article is a landmark of the highest quality, thanks to genuinely collaborative interactions between authors and review editors, who include some of the world's best academicians. Research must be certified by peers before entering a stream of knowledge that may eventually reach the public - and shape society; therefore, Frontiers only applies the most rigorous and unbiased reviews.
Frontiers revolutionizes research publishing by freely delivering the most outstanding research, evaluated with no bias from both the academic and social point of view. By applying the most advanced information technologies, Frontiers is catapulting scholarly publishing into a new generation.

What are Frontiers Research Topics?

Frontiers Research Topics are very popular trademarks of the Frontiers Journals Series: they are collections of at least ten articles, all centered on a particular subject. With their unique mix of varied contributions from Original Research to Review Articles, Frontiers Research Topics unify the most influential researchers, the latest key findings and historical advances in a hot research area! Find out more on how to host your own Frontiers Research Topic or contribute to one as an author by contacting the Frontiers Editorial Office: frontiersin.org/about/contact

INSIGHTS IN MICROBIAL PHYSIOLOGY AND METABOLISM: 2021

Topic Editors:
Biswarup Mukhopadhyay, Virginia Tech, United States
Ulrike Kappler, The University of Queensland, Australia
Katarzyna Potrykus, University of Gdansk, Poland
Xinqing Zhao, Shanghai Jiao Tong University, China
Cornelia Welte, Radboud University, Netherlands

Citation: Mukhopadhyay, B., Kappler, U., Potrykus, K., Zhao, X., Welte, C., eds. (2022). Insights in Microbial Physiology and Metabolism: 2021. Lausanne: Frontiers Media SA. doi: 10.3389/978-2-83250-180-1

Table of Contents

frontiers | Frontiers in Microbiology

TYPE Editorial
PUBLISHED 26 October 2022
DOI 10.3389/fmicb.2022.1054708

OPEN ACCESS

EDITED BY
Marina G. Kalyuzhanaya,
San Diego State University,
United States

REVIEWED BY
Christian Dusny,
Helmholtz Association of German Research Centres (HZ), Germany
Lisa Y. Stein,
University of Alberta, Canada
Shawn E. McGlynn,
Tokyo Institute of Technology, Japan

*CORRESPONDENCE
Ulrike Kappler
u.kappler@uq.edu.au

SPECIALTY SECTION
This article was submitted to Microbial Physiology and Metabolism, a section of the journal Frontiers in Microbiology

RECEIVED 27 September 2022
ACCEPTED 13 October 2022
PUBLISHED 26 October 2022

CITATION
Mukhopadhyay B, Kappler U, Potrykus K, Zhao X and Welte C (2022) Editorial: Insights in microbial physiology and metabolism: 2021. *Front. Microbiol.* 13:1054708. doi: 10.3389/fmicb.2022.1054708

Editorial: Insights in microbial physiology and metabolism: 2021

Biswarup Mukhopadhyay[1], Ulrike Kappler[2]*, Katarzyna Potrykus[3], Xinqing Zhao[4] and Cornelia Welte[5]

[1]Department of Biochemistry, Virginia Tech, Blacksburg, VA, United States, [2]School of Chemistry and Molecular Biosciences, The University of Queensland, Brisbane, QLD, Australia, [3]Department of Bacterial Molecular Genetics, University of Gdansk, Gdansk, Poland, [4]School of Life Sciences, Shanghai Jiao Tong University, Shanghai, China, [5]Department of Microbiology, Radboud University, Nijmegen, Netherlands

KEYWORDS

microorganisms, microbial metabolism, microbial physiology, enzymes, host—bacteria interaction, biotechnology

Editorial on the Research Topic
Insights in microbial physiology and metabolism: 2021

Metabolic diversity and adaptability drive the ability of microorganisms to colonize virtually all habitats on earth and to cope with extreme temperatures, pH and highly diverse nutrient profiles. Microbial metabolism is also emerging as a critical determinant of interactions between host organisms and the microorganisms associated with them and is also a resource as well as a tool in synthetic biology and biotechnology.

Despite decades of research, however, there is significant "metabolic dark matter" to be discovered in microorganisms, and this Research Topic highlights the many recent achievements in expanding our knowledge of microbial metabolism and discovering previously unknown, novel metabolic functions.

Several contributions to this Research Topic explore fundamental aspects of microbial physiology. Maintenance of cellular protein homeostasis is an essential process for all living cells. Matavacas and von Wachenfeldt have compiled the latest insights into this essential cellular function in the model bacterium *Bacillus subtilis* and also explore how protein homeostasis contributes to the survival of Gram-positive pathogens. Equally fundamental is the correct partitioning of DNA during cell division, where Mishra and Srinivasan have reviewed the latest insights into ParA ATPases.

Many bacterial cells contain alpha or beta carboxysomes, assemblies that contain Rubisco, carbonic anhydrase, and shell proteins and contribute to CO_2 fixation *via* Rubisco. USF Genomics Class 2020 have provided a review of "atypical carboxysomes" that lack one or more of these conserved carboxysome components. The evidence suggests that atypical carboxysomes give rise to functional assemblies and most likely have altered functions that depend on the elements present.

Stress responses are essential for the ability of bacteria to survive in adverse environments, and responses to nutrient starvation are central to bacterial survival. Hagberg et al. have investigated how *Sinorhizobium meliloti* responds to phosphate and

nitrogen stress while also aiming to understand the overlap of these two starvation responses. Interestingly, their work revealed only minimal overlap and concluded that phosphate stress is a more significant metabolic burden and takes precedence over nitrogen stress.

Stress responses were also a topic in papers that focussed on the physiology and metabolism of bacterial pathogens. Edelmann et al. have analyzed available data on the TisB Toxin/Antitoxin system that is thought to induce the formation of "persister cells" that are resistant to, for example, antibiotic-induced stress. The authors conclude that the expression pattern for TisB is more consistent with a different role, namely maintenance of the persister state once other cellular stress responses have induced it.

In an investigation of the role of metabolism in pathogenesis, Yue et al. found that in the human pathogen *Francisella tularensis*, arginine and polyamine metabolism positively correlate with hypervirulence, while less virulent strains frequently lacked relevant genes.

The largest groups of papers focused on metabolic and physiological processes relevant to biotechnology or organisms used in biotechnological applications.

Du et al. reveal a previously unrecognized function of the known carbon metabolism regulator Cat8 from budding yeast *Saccharomyces cerevisiae* in controlling nitrogen utilization. These results provide novel insights in regulation of yeast nutrient metabolism networks. Focussing on a specific group of enzymes, Bhayani et al. provide new insights into the catalytic properties of ADP-glucose pyrophosphorylases from various bacterial species and their reactivity toward glucosamine, which is relevant to future applications.

Commercial applications are often limited by the efficiency of relevant enzyme systems or a lack of access to substrates. A review by Henrion et al. examines the applicability of control theory in a broader context, namely to population-wide gene expression and phenotypic changes. Tao et al. analyzed the role of X2 modules in the cellulosomal scaffolding protein in *Clostridium cellulolyticum* where X2 modules play a role in elevating cellulose degradation efficiency by stimulating the binding between the bacterial cells and cellulose. This provides new perspectives for microbial engineering to improve degradation of cellulose.

Gendron and Allen reviewed current knowledge of archaeal methyl-/alkyl-coenzyme M reductases and provided insights into how heterologous expression of these enzymes could enhance biotechnological applications in methane and alkane production.

Lastly, the removal of CO or CO_2 from the atmosphere and their conversion into value-added products *via* the Wood-Ljungdahl pathway was reviewed by Lee et al.. The authors explore strategies to overcome growth limitations of acetogenic bacteria by engineering the Wood-Ljungdahl pathway, thereby increasing their potential as catalysts of this valuable transformation.

This collection of articles is an excellent example of the breadth and depth of research into microbial physiology and metabolism and its relevance to issues that range from fundamental research to applications in the health sciences and biotechnology industry. The research progress in this Research Topic provides novel insights in the field of microbial metabolism and metabolic engineering.

Author contributions

UK drafted the editorial text. All authors contributed to the editing of the draft and generation of the final editorial. All authors contributed to the article and approved the submitted version.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

frontiers in Microbiology

ORIGINAL RESEARCH
published: 27 January 2022
doi: 10.3389/fmicb.2022.800146

The *Sinorhizobium meliloti* Nitrogen Stress Response Changes Radically in the Face of Concurrent Phosphate Stress

*Kelly L. Hagberg[1], Jason P. Price[1], Svetlana N. Yurgel[1,2†] and Michael L. Kahn[1]**

[1] School of Molecular Biosciences, Institute of Biological Chemistry, Washington State University, Pullman, WA, United States, [2] Department of Plant, Food and Environmental Sciences, Dalhousie University, Truro, NS, Canada

OPEN ACCESS

Edited by:
Biswarup Mukhopadhyay, Virginia Tech, United States

Reviewed by:
Mario Alberto Martínez Núñez, Universidad Nacional Autónoma de México, Mexico
Lourdes Girard, National Autonomous University of Mexico, Mexico
Michael Frederick Dunn, National Autonomous University of Mexico, Mexico

***Correspondence:**
Michael L. Kahn
Kahn@wsu.edu

†Present address:
Svetlana N. Yurgel, Grain Legume Genetics and Physiology Research Unit, USDA, ARS, Prosser, WA, United States

Specialty section:
This article was submitted to Microbial Physiology and Metabolism, a section of the journal Frontiers in Microbiology

Received: *22 October 2021*
Accepted: *06 January 2022*
Published: *27 January 2022*

Citation:
Hagberg KL, Price JP, Yurgel SN and Kahn ML (2022) The Sinorhizobium meliloti Nitrogen Stress Response Changes Radically in the Face of Concurrent Phosphate Stress. Front. Microbiol. 13:800146. doi: 10.3389/fmicb.2022.800146

Expression of hundreds of *S. meliloti* genes changed more than two-fold in response to either nitrogen or phosphate limitation. When these two stresses were applied together, stress responsive gene expression shifted dramatically. In particular, the nitrogen stress response in the presence of phosphate stress had only 30 of about 350 genes in common with the 280 genes that responded to nitrogen stress with adequate phosphate. Expression of sRNAs was also altered in response to these stresses. 82% of genes that responded to nitrogen stress also responded to phosphate stress, including 20 sRNAs. A subset of these sRNAs is known to be chaperoned by the RNA binding protein, Hfq. Hfq had previously been shown to influence about a third of the genes that responded to both nitrogen and phosphate stresses. Phosphate limitation influenced changes in gene expression more than nitrogen limitation and, when both stresses were present, phosphate stress sometimes reversed the direction of some of the changes induced by nitrogen stress. These nutrient stress responses are therefore context dependent.

Keywords: ***Sinorhizobium***, **nitrogen stress, phosphate stress, stress responses, P_{II} proteins**

INTRODUCTION

Soils vary substantially in characteristics that influence bacterial growth and survival, such as nutrient or water availability, salt content, pH, and the abundance of competing and predatory organisms (Zahran, 1999; Vitousek et al., 2010; Fanin et al., 2016). Bacteria have developed pathways to cope with environmental changes and can respond to various growing conditions in ways believed to optimize their success under stressful conditions. Low levels of usable nitrogen and phosphate are the most commonly limiting nutrients (Vitousek et al., 2010; Fanin et al., 2016) and the ability to cope with nitrogen or phosphate starvation is important. Rhizobia are best known for their nitrogen-fixing symbiotic relationships with legumes in which the bacteria convert atmospheric dinitrogen into ammonium, which can then be assimilated by the legumes and support plant growth. This chemically available nitrogen is an important input in natural and agricultural ecosystems (Bohlool et al., 1992; Mylona et al., 1995). Free-living bacteria initiate these symbioses through infection of roots, making persistence of rhizobia in diverse soils an important precondition for success of the relationships. Since several stresses can occur simultaneously in

natural ecosystems, understanding how stress response mechanisms affect each other is important in understanding bacterial stress adaptation.

Historically, the nitrogen stress response (NSR) and the phosphate stress response (PSR) have been characterized separately by limiting only the designated nutrient in an otherwise complete growth medium. The general concept underlying these studies is that gene expression changes observed in response to the nutrient stress help the organism cope with low levels of the limiting nutrient. The experimental design of single nutrient limitation also helps to define the regulatory circuits that mediate the stress response. Crosstalk between stress responses has been seen when they share common regulatory elements such as RNA polymerase sigma subunits or, in the case of two-component regulatory systems, when the histidine kinase sensor proteins are able to partially modify non-cognate response regulators (Mike et al., 2014; Choudhary et al., 2020).

Sinorhizobium meliloti is a symbiont of alfalfa and other *Medicago* species. Its nitrogen and phosphate stress responses have been studied individually and key proteins and genes have been characterized for each pathway (Al-Niemi et al., 1997; Voegele et al., 1997; Patriarca et al., 2002; Krol and Becker, 2004; Yurgel et al., 2012). The primary sensor for the NSR is GlnD, a uridylyltransferase that modifies the P_{II} proteins GlnB and GlnK in response to nitrogen availability (Merrick and Edwards, 1995; Ikeda et al., 1996; Jiang et al., 2012). P_{II} proteins are common among bacteria and their modification state affects how they regulate several NSR components, including phosphorylation of the NtrB/NtrC two-component regulatory system. NtrC-P works with the RpoN sigma factor to alter expression of genes involved in enhancing nitrogen uptake and efficiently utilizing available nitrogen (Merrick and Edwards, 1995; Arcondeguy et al., 2001; Leigh and Dodsworth, 2007). The activity of glutamine synthetase (GS), which can assimilate low levels of ammonium into amino acids by coupling assimilation to ATP hydrolysis, is well characterized and is often used as a marker of the activation state of the NSR. GS is abundant and highly active during nitrogen limitation and has lower abundance and specific activity when nitrogen levels are sufficient (Yurgel et al., 2010). *S. meliloti* has two major GS enzymes: GSI, which is encoded by *glnA* and regulated primarily through posttranslational modification (Atkinson and Ninfa, 1998; Yurgel et al., 2010), and GSII, which is encoded by *glnII* and primarily regulated through transcriptional control by the action of the NtrB/NtrC two-component regulatory system (Arcondéguy et al., 1997).

The *S. meliloti* phosphate stress response (PSR) is regulated by the PhoR/PhoB two-component regulatory system. When phosphate is limiting, the phosphorylated form of the PhoB response regulator (PhoB-P) works with the RpoD sigma factor to alter target gene expression with the apparent goal of enhancing uptake and efficient use of available phosphate while downregulating less critical processes that consume phosphate (Makino et al., 1993; Al-Niemi et al., 1997). Alkaline phosphatase (AP), which releases phosphate from phosphate esters, is often used as an indicator of PSR activity, similar to using GS as an index of the NSR. In *S. meliloti,* a major AP is encoded by *phoX,* and its expression depends on PhoB-P binding to the *phoX* promoter region (Zaheer et al., 2009). *S. meliloti* has several phosphate transport systems (Pst, Pho, and Pit) that are regulated by the PSR and differ in their affinity and capacity for phosphate uptake (Voegele et al., 1997; Yuan et al., 2006a).

Adding nitrogen or phosphate can improve growth of microorganisms within an environment and can change the microbiome population dynamics to favor either bacteria or fungi (Vitousek et al., 2010; Fanin et al., 2016). Adding both N and P has a synergistic effect on autotrophic microorganism growth (Elser et al., 2007), supporting the idea that stress responses are coordinated and that the responses to multiple nutrient limitations are integrated. This idea contrasts with a model in which the stress responses are independent pathways that are activated in response to single nutrients. In the latter model, a set of stress response genes would be regulated by the primary nutrient regardless of other stresses, but a handful of additional genes might be affected by imposing a second stress.

Recent studies in *S. meliloti* identified interactions between the NSR and PSR (Yurgel et al., 2013; Hagberg et al., 2016) as well as environmental signals affecting two-component systems including PhoB (Choudhary et al., 2020). A microarray experiment identifying nitrogen responsive genes revealed that deleting the *glnB* and *glnK* genes, which encode the P_{II} proteins, influenced expression of many more genes than are controlled by the NSR, including genes known to respond to phosphate limitation (Yurgel et al., 2013). Previous data also show that GS is regulated by phosphate in addition to nitrogen (Hagberg et al., 2016). In contrast, AP is regulated by phosphate, but nitrogen availability did not substantially influence AP activity or expression (Hagberg et al., 2016). In proteobacteria like *S. meliloti,* the cellular response to nitrogen is integrated with carbon metabolism and storage as well. The phosphotransferase system (PTS) regulates nitrogen and carbon pools to maintain balance; for example, when nitrogen is limiting, the PTS system will work with the NSR to enhance nitrogen uptake and metabolism while also altering carbon metabolism and storage (Sánchez-Cañizares et al., 2020). sRNAs have recently been shown to play a role in fine-tuning the balance of the nitrogen and carbon pools (Prasse and Schmitz, 2018).

The *E. coli* nitrogen and phosphate stress response pathways and regulatory proteins are very similar to the *S. meliloti* pathways described above. PhoB is stimulated by phosphate stress, but there is also stimulation by cell envelope stress which highlights the cross-regulation of the well-characterized PhoB regulatory protein (Choudhary et al., 2020). Recently, Sastry et al. (2019) described nitrogen and phosphate "imodulons" in *E. coli* as groups of genes whose expression moves in similar patterns. They speculated that bacteria respond to nutrient stresses in ways that integrate two major drivers of bacterial behavior: greed and fear. "Greed" describes the bacterial motivation to grow and includes changing expression of genes to favor reproduction in response to limited amounts of available nutrient. "Fear" describes the bacterial motivation to survive and includes changing expression of genes to allocate resources to essential and protective functions in order to avoid cell death. So, for example, limiting the amount of ammonia available to the culture may induce genes that allow it to use alternative forms of

nitrogen and, if successful, would allow the culture to resume growing using this alternate nutrient. Activating a generalized shutdown of related metabolic pathways, like the stringent response induced by amino acid starvation, might be considered a "fear" response that blocks immediate resource use in favor of a more conservative allocation.

The cross-regulation observed in *E. coli* and the response of GS expression and activity to both nitrogen and phosphate in *S. meliloti* led us to ask whether other genes, NSR-related or not, also respond to both nutrients. In addition, we were interested in how *S. meliloti* prioritized its responses to nitrogen and phosphate availability and to see if the speculated fear vs. greed priorities change as nutrient stresses stack on one another. An RNA-Seq experiment was designed to explore global changes in mRNA and sRNA abundance in response to limited nitrogen, limited phosphate, or both.

The results below demonstrate that, at least for the nitrogen stress response, the set of genes that responded to nitrogen stress in the presence of high phosphate was very different from the set that responded in a low phosphate environment, and that only a small fraction of the large number of "nitrogen-responsive genes" found could be characterized as nitrogen-responsive under both phosphate conditions. A similar pattern was observed with phosphate responsive genes in the presence or absence of nitrogen stress, but the overlap of the two groups of phosphate-responsive genes was larger. We briefly discuss genes that were both nitrogen and phosphate responsive, for which responding to phosphate limitation appeared to have higher priority than responding to nitrogen limitation. We speculate about the existence of two different stress responses to a single nutrient under the different stress conditions.

MATERIALS AND METHODS

Bacterial Strains and Media

Sinorhizobium meliloti Rm1021 *pstC*+ and Rm1021*glnBK pstC*+ strains (Hagberg et al., 2016) were used. Cultures were grown at 30°C in a shaking incubator in Minimal MOPS media with 1% mannitol as the carbon source, 0.4% NH_4^+ (high N) or 0.04% glutamate (low N) as nitrogen sources and 2 mM KH_2PO_4 (high P) or 50 μM KH_2PO_4 (low P) as phosphate sources. Induction of NSR and PSR by the low concentrations of these nutrients has been previously studied (Al-Niemi et al., 1997; Hagberg et al., 2016).

RNA Isolation, rRNA Depletion, cDNA Library Formation, and RNA-Seq

Cultures were grown in supplemented Min MOPS (Hagberg et al., 2016) medium to an A_{600} of 0.500 ± 0.026. Cells were harvested by centrifugation at 20,000 *g* for 10 min at 4°C. Cell pellets from 36 mL of culture were suspended in 200 μL lysis buffer [1× phosphate buffered saline (137 mM NaCl, 2.7 mM KCl, 7.75 mM Na_2HPO_4, 1.47 mM KH_2PO_4), 1 mM EDTA, and 0.1% Triton-X]. Cells were sonicated for 10 sec, placed on ice for 1 min, sonicated for 15 s, iced for one min, and sonicated for 15 s. Lysates were spun at 20,000 *g* for 15 min at 4°C and the supernatant was added to 1 mL TRIzol reagent (Thermo Fisher Scientific, Waltham, MA, United States). Standard procedures were followed for phase separation, RNA precipitation and RNA washing with TRIzol. RNA was further purified using the QIAGEN RNeasy kit. RNA was processed by using the Illumina TruSeq Stranded Total RNA kit following standard procedures at half reaction strength per sample to deplete the samples of rRNA and generate and enrich cDNA libraries with indexed adaptors. The cDNA libraries were validated by DNA fragment analysis. Standard Qubit 2.0 procedures were used to measure library sample concentrations which were normalized by qPCR prior to pooling them. Samples were sequenced with 50 bp, single end read conditions using an Illumina 2500 at the Washington State University (WSU) Genomics core facility in Spokane, WA, United States.

Data Analysis

FASTQ files from the RNA sequencing were analyzed using the Tuxedo Suite of programs. Reads were mapped with Tophat (Trapnell et al., 2012) to the closely related *S. meliloti* 2011 genome, which has a more complete annotation than the *S. meliloti* 1021 genome and includes non-coding sRNAs, and was then analyzed further with Cuffdiff, which allowed comparison of samples to determine significant expression changes between medium conditions (Trapnell et al., 2013). Mapped reads were analyzed with the R bioinformatics software WGCNA for information about weighted co-expression of genes (Langfelder and Horvath, 2008). The weighted co-expression data was visualized using Cytoscape to generate a co-expression network map (Shannon et al., 2003). The heatmap3 R package was used to generate a heatmap of significant differentially expressed genes (Zhao et al., 2014).

Experimental Design and Analytical Comparisons

To test the interactions between bacterial nitrogen and phosphate stress responses, four RNA-seq determinations were made in triplicate to allow pairwise comparisons of nitrogen stress under similar phosphate conditions and of phosphate stress under similar nitrogen conditions; these experiments are labeled as described in **Figure 1** and **Supplementary Figure 1** with, for example, NSRP designating the Nitrogen Stress Response under high phosphate and NSRp designating the Nitrogen Stress Response under low phosphate. Stress responses were defined as the sets of genes where expression changed two-fold or more between the stress condition and the non-stress condition. As indicated in **Supplementary Figure 1B**, different numbers of potential stress responsive genes might respond to the "same" stress under different experimental conditions.

Specifically, we measured RNA levels using RNA-seq under four conditions, minimal MOPS medium containing ammonium/2mM phosphate [NP], ammonium/50 μM phosphate [Np], glutamate/2mM phosphate [nP], and glutamate/50 μM phosphate [np]. Pairwise comparisons were made (**Supplementary Table 1**) to identify genes with significant ($q \leq 0.01$) two-fold changes in abundance of RNA

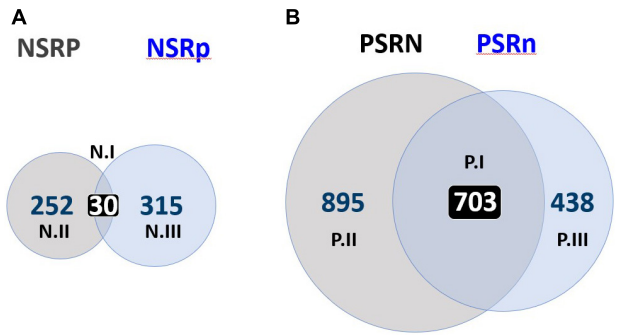


FIGURE 1 | Overview of gene expression changes under Nitrogen and Phosphate stress. Venn diagrams of genes that respond to nitrogen or phosphate stress in the presence or absence of the other stress. Stress responses are labeled by indicating primary nutrient comparison (i.e., NSR, nitrogen stress response or PSR, phosphate stress response) and then adding an upper- or lower-case letter for the other nutrient involved (i.e., NSRP for nitrogen stress response under high phosphate, NSRp for nitrogen stress response under low phosphate). **(A)** Venn diagram showing genes where mRNA abundance changed two-fold ($p \leq 0.01$) when comparing ammonia-grown cultures to glutamate–grown cultures under sufficient phosphate (gray = NSRP) and under limiting phosphate (blue = NSRp). **(B)** Genes where mRNA abundance changed two-fold ($p \leq 0.01$) when comparing 2 mM phosphate-grown cultures to 50 μM phosphate-grown cultures with ammonium as nitrogen source (gray = PSRN) or glutamate as the nitrogen source (blue = PSRn). Each section of the Venn diagram is labeled with the cluster name as defined in **Table 2** and the number of genes included in that cluster.

in response to nitrogen or phosphate availability. Genes were grouped based on their change in response to the primary nutrient, either nitrogen or phosphate, and then sorted further based on the availability of the second nutrient. NSRP or PSRN groups contained genes where there were significant changes in expression in response to limitation of one nutrient when the other nutrient was sufficient. NSRp or PSRn groups contained genes that changed in response to a nutrient limitation when the other nutrient was also limited. "Core NSR" or "Core PSR" groups contained genes where expression changed in response to one nutrient independent of the status of the other nutrient, e.g., induced >2× by nitrogen stress under *both* phosphate conditions. The results of this analysis are shown graphically in **Figure 1A** for the nitrogen stress responses NSRP and NSRp and in **Figure 1B** for the phosphate stress responses PSRN and PSRn. Genes that were upregulated when both nitrogen and phosphate were limited could include genes that are integral to the NSR or the PSR and specifically necessary to cope with either stress, but they could also include genes that are important for coping with a generalized reaction, such as decreased growth rate or a response to generalized nutrient stress.

RESULTS AND DISCUSSION

Overview

Overall, 597 genes had significant expression changes in response to nitrogen limitation (**Supplementary Table 2**) and 2036 genes had significant expression changes in response to phosphate limitation (**Supplementary Table 3**). When comparing the nitrogen and phosphate groupings, 493 genes were in common and had significantly changed expression in response to limitation of each nutrient (**Supplementary Table 4**). With respect to the direction of gene regulation, 53% of nitrogen responsive genes were upregulated in response to nitrogen deficient conditions and 47% were downregulated, whereas only 46% of phosphate responsive genes were upregulated in response to phosphate deficiency and 54% were downregulated. A heat map of the replicate experiments that led to these conclusions illustrates the variation in expression observed (**Supplementary Figure 2**).

In a microarray transcriptome study by Yurgel et al. (2013), 132 genes in wild type Rm1021 were identified as responding to nitrogen using the criterion of a two-fold change at $p \leq 0.05$. Of those, 49 were identified in this study. A microarray study by Krol and Becker (2004) identified 234 genes that were either phosphate- or PhoB-dependent using the criterion of a three-fold change at $p \leq 0.05$. 167 of these were also identified in this study. The appearance of previously identified genes in this study indicated that the bacteria responded to the nutrient stress in a consistent manner, while the inability to identify all the genes from previous studies most likely was due to experimental differences and a more stringent statistical limit. An important difference could be in the strain genetics. This study used Rm1021 *pstC*+ whereas Yurgel et al. used Rm1021 for their nitrogen stress transcriptome. Rm1021 contains a truncated *pstC* and cannot use the PstSCAB, a high-affinity, high-velocity phosphate transport system found in related strains (Yuan et al., 2006a). Krol and Becker's phosphate stress microarray involved Rm2011, a closely related *pstC*+ strain (Wais et al., 2002). Other differences between the studies included the media used, culture density at time of collection, and factors such as fold change cutoff and Q value. The novelty in this transcriptome compared to the previous studies is that not only were nitrogen or phosphate responsive genes identified, but their sensitivity to the other nutrient was also characterized, enabling analysis of genome-wide changes as a function of two variables and identification of genes or processes that were coordinately regulated by both stresses.

In total, 597 genes responded to nitrogen limitation under some condition (**Figure 1A**). The typically measured "NSR," NSRP, consisting of genes whose expression responded to nitrogen limitation when cells had sufficient phosphate, contained 282 genes or 47% of the total NSR genes. The NSRp group, which responded to limited nitrogen when phosphate was also limited, contained 345 or 58% of total nitrogen responsive genes. 30 genes (5%) had significant expression changes under both phosphate sufficient and phosphate deficient conditions and made up the "Core NSR" group described above. Of the 2036 genes that responded to phosphate limitation, 1598 genes or 78% responded to phosphate when nitrogen was sufficient and were placed in the PSRN group. The remaining 1141 genes or 56% were assigned to the PSRn because they responded to phosphate limitation when nitrogen was also limiting (**Figure 1B**). 703

TABLE 1 | Distribution of Nitrogen and Phosphate responsive genes on the *S. meliloti* replicons.

	# annotated genes including ncRNAs	% total	NSR total	% of NSR total	NSR up	% of NSR up	NSR down	% of NSR down	PSR total	% of PSR total	PSR up	% of PSR up	PSR down	% of PSR down
Chromosome	4446	54	368	62	235	72	147	50	1263	62	444	49	828	73
pSymB	1977	24	139	23	61	19	81	28	455	22	297	33	159	14
pSymA	1830	22	90	15	30	9	63	22	318	16	170	19	149	13
Totals	8253		597		326		291		2036		911		1136	

For each replicon, the number of annotated genes including non-coding RNAs is shown along with the percentage of the total # of genes that they represent. The total number of genes responding to nitrogen stress are indicated as well as the percentage of this group on each replicon. The number and percentages of nitrogen stress induced (NSR up) genes and nitrogen stress repressed (NSR down) genes are shown. Similar data for the phosphate stress responsive genes are also shown. When a gene was induced under one condition and repressed under another, it was counted twice; however, genes are only counted once for the NSR or PSR totals.

genes or 34% were phosphate responsive during both nitrogen conditions and were included in the "Core PSR." Most notable in both analyses was the number of genes that responded to stress under one condition but not the other. For phosphate stress, the percentage of the 2036 genes that responded under high nitrogen but not low nitrogen or vice versa was 1336, or 66% of the total. For nitrogen stress, 567 of the 597 genes or 95%, responded under one but not both conditions. Defining a gene as nitrogen stress responsive seems therefore to have little predictive power when applied to these different contexts.

In these experiments, phosphate deficiency influenced more than three times as many genes as nitrogen deficiency. The data indicate that expression of a substantial fraction of genes were influenced by both nutrients as the primary stress condition, identifiable by the presence of these genes in both the nitrogen-responsive and phosphate-responsive tables (**Supplementary Table 4**). This implies either that many genes can be part of both an NSR and a PSR, or that stress itself influenced their expression as part of a global response. We observed that phosphate limitation led to many changes in translational proteins, a systemic response that was less prominent during the nitrogen limitation.

Gene Expression Patterns Differ on the Three *S. meliloti* Replicons

Sinorhizobium meliloti Rm1021 contains three large replicons: a chromosome (3.7 Mb) and two megaplasmids, pSymA (1.4 Mb) and pSymB (1.7 Mb) (Galibert et al., 2001; **Table 1**). **Table 1** and **Figure 2** show an analysis of the number of genes on each replicon that are induced or repressed by nitrogen stress (**Figure 2B**) and phosphate stress (**Figure 2C**). The three replicons contain 54, 22, and 24% of the total annotated genes, but the genes responding to the stresses and the way that they respond are not proportionately distributed. In particular, the genes that respond to nitrogen stress are more common on the chromosome and less common on the pSymA megaplasmid. This is most pronounced for the genes induced by nitrogen stress, as can be seen in **Table 1** where the proportion of N-inducible genes on pSymA is less than half of what might be expected based on random distribution throughout the replicons. A similar trend is not seen for pSymA genes repressed by nitrogen stress. In contrast, the fraction of genes on both plasmids that are repressed during phosphate stress is lower than expected.

Nitrogen Responsive Genes

Supplementary Table 2 contains the list of 597 significantly changed nitrogen responsive genes. The genes with the highest induction ratios in response to nitrogen deficiency had all been identified previously (Yurgel et al., 2013), but measuring the phosphate modulation of these nitrogen responsive genes was novel and provided insight into how the stress response pathways were sensitive to the status of the other nutrient. The discussion of specific genes below is structured around the Venn diagrams in **Figure 1** and the clusters defined in **Table 2**. For both the NSR and PSR, we define a set of Core genes (N.I and P.I), and two sets of genes that were included in the stress response under one condition but not the other, i.e., N.II corresponds to the NSRP, while N.III corresponds to the NSRp. It is important to realize that, as strictly defined as the Venn diagrams make these distinctions appear, these boundaries are numerical and not conceptual. Within each cluster we have tried to focus on genes that seemed most interesting in terms of their previously described regulation by nitrogen or phosphate limitation, and to illustrate how the availability of one nutrient might affect expression of genes also controlled by the other. **Supplementary Table 5** shows the number of genes in major classification categories (as seen in **Supplementary Tables 2**, **3** in the "classification" column) for the NSR and PSR and the number of genes in each category that are up- or downregulated. For the set of genes described as N.III, which corresponds to the NSRp, it is important to recognize that these genes, which comprise the largest of the three sets, had not generally been defined as NSR genes in previous analyses that did not include the additional dimension of phosphate limitation. Our analysis thus evaluated stress response pathways individually, but also links the responses, demonstrating the large multi-dimensional aspect of gene regulation when more than one stress was applied.

Core Nitrogen Stress Response Genes

We defined the Core NSR group N.I as 30 genes that had significant expression changes in response to nitrogen limitation under both phosphate sufficient and phosphate limiting conditions. Given the limits of the experimental design, we consider these to "always" be part of the NSR (**Figure 1**, **Table 2**, and **Supplementary Table 2**). *glnII* and *dctA* were both highly induced. *glnII* is well known as a nitrogen-regulated gene although occurrence of induction during conditions of either

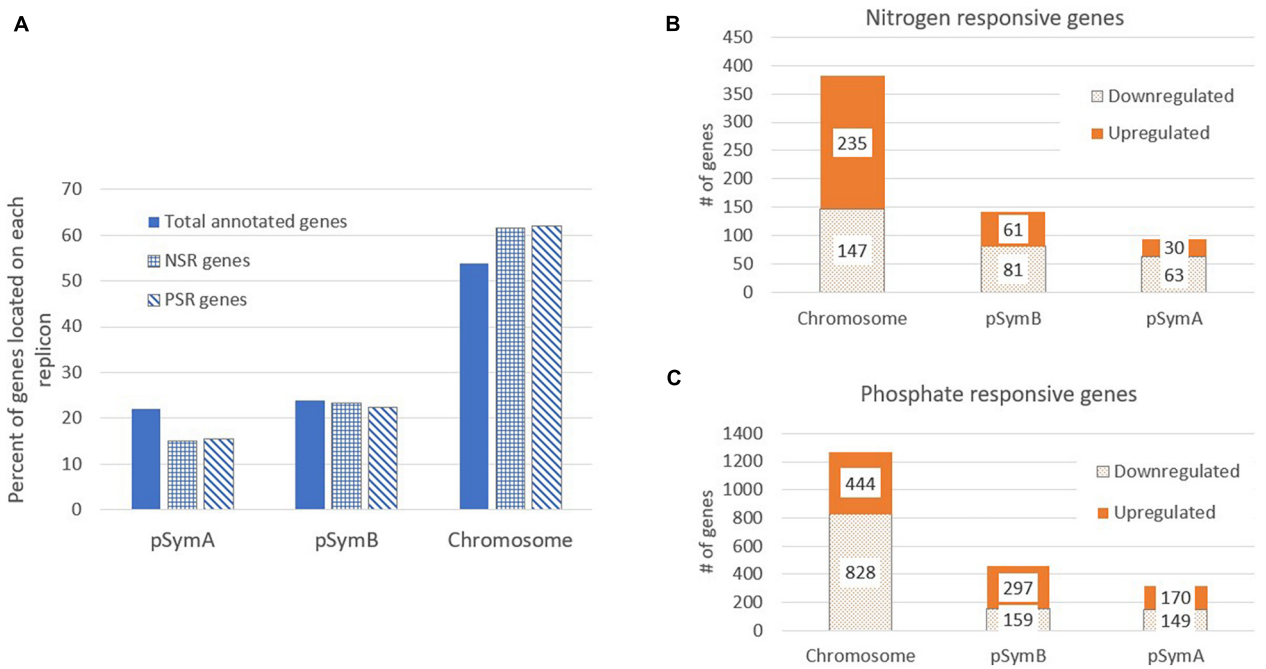


FIGURE 2 | Distribution of Nitrogen and Phosphate responsive genes on the *S. meliloti* replicons. **(A)** The percent of *S. meliloti* genes located on each replicon is shown for total annotated genes (solid bars), the nitrogen responsive genes (NSR) (checkered bars) and the phosphate responsive genes (PSR) (striped bars). The number of genes with significant expression changes in response to nitrogen stress **(B)** or phosphate stress **(C)** are indicated for each replicon. Each bar is split and shows the number of upregulated (solid bars) and downregulated (dotted bars) genes per replicon.

TABLE 2 | Nitrogen and Phosphate gene groupings.

Cluster	Cluster Definition	Genes in cluster
N.I	Responds to N stress during both P stress and non-stress conditions. NSRP AND NSRp	30
N.II	Responds to N stress only during P non-stress conditions. NSRP AND NOT NSRp	252
N.III	Responds to N stress only during P stress conditions. NOT NSRP AND NSRp	315
P.I	Responds to P stress during both N stress and non-stress conditions. PSRN AND PSRn	703
P.II	Responds to P stress only during N non-stress conditions. PSRN AND NOT PSRn	895
P.III	Responds to P stress only during N stress conditions. NOT PSRN AND PSRn	438

Each cluster is defined and the number of genes assigned to each cluster is shown.

phosphate sufficiency or deficiency had not been reported. The induction of *glnII* may indicate that it is very important during nitrogen limitation and needs to be expressed regardless of the phosphate status, although phosphate still had a substantial effect on the level of induction— *glnII* was induced 28-fold in high phosphate and only 2.9-fold in limiting phosphate. In a pattern that probably contributes further to the modulation of the GSII protein in response to low nitrogen, *gstI*, an inhibitor of glutamine synthetase II translation (Napolitani et al., 2004), was repressed four-fold by low nitrogen, but only when phosphate was sufficient. It was somewhat surprising to find *dctA* in this cluster. DctA is a dicarboxylic acid transporter that is needed for growth on C4 dicarboxylates like malate, succinate, fumarate and aspartate (Yurgel and Kahn, 2004, 2005). DctA is essential for nitrogen fixation during symbiosis, probably to acquire the carbon compounds needed for generating energy and reductant (Yurgel and Kahn, 2004). Like *glnII*, *dctA* expression was much higher under phosphate sufficiency (12.8×) than in limiting phosphate (2.6×).

Nitrogen Stress Response Genes During Phosphate Sufficiency

The NSR group N.II contained 252 genes that responded to nitrogen limitation only when phosphate was sufficient (**Figure 1**, **Table 2**, and **Supplementary Table 2**). Under high phosphate, 88 of these genes were upregulated by nitrogen stress and 164 were downregulated. Interestingly, genes like *glnK* and *amtB,* which were highly induced by nitrogen limitation during phosphate sufficiency (>100×), were induced less than 2× when cells were grown on glutamate during phosphate limitation. *amtB* codes for an ammonium uptake transporter that is often found in an operon with *glnK* and both GlnK and GlnB P_{II} proteins can activate *amtB* expression. *glnK* and *amtB* had previously been shown to be induced during nitrogen stress

(Yurgel et al., 2013). Identification of these genes as highly induced by nitrogen stress in high phosphate is consistent with other observations in the literature but the impact of phosphate stress on their expression had not been described. Other highly induced genes in this category included *nrtA,* encoding a nitrate reductase, and *rhtA*, encoding a protein involved in rhizobactin regulation. Rhizobactin is a siderophore that can bind iron and other metals used as cofactors for various proteins, suggesting that metal metabolism and acquisition may be affected by nitrogen availability.

Changes in expression of N.II genes occurred only during phosphate sufficiency suggesting that the additional phosphate stress rendered these changes less important to the cell even though nitrogen was still limited. Many of these genes have transport-related functions. For example, the genes in *hmuPSTUV,* an operon that codes for a hemin import system, were upregulated, as was *smhR*, an outer membrane hemin binding protein. In contrast, the unlinked *hemA, hemB, hemE, hemF, hemH* and *ctaA* genes, which code for proteins involved in heme biosynthesis, were downregulated. The upregulation of hemin transport and downregulation of heme biosynthesis suggests that under these conditions, the cell prefers to import heme rather than to generate it *de novo*. This example illustrates how a cell can conserve energy and resources by repressing biosynthesis of certain molecules, but still tries to obtain the molecules from its environment, potentially satisfying both the greed and fear motivators. The *livH-livK* operon, which codes for a branched chain amino acid transporter, was highly induced by nitrogen stress. The corresponding *bra* operon is important in the symbiosis between *Rhizobium leguminosarum* and the pea plant and when inactivated results in symbiotic auxotrophy. In the symbiosis between *S. meliloti* and alfalfa, the symbiotic auxotrophy phenotype is not observed when these transporters are inactivated (Prell et al., 2010). Induction of these transporters in this context suggests that these systems are beneficial during nitrogen stress, but phosphate is necessary to use them effectively, potentially because resources are needed elsewhere or to be consistent with other changes imposed by the additional phosphate stress. The *SMc02344-SMc02346* genes, which code for components of a glycine-betaine and choline ABC transporter, were also among the most highly induced genes. This could indicate that these compounds were a preferred source of nitrogen for the cell during nitrogen limitation, or that they were induced to take advantage of the possibility that these osmolytes could play a larger role in adaptation during nitrogen stress.

Nitrogen Stress Response Genes During Phosphate Deficiency

The N.III group contained 315 genes that showed significant expression changes in response to nitrogen limitation only when phosphate was also limiting (**Figure 1**, **Table 2**, and **Supplementary Table 2**). Of these, 209 were upregulated by low nitrogen and 106 were downregulated. Most genes in this group had not been identified as part of an NSR. Showing that there is a nitrogen stress response under low phosphate conditions that is distinct from the classical NSR is significant because not only does it include additional genes in the bacterial NSR, but it shows that defining an NSR is very context dependent.

One of the upregulated genes in the N.III group is a putative alkaline phosphatase, *SMc03243*. The presence of a putative alkaline phosphatase that responded to nitrogen limitation was interesting but not entirely unexpected because previous studies had shown that P_{II} proteins can influence expression of genes in the phosphate regulon (Hagberg et al., 2016). While our earlier work had not identified *SMc03243* as significantly responsive to nitrogen (Yurgel et al., 2013), the added P stress and strain differences, specifically the introduction of the *pstC*+ allele into the Rm1021 *pstC*+ strain used here, may have allowed *SMc03243* to respond differently than it does in Rm1021, which has the truncated *pstC* gene.

Phosphate Responsive Genes

In total, 2036 genes, or about one third of all *S. meliloti* genes, had significant expression changes when comparing low and high phosphate conditions (**Supplementary Table 3**). This was nearly four times the number of genes that responded to limited nitrogen. The 167 phosphate-responsive genes that were also found in the microarray experiments of Krol and Becker (2004) had various patterns of expression with respect to the influence of nitrogen. Choudhary et al. (2020) defined the regulon of PhoR/PhoB in *E. coli* which includes 170 genes. Homologous genes in *S. meliloti* responded to phosphate which was expected due to the conserved PhoRB two-component system involved in regulating the PSR in both bacteria. These genes also had various patterns of expression with respect to nitrogen influence, but the majority were in the Core PSR or PSRN groups.

Core Phosphate Stress Response Genes

The Core PSR group P.I contained 703 genes whose expression significantly changed in response to phosphate limitation under both nitrogen conditions (**Figure 1**, **Table 2**, and **Supplementary Table 3**). This large number contrasts with the 30 genes in the Core N.I NSR group. The difference suggests that phosphate limitation may be a greater strain on the cell than nitrogen limitation, and that a larger set of genes must be either induced or repressed to maintain critical cell functions or to restructure processes to use the available phosphate most efficiently. Of the 703 genes, 351 were upregulated and 343 were downregulated. To our surprise, nine mRNAs were significantly induced in one nitrogen condition and repressed in the other – a result that met the criteria for classification to the Core PSR but was unusual when considered from a viewpoint in which the PSR describes sets of "phosphate induced" or "phosphate repressed" genes that act to counter the cellular effect of phosphate stress.

Of the 100 genes most highly induced or repressed in response to phosphate stress, about 70% belong to the Core PSR. Many of these had been previously characterized as core components of the PSR. The *pho, pst,* and *phn* operons are involved in phosphate transport or phosphonate uptake and metabolism and are known to be induced by phosphate stress (Voegele et al., 1997; Parker et al., 1999; Yuan et al., 2006a; Choudhary et al., 2020). Induction of these operons allows cells to

catabolize phosphorous-containing molecules in order to transfer phosphate into molecules with a higher priority.

phoB, phoX, and *phoU* were among the most highly induced. The PhoB response regulator controls activation of the PSR in association with its sensor kinase, PhoR. Induction of *phoB* should permit a greater dynamic range in transducing the phosphate deficiency signal into responsive gene and protein expression changes. PhoX is an alkaline phosphatase whose gene was induced 254-fold by phosphate stress in these experiments. PhoX is one of several extracellular phosphatases that degrade phosphate esters to liberate phosphate. *SMc03243,* which codes for a putative alkaline phosphatase, was among the most induced genes. Its strong induction during phosphate limitation is consistent with its annotation as a putative alkaline phosphatase and indicates that the cell uses this phosphatase in addition to PhoX. Also in this group are genes that could be used in releasing phosphate from phosphonate and phytate. PhoU is thought to be a PSR regulatory protein but its precise role is unclear (Lubin et al., 2016). Among the most repressed genes was *pit1* which encodes a low affinity phosphate transporter, presumably reflecting its lack of utility under these conditions.

Other genes that were highly induced in the Core PSR P.I group were *SMc04316, SMc04317, wgeA, btaA* and *btaB. SMc04316* and *SMc04317* are in an operon that has a Pho box in its promoter region and are predicted to encode iron transport proteins (Yuan et al., 2006b). *wgeA* codes for a galactoglucan (EPSII) biosynthetic protein that is regulated in part by PhoB and induced under low phosphate conditions (McIntosh et al., 2008). *btaA* and *btaB* encode proteins that produce diacylglycerol-*N,N,N*-trimethylhomoserine (DGTS), a phosphorous-free lipid that can be incorporated into the membrane during phosphate stress (López-Lara et al., 2005). DGTS and related lipids are a major fraction of lipids induced by phosphate starvation (Geiger et al., 1999). Inducing *wgeA, btaA* and *btaB* may help restructure the cell envelope by replacing phosphate-containing molecules, like phospholipids, with phosphate-free molecules or to protect the cell during stress via increased polysaccharide production.

Of the 36 messages in the Core PSR coding for ribosomal proteins (*rps, rpm, and rpl*), all were repressed, with 30 downregulated more than 4×. Ten tRNA synthetases were repressed between 2.5- and 5-fold. This suggests phosphate stress significantly downregulates protein synthesis by interfering with protein translation.

Phosphate Stress Response Genes During Nitrogen Sufficiency

The P.II group contained 895 genes that had significant expression changes in response to phosphate stress only when nitrogen was sufficient (**Figure 1**, **Table 2**, and **Supplementary Table 3**). In particular, RNAs corresponding to the protein synthesis apparatus, including six Rpm proteins, seven Rps proteins, four Rpl proteins, and several other proteins annotated as important in ribosome function are in this group. Three tRNA molecules, gly_TCC, thr_CGT, and asp_GTC are repressed 9.8, 9.5, and 5.1× respectively. The glycine codon is not rare in *S. meliloti* and the threonine and aspartate codons are used extensively[1]. The P.II group also includes seven downregulated tRNA synthetases (including *thrS* and *glyQ*) and six tRNA modification activities, including *gatB*, which amidates aspartate and glutamate on charged tRNAs to yield asparagine and glutamine tRNAs. Thus, phosphate stress in the presence of ammonium has the potential to downregulate protein synthesis through direct interference with the translational machinery.

Amino acid biosynthetic gene expression also changed. *ilvA, ilvD1, ilvD2, and ilvE1, leuA1, leuB, leuD, and leuS* and *hisBHF, hisC2, hisD1 and hisZS* are independently regulated genes or parts of operons that are involved in branched chain amino acid, leucine, and histidine synthesis, respectively. These were all significantly downregulated in response to phosphate limitation only when nitrogen was sufficient. The downregulation of these amino acid-related genes by phosphate limitation was consistent with previous transcriptome data in *S. meliloti* (Krol and Becker, 2004) and with phosphorous limitation experiments in *E. coli* (Van Bogelen et al., 1996) and might reflect reduced demand for these amino acids in phosphate limited cells. Downregulation of these genes may potentially be metabolically complemented by induction of transporters such as *Smc02832* which is categorized as a putative taurine, valine, isoleucine and leucine ABC transporter and also in the P.II group.

The *nuoA1B1C1D1E1, nuoG1I, nuoK1LN* genes, which code for various subunits of an NADH-quinone oxidoreductase and are clustered in a regulon on the chromosome, and the *nuoI2* gene, which is in a separate cluster on pSymA, were downregulated during phosphate stress when nitrogen was sufficient. *ctaCDBGE*, an operon that codes for synthesis and assembly of a cytochrome c oxidase, was also downregulated. These changes are consistent with a change in the energy and carbon metabolism of the cell during phosphate stress.

Phosphate Stress Response Genes During Nitrogen Deficiency

The P.III group contained 438 genes whose expression changed at least two-fold in low phosphate, but only when nitrogen was also limited (**Figure 1**, **Table 2**, and **Supplementary Table 3**). 320 of these genes were upregulated and 118 were downregulated by phosphate stress. These genes are "new" phosphate stress response genes, in the sense that they add to the traditional PSR which only considered genes whose expression changed under the standard condition of limiting phosphate in an otherwise complete medium. One interpretation of the appearance of this new set is that when nitrogen is sufficient their baseline expression does not need to be modified substantially to adapt.

phoR, phoH, and *phnN* were among the most highly induced genes in response to phosphate stress, but these were induced only during nitrogen limitation. It is expected that these genes would be induced based on their role in phosphate transport and regulation. A few other genes also responding to phosphate stress during nitrogen deficiency were *katB* and the cold shock family RNA binding protein genes *cspA1, cspA3,* and *cspA7.* KatB encodes one of *S. meliloti*'s catalase/peroxidase enzymes and helps detoxify oxidative stress molecules. The cold shock family of

[1]http://www.kazusa.or.jp/codon/cgi-bin/showcodon.cgi?species=266834

stress proteins have protective roles during various stress events, such as low temperature and salt stress, and they also have roles in symbiotic development (O'Connell and Thomashow, 2000; Price, 2017). Increased expression of these genes indicates that cell protection and detoxification are important processes during nutrient stress and suggests the cell is reducing non-essential processes that may leave the cells vulnerable to other stresses imposed by predators, competing microorganisms, or toxic molecules in addition to scavenging for nutrients.

Genes Responding to Both Nitrogen Stress and Phosphate Stress

In total, 493 genes are present in both the nitrogen-responsive gene table and the phosphate-responsive gene table (**Supplementary Table 4**) and will be referred to as "dual-responsive genes." In some cases, a gene was upregulated or downregulated in response to both nutrient limitations, but 289 of these genes, or 58%, were upregulated in response to limiting one nutrient and downregulated in response to limiting the other nutrient. Of these, nearly 2/3 of the genes were upregulated during nitrogen limitation and downregulated by phosphate limitation, demonstrating that the changes in GS mRNA abundance as a function of both nitrogen and phosphate availability, as previously discussed in the introduction and documented in Hagberg et al. (2016) are not unique.

One gene that responded significantly to both nitrogen and phosphate stress was *SMc03243,* the putative alkaline phosphatase discussed previously. *SMc03243* was highly induced by phosphate limitation under both nitrogen conditions, but the induction was two-fold higher when nitrogen was limited compared to when it was sufficient. The *SMc03243* expression data and analysis illustrates how genes can be "turned on" to various degrees based on the number and type of stresses, further indicating an integrated response.

glnII expression was also dual responsive. Upregulation of *glnII* during nitrogen stress was expected (Yurgel et al., 2010). By using reporter fusions and immunoblot assays, we showed earlier that *glnII* induction by nitrogen limitation was lower when phosphate was limited (Hagberg et al., 2016). The data here showed that *glnII* mRNA was induced during nitrogen stress in both phosphate conditions, but the induction ratio was seven times greater when phosphate was sufficient compared to deficient. *gstI,* which codes for a GSII translational inhibitor located adjacent to *glnII* and transcribed divergently (Spinosa et al., 2000), was regulated in a way that counters expression of *glnII*, i.e., it was repressed by low nitrogen. The reciprocal regulation of these two genes is thought to enhance the range of GSII regulation (Spinosa et al., 2000; Napolitani et al., 2004). The response of *glnII* and *gstI* to phosphate stress was also inverted—during nitrogen stress, *glnII* expression was reduced by concurrent phosphate stress whereas *gstI* expression was enhanced by phosphate stress. The coordinated and reciprocal expression patterns of *glnII* and *gstI* indicate that phosphate stress decreased induction of GSII, a central component of the NSR, presumably to direct resources to both stress response pathways. Another interpretation is that phosphate stress has a more detrimental impact on the cell, and the standard NSR response pathway is downregulated to allow the cell to prioritize its response to phosphate stress.

Expression of a heme transport system encoded by an operon containing *hmuS* and *hmuT,* was upregulated by nitrogen stress, but downregulated when phosphate stress was added. *SMc01515* and *SMc01517,* uncharacterized genes situated just upstream of the *hmu* operon, were also upregulated by nitrogen stress and downregulated with simultaneous phosphate stress, suggesting that this region was regulated as a unit. In response to nitrogen limitation, *hmuS* was induced 2.2-fold when phosphate was sufficient but was repressed 3.6-fold when phosphate was limited. This pattern shows that phosphate stress can not only suppress the level of induction or repression that is caused by nitrogen stress but can even reverse the direction of the change from induction to repression or vice versa. 36 other dual-responsive genes had a reverse in direction of the induced change. Many of the genes that followed this pattern encoded transporters, suggesting that when nitrogen is hard to find, importing metabolites has a relatively higher priority than biosynthesis. However, the general downregulation of these transporters during phosphate stress suggests that this strategy is not as appropriate when phosphate is scarce. This could be because the two stresses are usually found in different contexts; nitrogen stress in a situation where other nutrients may still be available but phosphate stress when there is a general lack of external resources. Another possibility is that the demand for molecules decreases when the cell is scavenging for phosphate to maintain critical cell functions, and this impacts processes like import.

Members of the *liv* operon, which encode a branched chain (leucine, isoleucine, and valine) ABC transporter, were upregulated by nitrogen stress and downregulated by phosphate stress. The *liv* genes are known to be induced by nitrogen limitation, and Yurgel et al. (2013) suggested that they were regulated by GlnB/GlnK and another, unknown regulatory component. Sobrero et al. (2012) demonstrated that the Liv system and other amino acid transporters are negatively controlled by the RNA chaperone Hfq. Expression of *SMc02344-Smc02346,* encoding a putative glycine-betaine and choline transporter, also involves Hfq-dependent riboregulation (Barra-Bily et al., 2010; Sobrero et al., 2012). Like the branched chain amino acid transporters, this operon was upregulated during nitrogen stress and downregulated during phosphate stress. The identification of these nitrogen and phosphate responsive operons as being influenced by Hfq suggests that Hfq may be a major regulatory component that is involved in integrating the nitrogen and phosphate stress signals, perhaps by mediating sRNA-mRNA interactions.

sRNAs and Hfq

Hfq is an RNA chaperone that interacts with sRNAs used to regulate translation of expressed genes. Hfq often affects sRNA binding to the 5′ UTR of mRNAs, thereby affecting 30S ribosomal subunit binding to this region during the initiation of translation (Vytvytska et al., 2000; Soper et al., 2010). The abundance of 45 sRNAs was significantly altered in our experiment, and 14 of these are known to interact with Hfq

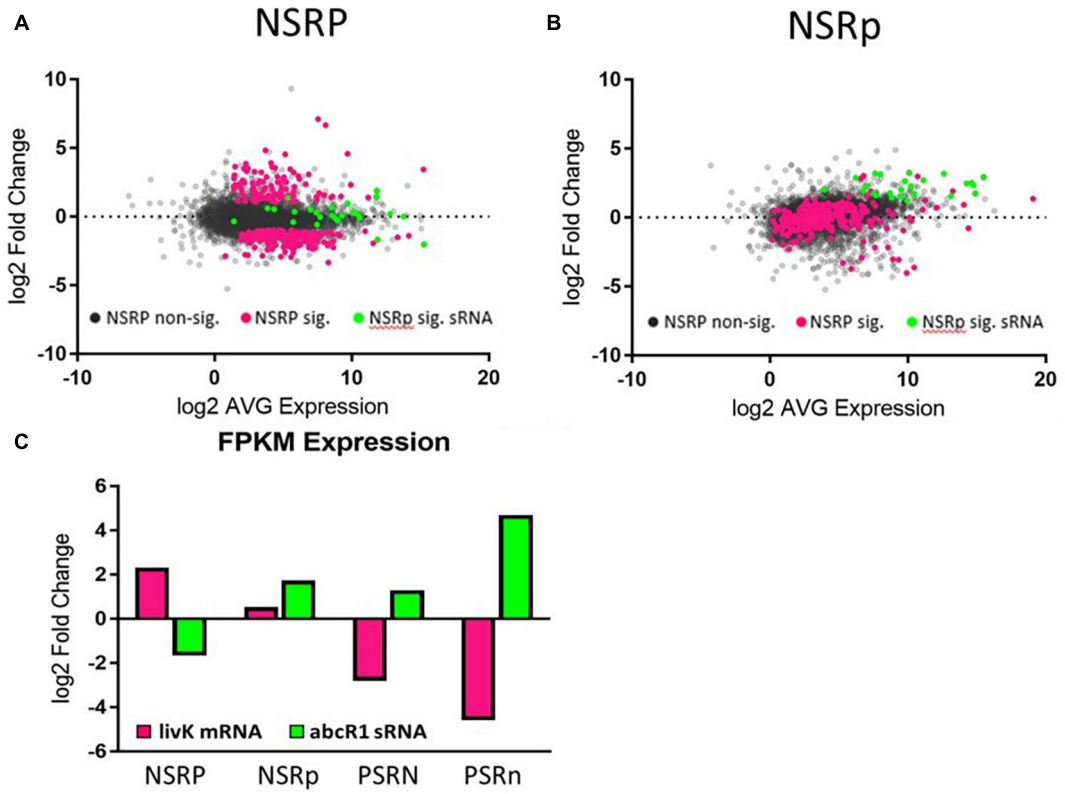


FIGURE 3 | Nitrogen stress affects sRNA expression differently under high and low phosphate. Graphs showing the changes in expression of genes in the **(A)** NSRP and **(B)** NSRp. Gray dots represent non-significant gene expression changes. Pink dots represent the set of genes that were significantly changed in the NSRP. Green dots represent the sRNAs that were significantly changed in the NSRp. The expression patterns of each group can be seen changing between the high P condition **(A)** and the low P condition **(B)**. **(C)** The levels of *livK* and *abcR1* expression under the various stress conditions demonstrate the different expression patterns of an mRNA and its regulatory sRNA.

(Torres-Quesada et al., 2014). These sRNAs had some of the largest fold changes between the stressed and non-stressed conditions, with levels of induction in the range of the most highly induced protein-encoding genes associated with nitrogen or phosphate stress. This suggests that these sRNAs play a major role in regulating translation of transcripts sensitive to nutrient stress. The abundance of a majority (19) of sRNAs responding to nitrogen were significantly changed only in the phosphate deficient condition (N.III) (**Figure 3**). Only 5 sRNAs responded to nitrogen during phosphate sufficiency (N.II). An additional 4 sRNAs responded to nitrogen during both phosphate sufficiency and deficiency (N.I). Phosphate stress altered the expression of 13 sRNAs during both nitrogen conditions (P.I). Only 8 sRNAs responded to phosphate limitation during nitrogen sufficiency (P.II) while 16 sRNAs were phosphate responsive during nitrogen deficiency (P.III). These data indicate phosphate stress influences sRNA abundance more than nitrogen stress does, consistent with the data observed for mRNA abundance changes.

One highly induced sRNA was made from *abcR1* and was induced during nitrogen limitation when phosphate was also limited. *abcR1* is known to negatively regulate expression of the *livK* gene with the aid of Hfq (Torres-Quesada et al., 2014). *livK* expression was induced by nitrogen limitation, but then strongly repressed by the addition of phosphate stress (**Figure 3C**). The induction of the *abcR1* sRNA during the double stressed condition could act to inhibit translation of the *livK* transcripts, which were also being repressed, suggesting that while LivK is useful during nitrogen stress, it might not be needed or was inefficient during phosphate stress, and that repression of LivK transcription and translation by low phosphate allows resources to be directed elsewhere.

In addition to the *livK* example linking a nutrient stress-induced sRNA with Hfq, regulation of the *E. coli rpoS* gene by both PhoB and Hfq has been observed (Ruiz and Silhavy, 2003). In this case, *rpoS* translation is increased during phosphate limitation by a PhoB-induced sRNA. Hfq mediates the sRNA-*rpoS* mRNA interaction, enhancing translation of the *rpoS* message, and demonstrates how PhoB and Hfq can both be involved in regulating a gene. There is no *rpoS* in *S. meliloti,* but Hfq has been linked to numerous changes involved in the metabolism of nitrogen-containing compounds, transporter systems, general stress, mobility, and membrane components (Gao et al., 2010; Torres-Quesada et al., 2014). One third of the dual-responsive genes in this study were previously identified as having significant expression changes in an *S. meliloti hfq* mutant compared to wild type (Barra-Bily et al., 2010; Gao et al., 2010;

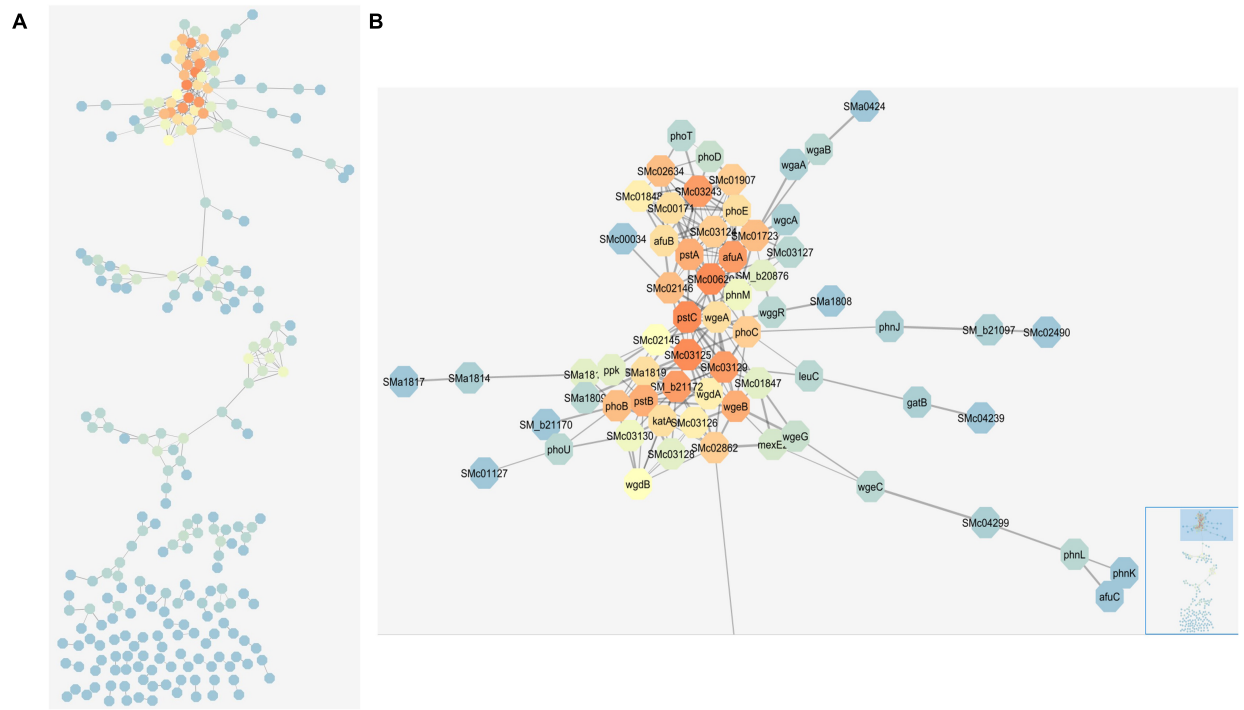


FIGURE 4 | Network map of connected genes based on weighted co-expression analysis. Network map created in Cytoscape depicting genes that are linked based on their expression patterns and correlation with each other. **(A)** Shows the full network map. **(B)** Shows the large group at the top of the map including many core phosphate stress response genes. Colors represent degrees and depict how connected a gene is to other genes in the map. The color scheme starts with blue as the least connected (fewest degrees) and flows through green to yellow to orange as the most connected (highest degrees). The bottom right inset in panel **(B)** highlights the area of the full network map that is enlarged.

Torres-Quesada et al., 2010, 2014; Sobrero et al., 2012). The possibility that Hfq-mediated sRNAs have a role in these processes correlates with the large number of transport genes identified as being dual-responsive to nitrogen and phosphate stresses. Hfq may be interacting with sRNAs to alter translation of a class of target transcripts and help integrate and fine-tune the response to these stresses.

Mapped Networks

Weighted co-expression analysis using R bioinformatics software and visualization with Cytoscape (Shannon et al., 2003) allowed us to construct a map of co-expressing genes (**Figure 4** and **Supplementary Figure 3**). The co-expression analysis was able to connect genes that had similar patterns of expression change across the four media conditions. For example, core components of the PSR, such as *phoB* and the *pst* operon, respond similarly to the four conditions and were connected in the network. The largest group in the network included *phoB* and many of the highly induced phosphate responsive genes described above, but also genes that met the covariance criteria which did not stand out in the pairwise analysis, such as *SMc00620, SMc01723,* and *SMc02862* (**Figure 4B**). The second largest network featured chemotaxis and mobility genes, which also form a well-documented regulon that responded coordinately to these nutrient stresses. In *E. coli,* the PhoB regulon also contains the chemotaxis genes along with the expected phosphate transport systems (Choudhary et al., 2020).

The visualized connections between the large group containing *phoB, phoCDET, pstSCAB, phoU*, and other genes involved in phosphate stress represent some of the most upregulated genes during phosphate stress with no significant up or down regulation during nitrogen stress. The highly connected nature of these genes was expected—they have been consistently identified and characterized as being part of the PSR (Al-Niemi et al., 1997; Krol and Becker, 2004). Also connected in this group were some regulators of exopolysaccharide synthesis. Exopolysaccharide synthesis is changed based on phosphate availability, so the connection of these genes with other phosphate genes further validates the network map.

Surprisingly, given the central importance assigned to the *phoBR* two-component regulatory system in phosphate stress regulation, these genes were not the "most connected" in the network as indicated by the degrees parameter. This distinction went to *SMc00620*, with 20 connections, many more than the 12 for *phoB*. *SMc00620* is a hypothetical protein which a BLAST search indicates is likely to be a phytase type of phosphodiesterase. Because of its large number of direct connections and co-expressing genes, *SMc00620* may be worth exploring further by mutation to determine whether it regulates or is regulated by this unusually large number of other PSR-associated genes. If this gene does produce a phytase, it is

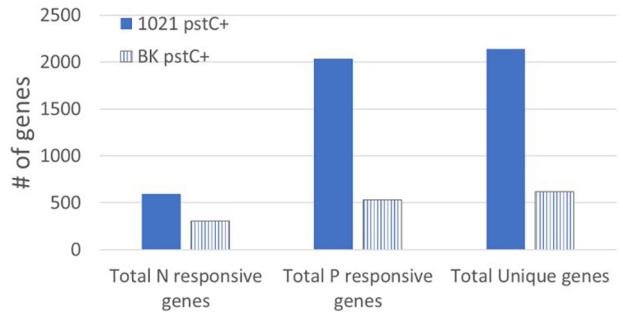


FIGURE 5 | Genes responding to nitrogen and phosphate stress in the BK *pstC*$^{+}$ strain. The total number of genes responding to nitrogen stress (left) and phosphate stress (center) in the 1021 *pstC*+ (solid bars) and BK *pstC*+ (striped bars) strains. The total number of unique genes (right) that changed in response to at least one nutrient stress is also shown for each strain.

conceivable that it is upregulated during phosphate stress to release phosphate from plant-produced phytate compounds that may be in the rhizosphere. Phytases may not be common in enteric bacteria due to the lack of phytate in their natural environments and thus, not commonly observed in traditional PSRs of model bacteria like *E. coli*.

The network map brought attention to genes with many connections that may have been overlooked in the pairwise analysis due to their uncharacterized or putative nature and it provides candidates for further exploration within the stress response pathways. *SMc02862,* which is annotated as a *pit* transport system accessory protein, also has more links to other genes in this cluster than *phoB* (14 > 12). The *pit* gene, *SMc02861*, was downregulated by phosphate stress, but only *SMc02862* was included in the co-expression network. The Pit transport system is a low-affinity phosphate transport system whose regulation by phosphate stress is known to be reciprocal to that of the high-affinity Pst and Pho transporters (Voegele et al., 1997).

The BK *pstC*+ Strain Has Far Fewer Significant Expression Changes Than 1021 *pstC*+

The P_{II} proteins, GlnB and GlnK, were originally identified as important components of the NSR signal transduction pathway, influencing both transcriptional and post-translational responses to low nitrogen. We analyzed the *glnB*$^{-}$*glnK*$^{-}$ strain, BK *pstC*+, in addition to the 1021 *pstC*+ strain for gene expression changes that resulted from nitrogen and/or phosphate stress. The most striking observation was that expression of only 620 genes was altered significantly by nitrogen stress, phosphate stress, or both (**Figure 5** and **Supplementary Table 6**). This number, while substantial, was far less than the 2140 stress responsive genes identified in 1021 *pstC*+ where the P_{II} proteins are intact. The lack of expression changes in BK *pstC*+ compared to 1021 *pstC*+ demonstrates the vital regulatory role that the P_{II} proteins play within the cell and reinforces the idea that their role is not limited to the NSR. Two scenarios could explain the lower number of gene expression changes in BK *pstC*+. One is that the slow growth of the BK *pstC*+ strain (Yurgel et al., 2010) may allow it to cope with stress by using smaller changes in gene expression due to a reduction in immediate demand for resources. In this scenario, the lack of the P_{II} proteins causes a change in growth rate that reduces the apparent stress, but the change is not directly coupled to the regulatory role of the P_{II} proteins.

A second scenario is that the mutant does not react normally to stress because of the lack of the P_{II} proteins and the direct disruption of stress signal transduction. The BK strains grew more slowly and had altered NSR regulation, as indicated by higher GS activity and expression than in wild type Rm1021 *pstC*+ during nitrogen sufficiency. However, the induction of GS activity and expression in the BK *pstC*+ deletion may not be the result of NtrC-P activity, but rather, mediated through PhoB-P, since a mutation in *phoB* lowers GS expression in the BK background (Hagberg et al., 2016). A link between GlnB/GlnK and PhoB has been observed, but the mechanism is unclear. Based on these observations, the nitrogen responsive changes observed in the BK *pstC*+ strain may not be caused by the classical GlnD→GlnB/GlnK→NtrB/NtrC pathway but may instead result from aberrant PhoB-P regulation or the involvement of some other regulatory circuit. The large number of genes in 1021 *pstC*+ that are phosphate responsive suggests that many of these are linked to a general stress response triggered by low phosphate but that this connection may not be phosphate specific. In the BK *pstC*+ strain, the general stress response may not be activated if GlnB and GlnK are indeed involved in regulation of general stress, resulting in slow growth and small colonies as the cell is extremely inefficient in coping with stress. In this interpretation, many of the "phosphate responsive genes" would not be changing as a function of phosphate limitation since they are inactive due to the lack of the P_{II} proteins and their probable role as part of the cell's general stress response.

CONCLUSION

Sinorhizobium meliloti is a model for an important group of alpha-proteobacteria that reduce nitrogen during symbiosis with legumes and contribute a substantial amount of fixed nitrogen to terrestrial ecosystems. To be able to initiate the symbiosis by infecting the plant roots, rhizobia must be able to survive in soil. This requires that they deal successfully with various stresses in the soil environment—stresses that can change with time and can differ from site to site. Bacterial stress response is a term that describes coordinated changes in gene expression that help the organisms deal with a stress. Historically, stress responses have been studied individually, which is a useful starting point but probably not representative of more natural conditions.

This work was aimed at characterizing the interaction and coordination of responses to limited nitrogen and limited phosphate, individually and together. The results show that gene expression in response to one stress may change drastically if another stress is applied at the same time. The change may be quantitative, where the degree of response differs in a pairwise comparison, but the direction of the response can reverse, and our analysis included genes where induction by N

limitation with adequate P shifted to repression by N limitation in the deficient P condition and other repression/induction combinations. Reversal implies that a change useful under one condition might be a problem under the other. Phosphate stress generated a larger number of gene expression changes than nitrogen stress. Moreover, the response to phosphate limitation appeared to be dominant—when coping with both stresses, a set of genes that were induced during nitrogen stress were subsequently repressed by an additional phosphate stress. We interpret this to indicate that the phosphate stress used here was a larger burden on *S. meliloti*, and that dealing with the phosphate limitation had priority over dealing with the nitrogen limitation.

While the analysis of the data presented here depends to some extent on certain parameters used in the analysis, like the level of induction or repression used to group genes, we have carried out similar analyses using other parameters and the general conclusions are robust. Most notable was a striking and unexpected difference between the genes involved in the high phosphate NSRP and the low phosphate NSRp. 597 genes had two-fold induction or repression under at least one of these conditions representing about 10% of the genome. 252 genes were in the high phosphate NSRP and 315 in the low phosphate NSRp with only 30 genes common to both NSRs. Thus, while an NSR can be defined under each phosphate condition, the two NSRs are very different. It follows that designating a gene as nitrogen stress responsive is not as simple or useful as we thought it would be in designing this experiment. It also means that we are describing a major new class of NSR genes, those in NSRp, that were missed in previous analyses.

We do not know how generalizable the observations here may be. There are clearly some interactions between the NSRs and the PSRs and these involve regulation of genes that could be involved in other stress responses, like the iron stress response. Experiments designed to test the impact of limiting other nutrients would need to be done to assess these or to determine if the interactions are binary or related to some more general control pattern. Cross-regulation by regulatory proteins and the identification of imodulons or regulons composed of genes that are coordinately regulated, is a focus of recent research in other model organisms like *E. coli* (Sastry et al., 2019; Choudhary et al., 2020). Homologs of regulatory circuits in *S. meliloti* are often found in enteric bacteria, and our data are consistent with regulons found in *E. coli*. We would be surprised if the interactions between the NSR and the PSR are unique to *S. meliloti*.

The identification of sRNAs and mRNAs that are dual responsive to both nitrogen and phosphate stress contribute to an understanding of the coordination of stress responses. Our data show that regulation of individual genes in response to one stress can be altered drastically by imposing a second stress. Not only were novel genes identified as responding to each individual stress, but a group of genes were identified that only changed when both stresses were present. In addition, there was a subset of genes for which the direction of induction/repression was changed by the presence of a second stress, demonstrating the interconnected nature of the stress responses and how a single gene can have multiple layers of regulation that respond to the availability of more than one nutrient. Our data suggest that *S. meliloti* can prioritize stress pathways or specific processes based on nitrogen or phosphate availability.

An interesting group of genes was identified that responded significantly to both nitrogen and phosphate stress; the expression of many of these genes had been shown to be influenced by Hfq, an RNA binding protein and RNA chaperone. This association suggests that Hfq is involved in integrating the responses to various stress signals through its interaction with sRNAs and its role in translating mRNAs. Future experiments are needed to determine how the rhizobial proteome changes in response to similar growth conditions to determine if the mRNA responses to nitrogen and phosphate stress are reflected in protein abundance.

We demonstrated that the nitrogen and phosphate stress response pathways were not activated or inactivated independently or in only one dimension, but that they were instead connected with some groups of genes responding differently based on the status of a secondary stress. A key question this raises is how other nutrient stresses would affect the NSRs and PSRs. Cell envelope stress has been demonstrated to affect the PhoB regulon in *E. coli* (Choudhary et al., 2020). Changes in expression of heme biosynthesis and heme transport-related genes, and rhizobactin-related genes, suggest that iron stress regulation may be connected to nitrogen and phosphate stress regulation. While there might be interactions between nitrogen or phosphate stress and other stressors such as DNA damaging agents and oxidizing agents, there were only a few genes of this type, like *katA*, that were involved in the web of interactions we observed (**Figure 4**).

The multi-dimensional response to stresses is important in developing system-level knowledge of bacterial stress adaptation. This schematic of complex, inter-related stress responses might allow the development of bacteria able to function better under the relatively constrained conditions of industrial fermenter production or in naturally constrained environments. For the rhizobia, this could lead to increased survival in the soil or an increase in symbiotic competency, a desired outcome in areas where nitrogen fertilizer is not readily accessible or a financially feasible input for farmers. In addition to the development of rhizobia that are better suited to various soils and ultimately a better inoculum for legume crops, the results presented here suggest that stress interactions may be important in other contexts and suggests that measuring bacterial stress responses as a coordinated system could be important in understanding other bacteria that interact with specialized environments. The limited availability of nitrogen and phosphate in both terrestrial and aquatic environments suggests that coordination of stress responses may be common.

DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and

accession number(s) can be found below: https://www.ncbi.nlm.nih.gov/, PRJNA773132.

AUTHOR CONTRIBUTIONS

KH, SY, and MK contributed to conception and design of the study. KH acquired the data and wrote the first draft of the manuscript. KH, JP, and SY analyzed the data. KH, JP, SY, and MK contributed to data interpretation. KH, JP, and MK contributed figures. All authors contributed to manuscript revisions, and read and approved the submitted version.

FUNDING

This research was supported by the Agriculture Research Center at Washington State University and grants DE-FG03-96ER20225 from the United States Department of Energy–Basic Energy Sciences–Physical Biosciences Program and IOS-1645590 from the United States National Science Foundation.

ACKNOWLEDGMENTS

We would like to thank the WSU Genomics core facility in Spokane, WA for their help and for sequencing the RNA samples. We would also like to thank Jason Price for the cited research described in his dissertation thesis regarding the role of cold shock proteins in *S. meliloti*. We also thank Aaron Ogden, Katie Adolphsen, and Michael Mortimer for helpful discussions.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmicb.2022.800146/full#supplementary-material

REFERENCES

Al-Niemi, T. S., Summers, M. L., Elkins, J. G., Kahn, M. L., and McDermott, T. R. (1997). Regulation of the phosphate stress response in rhizobium meliloti by PhoB. *Appl. Environ. Microbiol.* 63, 4978–4981. doi: 10.1128/aem.63.12.4978-4981.1997

Arcondéguy, T., Huez, I., Tillard, P., Gangneux, C., de Billy, F., Gojon, A., et al. (1997). The *Rhizobium meliloti* PII protein, which controls bacterial nitrogen metabolism, affects alfalfa nodule development. *Genes Dev.* 11, 1194–1206. doi: 10.1101/gad.11.9.1194

Arcondeguy, T., Jack, R., and Merrick, M. (2001). PII signal transduction proteins, pivotal players in microbial nitrogen control. *Microbiol. Mol. Biol. Rev.* 65, 80–105. doi: 10.1128/MMBR.65.1.80-105.2001

Atkinson, M. R., and Ninfa, A. J. (1998). Role of the GlnK signal transduction protein in the regulation of nitrogen assimilation in *Escherichia coli*. *Mol. Microbiol.* 29, 431–447. doi: 10.1046/j.1365-2958.1998.00932.x

Barra-Bily, L., Fontenelle, C., Jan, G., Flechard, M., Trautwetter, A., Pandey, S. P., et al. (2010). Proteomic alterations explain phenotypic changes in *Sinorhizobium meliloti* lacking the RNA chaperone Hfq. *J. Bacteriol.* 192, 1719–1729. doi: 10.1128/JB.01429-1429

Bohlool, B. B., Ladha, J. K., Garrity, D. P., and George, T. (1992). Biological nitrogen fixation for sustainable agriculture: a perspective. *Plant Soil* 141, 1–11.

Choudhary, K. S., Kleinmanns, J. A., Decker, K., Sastry, A. V., Gao, Y., Szubin, R., et al. (2020). Elucidation of regulatory modes for five two-component systems in *Escherichia coli* reveals novel relationships. *mSystems* 5:e00980-20. doi: 10.1128/mSystems.00980-920

Elser, J. J., Bracken, M. E. S., Cleland, E. E., Gruner, D. S., Harpole, W. S., Hillebrand, H., et al. (2007). Global analysis of nitrogen and phosphorus limitation of primary producers in freshwater, marine and terrestrial ecosystems. *Ecol. Lett.* 10, 1135–1142. doi: 10.1111/j.1461-0248.2007.01113.x

Fanin, N., Hättenschwiler, S., Chavez Soria, P. F., and Fromin, N. (2016). (A)synchronous availabilities of N and P regulate the activity and structure of the microbial decomposer community. *Terr. Microbiol.* 6:1507. doi: 10.3389/fmicb.2015.01507

Galibert, F., Finan, T. M., Long, S. R., Puhler, A., Abola, P., Ampe, F., et al. (2001). The composite genome of the legume symbiont *Sinorhizobium meliloti*. *Science* 293, 668–672. doi: 10.1126/science.1060966

Gao, M., Barnett, M. J., Long, S. R., and Teplitski, M. (2010). Role of the *Sinorhizobium meliloti* global regulator Hfq in gene regulation and symbiosis. *Mol. Plant. Microbe Interact.* 23, 355–365. doi: 10.1094/MPMI-23-4-0355

Geiger, O., Röhrs, V., Weissenmayer, B., Finan, T. M., and Thomas-Oates, J. E. (1999). The regulator gene phoB mediates phosphate stress-controlled synthesis of the membrane lipid diacylglyceryl-N,N,N-trimethylhomoserine in Rhizobium (*Sinorhizobium meliloti*). *Mol. Microbiol.* 32, 63–73. doi: 10.1046/j.1365-2958.1999.01325.x

Hagberg, K. L., Yurgel, S. N., Mulder, M., and Kahn, M. L. (2016). Interaction between nitrogen and phosphate stress responses in *Sinorhizobium meliloti*. *Front. Microbiol.* 7:1928. doi: 10.3389/fmicb.2016.01928

Ikeda, T. P., Shauger, A. E., and Kustu, S. (1996). *Salmonella typhimurium* apparently perceives External nitrogen limitation as internal glutamine limitation. *J. Mol. Biol.* 259, 589–607. doi: 10.1006/jmbi.1996.0342

Jiang, P., Ventura, A. C., and Ninfa, A. J. (2012). Characterization of the reconstituted UTase/UR-PII-NRII-NRI bicyclic signal transduction system that controls the transcription of nitrogen-regulated (Ntr) genes in *Escherichia coli*. *Biochemistry* 51, 9045–9057. doi: 10.1021/bi300575j

Krol, E., and Becker, A. (2004). Global transcriptional analysis of the phosphate starvation response in *Sinorhizobium meliloti* strains 1021 and 2011. *Mol. Genet. Genomics MGG* 272, 1–17. doi: 10.1007/s00438-004-1030-8

Langfelder, P., and Horvath, S. (2008). WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics* 9:559. doi: 10.1186/1471-2105-9-559

Leigh, J. A., and Dodsworth, J. A. (2007). Nitrogen regulation in bacteria and archaea. *Annu. Rev. Microbiol.* 61, 349–377. doi: 10.1146/annurev.micro.61.080706.093409

López-Lara, I. M., Gao, J.-L., Soto, M. J., Solares-Pérez, A., Weissenmayer, B., Sohlenkamp, C., et al. (2005). Phosphorus-Free membrane lipids of sinorhizobium meliloti are not required for the symbiosis with alfalfa but contribute to increased cell yields under phosphorus-limiting conditions of growth. *Mol. Plant. Microbe Interact.* 18, 973–982. doi: 10.1094/MPMI-18-0973

Lubin, E. A., Henry, J. T., Fiebig, A., Crosson, S., and Laub, M. T. (2016). Identification of the PhoB regulon and role of PhoU in the phosphate starvation response of *Caulobacter crescentus*. *J. Bacteriol.* 198, 187–200. doi: 10.1128/JB.00658-615

Makino, K., Amemura, M., Kim, S. K., Nakata, A., and Shinagawa, H. (1993). Role of the sigma 70 subunit of RNA polymerase in transcriptional activation by activator protein PhoB in *Escherichia coli*. *Genes Dev.* 7, 149–160. doi: 10.1101/gad.7.1.149

McIntosh, M., Krol, E., and Becker, A. (2008). Competitive and cooperative effects in quorum-sensing-regulated galactoglucan biosynthesis in *Sinorhizobium meliloti*. *J. Bacteriol.* 190, 5308–5317. doi: 10.1128/JB.00063-68

Merrick, M. J., and Edwards, R. A. (1995). Nitrogen control in bacteria. *Microbiol. Rev.* 59, 604–622. doi: 10.1128/mr.59.4.604-622.1995

Mike, L. A., Choby, J. E., Brinkman, P. R., Olive, L. Q., Dutter, B. F., Ivan, S. J., et al. (2014). Two-Component system cross-regulation integrates bacillus anthracis response to heme and cell envelope stress. *PLoS Pathog* 10:e1004044. doi: 10.1371/journal.ppat.1004044

Mylona, P., Pawlowski, K., and Bisseling, T. (1995). Symbiotic nitrogen fixation. *Plant Cell* 7, 869–885.

Napolitani, C., Mandrich, L., Riccio, A., Lamberti, A., Manco, G., and Patriarca, E. J. (2004). Mutational analysis of GstI protein, a glutamine synthetase translational inhibitor of *Rhizobium leguminosarum*. *FEBS Lett.* 558, 45–51. doi: 10.1016/S0014-5793(03)01511-4

O'Connell, K. P., and Thomashow, M. F. (2000). Transcriptional organization and regulation of a polycistronic cold shock operon in *Sinorhizobium meliloti* RM1021 encoding homologs of the *Escherichia coli* major cold shock gene cspA and ribosomal protein gene rpsU. *Appl. Environ. Microbiol.* 66, 392–400. doi: 10.1128/AEM.66.1.392-400.2000

Parker, G. F., Higgins, T. P., Hawkes, T., and Robson, R. L. (1999). Rhizobium (*Sinorhizobium meliloti*) phn genes: characterization and identification of their protein products. *J. Bacteriol.* 181, 389–395. doi: 10.1128/JB.181.2.389-395.1999

Patriarca, E. J., Tatè, R., and Iaccarino, M. (2002). Key role of bacterial NH(4)(+) metabolism in Rhizobium-plant symbiosis. *Microbiol. Mol. Biol. Rev. MMBR* 66, 203–222. doi: 10.1128/MMBR.66.2.203-222.2002

Prasse, D., and Schmitz, R. A. (2018). Small RNAs involved in regulation of nitrogen metabolism. *Microbiol. Spectr.* 6.

Prell, J., Bourdès, A., Kumar, S., Lodwig, E., Hosie, A., Kinghorn, S., et al. (2010). Role of symbiotic auxotrophy in the rhizobium-legume symbioses. *PLoS One* 5:e13933. doi: 10.1371/journal.pone.0013933

Price, J. (2017). *Control of RNA Structure by CspA Proteins in Rhizobia*. Morrisville, NC: Proquest.

Ruiz, N., and Silhavy, T. J. (2003). Constitutive activation of the *Escherichia coli* Pho regulon upregulates rpoS translation in an Hfq-Dependent fashion. *J. Bacteriol.* 185, 5984–5992. doi: 10.1128/JB.185.20.5984-5992.2003

Sánchez-Cañizares, C., Prell, J., Pini, F., Rutten, P., Kraxner, K., Wynands, B., et al. (2020). Global control of bacterial nitrogen and carbon metabolism by a PTSNtr-regulated switch. *Proc. Natl. Acad. Sci. U S A.* 117, 10234–10245. doi: 10.1073/pnas.1917471117

Sastry, A. V., Gao, Y., Szubin, R., Hefner, Y., Xu, S., Kim, D., et al. (2019). The *Escherichia coli* transcriptome mostly consists of independently regulated modules. *Nat. Commun.* 10:5536. doi: 10.1038/s41467-019-13483-w

Shannon, P., Markiel, A., Ozier, O., Baliga, N. S., Wang, J. T., Ramage, D., et al. (2003). Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res.* 13, 2498–2504. doi: 10.1101/gr.1239303

Sobrero, P., Schlüter, J.-P., Lanner, U., Schlosser, A., Becker, A., and Valverde, C. (2012). Quantitative proteomic analysis of the Hfq-Regulon in *Sinorhizobium meliloti* 2011. *PLoS One* 7:e48494. doi: 10.1371/journal.pone.0048494

Soper, T., Mandin, P., Majdalani, N., Gottesman, S., and Woodson, S. A. (2010). Positive regulation by small RNAs and the role of Hfq. *Proc. Natl. Acad. Sci. U S A.* 107, 9602–9607. doi: 10.1073/pnas.1004435107

Spinosa, M., Riccio, A., Mandrich, L., Manco, G., Lamberti, A., Iaccarino, M., et al. (2000). Inhibition of glutamine synthetase II expression by the product of the gstI gene. *Mol. Microbiol.* 37, 443–452. doi: 10.1046/j.1365-2958.2000.02018.x

Torres-Quesada, O., Oruezabal, R. I., Peregrina, A., Jofré, E., Lloret, J., Rivilla, R., et al. (2010). The *Sinorhizobium meliloti* RNA chaperone Hfq influences central carbon metabolism and the symbiotic interaction with alfalfa. *BMC Microbiol.* 10:71. doi: 10.1186/1471-2180-10-71

Torres-Quesada, O., Reinkensmeier, J., Schlüter, J.-P., Robledo, M., Peregrina, A., Giegerich, R., et al. (2014). Genome-wide profiling of Hfq-binding RNAs uncovers extensive post-transcriptional rewiring of major stress response and symbiotic regulons in *Sinorhizobium meliloti*. *RNA Biol.* 11, 563–579. doi: 10.4161/rna.28239

Trapnell, C., Hendrickson, D. G., Sauvageau, M., Goff, L., Rinn, J. L., and Pachter, L. (2013). Differential analysis of gene regulation at transcript resolution with RNA-seq. *Nat. Biotechnol.* 31, 46–53. doi: 10.1038/nbt.2450

Trapnell, C., Roberts, A., Goff, L., Pertea, G., Kim, D., Kelley, D. R., et al. (2012). Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. *Nat. Protoc.* 7, 562–578. doi: 10.1038/nprot.2012.016

Van Bogelen, R. A., Olson, E. R., Wanner, B. L., and Neidhardt, F. C. (1996). Global analysis of proteins synthesized during phosphorus restriction in *Escherichia coli*. *J. Bacteriol.* 178, 4344–4366. doi: 10.1128/jb.178.15.4344-4366.1996

Vitousek, P. M., Porder, S., Houlton, B. Z., and Chadwick, O. A. (2010). Terrestrial phosphorus limitation: mechanisms, implications, and nitrogen-phosphorus interactions. *Ecol. Appl. Publ. Ecol. Soc. Am.* 20, 5–15. doi: 10.1890/08-0127.1

Voegele, R. T., Bardin, S., and Finan, T. M. (1997). Characterization of the Rhizobium (*Sinorhizobium meliloti*) high- and low-affinity phosphate uptake systems. *J. Bacteriol.* 179, 7226–7232. doi: 10.1128/jb.179.23.7226-7232.1997

Vytvytska, O., Moll, I., Kaberdin, V. R., von Gabain, A., and Bläsi, U. (2000). Hfq (HF1) stimulates ompA mRNA decay by interfering with ribosome binding. *Genes Dev.* 14, 1109–1118.

Wais, R. J., Wells, D. H., and Long, S. R. (2002). Analysis of differences between sinorhizobium meliloti 1021 and 2011 strains using the host calcium spiking response. *Mol. Plant-Microbe Interactions* 15, 1245–1252. doi: 10.1094/MPMI.2002.15.12.1245

Yuan, Z.-C., Zaheer, R., and Finan, T. M. (2006a). Regulation and properties of PstSCAB, a high-affinity, high-velocity phosphate transport system of *Sinorhizobium meliloti*. *J. Bacteriol.* 188, 1089–1102. doi: 10.1128/JB.188.3.1089-1102.2006

Yuan, Z.-C., Zaheer, R., Morton, R., and Finan, T. M. (2006b). Genome prediction of PhoB regulated promoters in *Sinorhizobium meliloti* and twelve proteobacteria. *Nucleic Acids Res.* 34, 2686–2697. doi: 10.1093/nar/gkl365

Yurgel, S. N., and Kahn, M. L. (2004). Dicarboxylate transport by rhizobia. *FEMS Microbiol. Rev.* 28, 489–501. doi: 10.1016/j.femsre.2004.04.002

Yurgel, S. N., and Kahn, M. L. (2005). Sinorhizobium meliloti dctA mutants with partial ability to transport dicarboxylic acids. *J. Bacteriol.* 187, 1161–1172. doi: 10.1128/JB.187.3.1161-1172.2005

Yurgel, S. N., Rice, J., and Kahn, M. L. (2012). Nitrogen metabolism in *Sinorhizobium meliloti*-alfalfa symbiosis: dissecting the role of GlnD and PII proteins. *Mol. Plant. Microbe Interact.* 25, 355–362. doi: 10.1094/MPMI-09-11-0249

Yurgel, S. N., Rice, J., and Kahn, M. L. (2013). Transcriptome analysis of the role of GlnD/GlnBK in nitrogen stress adaptation by *Sinorhizobium meliloti* Rm1021. *PLoS One* 8:e58028. doi: 10.1371/journal.pone.0058028

Yurgel, S. N., Rice, J., Mulder, M., and Kahn, M. L. (2010). GlnB/GlnK PII proteins and regulation of the *Sinorhizobium meliloti* Rm1021 nitrogen stress response and symbiotic function. *J. Bacteriol.* 192, 2473–2481. doi: 10.1128/JB.01657-1659

Zaheer, R., Morton, R., Proudfoot, M., Yakunin, A., and Finan, T. M. (2009). Genetic and biochemical properties of an alkaline phosphatase PhoX family protein found in many bacteria. *Environ. Microbiol.* 11, 1572–1587. doi: 10.1111/j.1462-2920.2009.01885

Zahran, H. H. (1999). Rhizobium-legume symbiosis and nitrogen fixation under severe conditions and in an arid climate. *Microbiol. Mol. Biol. Rev.* 63, 968–989. doi: 10.1128/MMBR.63.4.968-989.1999

Zhao, S., Guo, Y., Sheng, Q., and Shyr, Y. (2014). Advanced heat map and clustering analysis using heatmap3. *BioMed. Res. Int.* 2014:986048. doi: 10.1155/2014/986048

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Microbiology

MINI REVIEW
published: 08 March 2022
doi: 10.3389/fmicb.2022.865141

Update on the Protein Homeostasis Network in *Bacillus subtilis*

*Judith Matavacas and Claes von Wachenfeldt**

Department of Biology, Lund University, Lund, Sweden

Protein homeostasis is fundamental to cell function and survival. It relies on an interconnected network of processes involving protein synthesis, folding, post-translational modification and degradation as well as regulators of these processes. Here we provide an update on the roles, regulation and subcellular localization of the protein homeostasis machinery in the Gram-positive model organism *Bacillus subtilis*. We discuss emerging ideas and current research gaps in the field that, if tackled, increase our understanding of how Gram-positive bacteria, including several human pathogens, maintain protein homeostasis and cope with stressful conditions that challenge their survival.

OPEN ACCESS

Edited by:
Katarzyna Potrykus, University of Gdańsk, Poland

Reviewed by:
Pierre Genevaux, FR3743 Centre de Biologie Intégrative (CBI), France
Erhard Bremer, University of Marburg, Germany

***Correspondence:**
Claes von Wachenfeldt
claes.von_wachenfeldt@biol.lu.se
orcid.org/0000-0001-8950-3218

Specialty section:
This article was submitted to Microbial Physiology and Metabolism, a section of the journal Frontiers in Microbiology

Received: *29 January 2022*
Accepted: *15 February 2022*
Published: *08 March 2022*

Citation:
Matavacas J and von Wachenfeldt C (2022) Update on the Protein Homeostasis Network in Bacillus subtilis. Front. Microbiol. 13:865141. doi: 10.3389/fmicb.2022.865141

Keywords: chaperone, protease, degradation tags, protein quality control, protein aggregation, proteotoxic stress

INTRODUCTION

Native proteins typically fold into well-defined three-dimensional structures. To function properly, all cells need to contain correctly folded proteins and have mechanisms to prevent accumulation of unneeded or aberrant proteins. The folded state of most proteins is marginally more stable than the unfolded state. Therefore, small changes of environmental conditions may affect the equilibrium between the folded and unfolded state. Protein homeostasis (proteostasis) is crucial to achieve a "healthy" proteome, and refers to the dynamic balance between synthesis, folding, post-translational modification, transport, and degradation of proteins (**Figure 1**; Powers et al., 2009; Richter et al., 2010; Schramm et al., 2020). The main components of the proteostasis network are the ancient and evolutionary conserved chaperones and proteases, which assist in protein folding and degrade specific protein substrates, respectively (Powers and Balch, 2013; Balchin et al., 2016; Olivares et al., 2016). Surprisingly, even though eukaryotic proteomes are typically much larger and complex and contain more aggregation-prone proteins than those of prokaryotes, no new core chaperones appear to have emerged during billion years of evolution (Rebeaud et al., 2021). Instead, the core chaperones and their relative abundance have remained invariant across the domains of life. Maintaining integrity of a more complex and unstable proteome has been dealt with by increasing cellular chaperone levels, as well as promoting cooperation between them (Rebeaud et al., 2021).

Understanding how cells maintain proteostasis is an important topic to address, not only because proteome integrity is crucial for the correct cellular function, but also because accumulation of protein aggregates – which mainly results from dysregulation of proteostasis – has been linked to aging and to human diseases, such as Parkinson's and Alzheimer's, and to defects in growth and survival in prokaryotes (Balchin et al., 2016; Cheng et al., 2018). In addition, the presence of protein aggregates is strongly correlated with dormant antibiotic-resistant cells, called persisters (Leszczynska et al., 2013; Pu et al., 2019; Yu et al., 2019; Dewachter et al., 2021;

Huemer et al., 2021). Because of this and other data, disruption of proteostasis has been suggested as an anti-bacterial strategy (Khodaparast et al., 2021).

Most information on chaperones and proteases in bacteria derives from studies in *Escherichia coli*, but there are many host-specific regulation mechanisms concerning proteostasis to unravel. *B. subtilis* is adapted to rapid intracellular and environmental fluctuations that challenge the stability of its proteome. Thus, it is a suitable model organism to study not only from the view of fundamental principles of proteostasis, but also regarding proteostasis maintenance in other Gram-positive bacteria, that includes several human pathogens. Here we provide an updated description of the main components of the *B. subtilis* proteostasis network (**Figure 1**), and address known and emerging mechanisms for its regulation during proteotoxic stress.

THE MAJOR CLASSES OF MOLECULAR CHAPERONES: CONSERVED MECHANISMS OF ACTION AND ROLES IN *B. subtilis*

Molecular chaperones are central to proteostasis by ensuring that proteins are correctly folded, and preventing protein misfolding and aggregation (Mogk et al., 2011; Balchin et al., 2016). The ancient and evolutionary conserved DnaK (Hsp70), GroEL (Hsp60), and trigger factor (TF) are three important abundant cytosolic chaperones in *B. subtilis* (Moliere and Turgay, 2009).

DnaK functions as a monomer and consists of an N-terminal ATPase domain and a C-terminal peptide-binding domain composed of a β-sandwich and an α-helical lid (Zhu et al., 1996; Perales-Calvo et al., 2018). Together with its co-chaperone DnaJ (Hsp40) and the nucleotide exchange factor GrpE, and through ATP hydrolysis, the α-helical lid closes over the β-sandwich, allowing tight binding of unfolded substrates (Liberek et al., 1991; Zhu et al., 1996). DnaK typically recognizes exposed hydrophobic peptide segments (∼5–7 residues) of client proteins that are prone to aggregate during folding (Rudiger et al., 1997; Mogk et al., 1999; Calloni et al., 2012); substrate binding and release cycles decrease the folding rate, and prevent non-native protein species from folding prematurely in a misfolded state or from aggregating (Szabo et al., 1994). Interestingly, some bacteria like *E. coli* contain more than one Hsp70 and Hsp40 homolog. For instance, apart from DnaK, *E. coli* harbors two additional Hsp70 proteins – Hsc66 (Seaton and Vickery, 1994) and Hsc62 (Yoshimune et al., 1998) – and five additional DnaJ-like proteins such as Hsc20 and CbpA (Ueguchi et al., 1994; Lelivelt and Kawula, 1995; Clarke et al., 1996; Itoh et al., 1999; Yoshimune et al., 2002). In contrast, *B. subtilis* appears to have only DnaK and DnaJ. Oligomeric GroEL is composed of two stacked heptameric rings, each forming large cylindrical cavities in which misfolded protein substrates can be enclosed (Langer et al., 1992; Mayhew et al., 1996). The GroES (Hsp10) heptameric co-chaperonin caps GroEL cavities, and through ATP hydrolysis allows complete substrate encapsulation, thus providing a "protected" folding environment (Mayhew et al., 1996). Finally, TF is a ribosome-associated chaperone comprised of three domains adopting an overall elongated shape (Stoller et al., 1995; Zarnt et al., 1997). An N-terminal ribosome-binding domain is followed by a peptidyl-prolyl isomerase domain linked to the C-terminal substrate-binding domain. The substrate-binding domain has two helical arms that form a promiscuous clamp-like structure, providing a shielded environment to nascent polypeptides as translation proceeds, and also slowing the folding rate preventing aggregation (Agashe et al., 2004; Ferbitz et al., 2004; Singhal et al., 2015). The clamp, together with the TF structural flexibility, allows TF to function with a wide range of emerging substrates (Martinez-Hackert and Hendrickson, 2009; Saio et al., 2014). While the monomeric form of TF is bound to the ribosome, its dimeric form exists mainly in the cytosol, and apart from assisting several proteins in their folding, it has anti-aggregation activity (Saio et al., 2018).

In addition to assisting in the folding of cytosolic proteins, DnaK/DnaJ/GrpE (Wild et al., 1992, 1993, 1996), GroEL/GroES (Kusukawa et al., 1989), and TF (Lee and Bernstein, 2002; Genevaux et al., 2004; Ullers et al., 2007; Oh et al., 2011) are also involved in protein secretion by preventing premature folding and aggregation of presecretory proteins in the cytosol. Most proteins are translocated in an unfolded state via the general secretion (Sec) pathway. Typically, the secretion-specific chaperone SecB binds newly synthesized presecretory proteins and targets them for SecA-driven protein translocation. However, SecB is absent in many Gram-positive bacteria, and in *B. subtilis* CsaA has been suggested to play a similar role since, among other evidence, it interacts with SecA as well as several presecretory proteins (Müller et al., 2000; Linde et al., 2003).

Even though the general mechanisms of action and structures of DnaK, GroEL, and TF chaperones are widely conserved among organisms, their specific roles and their contributions to proteostasis maintenance differ between bacteria. A clear example of such divergence is displayed by the phenotypic differences between *E. coli* and *B. subtilis* chaperone-deficient mutants. Although in both organisms *groES* and *groEL* are essential genes at all temperatures (Fayet et al., 1989; Commichau et al., 2013), single and double deletions of *dnaK* and *tig* (encoding DnaK and TF) give different effects. In *E. coli*, DnaK is essential for growth at high or low temperature (Paek and Walker, 1987; Bukau and Walker, 1989), and it plays a crucial role under both optimal and proteotoxic stress conditions. Its absence causes cellular defects such as reduced growth rates, dysregulation of the heat-shock genes, and abnormal cell division (Paek and Walker, 1987; Bukau and Walker, 1989, 1990). DnaK is not only needed for the folding of a large number of *E. coli* "thermolabile" proteins (Mogk et al., 1999), it also regulates the heat shock sigma factor σ^{32} (Gamer et al., 1992; Liberek et al., 1992).

Deletion of *tig* in *E. coli* also leads to cellular defects: it reduces the cell's tolerance to SDS/EDTA and vancomycin, thereby reducing outer membrane integrity (Oh et al., 2011), and induces the heat shock response (Deuerling et al., 2003). In *E. coli*, DnaK and TF possess overlapping functions in protein folding, and their shared role seems to be crucial for maintaining proteostasis, even under typical

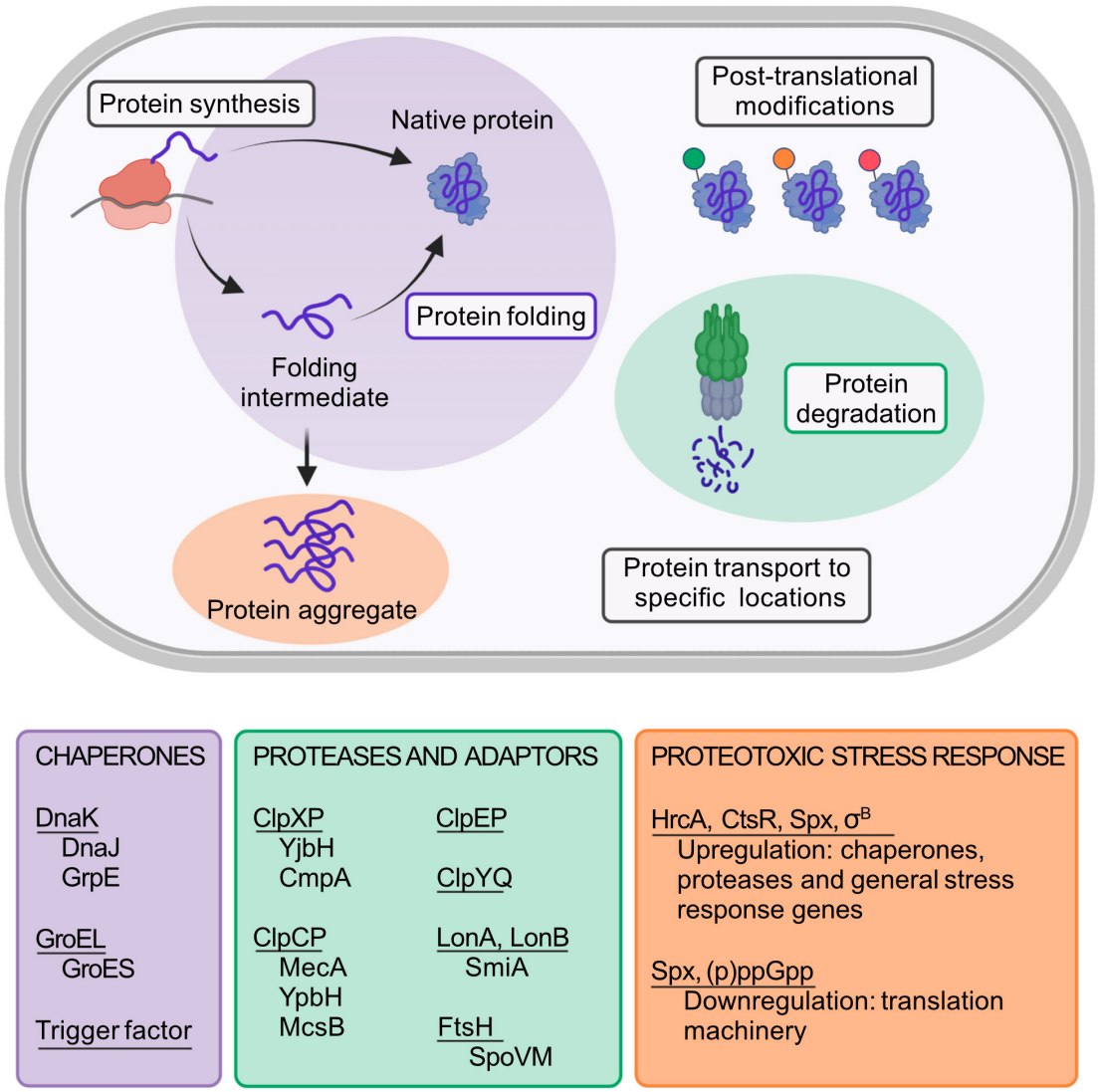


FIGURE 1 | Schematic representation of the *B. subtilis* proteostasis network and its regulators. Protein folding and degradation are two key cellular processes involved in proteostasis maintenance. While some proteins can spontaneously fold into their functional native state, many others need the assistance of molecular chaperones to do so. Underlined in the purple box are the main *B. subtilis* chaperones. For its full activity, DnaK requires the co-chaperone DnaJ and the nucleotide exchange factor GrpE. Similarly, GroEL requires the co-chaperonin GroES. Unneeded, misfolded or damaged proteins are eliminated from the cell by proteases (underlined in the green box) and its respective adaptor proteins (shown below each protease complex). In addition to protein folding and degradation, other key processes that affect proteostasis maintenance are protein synthesis, post-translational modifications, and the transport of proteins to specific locations. Failure in proteostasis maintenance often leads to the formation and accumulation of misfolded and aggregated proteins, a condition termed as proteotoxic stress. The main regulators that are known to be involved in the *B. subtilis* proteotoxic stress response are shown in the orange text box. Created with BioRender.com.

growth temperatures (Deuerling et al., 1999, 2003; Teter et al., 1999; Genevaux et al., 2004; Calloni et al., 2012). This redundancy might explain why a double deletion of *dnaK* and *tig* is synthetically lethal at temperatures above 30°C (Deuerling et al., 1999; Teter et al., 1999; Genevaux et al., 2004). Another example of chaperone collaboration is found between Hsp70 and the ATP-dependent Hsp90 chaperone, which in eukaryotes are well known to function together to orchestrate the proteostasis network (Schopf et al., 2017). In bacteria, the function of the Hsp90 homolog HtpG is not as well-characterized [see Wickner et al. (2021) for the latest review], but it has been shown that HtpG and DnaK systems also collaborate during the protein folding process in *E. coli* (Genest et al., 2011). DnaK-HtpG interaction involves the DnaJ-like protein CbpA (Genest et al., 2015). Unlike in eukaryotes, HtpG is not essential for the growth of many bacteria including *E. coli* and *B. subtilis*, and deletion of *htpG* causes minor growth defects after temperature upshifts in both organisms (Bardwell and Craig, 1988; Thomas and Baneyx, 1998; Versteeg et al., 1999). In addition, recent proteomic studies suggest that

HtpG enhances degradation of DnaK/DnaJ client substrates (Fauvet et al., 2021).

While the implications and roles of chaperones in proteostasis are well-characterized in *E. coli*, far less is known for other bacteria including *B. subtilis*. In stark contrast to *E. coli*, the absence of DnaK, TF or both proteins in *B. subtilis* does not affect cell viability in the 16–52°C temperature range (Schulz et al., 1995; Gothel et al., 1998; Reyes and Yoshikawa, 2002). Apart from a very short study (Reyes and Yoshikawa, 2002), no characterization of the effects of a *dnaK tig* double deletion in *B. subtilis* has been published. *B. subtilis dnaK tig* double mutants are viable below 53°C, suggesting that folding of nascent peptide chains is assisted also by other proteins than DnaK and TF (Reyes and Yoshikawa, 2002).

B. subtilis DnaK and TF are subjected to regulation by phosphorylation. Tyrosine residue 601 in the C-terminal region of DnaK can be phosphorylated by the PtkA kinase and dephosphorylated by the PtpZ phosphatase, influencing its chaperone activity and survival of the cell upon heat-shock (Shi et al., 2016). In the case of TF, phosphorylation of Arg45 by the McsB kinase negatively influences its association with the ribosome (Zhou et al., 2019). Interestingly, spore germination requires Arg45 to be dephosphorylated by the YwlE phosphatase, since this licenses TF to interact with ribosomes and resume translation (Zhou et al., 2019). These observations are in line with the notion that chaperones possess host-specific roles, in addition to their conserved functions.

ROLE OF AAA+ PROTEASES, ADAPTORS, AND DEGRADATION TAGS IN CLEARANCE OF ABERRANT PROTEINS

Unfinished, damaged, misfolded, or unneeded proteins are eliminated from the cell to maintain proteome integrity. In *B. subtilis*, degradation of most cytoplasmic proteins is performed by the conserved AAA+ family of intracellular proteases (AAA+; ATPases associated with a variety of cellular activities), which recognize, unfold, and degrade specific protein substrates (Sauer and Baker, 2011). *B. subtilis* has seven AAA+ proteases: ClpCP, ClpEP, ClpXP, ClpYQ, LonA, LonB, and FtsH (Elsholz et al., 2017), whose mechanisms of action have been described in numerous reviews (Sauer et al., 2004; Kirstein et al., 2009; Sauer and Baker, 2011; Olivares et al., 2016; Elsholz et al., 2017). The Clp complexes consist of an AAA+ unfoldase coupled to an ATP-dependent serine protease, whereas LonA, LonB, and FtsH have both unfoldase and protease domains within a single polypeptide (Elsholz et al., 2017).

Proteolysis can be regulated by adaptor proteins, which provide substrate specificity to proteases, usually by interacting with both substrate and protease. Several adaptor proteins have been characterized in *B. subtilis*, such as the ClpCP adaptor proteins MecA, YpbH, and McsB, and the ClpXP adaptor proteins YjbH and CmpA (Elsholz et al., 2017). The mechanism of adaptors often involves tethering the substrate to the protease to increase the local substrate concentration to facilitate proteolysis (Battesti and Gottesman, 2013). Less common mechanisms of adaptors are also known. For instance, the ClpXP adaptor protein YjbH does not appear to directly interact with ClpX (Chan et al., 2012), but enhances degradation of the stress-responsive regulator Spx by binding and stabilizing it, promoting its recognition by ClpXP (Awad et al., 2019).

AAA+ proteases or their respective adaptors recognize short degradation signals (degrons) located at the N-terminal, internal, or C-terminal position of protein substrates (Kirstein et al., 2009). A degron with relevance in protein homeostasis maintenance is the C-terminal SsrA degradation tag, which is added co-translationally by the transfer-messenger RNA (tmRNA) system to unfinished polypeptides when ribosomes stall (Keiler et al., 1996; Moore and Sauer, 2007). Truncated polypeptides challenge the stability of the proteome, and it is important that they are eliminated from the cell. SsrA-tagged polypeptides are typically degraded by the ClpXP protease complex (Sauer and Baker, 2011), although in *E. coli*, ClpAP and FtsH proteases can also recognize and degrade SsrA-tagged proteins (Gottesman et al., 1998; Herman et al., 1998). Cryo-EM studies have provided a detailed molecular mechanism of SsrA-tagged substrate recognition by ClpXP. Specific binding of the SsrA degron to ClpX triggers a ClpX conformational change from a "closed-pore" conformation to an "open-pore" conformation, allowing substrate translocation through the channel and subsequent non-specific interactions of the unfolded substrate with inner channel residues (Fei et al., 2020).

Interestingly, a novel ClpXP proteolytic mechanism for degradation of unfinished polypeptides in *B. subtilis* that is redundant with the SsrA tagging has been uncovered (Lytvynenko et al., 2019). Here, the *B. subtilis* RqcH recognizes stalled ribosomes and recruits $tRNA^{Ala}$ to mark aberrant nascent chains for degradation with C-terminal poly-alanine tails, which are recognized by ClpXP (Lytvynenko et al., 2019). Because the ALAA motif of the SsrA tag and poly-alanine tails are similar, it would be no surprise if their recognition and degradation mechanism would be similar. However, whether the poly-alanine tagged proteins are also degraded by ClpAP and FtsH in *B. subtilis* remains to be answered.

Another degradation tag of vast importance in *B. subtilis* proteostasis is the phospho-arginine (pArg) tag introduced by McsB (Trentini et al., 2016). McsB is conserved among Gram-positive bacteria and functions both as an adaptor protein for ClpCP (Kirstein et al., 2007) and as an arginine kinase with a major role in eliminating hundreds of damaged proteins from the cytoplasm, particularly under proteotoxic stress conditions (Elsholz et al., 2012; Trentini et al., 2016). McsB phosphorylates arginine residues, marking proteins for degradation by ClpCP (Trentini et al., 2016). Among the McsB phosphorylated proteins are the protein quality control members CtsR, HrcA, GroEL, TF, ClpC, and ClpP (Schmidt et al., 2014). In the case of the transcriptional repressors CtsR and HrcA, phosphorylation of residues in their DNA-binding domains greatly contributes to induction of the proteotoxic stress response (Kirstein et al., 2005; Fuhrmann et al., 2009; Schmidt et al., 2014). A recent study uncovered the molecular mechanism of McsB targeting

(Hajdusits et al., 2021). McsB assemble into octamers, stabilized by auto-phosphorylation to form a molecular chamber-like structure, with the kinase active site buried inside (Hajdusits et al., 2021). McsB octamers are formed upon proteotoxic stress conditions, when McsB levels increase, and possess high selectivity for phosphorylation of unfolded proteins, which are able to access the kinase chamber through a narrow entrance. The phosphorylated proteins are thus targeted for degradation by ClpCP (Hajdusits et al., 2021).

Although it is likely that *B. subtilis* Lon proteins and FtsH play some role in protein quality control, experimental evidence is lacking. However, roles in degradation of regulatory proteins have been reported (Thi Nguyen and Schumann, 2012; Bradshaw and Losick, 2015; Mukherjee et al., 2015).

SUBCELLULAR LOCALIZATION OF THE PROTEIN QUALITY CONTROL MACHINERY

B. subtilis ClpC, ClpE, ClpX, and ClpP have been shown to co-localize with heat-induced protein aggregates or PorA inclusion bodies (Kruger et al., 2000; Jurgen et al., 2001; Miethke et al., 2006). In non-stressed cells, without aggregates, ClpP appears in the cytoplasm, while ClpC and ClpX are found both in the cytoplasm and associated with the membrane (Kruger et al., 2000; Jurgen et al., 2001). Three independent publications reported in 2008 that *B. subtilis* GFP-tagged Clp proteins, such as ClpX and ClpP, form foci with a cell polar localization pattern (Kain et al., 2008; Kirstein et al., 2008; Simmons et al., 2008). The subcellular localization of Lon seems to be developmentally regulated: *B. subtilis* LonA-GFP associates with the nucleoid under normal growth, and with the forespore during sporulation (Simmons et al., 2008).

It is important to note that used fluorescent protein tags lead to clustering artifacts when fused to homo-oligomers such as Clp proteins, and that, at least in *E. coli*, Clp proteins are homogenously distributed in the cell (Landgraf et al., 2012). Therefore, the native sub-cellular localization of the *B. subtilis* proteolytic machinery should be revaluated.

Localization of chaperones seems to be conditional to stress in several bacteria. For example, in *E. coli* the co-chaperone DnaJ mediates ATP-DnaK binding to protein aggregates (Acebron et al., 2008). Moreover, large heat-induced protein aggregates localize at the cell poles, and such a localization requires DnaK and DnaJ, as well as ATP synthesis and the membrane proton motive force (Rokney et al., 2009). DnaK and the ClpB disaggregase are essential for dissolving polar aggregates (Rokney et al., 2009), but it is not clear whether the polar localization of aggregates in *E. coli* is energy-dependent, since other studies claim it to be a passive process, driven by the molecular crowding in the nucleoid region (Winkler et al., 2010; Coquel et al., 2013; Gupta et al., 2014; Neeli-Venkata et al., 2016; Oliveira et al., 2016). Large, polar localized, protein aggregates are asymmetrically inherited in *E. coli*, as division generates cells with aggregates at the old cell poles (Lindner et al., 2008; Winkler et al., 2010). In *B. subtilis*, protein aggregates have also been shown to locate at cell poles (Kirstein et al., 2008; Runde et al., 2014; Stannek et al., 2014; Hantke et al., 2019; Schafer et al., 2019), but their inheritance after cell division has not been studied.

To our knowledge, chaperone localization in *B. subtilis* has been only addressed in few studies. *B. subtilis* GFP-DnaK localizes as multiple discrete foci proximal to the membrane (Meile et al., 2006). In response to short-term ethanol stress, phosphorylated DnaK and GroEL chaperones are recruited to the *B. subtilis* cytoplasmic membrane (Seydlova et al., 2012).

Collectively, it seems that to cope with proteotoxic stress the cell redirects the protein quality control machinery to sub-cellular areas, containing protein aggregates, such as the cell poles. Interestingly, protein aggregates are typically associated with detrimental effects for cellular fitness (Ross and Poirier, 2004; Lindner et al., 2008; Mortier et al., 2019), but their presence has been reported to pre-adapt lineages to subsequent proteotoxic stress (Govers et al., 2014; Mortier et al., 2019). Such pre-adaptation may arise from the increased levels of protein quality control agents such as proteases and chaperones that co-localize with protein aggregates (Kruger et al., 2000; Jurgen et al., 2001; Acebron et al., 2008; Govers et al., 2014; Mortier et al., 2019).

REGULATORS OF THE PROTEOTOXIC STRESS RESPONSE

During proteotoxic stress, response mechanisms are activated which help the bacterium adapt to the new cellular or environmental condition. Of particular importance in *B. subtilis* are the HrcA and CtsR regulators. HrcA represses transcription of the *hrcA-grpE-dnaK-dnaJ-yqeT-yqeU-yqeV* and the *groES-groEL* operons (Schumann, 2016), and thus regulates the synthesis of chaperones. HrcA levels are depleted upon proteotoxic stress by a feedback mechanism involving the GroEL-GroES chaperone complex (Mogk et al., 1997; Schumann, 2016). CtsR, a regulator of protein degradation, represses transcription of the *ctsR-mcsA-mcsB-clpC-radA-disA* operon, and the *clpP* and *clpE* genes (Derre et al., 2000; Kruger et al., 2001; Elsholz et al., 2010). Regulation by CtsR involves a complex regulatory network, where McsB, McsA, and ClpCP play important roles in derepressing the CtsR regulon upon proteotoxic stress conditions [for a review, see Elsholz et al. (2017)].

Another player in the proteotoxic stress response is the Spx protein. Spx was initially characterized as a global regulator of the thiol-specific oxidative stress response (Nakano et al., 2003), controlling ∼144 transcriptional units (Rochat et al., 2012). However, an increasing number of studies have reported the involvement of Spx in the response to other stress conditions, such as heat shock and compounds targeting the cell wall (Runde et al., 2014; Rojas-Tapias and Helmann, 2018; Schafer and Turgay, 2019). The view that Spx is an important regulator of the proteotoxic stress response is becoming established (Rojas-Tapias and Helmann, 2019a,b; Schafer et al., 2019). Spx interacts with the C-terminal domain of the α-subunit (αCTD) of the RNA polymerase (RNAP), activating or repressing target genes in order to cope with the stress (Zuber, 2004; Newberry et al., 2005; Reyes and Zuber, 2008; Lamour et al., 2009; Nakano et al., 2010;

Rochat et al., 2012). As revealed by structural studies, redox activated Spx with a disulfide bond between the two cysteine residues (Cys10 and Cys13) interacts both with the αCTD and σ^A in the holo RNAP, and this complex binds to the −44 position of promoter DNA to enhance transcription activation (Shi et al., 2021). Among the Spx-induced genes are *trxA* (thioredoxin) and *trxB* (thioredoxin reductase), as well as the *clpX*, *clpE*, and *clpC* genes, and putatively *clpP* (Nakano et al., 2003; Rochat et al., 2012). Spx also induces the *ctsR* operon (Rojas-Tapias and Helmann, 2019a).

Control of the cellular level and activity of Spx involves many layers of regulation that are fine-tuned depending on the type of stress [reviewed in Rojas-Tapias and Helmann (2019b)]. The most important layer of Spx regulation seems to be through ClpXP proteolysis. Efficient degradation of Spx under normal conditions requires the ClpXP adaptor protein YjbH, which aggregates upon proteotoxic stress conditions and causes a decrease in Spx proteolysis (Nakano et al., 2003; Larsson et al., 2007; Garg et al., 2009; Engman and von Wachenfeldt, 2015). By hydrogen-deuterium exchange mass spectrometry it was determined that binding to YjbH decreases the Spx dynamics, reducing the conformational entropy and probably allowing a more efficient recognition of its C-terminal end, needed for ClpXP degradation (Awad et al., 2019). ClpCP and its adaptor McsB are also involved in Spx degradation, although to a lesser extent (Rojas-Tapias and Helmann, 2019a). Interestingly, the Spx paralog MgsR, which is also involved in the oxidative stress response (Reder et al., 2008), has been shown to interact with McsB upon ethanol stress, and McsB enhances MgsR degradation by ClpXP *in vivo* (Lilge et al., 2020).

Proteotoxic stress conditions also induce the general stress response, which is governed by the alternative sigma factor σ^B and is one of the most important non-specific stress response mechanisms of *B. subtilis*. The σ^B regulon is induced through a signal transduction cascade, involving the RsbV, RsbW, and RsbX regulators, and comprises about 200 genes, defined as class II heat-shock genes (Schumann, 2003; Nannapaneni et al., 2012). Other than heat, the general stress response is triggered by a wide range of stresses (Hecker et al., 2007). Genes included in the σ^B regulon are, for example, genes that protect against elevated temperatures, such as *clpP* and *clpC*, and against oxidative stress, such as thioredoxin (*trxA*), peroxidase (*ohrA*), and superoxide dismutase (*sodA*) (Helmann et al., 2001; Petersohn et al., 2001; Price et al., 2001; Nannapaneni et al., 2012) and regulators (e.g., CtsR and Spx) (Hecker et al., 2007).

MECHANISMS TO DOWNREGULATE THE TRANSLATION MACHINERY DURING PROTEOTOXIC STRESS

Proteotoxic stress and other physiological demands on proteostasis may lead to insufficient protein folding capacity, resulting in accumulation of aberrant proteins. Reducing the rate of translation lowers the protein load, preventing further protein damage, and may help maintenance of proteostasis. Under proteotoxic stress conditions, such as heat or oxidative stress, *B. subtilis* downregulates transcription of translation-related genes, including ribosomal-protein encoding genes (*rplD*, *rpsC*, *rplW*, and *rpsJ*), and ribosomal RNA (rRNA) genes (Price et al., 2001; Leichert et al., 2003; Mostertz et al., 2004; Rochat et al., 2012; Schafer et al., 2019). In line with this notion, translation-related proteins have been found in protein aggregates in *E. coli* (Kwiatkowska et al., 2008; Dewachter et al., 2021). Moreover, the translation rate is reduced in *E. coli* cells containing a dysfunctional GroEL (Chapman et al., 2006), and protein folding is enhanced by slowing down translation rates in *E. coli* cells harboring mutant ribosomes (Siller et al., 2010).

Spx, apart from inducing transcription of stress-responsive genes, is also capable of repressing expression of genes for ribosomal proteins and rRNA (Nakano et al., 2003; Rochat et al., 2012; Schafer et al., 2019). Among the Spx downregulated genes are *rpoA* and *rpoC*, encoding for RNAP core subunits, and *lepA*, encoding for elongation factor 4 (EF4/LepA; Schafer et al., 2019), a paralog of the canonical elongation factor EF-G (Evans et al., 2008).

Downregulation of the translation machinery is also observed in cells lacking Spx (Schafer et al., 2019) suggesting that, at least under the stress conditions investigated (heat, oxidative, and cell wall stress), there are additional mechanisms to reduce translation in *B. subtilis*.

A downregulation of translation is frequently linked to the second messengers of nutrient starvation, ppGpp and pppGpp [collectively referred to as (p)ppGpp], which are synthesized by RelA (Rel in *B. subtilis*) and mediate the stringent response (Roghanian et al., 2021). Recently, a role of (p)ppGpp in slowing down the translation rate upon proteotoxic stress in *B. subtilis* has been suggested (Schafer et al., 2020).

McsB, which as mentioned before targets proteins for degradation by ClpCP, also has a potential role in regulating translation upon heat or oxidative stress, since it targets proteins related to translational control (Schmidt et al., 2014).

Inactivation of the *B. subtilis* methionine synthase MetE could also contribute to limit the translation rate by depleting the biosynthesis of the precursor of the initiation codon formyl-Met (Chi et al., 2011). Indeed, oxidative stress has been shown to inactivate *E. coli* MetE (Hondorp and Matthews, 2004), and oxidized MetE appears in the aggregate fraction of GroEL-defective mutants cells (Chapman et al., 2006). In response to the oxidative stress agents diamide or sodium hypochlorite, specific cysteine residues of *B. subtilis* MetE become either S-cysteinylated (Hochgrafe et al., 2007) or S-bacillithiolated, leading to its enzymatic inactivation (Chi et al., 2011). Moreover, MetE is among 108 identified S-thioallylated proteins caused by garlic sulfur compounds, and such compounds were found to induce the HrcA, CtsR, and Spx regulons (Chi et al., 2019).

CONCLUDING REMARKS

B. subtilis is tolerant to drastic and rapid environmental changes by having a network of regulators and mechanisms controlling the synthesis, folding, post-translational

modifications, sub-cellular localization and clearance of proteins to ensure proteostasis. Several of these processes are controlled differently in *E. coli* and *B. subtilis*, reflecting a long history of adaptations of the two model organisms to different niches.

AUTHOR CONTRIBUTIONS

Both authors listed have made a substantial, direct, and intellectual contribution to the work, and approved it for publication.

FUNDING

Work in the laboratory of CW was supported by the Swedish Research Council Grant 2019-05578_3.

ACKNOWLEDGMENTS

We thank L. Hederstedt for critical reading of the manuscript.

REFERENCES

Acebron, S. P., Fernandez-Saiz, V., Taneva, S. G., Moro, F., and Muga, A. (2008). DnaJ recruits DnaK to protein aggregates. *J. Biol. Chem.* 283, 1381–1390. doi: 10.1074/jbc.M706189200

Agashe, V. R., Guha, S., Chang, H. C., Genevaux, P., Hayer-Hartl, M., Stemp, M., et al. (2004). Function of trigger factor and DnaK in multidomain protein folding: increase in yield at the expense of folding speed. *Cell* 117, 199–209. doi: 10.1016/s0092-8674(04)00299-5

Awad, W., Al-Eryani, Y., Ekstrom, S., Logan, D. T., and von Wachenfeldt, C. (2019). Structural basis for YjbH adaptor-mediated recognition of transcription factor Spx. *Structure* 27, 923–936.e6. doi: 10.1016/j.str.2019.03.009

Balchin, D., Hayer-Hartl, M., and Hartl, F. U. (2016). *In vivo* aspects of protein folding and quality control. *Science* 353:aac4354. doi: 10.1126/science.aac4354

Bardwell, J. C., and Craig, E. A. (1988). Ancient heat shock gene is dispensable. *J. Bacteriol.* 170, 2977–2983. doi: 10.1128/jb.170.7.2977-2983.1988

Battesti, A., and Gottesman, S. (2013). Roles of adaptor proteins in regulation of bacterial proteolysis. *Curr. Opin. Microbiol.* 16, 140–147. doi: 10.1016/j.mib.2013.01.002

Bradshaw, N., and Losick, R. (2015). Asymmetric division triggers cell-specific gene expression through coupled capture and stabilization of a phosphatase. *eLife* 4:e08145. doi: 10.7554/eLife.08145

Bukau, B., and Walker, G. C. (1989). Cellular defects caused by deletion of the Escherichia coli dnaK gene indicate roles for heat shock protein in normal metabolism. *J. Bacteriol.* 171, 2337–2346. doi: 10.1128/jb.171.5.2337-2346.1989

Bukau, B., and Walker, G. C. (1990). Mutations altering heat shock specific subunit of RNA polymerase suppress major cellular defects of *E. coli* mutants lacking the DnaK chaperone. *EMBO J.* 9, 4027–4036. doi: 10.1002/j.1460-2075.1990.tb07624.x

Calloni, G., Chen, T., Schermann, S. M., Chang, H. C., Genevaux, P., Agostini, F., et al. (2012). DnaK functions as a central hub in the *E. coli* chaperone network. *Cell Rep.* 1, 251–264. doi: 10.1016/j.celrep.2011.12.007

Chan, C. M., Garg, S., Lin, A. A., and Zuber, P. (2012). *Geobacillus thermodenitrificans* YjbH recognizes the C-terminal end of *Bacillus subtilis* Spx to accelerate Spx proteolysis by ClpXP. *Microbiology* 158, 1268–1278. doi: 10.1099/mic.0.057661-0

Chapman, E., Farr, G. W., Usaite, R., Furtak, K., Fenton, W. A., Chaudhuri, T. K., et al. (2006). Global aggregation of newly translated proteins in an *Escherichia coli* strain deficient of the chaperonin GroEL. *Proc. Natl. Acad. Sci. U.S.A.* 103, 15800–15805. doi: 10.1073/pnas.0607534103

Cheng, J., North, B. J., Zhang, T., Dai, X., Tao, K., Guo, J., et al. (2018). The emerging roles of protein homeostasis-governing pathways in Alzheimer's disease. *Aging Cell* 17:e12801. doi: 10.1111/acel.12801

Chi, B. K., Gronau, K., Mader, U., Hessling, B., Becher, D., and Antelmann, H. (2011). S-bacillithiolation protects against hypochlorite stress in *Bacillus subtilis* as revealed by transcriptomics and redox proteomics. *Mol. Cell Proteomics* 10:M111.009506. doi: 10.1074/mcp.M111.009506

Chi, B. K., Huyen, N. T. T., Loi, V. V., Gruhlke, M. C. H., Schaffer, M., Mader, U., et al. (2019). The disulfide stress response and protein S-thioallylation caused by allicin and diallyl polysulfanes in *Bacillus subtilis* as revealed by transcriptomics and proteomics. *Antioxidants* 8:605. doi: 10.3390/antiox8120605

Clarke, D. J., Jacq, A., and Holland, I. B. (1996). A novel DnaJ-like protein in *Escherichia coli* inserts into the cytoplasmic membrane with a type III topology. *Mol. Microbiol.* 20, 1273–1286. doi: 10.1111/j.1365-2958.1996.tb02646.x

Commichau, F. M., Pietack, N., and Stulke, J. (2013). Essential genes in *Bacillus subtilis*: a re-evaluation after ten years. *Mol. Biosyst.* 9, 1068–1075. doi: 10.1039/c3mb25595f

Coquel, A. S., Jacob, J. P., Primet, M., Demarez, A., Dimiccoli, M., Julou, T., et al. (2013). Localization of protein aggregation in *Escherichia coli* is governed by diffusion and nucleoid macromolecular crowding effect. *PLoS Comput. Biol.* 9:e1003038. doi: 10.1371/journal.pcbi.1003038

Derre, I., Rapoport, G., and Msadek, T. (2000). The CtsR regulator of stress response is active as a dimer and specifically degraded *in vivo* at 37 degrees C. *Mol. Microbiol.* 38, 335–347. doi: 10.1046/j.1365-2958.2000.02124.x

Deuerling, E., Patzelt, H., Vorderwulbecke, S., Rauch, T., Kramer, G., Schaffitzel, E., et al. (2003). Trigger Factor and DnaK possess overlapping substrate pools and binding specificities. *Mol. Microbiol.* 47, 1317–1328. doi: 10.1046/j.1365-2958.2003.03370.x

Deuerling, E., Schulze-Specking, A., Tomoyasu, T., Mogk, A., and Bukau, B. (1999). Trigger factor and DnaK cooperate in folding of newly synthesized proteins. *Nature* 400, 693–696. doi: 10.1038/23301

Dewachter, L., Bollen, C., Wilmaerts, D., Louwagie, E., Herpels, P., Matthay, P., et al. (2021). The dynamic transition of persistence toward the viable but nonculturable state during stationary phase is driven by protein aggregation. *mBio* 12:e0070321. doi: 10.1128/mBio.00703-21

Elsholz, A. K., Gerth, U., and Hecker, M. (2010). Regulation of CtsR activity in low GC. Gram+ bacteria. *Adv. Microb. Physiol.* 57, 119–144. doi: 10.1016/B978-0-12-381045-8.00003-5

Elsholz, A. K., Turgay, K., Michalik, S., Hessling, B., Gronau, K., Oertel, D., et al. (2012). Global impact of protein arginine phosphorylation on the physiology of *Bacillus subtilis*. *Proc. Natl. Acad. Sci. U.S.A.* 109, 7451–7456. doi: 10.1073/pnas.1117483109

Elsholz, A. K. W., Birk, M. S., Charpentier, E., and Turgay, K. (2017). Functional diversity of AAA+ protease complexes in *Bacillus subtilis*. *Front. Mol. Biosci.* 4:44. doi: 10.3389/fmolb.2017.00044

Engman, J., and von Wachenfeldt, C. (2015). Regulated protein aggregation: a mechanism to control the activity of the ClpXP adaptor protein YjbH. *Mol. Microbiol.* 95, 51–63. doi: 10.1111/mmi.12842

Evans, R. N., Blaha, G., Bailey, S., and Steitz, T. A. (2008). The structure of LepA, the ribosomal back translocase. *Proc. Natl. Acad. Sci. U.S.A.* 105, 4673–4678. doi: 10.1073/pnas.0801308105

Fauvet, B., Finka, A., Castanie-Cornet, M. P., Cirinesi, A. M., Genevaux, P., Quadroni, M., et al. (2021). Bacterial Hsp90 facilitates the degradation of aggregation-prone Hsp70-Hsp40 substrates. *Front. Mol. Biosci.* 8:653073. doi: 10.3389/fmolb.2021.653073

Fayet, O., Ziegelhoffer, T., and Georgopoulos, C. (1989). The *groES* and *groEL* heat shock gene products of *Escherichia coli* are essential for bacterial growth at all temperatures. *J. Bacteriol.* 171, 1379–1385. doi: 10.1128/jb.171.3.1379-1385.1989

Fei, X., Bell, T. A., Barkow, S. R., Baker, T. A., and Sauer, R. T. (2020). Structural basis of ClpXP recognition and unfolding of ssrA-tagged substrates. *eLife* 9:e61496. doi: 10.7554/eLife.61496

Ferbitz, L., Maier, T., Patzelt, H., Bukau, B., Deuerling, E., and Ban, N. (2004). Trigger factor in complex with the ribosome forms a molecular cradle for nascent proteins. *Nature* 431, 590–596. doi: 10.1038/nature02899

Fuhrmann, J., Schmidt, A., Spiess, S., Lehner, A., Turgay, K., Mechtler, K., et al. (2009). McsB is a protein arginine kinase that phosphorylates and inhibits the heat-shock regulator CtsR. *Science* 324, 1323–1327. doi: 10.1126/science.1170088

Gamer, J., Bujard, H., and Bukau, B. (1992). Physical interaction between heat shock proteins DnaK, DnaJ, and GrpE and the bacterial heat shock transcription factor sigma 32. *Cell* 69, 833–842. doi: 10.1016/0092-8674(92)90294-m

Garg, S. K., Kommineni, S., Henslee, L., Zhang, Y., and Zuber, P. (2009). The YjbH protein of *Bacillus subtilis* enhances ClpXP-catalyzed proteolysis of Spx. *J. Bacteriol.* 191, 1268–1277. doi: 10.1128/JB.01289-08

Genest, O., Hoskins, J. R., Camberg, J. L., Doyle, S. M., and Wickner, S. (2011). Heat shock protein 90 from *Escherichia coli* collaborates with the DnaK chaperone system in client protein remodeling. *Proc. Natl. Acad. Sci. U.S.A.* 108, 8206–8211. doi: 10.1073/pnas.1104703108

Genest, O., Hoskins, J. R., Kravats, A. N., Doyle, S. M., and Wickner, S. (2015). Hsp70 and Hsp90 of *E. coli* directly interact for collaboration in protein remodeling. *J. Mol. Biol.* 427, 3877–3889. doi: 10.1016/j.jmb.2015.10.010

Genevaux, P., Keppel, F., Schwager, F., Langendijk-Genevaux, P. S., Hartl, F. U., and Georgopoulos, C. (2004). *In vivo* analysis of the overlapping functions of DnaK and trigger factor. *EMBO Rep.* 5, 195–200. doi: 10.1038/sj.embor.7400067

Gothel, S. F., Scholz, C., Schmid, F. X., and Marahiel, M. A. (1998). Cyclophilin and trigger factor from *Bacillus subtilis* catalyze *in vitro* protein folding and are necessary for viability under starvation conditions. *Biochemistry* 37, 13392–13399. doi: 10.1021/bi981253w

Gottesman, S., Roche, E., Zhou, Y., and Sauer, R. T. (1998). The ClpXP and ClpAP proteases degrade proteins with carboxy-terminal peptide tails added by the SsrA-tagging system. *Genes Dev.* 12, 1338–1347. doi: 10.1101/gad.12.9.1338

Govers, S. K., Dutre, P., and Aertsen, A. (2014). *In vivo* disassembly and reassembly of protein aggregates in *Escherichia coli*. *J. Bacteriol.* 196, 2325–2332. doi: 10.1128/JB.01549-14

Gupta, A., Lloyd-Price, J., Neeli-Venkata, R., Oliveira, S. M., and Ribeiro, A. S. (2014). *In vivo* kinetics of segregation and polar retention of MS2-GFP-RNA complexes in *Escherichia coli*. *Biophys. J.* 106, 1928–1937. doi: 10.1016/j.bpj.2014.03.035

Hajdusits, B., Suskiewicz, M. J., Hundt, N., Meinhart, A., Kurzbauer, R., Leodolter, J., et al. (2021). McsB forms a gated kinase chamber to mark aberrant bacterial proteins for degradation. *eLife* 10:e63505. doi: 10.7554/eLife.63505

Hantke, I., Schafer, H., Janczikowski, A., and Turgay, K. (2019). YocM a small heat shock protein can protect *Bacillus subtilis* cells during salt stress. *Mol. Microbiol.* 111, 423–440. doi: 10.1111/mmi.14164

Hecker, M., Pane-Farre, J., and Volker, U. (2007). SigB-dependent general stress response in *Bacillus subtilis* and related gram-positive bacteria. *Annu. Rev. Microbiol.* 61, 215–236. doi: 10.1146/annurev.micro.61.080706.093445

Helmann, J. D., Wu, M. F., Kobel, P. A., Gamo, F. J., Wilson, M., Morshedi, M. M., et al. (2001). Global transcriptional response of *Bacillus subtilis* to heat shock. *J. Bacteriol.* 183, 7318–7328. doi: 10.1128/JB.183.24.7318-7328.2001

Herman, C., Thevenet, D., Bouloc, P., Walker, G. C., and D'Ari, R. (1998). Degradation of carboxy-terminal-tagged cytoplasmic proteins by the *Escherichia coli* protease HflB (FtsH). *Genes Dev.* 12, 1348–1355. doi: 10.1101/gad.12.9.1348

Hochgrafe, F., Mostertz, J., Pother, D. C., Becher, D., Helmann, J. D., and Hecker, M. (2007). S-cysteinylation is a general mechanism for thiol protection of *Bacillus subtilis* proteins after oxidative stress. *J. Biol. Chem.* 282, 25981–25985. doi: 10.1074/jbc.C700105200

Hondorp, E. R., and Matthews, R. G. (2004). Oxidative stress inactivates cobalamin-independent methionine synthase (MetE) in *Escherichia coli*. *PLoS Biol.* 2:e336. doi: 10.1371/journal.pbio.0020336

Huemer, M., Mairpady Shambat, S., Bergada-Pijuan, J., Soderholm, S., Boumasmoud, M., Vulin, C., et al. (2021). Molecular reprogramming and phenotype switching in *Staphylococcus aureus* lead to high antibiotic persistence and affect therapy success. *Proc. Natl. Acad. Sci. U.S.A.* 118:e2014920118 doi: 10.1073/pnas.2014920118

Itoh, T., Matsuda, H., and Mori, H. (1999). Phylogenetic analysis of the third Hsp70 homolog in *Escherichia coli*; a novel member of the Hsc66 subfamily and its possible co-chaperone. *DNA Res.* 6, 299–305. doi: 10.1093/dnares/6.5.299

Jurgen, B., Hanschke, R., Sarvas, M., Hecker, M., and Schweder, T. (2001). Proteome and transcriptome based analysis of *Bacillus subtilis* cells overproducing an insoluble heterologous protein. *Appl. Microbiol. Biotechnol.* 55, 326–332. doi: 10.1007/s002530000531

Kain, J., He, G. G., and Losick, R. (2008). Polar localization and compartmentalization of ClpP proteases during growth and sporulation in *Bacillus subtilis*. *J. Bacteriol.* 190, 6749–6757. doi: 10.1128/JB.00589-08

Keiler, K. C., Waller, P. R., and Sauer, R. T. (1996). Role of a peptide tagging system in degradation of proteins synthesized from damaged messenger RNA. *Science* 271, 990–993. doi: 10.1126/science.271.5251.990

Khodaparast, L., Wu, G., Khodaparast, L., Schmidt, B. Z., Rousseau, F., and Schymkowitz, J. (2021). Bacterial protein homeostasis disruption as a therapeutic intervention. *Front. Mol. Biosci.* 8:681855. doi: 10.3389/fmolb.2021.681855

Kirstein, J., Dougan, D. A., Gerth, U., Hecker, M., and Turgay, K. (2007). The tyrosine kinase McsB is a regulated adaptor protein for ClpCP. *EMBO J.* 26, 2061–2070. doi: 10.1038/sj.emboj.7601655

Kirstein, J., Moliere, N., Dougan, D. A., and Turgay, K. (2009). Adapting the machine: adaptor proteins for Hsp100/Clp and AAA+ proteases. *Nat. Rev. Microbiol.* 7, 589–599. doi: 10.1038/nrmicro2185

Kirstein, J., Strahl, H., Moliere, N., Hamoen, L. W., and Turgay, K. (2008). Localization of general and regulatory proteolysis in *Bacillus subtilis* cells. *Mol. Microbiol.* 70, 682–694. doi: 10.1111/j.1365-2958.2008.06438.x

Kirstein, J., Zuhlke, D., Gerth, U., Turgay, K., and Hecker, M. (2005). A tyrosine kinase and its activator control the activity of the CtsR heat shock repressor in *B. subtilis*. *EMBO J.* 24, 3435–3445. doi: 10.1038/sj.emboj.7600780

Kruger, E., Witt, E., Ohlmeier, S., Hanschke, R., and Hecker, M. (2000). The clp proteases of *Bacillus subtilis* are directly involved in degradation of misfolded proteins. *J. Bacteriol.* 182, 3259–3265. doi: 10.1128/JB.182.11.3259-3265.2000

Kruger, E., Zuhlke, D., Witt, E., Ludwig, H., and Hecker, M. (2001). Clp-mediated proteolysis in Gram-positive bacteria is autoregulated by the stability of a repressor. *EMBO J.* 20, 852–863. doi: 10.1093/emboj/20.4.852

Kusukawa, N., Yura, T., Ueguchi, C., Akiyama, Y., and Ito, K. (1989). Effects of mutations in heat-shock genes *groES* and *groEL* on protein export in *Escherichia coli*. *EMBO J.* 8, 3517–3521. doi: 10.1002/j.1460-2075.1989.tb08517.x

Kwiatkowska, J., Matuszewska, E., Kuczynska-Wisnik, D., and Laskowska, E. (2008). Aggregation of *Escherichia coli* proteins during stationary phase depends on glucose and oxygen availability. *Res. Microbiol.* 159, 651–657. doi: 10.1016/j.resmic.2008.09.008

Lamour, V., Westblade, L. F., Campbell, E. A., and Darst, S. A. (2009). Crystal structure of the *in vivo*-assembled *Bacillus subtilis* Spx/RNA polymerase alpha subunit C-terminal domain complex. *J. Struct. Biol.* 168, 352–356. doi: 10.1016/j.jsb.2009.07.001

Landgraf, D., Okumus, B., Chien, P., Baker, T. A., and Paulsson, J. (2012). Segregation of molecules at cell division reveals native protein localization. *Nat. Methods* 9, 480–482. doi: 10.1038/nmeth.1955

Langer, T., Pfeifer, G., Martin, J., Baumeister, W., and Hartl, F. U. (1992). Chaperonin-mediated protein folding: GroES binds to one end of the GroEL cylinder, which accommodates the protein substrate within its central cavity. *EMBO J.* 11, 4757–4765. doi: 10.1002/j.1460-2075.1992.tb05581.x

Larsson, J. T., Rogstam, A., and von Wachenfeldt, C. (2007). YjbH is a novel negative effector of the disulphide stress regulator, Spx, in *Bacillus subtilis*. *Mol. Microbiol.* 66, 669–684. doi: 10.1111/j.1365-2958.2007.05949.x

Lee, H. C., and Bernstein, H. D. (2002). Trigger factor retards protein export in *Escherichia coli*. *J. Biol. Chem.* 277, 43527–43535. doi: 10.1074/jbc.M205950200

Leichert, L. I., Scharf, C., and Hecker, M. (2003). Global characterization of disulfide stress in *Bacillus subtilis*. *J. Bacteriol.* 185, 1967–1975. doi: 10.1128/JB.185.6.1967-1975.2003

Lelivelt, M. J., and Kawula, T. H. (1995). Hsc66, an Hsp70 homolog in *Escherichia coli*, is induced by cold shock but not by heat shock. *J. Bacteriol.* 177, 4900–4907. doi: 10.1128/jb.177.17.4900-4907.1995

Leszczynska, D., Matuszewska, E., Kuczynska-Wisnik, D., Furmanek-Blaszk, B., and Laskowska, E. (2013). The formation of persister cells in stationary-phase cultures of *Escherichia coli* is associated with the aggregation of endogenous proteins. *PLoS One* 8:e54737. doi: 10.1371/journal.pone.0054737

Liberek, K., Galitski, T. P., Zylicz, M., and Georgopoulos, C. (1992). The DnaK chaperone modulates the heat shock response of *Escherichia coli* by binding to the sigma 32 transcription factor. *Proc. Natl. Acad. Sci. U.S.A.* 89, 3516–3520. doi: 10.1073/pnas.89.8.3516

Liberek, K., Marszalek, J., Ang, D., Georgopoulos, C., and Zylicz, M. (1991). *Escherichia coli* DnaJ and GrpE heat shock proteins jointly stimulate ATPase activity of DnaK. *Proc. Natl. Acad. Sci. U.S.A.* 88, 2874–2878. doi: 10.1073/pnas.88.7.2874

Lilge, L., Reder, A., Tippmann, F., Morgenroth, F., Grohmann, J., Becher, D., et al. (2020). The involvement of the McsB arginine kinase in Clp-dependent degradation of the MgsR regulator in *Bacillus subtilis*. *Front. Microbiol.* 11:900. doi: 10.3389/fmicb.2020.00900

Linde, D., Volkmer-Engert, R., Schreiber, S., and Müller, J. P. (2003). Interaction of the *Bacillus subtilis* chaperone CsaA with the secretory protein YvaY. *FEMS Microbiol. Lett.* 226, 93–100. doi: 10.1016/S0378-1097(03)00578-0

Lindner, A. B., Madden, R., Demarez, A., Stewart, E. J., and Taddei, F. (2008). Asymmetric segregation of protein aggregates is associated with cellular aging and rejuvenation. *Proc. Natl. Acad. Sci. U.S.A.* 105, 3076–3081. doi: 10.1073/pnas.0708931105

Lytvynenko, I., Paternoga, H., Thrun, A., Balke, A., Muller, T. A., Chiang, C. H., et al. (2019). Alanine tails signal proteolysis in bacterial ribosome-associated quality control. *Cell* 178, 76–90.e22. doi: 10.1016/j.cell.2019.05.002

Martinez-Hackert, E., and Hendrickson, W. A. (2009). Promiscuous substrate recognition in folding and assembly activities of the trigger factor chaperone. *Cell* 138, 923–934. doi: 10.1016/j.cell.2009.07.044

Mayhew, M., da Silva, A. C., Martin, J., Erdjument-Bromage, H., Tempst, P., and Hartl, F. U. (1996). Protein folding in the central cavity of the GroEL-GroES chaperonin complex. *Nature* 379, 420–426. doi: 10.1038/379420a0

Meile, J. C., Wu, L. J., Ehrlich, S. D., Errington, J., and Noirot, P. (2006). Systematic localisation of proteins fused to the green fluorescent protein in *Bacillus subtilis*: identification of new proteins at the DNA replication factory. *Proteomics* 6, 2135–2146. doi: 10.1002/pmic.200500512

Miethke, M., Hecker, M., and Gerth, U. (2006). Involvement of *Bacillus subtilis* ClpE in CtsR degradation and protein quality control. *J. Bacteriol.* 188, 4610–4619. doi: 10.1128/JB.00287-06

Mogk, A., Homuth, G., Scholz, C., Kim, L., Schmid, F. X., and Schumann, W. (1997). The GroE chaperonin machine is a major modulator of the CIRCE heat shock regulon of *Bacillus subtilis*. *EMBO J.* 16, 4579–4590. doi: 10.1093/emboj/16.15.4579

Mogk, A., Huber, D., and Bukau, B. (2011). Integrating protein homeostasis strategies in prokaryotes. *Cold Spring Harb. Perspect. Biol.* 3:a004366. doi: 10.1101/cshperspect.a004366

Mogk, A., Tomoyasu, T., Goloubinoff, P., Rudiger, S., Roder, D., Langen, H., et al. (1999). Identification of thermolabile *Escherichia coli* proteins: prevention and reversion of aggregation by DnaK and ClpB. *EMBO J.* 18, 6934–6949. doi: 10.1093/emboj/18.24.6934

Moliere, N., and Turgay, K. (2009). Chaperone-protease systems in regulation and protein quality control in *Bacillus subtilis*. *Res. Microbiol.* 160, 637–644. doi: 10.1016/j.resmic.2009.08.020

Moore, S. D., and Sauer, R. T. (2007). The tmRNA system for translational surveillance and ribosome rescue. *Annu. Rev. Biochem.* 76, 101–124. doi: 10.1146/annurev.biochem.75.103004.142733

Mortier, J., Tadesse, W., Govers, S. K., and Aertsen, A. (2019). Stress-induced protein aggregates shape population heterogeneity in bacteria. *Curr. Genet.* 65, 865–869. doi: 10.1007/s00294-019-00947-1

Mostertz, J., Scharf, C., Hecker, M., and Homuth, G. (2004). Transcriptome and proteome analysis of *Bacillus subtilis* gene expression in response to superoxide and peroxide stress. *Microbiology* 150, 497–512. doi: 10.1099/mic.0.26665-0

Mukherjee, S., Bree, A. C., Liu, J., Patrick, J. E., Chien, P., and Kearns, D. B. (2015). Adaptor-mediated Lon proteolysis restricts *Bacillus subtilis* hyperflagellation. *Proc. Natl. Acad. Sci. U.S.A.* 112, 250–255. doi: 10.1073/pnas.1417419112

Müller, J. P., Ozegowski, J., Vettermann, S., Swaving, J., Van Wely, K. H., and Driessen, A. J. (2000). Interaction of *Bacillus subtilis* CsaA with SecA and precursor proteins. *Biochem. J.* 348, 367–373. doi: 10.1042/bj3480367

Nakano, M. M., Lin, A., Zuber, C. S., Newberry, K. J., Brennan, R. G., and Zuber, P. (2010). Promoter recognition by a complex of Spx and the C-terminal domain of the RNA polymerase alpha subunit. *PLoS One* 5:e8664. doi: 10.1371/journal.pone.0008664

Nakano, S., Kuster-Schock, E., Grossman, A. D., and Zuber, P. (2003). Spx-dependent global transcriptional control is induced by thiol-specific oxidative stress in *Bacillus subtilis*. *Proc. Natl. Acad. Sci. U.S.A.* 100, 13603–13608. doi: 10.1073/pnas.2235180100

Nannapaneni, P., Hertwig, F., Depke, M., Hecker, M., Mader, U., Volker, U., et al. (2012). Defining the structure of the general stress regulon of *Bacillus subtilis* using targeted microarray analysis and random forest classification. *Microbiology* 158, 696–707. doi: 10.1099/mic.0.055434-0

Neeli-Venkata, R., Martikainen, A., Gupta, A., Goncalves, N., Fonseca, J., and Ribeiro, A. S. (2016). Robustness of the process of nucleoid exclusion of protein aggregates in *Escherichia coli*. *J. Bacteriol.* 198, 898–906. doi: 10.1128/JB.00848-15

Newberry, K. J., Nakano, S., Zuber, P., and Brennan, R. G. (2005). Crystal structure of the *Bacillus subtilis* anti-alpha, global transcriptional regulator. Spx, in complex with the alpha C-terminal domain of RNA polymerase. *Proc. Natl. Acad. Sci. U.S.A.* 102, 15839–15844. doi: 10.1073/pnas.0506592102

Oh, E., Becker, A. H., Sandikci, A., Huber, D., Chaba, R., Gloge, F., et al. (2011). Selective ribosome profiling reveals the cotranslational chaperone action of trigger factor *in vivo*. *Cell* 147, 1295–1308. doi: 10.1016/j.cell.2011.10.044

Olivares, A. O., Baker, T. A., and Sauer, R. T. (2016). Mechanistic insights into bacterial AAA+ proteases and protein-remodelling machines. *Nat. Rev. Microbiol.* 14, 33–44. doi: 10.1038/nrmicro.2015.4

Oliveira, S. M., Neeli-Venkata, R., Goncalves, N. S., Santinha, J. A., Martins, L., Tran, H., et al. (2016). Increased cytoplasm viscosity hampers aggregate polar segregation in *Escherichia coli*. *Mol. Microbiol.* 99, 686–699. doi: 10.1111/mmi.13257

Paek, K. H., and Walker, G. C. (1987). *Escherichia coli dnaK* null mutants are inviable at high temperature. *J. Bacteriol.* 169, 283–290. doi: 10.1128/jb.169.1.283-290.1987

Perales-Calvo, J., Giganti, D., Stirnemann, G., and Garcia-Manyes, S. (2018). The force-dependent mechanism of DnaK-mediated mechanical folding. *Sci. Adv.* 4:eaaq0243. doi: 10.1126/sciadv.aaq0243

Petersohn, A., Brigulla, M., Haas, S., Hoheisel, J. D., Volker, U., and Hecker, M. (2001). Global analysis of the general stress response of *Bacillus subtilis*. *J. Bacteriol.* 183, 5617–5631. doi: 10.1128/JB.183.19.5617-5631.2001

Powers, E. T., and Balch, W. E. (2013). Diversity in the origins of proteostasis networks-a driver for protein function in evolution. *Nat. Rev. Mol. Cell Biol.* 14, 237–248. doi: 10.1038/nrm3542

Powers, E. T., Morimoto, R. I., Dillin, A., Kelly, J. W., and Balch, W. E. (2009). Biological and chemical approaches to diseases of proteostasis deficiency. *Annu. Rev. Biochem.* 78, 959–991. doi: 10.1146/annurev.biochem.052308.114844

Price, C. W., Fawcett, P., Ceremonie, H., Su, N., Murphy, C. K., and Youngman, P. (2001). Genome-wide analysis of the general stress response in *Bacillus subtilis*. *Mol. Microbiol.* 41, 757–774. doi: 10.1046/j.1365-2958.2001.02534.x

Pu, Y., Li, Y., Jin, X., Tian, T., Ma, Q., Zhao, Z., et al. (2019). ATP-dependent dynamic protein aggregation regulates bacterial dormancy depth critical for antibiotic tolerance. *Mol. Cell* 73, 143–156. doi: 10.1016/j.molcel.2018.10.022

Rebeaud, M. E., Mallik, S., Goloubinoff, P., and Tawfik, D. S. (2021). On the evolution of chaperones and cochaperones and the expansion of proteomes across the Tree of Life. *Proc. Natl. Acad. Sci. U.S.A.* 118:e2020885118. doi: 10.1073/pnas.2020885118

Reder, A., Hoper, D., Weinberg, C., Gerth, U., Fraunholz, M., and Hecker, M. (2008). The Spx paralogue MgsR (YqgZ) controls a subregulon within the general stress response of *Bacillus subtilis*. *Mol. Microbiol.* 69, 1104–1120. doi: 10.1111/j.1365-2958.2008.06332.x

Reyes, D. Y., and Yoshikawa, H. (2002). DnaK chaperone machine and trigger factor are only partially required for normal growth of *Bacillus subtilis*. *Biosci. Biotechnol. Biochem.* 66, 1583–1586. doi: 10.1271/bbb.66.1583

Reyes, D. Y., and Zuber, P. (2008). Activation of transcription initiation by Spx: formation of transcription complex and identification of a Cis-acting element required for transcriptional activation. *Mol. Microbiol.* 69, 765–779. doi: 10.1111/j.1365-2958.2008.06330.x

Richter, K., Haslbeck, M., and Buchner, J. (2010). The heat shock response: life on the verge of death. *Mol. Cell* 40, 253–266. doi: 10.1016/j.molcel.2010.10.006

Rochat, T., Nicolas, P., Delumeau, O., Rabatinova, A., Korelusova, J., Leduc, A., et al. (2012). Genome-wide identification of genes directly regulated by the

pleiotropic transcription factor Spx in *Bacillus subtilis*. *Nucleic Acids Res.* 40, 9571–9583. doi: 10.1093/nar/gks755

Roghanian, M., Van Nerom, K., Takada, H., Caballero-Montes, J., Tamman, H., Kudrin, P., et al. (2021). (p)ppGpp controls stringent factors by exploiting antagonistic allosteric coupling between catalytic domains. *Mol. Cell* 81, 3310–3322.e6. doi: 10.1016/j.molcel.2021.07.026

Rojas-Tapias, D. F., and Helmann, J. D. (2018). Induction of the Spx regulon by cell wall stress reveals novel regulatory mechanisms in *Bacillus subtilis*. *Mol. Microbiol.* 107, 659–674. doi: 10.1111/mmi.13906

Rojas-Tapias, D. F., and Helmann, J. D. (2019b). Roles and regulation of Spx family transcription factors in *Bacillus subtilis* and related species. *Adv. Microb. Physiol.* 75, 279–323. doi: 10.1016/bs.ampbs.2019.05.003

Rojas-Tapias, D. F., and Helmann, J. D. (2019a). Identification of novel Spx regulatory pathways in *Bacillus subtilis* uncovers a close relationship between the CtsR and Spx regulons. *J. Bacteriol.* 201, e00151–19. doi: 10.1128/JB.00151-19

Rokney, A., Shagan, M., Kessel, M., Smith, Y., Rosenshine, I., and Oppenheim, A. B. (2009). *E. coli* transports aggregated proteins to the poles by a specific and energy-dependent process. *J. Mol. Biol.* 392, 589–601. doi: 10.1016/j.jmb.2009.07.009

Ross, C. A., and Poirier, M. A. (2004). Protein aggregation and neurodegenerative disease. *Nat. Med.* 10(Suppl.), S10–S17. doi: 10.1038/nm1066

Rudiger, S., Buchberger, A., and Bukau, B. (1997). Interaction of Hsp70 chaperones with substrates. *Nat. Struct. Biol.* 4, 342–349. doi: 10.1038/nsb0597-342

Runde, S., Moliere, N., Heinz, A., Maisonneuve, E., Janczikowski, A., Elsholz, A. K., et al. (2014). The role of thiol oxidative stress response in heat-induced protein aggregate formation during thermotolerance in *Bacillus subtilis*. *Mol. Microbiol.* 91, 1036–1052. doi: 10.1111/mmi.12521

Saio, T., Guan, X., Rossi, P., Economou, A., and Kalodimos, C. G. (2014). Structural basis for protein antiaggregation activity of the trigger factor chaperone. *Science* 344:1250494. doi: 10.1126/science.1250494

Saio, T., Kawagoe, S., Ishimori, K., and Kalodimos, C. G. (2018). Oligomerization of a molecular chaperone modulates its activity. *eLife* 7:e35731. doi: 10.7554/eLife.35731

Sauer, R. T., and Baker, T. A. (2011). AAA+ proteases: ATP-fueled machines of protein destruction. *Annu. Rev. Biochem.* 80, 587–612. doi: 10.1146/annurev-biochem-060408-172623

Sauer, R. T., Bolon, D. N., Burton, B. M., Burton, R. E., Flynn, J. M., Grant, R. A., et al. (2004). Sculpting the proteome with AAA(+) proteases and disassembly machines. *Cell* 119, 9–18. doi: 10.1016/j.cell.2004.09.020

Schafer, H., Beckert, B., Frese, C. K., Steinchen, W., Nuss, A. M., Beckstette, M., et al. (2020). The alarmones (p)ppGpp are part of the heat shock response of *Bacillus subtilis*. *PLoS Genet.* 16:e1008275. doi: 10.1371/journal.pgen.1008275

Schafer, H., Heinz, A., Sudzinova, P., Voss, M., Hantke, I., Krasny, L., et al. (2019). Spx, the central regulator of the heat and oxidative stress response in *B. subtilis*, can repress transcription of translation-related genes. *Mol. Microbiol.* 111, 514–533. doi: 10.1111/mmi.14171

Schafer, H., and Turgay, K. (2019). Spx, a versatile regulator of the *Bacillus subtilis* stress response. *Curr. Genet.* 65, 871–876. doi: 10.1007/s00294-019-00950-6

Schmidt, A., Trentini, D. B., Spiess, S., Fuhrmann, J., Ammerer, G., Mechtler, K., et al. (2014). Quantitative phosphoproteomics reveals the role of protein arginine phosphorylation in the bacterial stress response. *Mol. Cell Proteomics* 13, 537–550. doi: 10.1074/mcp.M113.032292

Schopf, F. H., Biebl, M. M., and Buchner, J. (2017). The HSP90 chaperone machinery. *Nat. Rev. Mol. Cell Biol.* 18, 345–360. doi: 10.1038/nrm.2017.20

Schramm, F. D., Schroeder, K., and Jonas, K. (2020). Protein aggregation in bacteria. *FEMS Microbiol. Rev.* 44, 54–72. doi: 10.1093/femsre/fuz026

Schulz, A., Tzschaschel, B., and Schumann, W. (1995). Isolation and analysis of mutants of the *dnaK* operon of *Bacillus subtilis*. *Mol. Microbiol.* 15, 421–429. doi: 10.1111/j.1365-2958.1995.tb02256.x

Schumann, W. (2003). The *Bacillus subtilis* heat shock stimulon. *Cell Stress Chaperones* 8, 207–217. doi: 10.1379/1466-12682003008<0207:tbshss<2.0.co;2

Schumann, W. (2016). Regulation of bacterial heat shock stimulons. *Cell Stress Chaperones* 21, 959–968. doi: 10.1007/s12192-016-0727-z

Seaton, B. L., and Vickery, L. E. (1994). A gene encoding a DnaK/hsp70 homolog in *Escherichia coli*. *Proc. Natl. Acad. Sci. U.S.A.* 91, 2066–2070. doi: 10.1073/pnas.91.6.2066

Seydlova, G., Halada, P., Fiser, R., Toman, O., Ulrych, A., and Svobodova, J. (2012). DnaK and GroEL chaperones are recruited to the *Bacillus subtilis* membrane after short-term ethanol stress. *J. Appl. Microbiol.* 112, 765–774. doi: 10.1111/j.1365-2672.2012.05238.x

Shi, J., Li, F., Wen, A., Yu, L., Wang, L., Wang, F., et al. (2021). Structural basis of transcription activation by the global regulator Spx. *Nucleic Acids Res.* 49, 10756–10769. doi: 10.1093/nar/gkab790

Shi, L., Ravikumar, V., Derouiche, A., Macek, B., and Mijakovic, I. (2016). Tyrosine 601 of *Bacillus subtilis* DnaK undergoes phosphorylation and is crucial for chaperone activity and heat shock survival. *Front. Microbiol.* 7:533. doi: 10.3389/fmicb.2016.00533

Siller, E., DeZwaan, D. C., Anderson, J. F., Freeman, B. C., and Barral, J. M. (2010). Slowing bacterial translation speed enhances eukaryotic protein folding efficiency. *J. Mol. Biol.* 396, 1310–1318. doi: 10.1016/j.jmb.2009.12.042

Simmons, L. A., Grossman, A. D., and Walker, G. C. (2008). Clp and Lon proteases occupy distinct subcellular positions in *Bacillus subtilis*. *J. Bacteriol.* 190, 6758–6768. doi: 10.1128/JB.00590-08

Singhal, K., Vreede, J., Mashaghi, A., Tans, S. J., and Bolhuis, P. G. (2015). The trigger factor chaperone encapsulates and stabilizes partial folds of substrate proteins. *PLoS Comput. Biol.* 11:e1004444. doi: 10.1371/journal.pcbi.1004444

Stannek, L., Gunka, K., Care, R. A., Gerth, U., and Commichau, F. M. (2014). Factors that mediate and prevent degradation of the inactive and unstable GudB protein in *Bacillus subtilis*. *Front. Microbiol.* 5:758. doi: 10.3389/fmicb.2014.00758

Stoller, G., Rucknagel, K. P., Nierhaus, K. H., Schmid, F. X., Fischer, G., and Rahfeld, J. U. (1995). A ribosome-associated peptidyl-prolyl cis/trans isomerase identified as the trigger factor. *EMBO J.* 14, 4939–4948.

Szabo, A., Langer, T., Schroder, H., Flanagan, J., Bukau, B., and Hartl, F. U. (1994). The ATP hydrolysis-dependent reaction cycle of the *Escherichia coli* Hsp70 system DnaK. DnaJ, and GrpE. *Proc. Natl. Acad. Sci. U.S.A.* 91, 10345–10349. doi: 10.1073/pnas.91.22.10345

Teter, S. A., Houry, W. A., Ang, D., Tradler, T., Rockabrand, D., Fischer, G., et al. (1999). Polypeptide flux through bacterial Hsp70: DnaK cooperates with trigger factor in chaperoning nascent chains. *Cell* 97, 755–765. doi: 10.1016/s0092-8674(00)80787-4

Thi Nguyen, H. B., and Schumann, W. (2012). The sporulation control gene spo0M of *Bacillus subtilis* is a target of the FtsH metalloprotease. *Res. Microbiol.* 163, 114–118. doi: 10.1016/j.resmic.2011.10.011

Thomas, J. G., and Baneyx, F. (1998). Roles of the *Escherichia coli* small heat shock proteins IbpA and IbpB in thermal stress management: comparison with ClpA. ClpB, and HtpG *In vivo*. *J. Bacteriol.* 180, 5165–5172. doi: 10.1128/JB.180.19.5165-5172.1998

Trentini, D. B., Suskiewicz, M. J., Heuck, A., Kurzbauer, R., Deszcz, L., Mechtler, K., et al. (2016). Arginine phosphorylation marks proteins for degradation by a Clp protease. *Nature* 539, 48–53. doi: 10.1038/nature20122

Ueguchi, C., Kakeda, M., Yamada, H., and Mizuno, T. (1994). An analogue of the DnaJ molecular chaperone in *Escherichia coli*. *Proc. Natl. Acad. Sci. U.S.A.* 91, 1054–1058. doi: 10.1073/pnas.91.3.1054

Ullers, R. S., Ang, D., Schwager, F., Georgopoulos, C., and Genevaux, P. (2007). Trigger Factor can antagonize both SecB and DnaK/DnaJ chaperone functions in *Escherichia coli*. *Proc. Natl. Acad. Sci. U.S.A.* 104, 3101–3106. doi: 10.1073/pnas.0608232104

Versteeg, S., Mogk, A., and Schumann, W. (1999). The *Bacillus subtilis htpG* gene is not involved in thermal stress management. *Mol. Gen. Genet.* 261, 582–588. doi: 10.1007/s004380051004

Wickner, S., Nguyen, T. L., and Genest, O. (2021). The bacterial Hsp90 chaperone: cellular functions and mechanism of action. *Annu. Rev. Microbiol.* 75, 719–739. doi: 10.1146/annurev-micro-032421-035644

Wild, J., Altman, E., Yura, T., and Gross, C. A. (1992). DnaK and DnaJ heat shock proteins participate in protein export in *Escherichia coli*. *Genes Dev.* 6, 1165–1172. doi: 10.1101/gad.6.7.1165

Wild, J., Rossmeissl, P., Walter, W. A., and Gross, C. A. (1996). Involvement of the DnaK-DnaJ-GrpE chaperone team in protein secretion in *Escherichia coli*. *J. Bacteriol.* 178, 3608–3613. doi: 10.1128/jb.178.12.3608-3613.1996

Wild, J., Walter, W. A., Gross, C. A., and Altman, E. (1993). Accumulation of secretory protein precursors in *Escherichia coli* induces the heat shock response. *J. Bacteriol.* 175, 3992–3997. doi: 10.1128/jb.175.13.3992-3997.1993

Winkler, J., Seybert, A., Konig, L., Pruggnaller, S., Haselmann, U., Sourjik, V., et al. (2010). Quantitative and spatio-temporal features of protein aggregation in *Escherichia coli* and consequences on protein quality control and cellular ageing. *EMBO J.* 29, 910–923. doi: 10.1038/emboj.2009.412

Yoshimune, K., Yoshimura, T., and Esaki, N. (1998). Hsc62, a new DnaK homologue of *Escherichia coli*. *Biochem. Biophys. Res. Commun.* 250, 115–118. doi: 10.1006/bbrc.1998.9255

Yoshimune, K., Yoshimura, T., Nakayama, T., Nishino, T., and Esaki, N. (2002). Hsc62, Hsc56, and GrpE, the third Hsp70 chaperone system of *Escherichia coli*. *Biochem. Biophys. Res. Commun.* 293, 1389–1395. doi: 10.1016/S0006-291X(02)00403-5

Yu, J., Liu, Y., Yin, H., and Chang, Z. (2019). Regrowth-delay body as a bacterial subcellular structure marking multidrug-tolerant persisters. *Cell Discov.* 5:8. doi: 10.1038/s41421-019-0080-3

Zarnt, T., Tradler, T., Stoller, G., Scholz, C., Schmid, F. X., and Fischer, G. (1997). Modular structure of the trigger factor required for high activity in protein folding. *J. Mol. Biol.* 271, 827–837. doi: 10.1006/jmbi.1997.1206

Zhou, B., Semanjski, M., Orlovetskie, N., Bhattacharya, S., A.lon, S., Argaman, L., et al. (2019). Arginine dephosphorylation propels spore germination in bacteria. *Proc. Natl. Acad. Sci. U.S.A.* 116, 14228–14237. doi: 10.1073/pnas.1817742116

Zhu, X., Zhao, X., Burkholder, W. F., Gragerov, A., Ogata, C. M., Gottesman, M. E., et al. (1996). Structural analysis of substrate binding by the molecular chaperone DnaK. *Science* 272, 1606–1614. doi: 10.1126/science.272.5268.1606

Zuber, P. (2004). Spx-RNA polymerase interaction and global transcriptional control during oxidative stress. *J. Bacteriol.* 186, 1911–1918. doi: 10.1128/JB.186.7.1911-1918.2004

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

frontiers in Microbiology

PERSPECTIVE
published: 17 March 2022
doi: 10.3389/fmicb.2022.871699

A Shift in Perspective: A Role for the Type I Toxin TisB as Persistence-Stabilizing Factor

Daniel Edelmann and Bork A. Berghoff**

Institute for Microbiology and Molecular Biology, Justus Liebig University Giessen, Giessen, Germany

Bacterial persistence is a phenomenon that is founded by the existence of a subpopulation of multidrug-tolerant cells. These so-called persister cells endure otherwise lethal stress situations and enable restoration of bacterial populations upon return to favorable conditions. Persisters are especially notorious for their ability to survive antibiotic treatments without conventional resistance genes and to cause infection relapse. The persister state is typically correlated with reduction or inhibition of cellular activity. Early on, chromosomal toxin-antitoxin (TA) systems were suspected to induce the persister state in response to environmental stress. However, this idea has been challenged during the last years. Especially the involvement of toxins from type II TA systems in persister formation is put into question. For toxins from type I TA systems the debate has just started. Here, we would like to summarize recent knowledge gained for the type I TA system *tisB/istR-1* from *Escherichia coli*. TisB is a small, membrane-targeting toxin, which disrupts the proton motive force (PMF), leading to membrane depolarization. Based on experimental data, we hypothesize that TisB primarily stabilizes the persister state through depolarization and further, secondary effects. We will present a simple model that will provide a framework for future directions.

Keywords: toxin-antitoxin systems, *tisB/istR-1*, type I toxins, persistence, awakening, fluoroquinolones, SOS response

OPEN ACCESS

Edited by:
Ulrike Kappler,
The University of Queensland, Australia

Reviewed by:
Ralph Bertram,
Paracelsus Medical Private University, Nuremberg, Germany

****Correspondence:***
Daniel Edelmann
daniel.b.edelmann@bio.uni-giessen.de
Bork A. Berghoff
bork.a.berghoff@mikro.bio.uni-giessen.de

Specialty section:
This article was submitted to Microbial Physiology and Metabolism, a section of the journal Frontiers in Microbiology

Received: *08 February 2022*
Accepted: *28 February 2022*
Published: *17 March 2022*

Citation:
Edelmann D and Berghoff BA (2022) A Shift in Perspective: A Role for the Type I Toxin TisB as Persistence-Stabilizing Factor. Front. Microbiol. 13:871699. doi: 10.3389/fmicb.2022.871699

INTRODUCTION

Toxin-antitoxin (TA) systems are genetic modules that are frequently found on bacterial chromosomes and plasmids (Pandey and Gerdes, 2005; Leplae et al., 2011; Harms et al., 2018; Srivastava et al., 2021; Jurėnas et al., 2022). TA systems typically consist of two factors: a toxin protein that has an inhibitory or poisonous effect on the host organism and an antitoxin that either counteracts expression of the toxin or antagonizes toxin activity (Page and Peti, 2016; Harms et al., 2018). However, under particular stress conditions, the toxin may overcome the neutralizing effect of its cognate antitoxin, which eventually leads to a targeted inhibition or poisoning of an essential biosynthetic process. Various aspects of the bacterial lifestyle are associated with expression of toxins from TA systems including (i) genome stability, (ii) phage defense, (iii) stress adaptation, and (iv) persister formation (Wang and Wood, 2011; Page and Peti, 2016; Harms et al., 2018; Fraikin et al., 2020; Song and Wood, 2020; Jurėnas et al., 2022).

Different TA system types have been described over the years, and we would like to refer the reader to recent reviews on TA system biology (Srivastava et al., 2021; Jurėnas et al., 2022). For the scope of this *Perspective*, we would like to focus on type I TA systems. A hallmark of type I TA systems is the existence of an RNA antitoxin that confers translational inhibition on the toxin mRNA (Gerdes and Wagner, 2007; Brantl and Jahn, 2015; Berghoff and Wagner, 2017). Antitoxin and toxin transcripts are often *cis*-encoded by overlapping genes, while transcription depends on the activity of individual promoters (Fozo et al., 2008; Brantl and Jahn, 2015). Alternatively, some type I TA systems are arranged with non-overlapping genes, but the transcripts show sufficient complementarity for duplex formation when encoded in *trans*. *Via* duplex formation, antitoxin RNAs render the ribosome binding site (canonical or standby) of toxin mRNAs inaccessible and suppress translational (e.g., Thisted and Gerdes, 1992; Darfeuille et al., 2007; Jahn and Brantl, 2013; Wen et al., 2017). The duplex is typically cleaved and/or degraded by cellular RNases. In addition to control by RNA antitoxins, secondary structures in the 5′ untranslated region (UTR) render the toxin primary transcripts translationally inactive, resulting in transcription-translation uncoupling (Berghoff and Wagner, 2017; Masachis and Darfeuille, 2018). This sophisticated post-transcriptional regulation by both mRNA secondary structures and RNA antitoxins likely avoids toxin production during growth-promoting, non-stress conditions. However, elevated transcription of the toxin gene (e.g., upon stress), and subsequent processing of the primary transcript into a translationally active mRNA, might eventually shift the ratio between toxin mRNA and RNA antitoxin in favor of the toxin mRNA. Hypothetically, excess toxin mRNA only occurs in a fraction of cells, and only these cells will produce sufficient toxin amounts to be physiologically affected (Berghoff and Wagner, 2019). Therefore, the regulatory features of type I TA systems favor phenotypic heterogeneity of clonal populations (Berghoff et al., 2017a). Type I toxins are typically small (<50 amino acids), membrane-targeting proteins that are frequently associated with disruption of the proton motive force (PMF) and/or the cytoplasmic membrane (Brielle et al., 2016). The concomitant growth inhibition in a fraction of cells is clearly consistent with the phenomenon of bacterial persistence. Persistence is caused by a subpopulation of so-called persister cells that have reduced their cellular activity and, hence, entered a state of multidrug tolerance (Lewis, 2010; Brauner et al., 2016). It is commonly assumed that a persister cell has undergone a phenotypic switch due to altered gene expression. However, whether a transient genetic change, as sometimes observed in the case of genetic heteroresistance (Andersson et al., 2019; Dewachter et al., 2019), could account for the persister phenomenon is currently unclear. At least two type I toxins, TisB and HokB, were suggested to affect the formation of persisters, either through PMF disruption, ATP depletion, or both (Dörr et al., 2010; Verstraeten et al., 2015; Berghoff et al., 2017a; Wilmaerts et al., 2018; Edelmann et al., 2021).

A recent debate in the field of persister research is whether and, if so, how TA systems affect the physiology of these special cells? This debate is particularly focused on type II TA systems and their role in persister physiology is highly questioned (Harms et al., 2017; Goormaghtigh et al., 2018), as also substantiated by retraction of some key publications (Maisonneuve et al., 2011, 2013). The debate is further fueled by the fact that genetic deletions of single TA systems do not generally result in persister-associated phenotypes, i.e., reduced survival rates upon antibiotic treatments when compared to wild-type strains. However, this does not necessarily reject TA systems as important factors for persistence. On the one hand, redundancy among TA systems might obscure phenotypes of single-deletion strains. On the other hand, phenotypes might only be detected under more complex experimental conditions, e.g., in infection models (Helaine et al., 2014; Ronneau and Helaine, 2019).

In the case of type I TA systems, the *tisB/istR-1* locus of *Escherichia coli* has been studied extensively with regard to persistence. Most type I TA systems are narrowly distributed among bacteria, and TisB homologs were only found in the family of *Enterobacteriaceae* (Coray et al., 2017). TisB was linked to persister formation more than 10 years ago (Dörr et al., 2010). Since transcription of toxin gene *tisB* is strongly activated upon DNA damage as part of the SOS response (Courcelle et al., 2001; Vogel et al., 2004; Berghoff et al., 2017b), the influence of *tisB* on persister formation was mainly elucidated using DNA-damaging antibiotics, such as the fluoroquinolone ciprofloxacin. Deletion of *tisB* significantly reduced persister levels upon treatment with ciprofloxacin, a finding that has since been interpreted as drug-induced persister formation caused by *tisB* expression (Dörr et al., 2010; Wagner and Unoson, 2012; Balaban et al., 2019). Even though a *tisB* related persister phenotype was validated independently using ciprofloxacin (Berghoff et al., 2017a), another study failed to observe a comparable phenotype when using the related antibiotic ofloxacin (Goormaghtigh and Van Melderen, 2019). Here, we would like to push forward the idea that toxin TisB, and probably other type I toxins, do not necessarily initiate the persister formation process, but mainly stabilize the persister state and affect the duration of cellular inactivity, both by primary and secondary effects. Our *Perspective* aims to stimulate a different view on the physiological role of type I toxins, and might help to solve some of the controversy concerning the link between TA systems and persistence in bacteria.

DISRUPTION OF THE PMF IS THE MAJOR PRIMARY EFFECT OF TisB IN THE WILD-TYPE BACKGROUND

TisB has a length of 29 amino acids and integrates into the cytoplasmic membrane (Unoson and Wagner, 2008; Gurnev et al., 2012). Upon integration, TisB is believed to cause a breakdown of PMF, which is expected to deprive the cell of its foremost means to generate ATP (Unoson and Wagner, 2008; Gurnev et al., 2012; Edelmann et al., 2021). ATP depletion might ultimately link the action of TisB to growth inhibition and persister formation. While this hypothesis is apparently intuitive and can be corroborated for instance by chemical treatments with arsenate (Conlon et al., 2016; Shan et al., 2017), the evidence for TisB is mainly based on overexpression experiments (Unoson and Wagner, 2008; Edelmann et al., 2021), which might be a poor representative for *tisB* expression levels

from its native locus. Furthermore, overexpression experiments often tend to result in drastic phenotypes that are, however, rarely further substantiated with experiments in the wild-type background. Therefore, we seek to review the material published on toxin TisB and present our view of what are likely authentic effects of toxin activity, and what effects are probably attributable to non-physiological expression levels in the experiment.

In point of fact, breakdown of the PMF (i.e., depolarization of the cytoplasmic membrane) by TisB is the effect best documented upon TisB expression from its native locus. In *E. coli* MG1655 wild-type cultures, TisB-dependent depolarization can be observed in a fraction of cells upon prolonged ciprofloxacin treatments (~20% depolarized cells after 4 h and ~50% after 6 h), and deletion of *tisB* almost completely prevents depolarization under these conditions (Berghoff et al., 2017a). Importantly, phenotypes of a *tisB* deletion, including lack of depolarization and a reduced persister level, can be observed only after several hours of antibiotic treatment in comparison to the wild type (Dörr et al., 2010; Berghoff et al., 2017a). These observations suggest that, in wild-type cultures, physiologically relevant amounts of TisB protein accumulate only after long periods of SOS induction, which is congruent with strong post-transcriptional repression of *tisB* mRNA by its 5′ UTR structure and antitoxin IstR-1 (Vogel et al., 2004; Darfeuille et al., 2007; Berghoff et al., 2017a).

TisB-dependent membrane depolarization is expected to ultimately affect ATP synthesis. Indeed, our unpublished results show that intracellular ATP levels are reduced in *E. coli* wild-type cultures after prolonged ciprofloxacin treatments, while ATP levels do not decrease in a *tisB* deletion strain. Interestingly, the decline of ATP only occurs at later time points and is rather moderate (~1.5-fold reduction after 4–6 h of ciprofloxacin). This suggests that drastic ATP depletion, as detected in response to strong TisB expression (Unoson and Wagner, 2008; Edelmann et al., 2021), does probably not resemble the wild-type situation. Alternatively, severe ATP depletion may only occur in a fraction of cells, which is not resolved in bulk measurements and demands single-cell reporters for ATP (Manuse et al., 2021). Either way, whether TisB-dependent depolarization and subsequent ATP depletion are the direct cause for drug-induced persister formation is yet to be demonstrated. Nevertheless, lowering the intracellular ATP levels can be an effective means to reach or modulate a persistent state (Conlon et al., 2016; Shan et al., 2017), and might be the mode of action for other type I toxins, such as HokB, which confers persistence by direct leakage of ATP (Wilmaerts et al., 2018).

TisB EXPRESSION CAUSES DIVERSE SECONDARY EFFECTS

Besides a drop in ATP levels as ultimate consequence of PMF disruption, further secondary effects can be assigned to TisB expression. For instance, TisB was shown to cause the formation of reactive oxygen species (ROS) upon high doses of ciprofloxacin and during long-term treatments (Edelmann and Berghoff, 2019). The mechanism that causes the formation of ROS is unknown. However, the ROS formed in response to TisB expression is mainly superoxide (Edelmann and Berghoff, 2019). Since superoxide is primarily formed as a byproduct of respiration (Imlay, 2013), it is tempting to speculate that the disruption at the cytoplasmic membrane caused by TisB generates an increased flux through the respiratory chain, which in turn leads to the formation of superoxide. Aerobic microorganisms are well adapted to deal with oxidative stress and have powerful enzymes to detoxify superoxide called superoxide dismutases (Imlay, 2013). Deletion of superoxide-scavenging enzymes revealed that they are essential for TisB-dependent persisters, while superoxide detoxification is of minor significance for persisters generated by other factors (Edelmann and Berghoff, 2019).

Early reports on the function of TisB indicated that overexpression of the toxin interferes with essential cellular processes such as replication, transcription, and translation (Unoson and Wagner, 2008). These findings were confirmed by more recent findings that link the expression of toxin TisB with the interference of essential cellular processes also in the wild-type background. In particular, protein biosynthesis was found to be impeded in a toxin-dependent manner in wild-type cells exposed to long treatments with ciprofloxacin (Edelmann et al., 2021). Indeed, comparison of wild-type cultures with a *tisB* deletion strain proves to be the most reliable when overexpression artifacts are to be avoided. Though phenotypes are typically rather subtle in toxin deletion mutants.

The link between TisB expression and protein biosynthesis is quite intriguing since a cessation of translation, either by toxin expression or bacteriostatic agents, might be a critical factor to confer persistence (Kwan et al., 2013; Cheverton et al., 2016; Rycroft et al., 2018). However, it is currently unknown whether TisB-dependent translation inhibition—and other secondary effects—influence the persister formation process.

TisB HAS THE POTENTIAL TO STABILIZE THE PERSISTER STATE

The *tisB* gene is one of the most highly induced genes of the SOS response, and transcripts can be detected within minutes after treatment initiation (Vogel et al., 2004; Berghoff et al., 2017b). As outlined above, post-transcriptional repression of *tisB* is expected to efficiently prevent immediate translation. Unfortunately, our attempts to directly detect TisB protein expressed from its native locus were not successful. We can, however, deduce some information from depolarization measurements. In wild-type cultures, depolarization is not observed after 2 h of ciprofloxacin treatment, whereas 20% of cells are depolarized after 4 h (Berghoff et al., 2017a), suggesting a strong delay in TisB translation and/or accumulation (**Figure 1**). Hence, only long-term treatments are relevant in order to observe TisB-dependent phenotypes. This certainly applies to depolarization as primary effect, but also to ATP depletion, ROS formation and translation inhibition as secondary effects (Edelmann and Berghoff, 2019; Edelmann et al., 2021).

The absence of measurable TisB-dependent effects at early time points does not necessarily challenge the assumption that the *tisB/istR-1* locus represents a regulatory module with a

potential to induce the persister state. However, rapid TisB synthesis and subsequent TisB-dependent persister formation might be a rare event. It is rather plausible that—in most cells—the persister state is induced by other means, and that the actual function of TisB is more likely to be found at later time points. Therefore, we suggest a reinterpretation of TisB as a potential persistence-stabilizing factor. In such a scenario, TisB accumulation would take place in both pre-existing persisters and persisters that were formed in response to the fluoroquinolone treatment (drug-induced persisters) by the action of TisB or probably other factors (**Figure 1**). Upon prolonged treatment, TisB-dependent depolarization and subsequent secondary effects prevent premature awakening, thereby reducing the probability of a too early and thus lethal resumption of cellular activity (**Figure 1**). Furthermore, we assume that TisB accumulation is heterogeneous and differentially modulates wake-up kinetics. Indeed, wild-type cultures contain a subpopulation of late awakening persisters, which is not observed in a *tisB* deletion strain, when cultures are treated with a high dose of ciprofloxacin for 6 h (our unpublished results).

Our model predicts that, at early time points during fluoroquinolone treatments, wild-type and *tisB* deletion strains have comparable persister levels, but that persister levels deviate at later time points. Indeed, Dörr et al. (2010) observed in their original publication that lower persister levels of the *tisB* deletion strain can be obtained only if the ciprofloxacin treatment lasts longer than 3 h. This time point is in perfect accordance with our model in **Figure 1**.

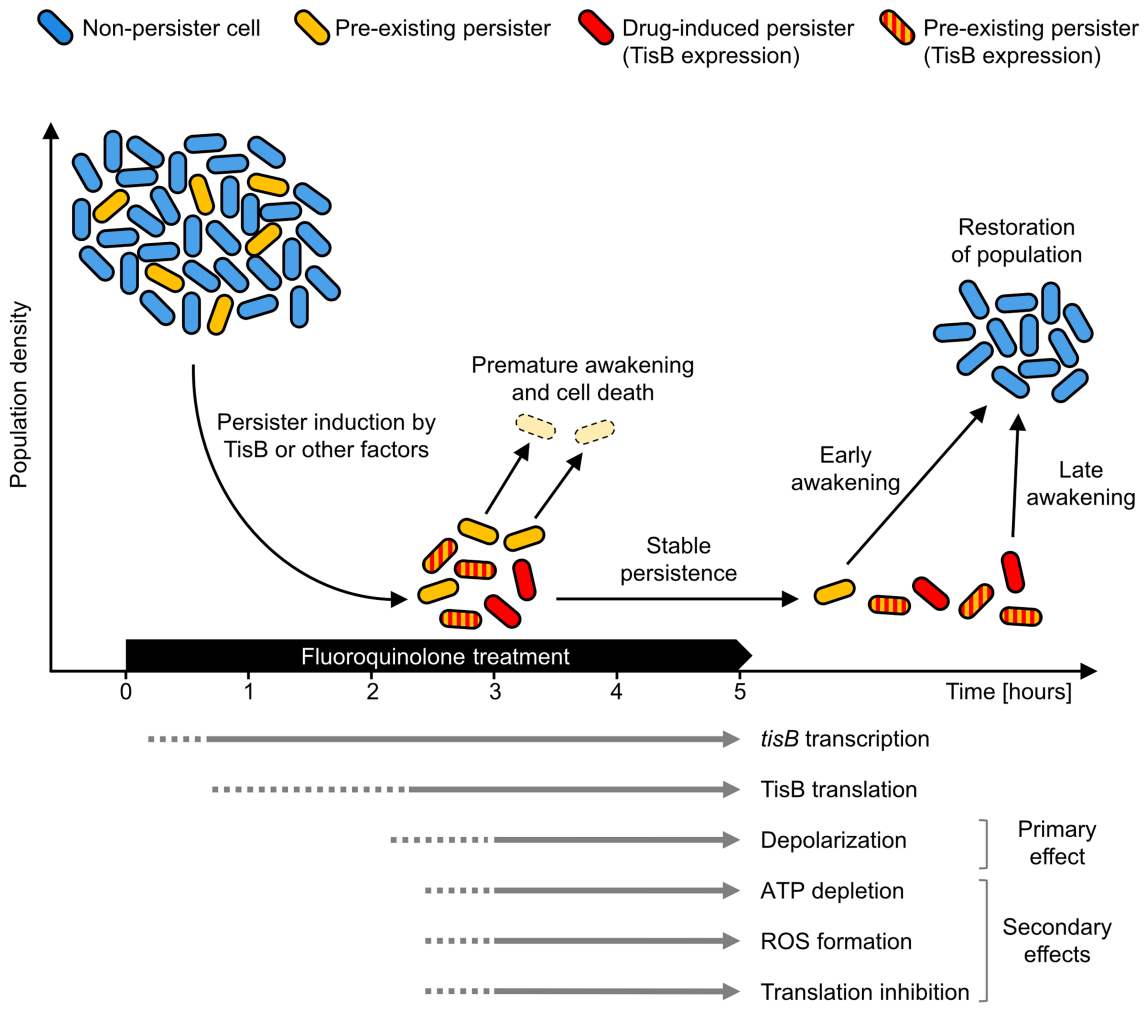


FIGURE 1 | TisB-dependent stabilization of the persister state. Persister cells arise from exposure to a fluoroquinolone antibiotic (drug-induced persisters, red) from non-persister cells (blue) by expression of TisB or by other factors. Persisters may also exist before treatment (yellow). During treatment, accumulation of the TisB toxin has a stabilizing effect on the persister state. This probably affects both drug-induced persisters (red) and pre-existing persisters formed by other factors (striped red/yellow). Survival probability increases as the persister state is prolonged beyond the duration of antibiotic treatment. Non-stabilized persisters (yellow) have a higher probability for premature awakening and will be eliminated (pale yellow). Only cells that remain sufficiently long in the persister state can successfully contribute to population recovery after treatment. The arrows at the bottom represent key events as well as primary and secondary effects of TisB expression in *Escherichia coli* wild-type cultures. A solid line indicates mostly confirmed time points for an event, while a dotted line indicates that an event is likely to have already occurred at that time point but no corresponding data are available.

CONCLUDING REMARKS

Based on our observations with the *tisB* deletion strain, we suggest that future studies should aim to decipher the precise role of type I toxins in persister survival. However, we assume that, due to the inherent heterogeneity of populations, it might be difficult to differentiate clearly between an initiating and stabilizing function of toxins by experimental means. Recent efforts to study persister awakening on the single-cell level might offer a valuable starting point (Goormaghtigh and Van Melderen, 2019; Svenningsen et al., 2019; Windels et al., 2019; Kaplan et al., 2021).

It came to our attention that a persistence-stabilizing function of type I toxins might be in support of the "dormancy continuum hypothesis" (Ayrapetyan et al., 2015, 2018). Similarly to what was suggested for protein aggregation (Pu et al., 2019; Dewachter et al., 2021), accumulation of TisB and its concomitant effects might not only stabilize the persister state and affect dormancy depth, but also enable transition from a persister to a viable but non-culturable (VBNC) state. Further experiments are clearly needed to elucidate the role that type I toxins play in bacterial dormancy in response to environmental stress. Another intriguing question is whether a persistence-stabilizing function also applies to type II toxins. Using alternative experimental setups and focusing on long-term treatments might answer these questions in the future.

DATA AVAILABILITY STATEMENT

The data analyzed in this study is subject to the following licenses/restrictions: The datasets are unpublished results. Requests to access these datasets should be directed to bork.a.berghoff@mikro.bio.uni-giessen.de.

AUTHOR CONTRIBUTIONS

DE and BB wrote the manuscript. All authors contributed to the article and approved the submitted version.

FUNDING

Work in the group of BB was supported by the German Research Council (DFG) in the framework of the SPP 2002 (BE 5210/3-1 and BE 5210/3-2) and by the University of Giessen (Research Grant).

ACKNOWLEDGMENTS

The authors are grateful to Florian Leinberger and other members of our group for experimental support and sharing unpublished data. We thank Gerhart Wagner (Uppsala University) for critical reading of the manuscript.

REFERENCES

Andersson, D. I., Nicoloff, H., and Hjort, K. (2019). Mechanisms and clinical relevance of bacterial heteroresistance. *Nat. Rev. Microbiol.* 17, 479–496. doi: 10.1038/s41579-019-0218-1

Ayrapetyan, M., Williams, T. C., and Oliver, J. D. (2015). Bridging the gap between viable but non-culturable and antibiotic persistent bacteria. *Trends Microbiol.* 23, 7–13. doi: 10.1016/j.tim.2014.09.004

Ayrapetyan, M., Williams, T., and Oliver, J. D. (2018). Relationship between the viable but nonculturable state and antibiotic persister cells. *J. Bacteriol.* 200, e00249–e00318. doi: 10.1128/JB.00249-18

Balaban, N. Q., Helaine, S., Lewis, K., Ackermann, M., Aldridge, B., Andersson, D. I., et al. (2019). Definitions and guidelines for research on antibiotic persistence. *Nat. Rev. Microbiol.* 17, 441–448. doi: 10.1038/s41579-019-0196-3

Berghoff, B. A., Hoekzema, M., Aulbach, L., and Wagner, E. G. H. (2017a). Two regulatory RNA elements affect TisB-dependent depolarization and persister formation. *Mol. Microbiol.* 103, 1020–1033. doi: 10.1111/mmi.13607

Berghoff, B. A., Karlsson, T., Källman, T., Wagner, E. G. H., and Grabherr, M. G. (2017b). RNA-sequence data normalization through in silico prediction of reference genes: the bacterial response to DNA damage as case study. *BioData Min.* 10:30. doi: 10.1186/s13040-017-0150-8

Berghoff, B. A., and Wagner, E. G. H. (2017). RNA-based regulation in type I toxin–antitoxin systems and its implication for bacterial persistence. *Curr. Genet.* 63, 1011–1016. doi: 10.1007/s00294-017-0710-y

Berghoff, B. A., and Wagner, E. G. H. (2019). "Persister formation driven by TisB-dependent membrane depolarization," in *Persister Cells and Infectious Disease*. ed. K. Lewis (Cham: Springer International Publishing), 77–97.

Brantl, S., and Jahn, N. (2015). sRNAs in bacterial type I and type III toxin-antitoxin systems. *FEMS Microbiol. Rev.* 39, 413–427. doi: 10.1093/femsre/fuv003

Brauner, A., Fridman, O., Gefen, O., and Balaban, N. Q. (2016). Distinguishing between resistance, tolerance and persistence to antibiotic treatment. *Nat. Rev. Microbiol.* 14, 320–330. doi: 10.1038/nrmicro.2016.34

Brielle, R., Pinel-Marie, M.-L., and Felden, B. (2016). Linking bacterial type I toxins with their actions. *Curr. Opin. Microbiol.* 30, 114–121. doi: 10.1016/j.mib.2016.01.009

Cheverton, A. M., Gollan, B., Przydacz, M., Wong, C. T., Mylona, A., Hare, S. A., et al. (2016). A *Salmonella* toxin promotes Persister formation through acetylation of tRNA. *Mol. Cell* 63, 86–96. doi: 10.1016/j.molcel.2016.05.002

Conlon, B. P., Rowe, S. E., Gandt, A. B., Nuxoll, A. S., Donegan, N. P., Zalis, E. A., et al. (2016). Persister formation in *Staphylococcus aureus* is associated with ATP depletion. *Nat. Microbiol.* 1:16051. doi: 10.1038/nmicrobiol.2016.51

Coray, D. S., Wheeler, N. E., Heinemann, J. A., and Gardner, P. P. (2017). Why so narrow: distribution of anti-sense regulated, type I toxin-antitoxin systems compared with type II and type III systems. *RNA Biol.* 14, 275–280. doi: 10.1080/15476286.2016.1272747

Courcelle, J., Khodursky, A., Peter, B., Brown, P. O., and Hanawalt, P. C. (2001). Comparative gene expression profiles following UV exposure in wild-type and SOS-deficient *Escherichia coli*. *Genetics* 158, 41–64. doi: 10.1093/genetics/158.1.41

Darfeuille, F., Unoson, C., Vogel, J., and Wagner, E. G. H. (2007). An antisense RNA inhibits translation by competing with standby ribosomes. *Mol. Cell* 26, 381–392. doi: 10.1016/j.molcel.2007.04.003

Dewachter, L., Bollen, C., Wilmaerts, D., Louwagie, E., Herpels, P., Matthay, P., et al. (2021). The dynamic transition of persistence toward the viable but nonculturable state during stationary phase is driven by protein aggregation. *MBio* 12:e0070321. doi: 10.1128/mBio.00703-21

Dewachter, L., Fauvart, M., and Michiels, J. (2019). Bacterial heterogeneity and antibiotic survival: understanding and combatting persistence and heteroresistance. *Mol. Cell* 76, 255–267. doi: 10.1016/j.molcel.2019.09.028

Dörr, T., Vulic, M., and Lewis, K. (2010). Ciprofloxacin causes persister formation by inducing the TisB toxin in *Escherichia coli*. *PLoS Biol.* 8:e1000317. doi: 10.1371/journal.pbio.1000317

Edelmann, D., and Berghoff, B. A. (2019). Type I toxin-dependent generation of superoxide affects the persister life cycle of *Escherichia coli*. *Sci. Rep.* 9:14256. doi: 10.1038/s41598-019-50668-1

Edelmann, D., Leinberger, F. H., Schmid, N. E., Oberpaul, M., Schäberle, T. F., and Berghoff, B. A. (2021). Elevated expression of toxin TisB protects persister cells against ciprofloxacin but enhances susceptibility to mitomycin C. *Microorganisms* 9:943. doi: 10.3390/microorganisms9050943

Fozo, E. M., Hemm, M. R., and Storz, G. (2008). Small toxic proteins and the antisense RNAs that repress them. *Microbiol. Mol. Biol. Rev.* 72, 579–589. doi: 10.1128/MMBR.00025-08

Fraikin, N., Goormaghtigh, F., and Van Melderen, L. (2020). Type II toxin-antitoxin systems: evolution and revolutions. *J. Bacteriol.* 202, e00763–e00819. doi: 10.1128/JB.00763-19

Gerdes, K., and Wagner, E. G. H. (2007). RNA antitoxins. *Curr. Opin. Microbiol.* 10, 117–124. doi: 10.1016/j.mib.2007.03.003

Goormaghtigh, F., Fraikin, N., Putrinš, M., Hallaert, T., Hauryliuk, V., Garcia-Pino, A., et al. (2018). Reassessing the role of type II toxin-antitoxin systems in formation of *Escherichia coli* type II persister cells. *MBio* 9, e00640–e00718. doi: 10.1128/mBio.00640-18

Goormaghtigh, F., and Van Melderen, L. (2019). Single-cell imaging and characterization of *Escherichia coli* persister cells to ofloxacin in exponential cultures. *Sci. Adv.* 5:eaav9462. doi: 10.1126/sciadv.aav9462

Gurnev, P. A., Ortenberg, R., Dörr, T., Lewis, K., and Bezrukov, S. M. (2012). Persister-promoting bacterial toxin TisB produces anion-selective pores in planar lipid bilayers. *FEBS Lett.* 586, 2529–2534. doi: 10.1016/j.febslet.2012.06.021

Harms, A., Brodersen, D. E., Mitarai, N., and Gerdes, K. (2018). Toxins, targets, and triggers: an overview of toxin-antitoxin biology. *Mol. Cell* 70, 768–784. doi: 10.1016/j.molcel.2018.01.003

Harms, A., Fino, C., Sørensen, M. A., Semsey, S., and Gerdes, K. (2017). Prophages and growth dynamics confound experimental results with antibiotic-tolerant persister cells. *MBio* 8, e01964–e01117. doi: 10.1128/mBio.01964-17

Helaine, S., Cheverton, A. M., Watson, K. G., Faure, L. M., Matthews, S. A., and Holden, D. W. (2014). Internalization of *Salmonella* by macrophages induces formation of nonreplicating persisters. *Science* 343, 204–218. doi: 10.1126/science.1244705

Imlay, J. A. (2013). The molecular mechanisms and physiological consequences of oxidative stress: lessons from a model bacterium. *Nat. Rev. Microbiol.* 11, 443–454. doi: 10.1038/nrmicro3032

Jahn, N., and Brantl, S. (2013). One antitoxin-two functions: SR4 controls toxin mRNA decay and translation. *Nucleic Acids Res.* 41, 9870–9880. doi: 10.1093/nar/gkt735

Jurėnas, D., Fraikin, N., Goormaghtigh, F., and Van Melderen, L. (2022). Biology and evolution of bacterial toxin-antitoxin systems. *Nat. Rev. Microbiol*. doi: 10.1038/s41579-021-00661-1 [Epub ahead of print].

Kaplan, Y., Reich, S., Oster, E., Maoz, S., Levin-Reisman, I., Ronin, I., et al. (2021). Observation of universal ageing dynamics in antibiotic persistence. *Nature* 600, 290–294. doi: 10.1038/s41586-021-04114-w

Kwan, B. W., Valenta, J. A., Benedik, M. J., and Wood, T. K. (2013). Arrested protein synthesis increases persister-like cell formation. *Antimicrob. Agents Chemother.* 57, 1468–1473. doi: 10.1128/AAC.02135-12

Leplae, R., Geeraerts, D., Hallez, R., Guglielmini, J., Drèze, P., and Van Melderen, L. (2011). Diversity of bacterial type II toxin–antitoxin systems: a comprehensive search and functional analysis of novel families. *Nucleic Acids Res.* 39, 5513–5525. doi: 10.1093/nar/gkr131

Lewis, K. (2010). Persister cells. *Annu. Rev. Microbiol.* 64, 357–372. doi: 10.1146/annurev.micro.112408.134306

Maisonneuve, E., Castro-Camargo, M., and Gerdes, K. (2013). (p)ppGpp controls bacterial persistence by stochastic induction of toxin-antitoxin activity. *Cell* 154, 1140–1150. doi: 10.1016/j.cell.2013.07.048

Maisonneuve, E., Shakespeare, L. J., Jørgensen, M. G., and Gerdes, K. (2011). Bacterial persistence by RNA endonucleases. *Proc. Natl. Acad. Sci. U. S. A.* 108, 13206–13211. doi: 10.1073/pnas.1100186108

Manuse, S., Shan, Y., Canas-Duarte, S. J., Bakshi, S., Sun, W.-S., Mori, H., et al. (2021). Bacterial persisters are a stochastically formed subpopulation of low-energy cells. *PLoS Biol.* 19:e3001194. doi: 10.1371/journal.pbio.3001194

Masachis, S., and Darfeuille, F. (2018). Type I Toxin-Antitoxin Systems: Regulating Toxin Expression *via* Shine-Dalgarno Sequence Sequestration and Small RNA Binding. *Microbiol. Spectr.* 6, 173–190. doi: 10.1128/microbiolspec.RWR-0030-2018

Page, R., and Peti, W. (2016). Toxin-antitoxin systems in bacterial growth arrest and persistence. *Nat. Chem. Biol.* 12, 208–214. doi: 10.1038/nchembio.2044

Pandey, D. P., and Gerdes, K. (2005). Toxin-antitoxin loci are highly abundant in free-living but lost from host-associated prokaryotes. *Nucleic Acids Res.* 33, 966–976. doi: 10.1093/nar/gki201

Pu, Y., Li, Y., Jin, X., Tian, T., Ma, Q., Zhao, Z., et al. (2019). ATP-dependent dynamic protein aggregation regulates bacterial dormancy depth critical for antibiotic tolerance. *Mol. Cell* 73, 143.e4–156.e4. doi: 10.1016/j.molcel.2018.10.022

Ronneau, S., and Helaine, S. (2019). Clarifying the link between toxin-antitoxin modules and bacterial persistence. *J. Mol. Biol.* 431, 3462–3471. doi: 10.1016/j.jmb.2019.03.019

Rycroft, J. A., Gollan, B., Grabe, G. J., Hall, A., Cheverton, A. M., Larrouy-Maumus, G., et al. (2018). Activity of acetyltransferase toxins involved in *Salmonella* persister formation during macrophage infection. *Nat. Commun.* 9:1993. doi: 10.1038/s41467-018-04472-6

Shan, Y., Brown Gandt, A., Rowe, S. E., Deisinger, J. P., Conlon, B. P., and Lewis, K. (2017). ATP-Dependent Persister Formation in *Escherichia coli*. *MBio* 8, e02267–e02316. doi: 10.1128/mBio.02267-16

Song, S., and Wood, T. K. (2020). A primary physiological role of toxin/antitoxin systems is phage inhibition. *Front. Microbiol.* 11:1895. doi: 10.3389/fmicb.2020.01895

Srivastava, A., Pati, S., Kaushik, H., Singh, S., and Garg, L. C. (2021). Toxin-antitoxin systems and their medical applications: current status and future perspective. *Appl. Microbiol. Biotechnol.* 105, 1803–1821. doi: 10.1007/s00253-021-11134-z

Svenningsen, M. S., Veress, A., Harms, A., Mitarai, N., and Semsey, S. (2019). Birth and resuscitation of (p)ppGpp induced antibiotic tolerant persister cells. *Sci. Rep.* 9:6056. doi: 10.1038/s41598-019-42403-7

Thisted, T., and Gerdes, K. (1992). Mechanism of post-segregational killing by the *hok/sok* system of plasmid R1: Sok antisense RNA regulates *hok* gene expression indirectly through the overlapping *mok* gene. *J. Mol. Biol.* 223, 41–54. doi: 10.1016/0022-2836(92)90714-U

Unoson, C., and Wagner, E. G. H. (2008). A small SOS-induced toxin is targeted against the inner membrane in *Escherichia coli*. *Mol. Microbiol.* 70, 258–270. doi: 10.1111/j.1365-2958.2008.06416.x

Verstraeten, N., Knapen, W. J., Kint, C. I., Liebens, V., Van den Bergh, B., Dewachter, L., et al. (2015). Obg and membrane depolarization are part of a microbial bet-hedging strategy that leads to antibiotic tolerance. *Mol. Cell* 59, 9–21. doi: 10.1016/j.molcel.2015.05.011

Vogel, J., Argaman, L., Wagner, E. G. H., and Altuvia, S. (2004). The small RNA IstR inhibits synthesis of an SOS-induced toxic peptide. *Curr. Biol.* 14, 2271–2276. doi: 10.1016/j.cub.2004.12.003

Wagner, E. G. H., and Unoson, C. (2012). The toxin-antitoxin system *tisB-istR1*: expression, regulation, and biological role in persister phenotypes. *RNA Biol.* 9, 1513–1619. doi: 10.4161/rna.22578

Wang, X., and Wood, T. K. (2011). Toxin-antitoxin systems influence biofilm and persister cell formation and the general stress response. *Appl. Environ. Microbiol.* 77, 5577–5583. doi: 10.1128/AEM.05068-11

Wen, J., Harp, J. R., and Fozo, E. M. (2017). The 5′ UTR of the type I toxin ZorO can both inhibit and enhance translation. *Nucleic Acids Res.* 45, 4006–4020. doi: 10.1093/nar/gkw1172

Wilmaerts, D., Bayoumi, M., Dewachter, L., Knapen, W., Mika, J. T., Hofkens, J., et al. (2018). The persistence-inducing toxin HokB forms dynamic pores that cause ATP leakage. *MBio* 9, e00744–e00818. doi: 10.1128/mBio.00744-18

Windels, E. M., Ben Meriem, Z., Zahir, T., Verstrepen, K. J., Hersen, P., Van den Bergh, B., et al. (2019). Enrichment of persisters enabled by a ß-lactam-induced filamentation method reveals their stochastic

single-cell awakening. *Commun. Biol.* 2:426. doi: 10.1038/s42003-019-0672-3

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

frontiers | Frontiers in Microbiology

ORIGINAL RESEARCH
published: 07 April 2022
doi: 10.3389/fmicb.2022.861549

In vivo Functional Characterization of Hydrophilic X2 Modules in the Cellulosomal Scaffolding Protein

Xuanyu Tao[1†], Jiantao Liu[1,2†], Megan L. Kempher[1], Tao Xu[1,3,4] and Jizhong Zhou[1,5*]*

[1] *Institute for Environmental Genomics and Department of Microbiology and Plant Biology, University of Oklahoma, Norman, OK, United States,* [2] *School of Life Sciences, Jiangxi Science & Technology Normal University, Nanchang, China,* [3] *Section on Pathophysiology and Molecular Pharmacology, Joslin Diabetes Center, Boston, MA, United States,* [4] *Department of Microbiology and Immunobiology, Harvard Medical School, Boston, MA, United States,* [5] *Earth Sciences Division, Lawrence Berkeley National Laboratory, Berkeley, CA, United States*

OPEN ACCESS

Edited by:
Xinqing Zhao,
Shanghai Jiao Tong University, China

Reviewed by:
Fu-Li Li,
Qingdao Institute of Bioenergy and Bioprocess Technology (CAS), China
Roberto Mazzoli,
University of Turin, Italy

***Correspondence:**
Tao Xu
tao.xu@joslin.harvard.edu
Jizhong Zhou
jzhou@ou.edu

[†] These authors have contributed equally to this work and share first authorship

Specialty section:
This article was submitted to Microbial Physiology and Metabolism, a section of the journal Frontiers in Microbiology

Received: *24 January 2022*
Accepted: *10 March 2022*
Published: *07 April 2022*

Citation:
Tao X, Liu J, Kempher ML, Xu T and Zhou J (2022) In vivo Functional Characterization of Hydrophilic X2 Modules in the Cellulosomal Scaffolding Protein. Front. Microbiol. 13:861549. doi: 10.3389/fmicb.2022.861549

As part of free cellulases or scaffolding proteins in cellulosomes, the hydrophilic non-catalytic X2 module is widely distributed in cellulolytic *Clostridia* or other *Firmicutes* bacteria. Previous biochemical studies suggest that X2 modules might increase the solubility and substrate binding affinity of X2-bearing proteins. However, their *in vivo* biological functions remain elusive. Here we employed CRISPR-Cas9 editing to genetically modify X2 modules by deleting the conserved motif (NGNT) from the CipC scaffoldin. Both single and double X2 mutants (X2-N: near the N terminus of CipC; X2-C: near the C terminus of CipC) presented similar stoichiometric compositions in isolated cellulosomes as the wildtype strain (WT). These X2 mutants had an elongated adaptation stage during growth on cellulose compared to cellobiose. Compared to WT, the double mutant ΔX2-NC reduced cellulose degradation by 15% and the amount of released soluble sugars by 63%. Since single X2 mutants did not present such obvious physiological changes as ΔX2-NC, there seems to be a functional redundancy between X2 modules in CipC. The *in vivo* adhesion assay revealed that ΔX2-NC decreased cell attachment to cellulose by 70% but a weaker effect was also overserved in single X2 mutants. These results highlight the *in vivo* biological role of X2 in increasing cellulose degradation efficiency by enhancing the binding affinity between cells and cellulose, which provides new perspectives for microbial engineering.

Keywords: cellulosome, X2 module, cellulose degradation, *Clostridium cellulolyticum*, motif deletion

INTRODUCTION

As a sustainable and carbon-neutral energy source, biofuels are one of the best alternative energy forms for replacing fossil fuels (Farrell et al., 2006; Naik et al., 2010). Compared to other biomass resources including food grade sources, municipal solid waste, or algae, non-edible lignocellulosic feedstock is a promising source of material for biofuel production as it is cheap, abundant and renewable (Naik et al., 2010; Liao et al., 2016). Although there are many platforms for bioconversion of lignocellulose into biofuels, such as non-isothermal simultaneous saccharification and simultaneous saccharification and co-fermentation, the common barrier for these bioconversion processes is the high cost for cellulase production (Balan, 2014;

Liao et al., 2016; Lynd, 2017). In order to reduce the cost of cellulase production and increase the efficiency of bioconversion of lignocellulose, the on-site saccharification strategies, which include consolidated bioprocessing (CBP) and consolidated bio-saccharification (CBS), have been proposed (Lynd et al., 2002, 2005; Liu et al., 2020).

As a model mesophilic clostridial species, *Clostridium cellulolyticum* can hydrolyze cellulose or hemicellulose and then ferment hydrolysis products to ethanol and other organic acids, which enables it as a potential candidate strain for CBP (Desvaux, 2005). Like other cellulolytic clostridia, *C. cellulolyticum* possesses an extracellular enzymatic complex termed the cellulosome, which allows *C. cellulolyticum* to efficiently degrade crystalline cellulose and enables it as a potential candidate strain for CBS (Gal et al., 1997; Shoham et al., 1999; Liu et al., 2020). The pivotal component of the cellulosome in *C. cellulolyticum* is a scaffolding protein/integrating protein encoded by the *cipC* gene, to which up to eight different catalytic cellulases can bind. Without catalytic activity, the CipC scaffoldin contains eight Cohesion modules (type I), one carbohydrate binding module (CBM) and two X2 modules (**Supplementary Figure 1**; Gal et al., 1997; Schwarz, 2001). The Cohesion module allows the binding from cellulases with type I dockerin (Fierobe et al., 1999). The CBM, belonging to the family IIIa (CBM3a), carries out binding between the entire cellulosome and cellulosic substrate (Boraston et al., 2004). The X2 modules are two hydrophilic domains, each of which contains 100 amino acid residues (Mosbah et al., 2000). In addition to being a part of the CipC scaffoldin, the X2 module is also found in other scaffoldin proteins from other cellulosome-producing bacteria, such as *Clostridium thermocellum* (Lamed and Bayer, 1988), *Clostridium cellulovorans* (Doi et al., 1994), and even in free cellulase enzymes from non-clostridial cellulolytic bacteria, such as *Lachnoclostridium phytofermentans* (Ravachol et al., 2015; Vita et al., 2019) and *Paenibacillus polymyxa* (Pasari et al., 2017). This suggests that X2 modules are widely distributed and may serve an important role in the biodegradation of lignocellulosic biomass. Based on the structure and *in vitro* biochemistry assays of X2 modules, it belongs to the immunoglobulin superfamily and has been predicted to be associated with the localization of the cellulosome, cellulose binding, cell wall binding, or increased enzymatic activities for free cellulase (Kosugi et al., 2004; Chanal et al., 2011; Ravachol et al., 2015; Pasari et al., 2017; Zhang et al., 2018; Tarraran et al., 2021). However, little is known about its *in vivo* biological functions and significance.

It is technically challenging to determine the *in vivo* significance of non-catalytic X2 modules in multi-modular proteins. With the development of CRISPR-Cas9 nickase-based genome editing method (Xu et al., 2015), we are now able to manipulate genes for protein module engineering. Here, using *C. cellulolyticum* as a model strain, we created genetically modified X2 modules in the cellulosomal scaffoldin CipC to study the *in vivo* biological function of X2 modules. We systematically characterized single and double X2 mutants from transcriptional, physiological, and biochemical aspects. Our results demonstrated that the conserved motif in X2 modules is important to cell growth on cellulose and cellulose hydrolysis, probably *via* mediating the binding affinity of cells to cellulose.

MATERIALS AND METHODS

Bacterial Strains and Plasmid Construction

Strains and plasmids used in this study are listed in **Table 1**. The ΔX2-N and ΔX2-C mutants with deletions of the conserved motif in the N-terminal and C-terminal X2 respectively, were constructed in our previous study (Xu et al., 2015). The dual X2 modules mutant (ΔX2-NC) was constructed using the pCas9n-X2-N-delete-donor to further delete the conserved motif of the X2-N module in ΔX2-C (Li et al., 2012). In brief, for each X2 module, its corresponding variant was created by deleting the 12-bp DNA sequence coding for the conserved Asn-Gly-Asn-Thr motif. To do so, the 23-bp target site for Cas9 nickase was partially overlapped with the deletion region such that the customized donor template could direct the nick repair in the genome to make intended changes (**Figure 1A**). Finally, the N-terminal and C-terminal X2 modules (X2-N and X2-C) were mutated precisely in the chromosome, generating ΔX2-N and ΔX2-C variants respectively (Xu et al., 2015). Sequential mutation of the N-terminal X2 module in ΔX2-C variant created a dual mutated X2 variant, named ΔX2-NC (**Figure 1A**). All mutations were verified by amplicon sequencing (**Figure 1B**).

E. coli DH5α strain (Invitrogen, Carlsbad, CA, United States) was used for cloning. Rosetta^{TM}(DE3) (Invitrogen, Carlsbad, CA, United States) was used for protein expression. Sequences of X2-C module and CBM3a, amplified from the genome of WT strain, were cloned into pET-28a vectors with a C-terminal 6 × His tag, yielding pET-X2-C and pET- CBM3a, respectively. The pET-ΔX2-C vector was generated from the pET-X2-C by Q5

TABLE 1 | List of plasmids and strains used in this study.

Strain or plasmid	Phenotype or genotype	Source or References
Strains		
Wild type of *C. cellulolyticum* H10	ATCC 35319	Boraston et al., 2004
ΔX2-N	Deletion of the conserved motif (NGNT) of X2-N	Xu et al., 2015
ΔX2-C	Deletion of the conserved motif (NGNT) in X2-C	This study
ΔX2-NC	Deletion of the conserved motif (NGNT) in both X2-C and X2-N	This study
Plasmids		
pFdCas9n-p4-pyrF w/2kbΔ	Cmp^r in *E.coli* and Tmp^r in *C. cellulolyticum* H10	Xu et al., 2015
pMS-RNA	spec^r in *E.coli*	Xu et al., 2015
pCR/8w p4-4 prom	spec^r in *E.coli*	Xu et al., 2015
pCas9n-X2-N-delete-donor	Cmp^r in *E.coli* and Tmp^r in *C. cellulolyticum* H10	This study

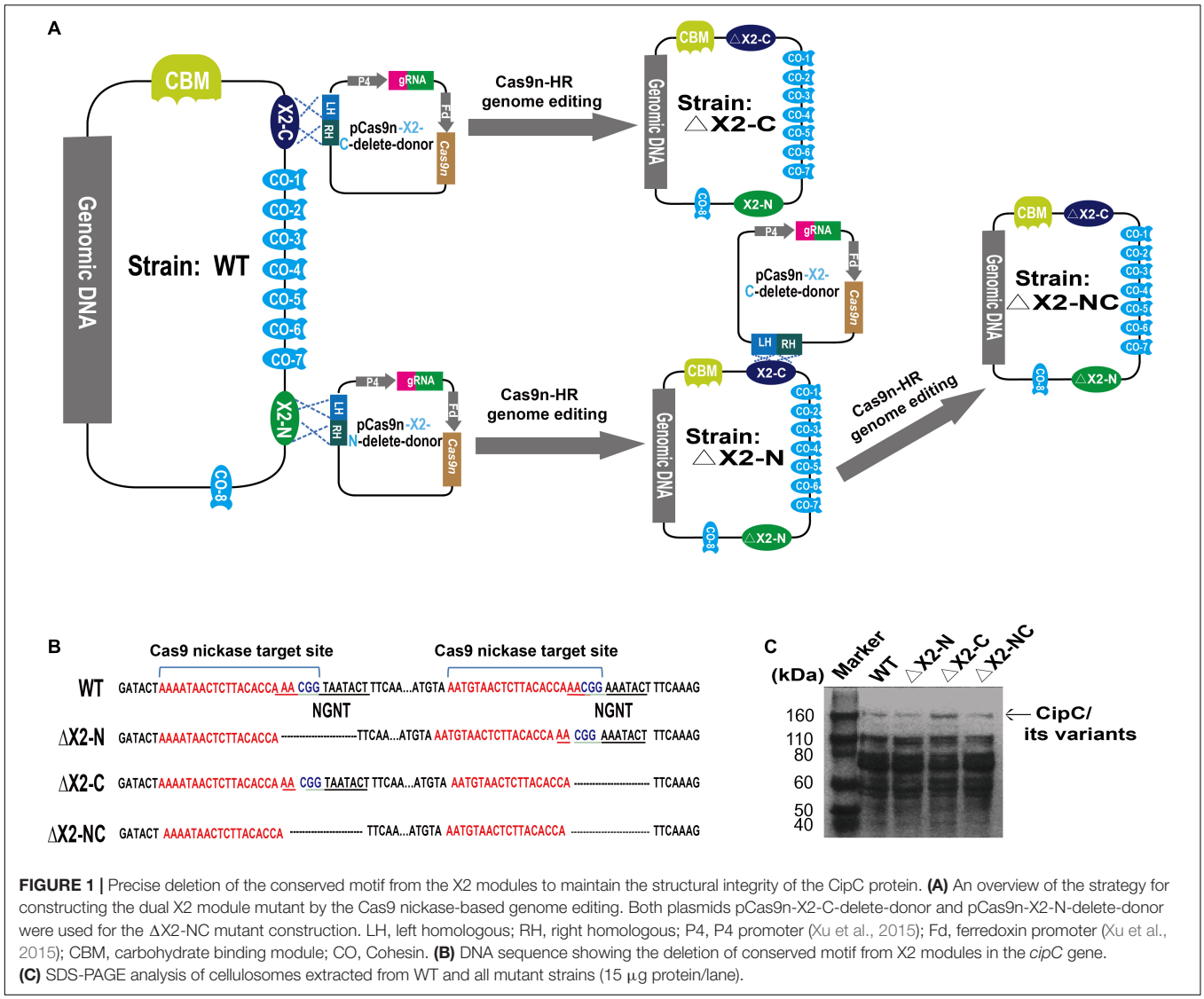


FIGURE 1 | Precise deletion of the conserved motif from the X2 modules to maintain the structural integrity of the CipC protein. **(A)** An overview of the strategy for constructing the dual X2 module mutant by the Cas9 nickase-based genome editing. Both plasmids pCas9n-X2-C-delete-donor and pCas9n-X2-N-delete-donor were used for the ΔX2-NC mutant construction. LH, left homologous; RH, right homologous; P4, P4 promoter (Xu et al., 2015); Fd, ferredoxin promoter (Xu et al., 2015); CBM, carbohydrate binding module; CO, Cohesin. **(B)** DNA sequence showing the deletion of conserved motif from X2 modules in the *cipC* gene. **(C)** SDS-PAGE analysis of cellulosomes extracted from WT and all mutant strains (15 μg protein/lane).

Site-Directed Mutagenesis Kit (New England Biolabs, Ipswich, MA, United States).

Media and Culture Conditions

LB medium with 35 μg/ml chloramphenicol or 50 μg/ml kanamycin was used for pCas9n-X2-N-delete-donor cloning or pET-X2C/pET-ΔX2C/pET-CBM3a cloning. Complex modified VM medium supplemented with 2.0 g/L yeast extract was used for reviving and transformation of ΔX2-C (Li et al., 2012). Defined modified VM medium containing necessary vitamin solution and mineral solution was used for growth determination and omics experiments (Higashide et al., 2011). All strains (ΔX2-N, ΔX2-C, ΔX2-NC, and WT) were cultured with 5 g/L cellobiose or 10 g/L Avicel PH101 crystalline cellulose (Sigma-Aldrich, St. Louis, MO, United States) at 34°C anaerobically. Solid VM medium with 1.0% (w/v) of Bacto agar (VWR) and 15 μg/ml thiamphenicol was used for selecting the ΔX2-NC mutant.

Transformation and Verification of Cellulosome Structure Integrity in Mutant Strains

The pCas9n-X2N-donor was firstly methylated by *Msp*I Methyltransferase (New England Biolabs, Ipswich, MA, United States) and then transformed into ΔX2-C by electroporation as described previously (Li et al., 2012). The pET-X2N, pET-ΔX2N, and pET-CBM3a were transformed into Rosetta™(DE3) according to the manufacturer's instructions. For each strain, the cellulosome fraction was isolated from 200 ml cellulose-grown culture at mid-exponential growth phase as previously reported (Tao et al., 2020).

Measurement of Cell Growth and Remaining Cellulose

Microbial growth (ΔX2-N, ΔX2-C, ΔX2-NC, and WT) on cellobiose was determined by optical density (OD) at 600 nm; on

cellulose, cell growth was estimated by total protein measurement as previously described (Li et al., 2012). Each strain had three biological replicates. The remaining cellulose and released soluble sugar in the medium were measured by the phenol-sulfuric acid method (Hemme et al., 2011).

Cell-Cellulose Adhesion Assay

All strains (ΔX2-N, ΔX2-C, ΔX2-NC, and WT) were grown on 5 g/L cellobiose to the same OD_{600}. Cells were incubated with or without Whatman filter paper (cellulose) on a tube rotator for 1 h. Then, the OD_{600} of the planktonic phase was determined to reversely infer the binding affinity of cells to cellulose. Each group contained three biological replicates. The relative cell adhesion capability for each strain was normalized to WT.

Microarray Analysis

All strains (ΔX2-N, ΔX2-C, ΔX2-NC, and WT) were grown in the defined VM medium with 10 g/L cellulose. Each strain was collected at the mid-exponential growth phase. RNA extraction, microarray hybridization, and microarray data analysis were performed as previously described (Tao et al., 2020). For each strain, four biological replicates were performed.

Expression and Purification of X2-N, ΔX2-N, and CBM3a

For the expression of X2-N or ΔX2-N module, Rosetta^{TM}(DE3) containing the corresponding vector was grown to OD_{600} 1.0–1.2 and then induced with 0.5 mM isopropyl-d-1-thiogalactopyranoside (IPTG) for 8 h at 37°C. Then, cells pellets were resuspended in the lysis buffer containing 20 mM Tris-HCl pH 8.0, 100 mM NaCl, 10 mM imidazole. After sonication and centrifugation at 12, 000 *g* for 30 min, the supernatant was loaded to Ni^{2+}-nitrilotriacetate affinity resin (Qiagen, Hilden, Germany) equilibrated with 20 mM Tris−HCl pH 8.0, 150 mM NaCl. The X2-N or ΔX2-N protein was eluted with 20 mM Tris−HCl pH 8.0, 150 mM NaCl, 350 mM imidazole and further purified by buffer exchange. For CBM3a expression, Rosetta^{TM}(DE3) with pET-CBM3a was grown to OD_{600} 0.6–0.8 and then induced with 0.2 mM IPTG for 20 h at 16°C. The remaining steps were the same as above. The structure of X2-N (PDB: 1EHX) was used as the template to construct the structure of X2-C by homology modeling in SWISS-MODEL[1].

The X2 Module-Cell Wall Fragments Binding Assay

A 20 ml of *E. coli* grown in LB medium, *Clostridium thermocellum* and *C. cellulolyticum* grown on 5 g/L cellobiose in VM medium were collected and centrifuged respectively. For each strain, the cell pellets were treated with NaN_3/Ca^{2+} and washed three times with 50 mM Phosphate-buffered saline (PBS) buffer (Sigma-Aldrich, St. Louis, MO, United States). Then, 50 μg purified X2-N, ΔX2-N, and CBM3a proteins were incubated with collected cell pellets respectively in the PBS buffer at 4°C for 12 h. After centrifugation, the pellet from each incubation was resuspended in the SDS-PAGE Protein Loading Buffer (Thermo Fisher, Waltham, MA, United States) and boiled for 10 min. Finally, the 6x-His Tag antibody (R930-25, Thermo Fisher, Waltham, MA, United States) was used to detect the binding between proteins and cell wall by Western blotting.

The X2 Module-Cellulose Binding Assay

A 50 μg of X2-N, CBM3a, and BSA (Thermo Fisher, Waltham, MA, United States) were separately incubated with 20 mg Avicel PH101 crystalline cellulose in the reaction buffer (20 mM Tris-HCl pH 8.0) at 4°C for 12 h. After centrifugation, the pellet from each incubation was washed with the reaction buffer three times. Then, the supernatant and pellet were analyzed by SDS-PAGE to detect binding between proteins and cellulose.

Isothermal Titration Calorimetry Assay

Titration calorimetry measurement was performed with a VP-Isothermal Titration Calorimetry (ITC) calorimeter (MicroCal, Northampton, MA, United States) as previously described (Duff et al., 2011; Brown et al., 2013). The buffer containing 20 mM Tris-HCl pH 8.0 and 100 mM NaCl was used for the assay. In brief, titration with buffer alone (background) and titration from X2-N protein into the buffer (control), were firstly performed as background and control groups. Then, 70 μM X2-N protein was titrated into 10 μM CBM3a protein with an injection volume of 70 μL and constant stirring at 25°C. Finally, a one-site binding model from the Origin 8.0 software (MicroCal) was used for data analysis.

RESULTS

Non-catalytic X2 Modules Contain a Highly Conserved Asn-Gly-Asn-Thr Motif

We first evaluated the prevalence of X2 modules (PF03442) in assembled bacterial genomes using HMMER tools[2] and found 1243 species that genetically encode proteins containing X2 modules including bacteria (1024), eukaryote (214), archaea (4), and unclassified (1). Among bacteria, more than 60% of identified X2 modules were found in the *Firmicutes* phylum including the classes *Bacilli* and *Clostridia*. Interestingly, at the protein level, more than 50% of X2 modules were found in multi-modular cellulases or cellulosomes, adjacent to the CBM module (X2-CBM). These architectural features may indicate an important role of X2 in insoluble carbon-associated microbial metabolism.

To test this hypothesis, we selected X2 modules found in the cellulosomal scaffolding protein CipC of *C. cellulolyticum* to interrogate further. These two X2 modules were named as X2-N (near the N terminus of CipC) and X2-C (near the C terminus of CipC) based on their loci in the CipC protein (**Supplementary Figure 1**). Since a truncated CipC led to the failure of cellulosome assembly (Maamar et al., 2004), a precise genome editing tool is required to mutate the X2 modules without sacrificing the overall architecture of CipC. It is interesting that an alignment of X2-like modules from different bacteria revealed a highly conserved short

[1] swissmodel.expasy.org

[2] https://www.ebi.ac.uk/Tools/hmmer/

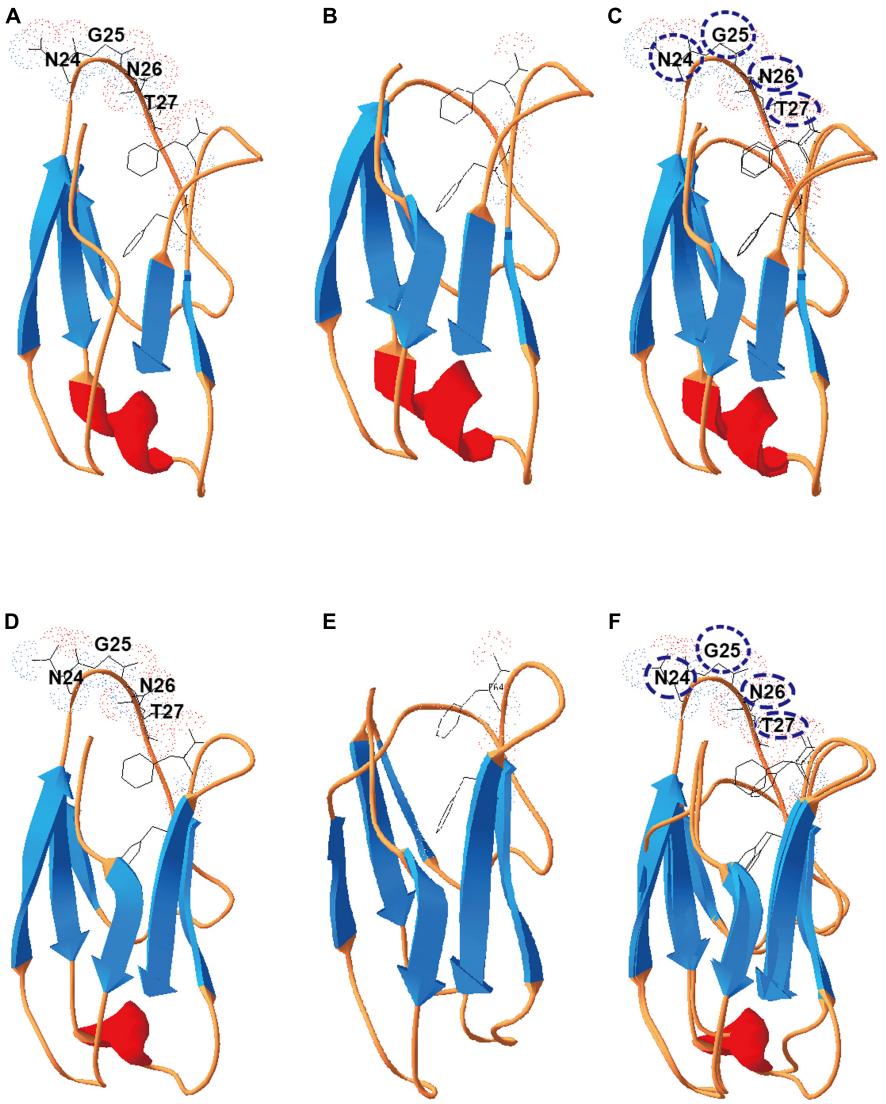


FIGURE 2 | Deletion of the conserved motif (NGNT) may lead to a conformational change of the X2 module. **(A)** The structure of the X2-N module protein; **(B)** the structure of ΔX2-N module in which the conserved NGNT residues were deleted. **(C)** Structures overlapping between X2-N and ΔX2-N modules. **(D)** The structure of the X2-C module protein; **(E)** the structure of ΔX2-C module in which the conserved NGNT residues were deleted; **(F)** structures overlapping between X2-C and ΔX2-C modules.

motif (Asn-Gly-Asn-Thr) (**Supplementary Figure 2**). This motif is structurally located in an exposed loop on X2-N (PDB: 1EHX) and X2-C (**Figure 2**). Deletion of this short motif would shorten the loop region and form a wider groove as shown by structure modeling (**Figure 2**), which would allow for investigation of X2 function in CipC with a minimal impact on CipC structure.

Single and Dual X2 Modifications in CipC Diminish Physiological Performance on Cellulose

We precisely deleted the conserved motif in the X2 modules of CipC by Cas9 nickase-based genome editing, yielding single (ΔX2-N and ΔX2-C) and double X2 mutants (ΔX2-NC) (**Figures 1A,B**). SDS-PAGE analysis of cellulosome fractions from cellulose-grown WT and mutant strains demonstrated no significant changes in gel patterns and band intensities (**Figure 1C**), which suggested the cellulosome architecture was not impeded by the X2 mutations.

To test if X2 was involved in insoluble carbon-dependent microbial metabolism, we monitored the growth of X2 mutants on different carbon sources. On cellobiose, all mutants (ΔX2-N, ΔX2-C, and ΔX2-NC) presented similar growth profiles as WT (**Figure 3A**). However, when grown on 10 g/L cellulose, ΔX2-N, ΔX2-C, and ΔX2-NC had a much longer adaptation stage than WT before achieving a similar growth rate and final biomass level as WT (**Figure 3B**), suggesting the importance of X2 modules in cell growth on insoluble cellulose. In addition, compared to WT, the release of soluble sugars in ΔX2-N, ΔX2-C,

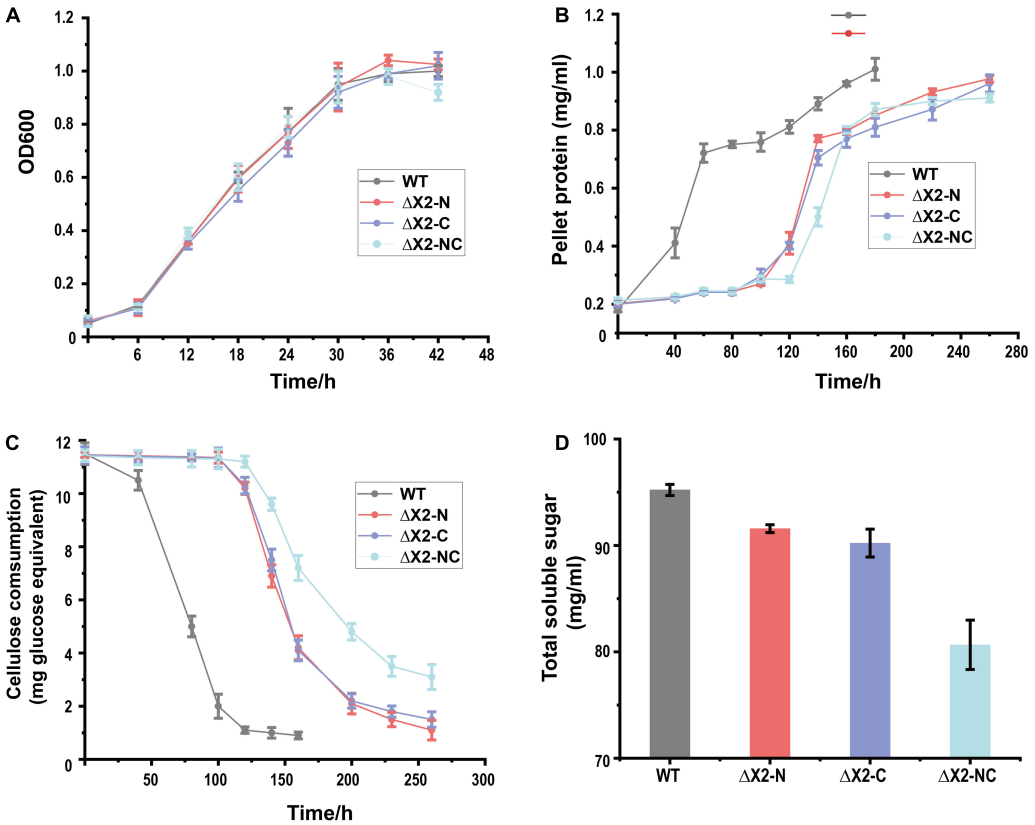


FIGURE 3 | Disruption of X2 modules increased the lag phase and decreased the cellulose degradation efficiency when mutants were grown on cellulose. **(A)** Growth profiles of WT, ΔX2-N, ΔX2-C, and ΔX2-NC grown on cellobiose. **(B)** Growth profiles of WT, ΔX2-N, ΔX2-C, and ΔX2-NC grown on cellulose. **(C)** Cellulose degradation profiles of WT, ΔX2-N, ΔX2-C, and ΔX2-NC. **(D)** Released total soluble sugars in supernatant of medium at final time point for each strain when grown on cellulose. Data are presented as the mean of three biological replicates and error bars represent standard deviation (SD).

and ΔX2-NC was decreased by 28, 40, and 63%, respectively. The amount of residual cellulose in ΔX2-NC was 15% higher than WT and approximately 10% higher than ΔX2-N and ΔX2-C. These results concordantly demonstrate that genetic mutation of X2 in CipC can reduce cellulose hydrolysis of the cellulosome (**Figures 3C,D** and **Supplementary Figure 3**). This is consistent with previous *in vitro* studies on X2 in *C. cellulovorans* (Kosugi et al., 2004). These data also solidify the importance of the conserved motifs in the X2 module for the first time.

Disrupted X2 Modules Reduce Cell Adhesion to Cellulose

To understand if a polar effect was introduced to downstream genes or the expressions of other genes were influenced by the X2 mutations (Maamar et al., 2004), we performed microarray-based transcriptomic analysis for the whole transcriptome of *C. cellulolyticum*. We found there were no differentially expressed genes present between WT and X2 mutants grown on cellulose during the exponential phase. Therefore, the disruption of the X2 modules seemed to have influenced cellulose degradation not by regulating gene expressions.

We then speculated that X2 modules might be key factors responsible for (i) increasing the binding affinity of cellulosomes to cellulose, (ii) increasing the localization and adhesion of cellulosomes to the cell surfaces, or (iii) increasing both binding affinity to cellulose and adhesion to cell surfaces. To determine whether the X2 mutations influenced the binding affinity between cells and cellulose, a cell-cellulose adhesion assay was performed. With cells collected at the early exponential stage, we found ΔX2-NC mutant decreased cell adhesion to cellulose by 50% when compared to WT. The ΔX2-C presented a more profound effect than ΔX2-N (**Figure 4A**). The decrease in cell adhesion was even more obvious when cells were at the late exponential stage (**Figure 4B**). Therefore, X2 modules, especially the X2-C module in CipC, play an important role in cell adhesion to cellulose.

X2 Modules Cannot Directly Bind to the Cell Surface of *Clostridium cellulolyticum*

To understand how X2 modules influence cell adhesion to cellulose, we performed *in vitro* X2 module-cell wall binding assay to test if the X2 modules mediate cellulosome localization on the cell surface. As ΔX2-C generated more significant physiological changes than ΔX2-N on cellulose, we expressed and purified X2-C module, CBM3a and ΔX2-C module (deletion of

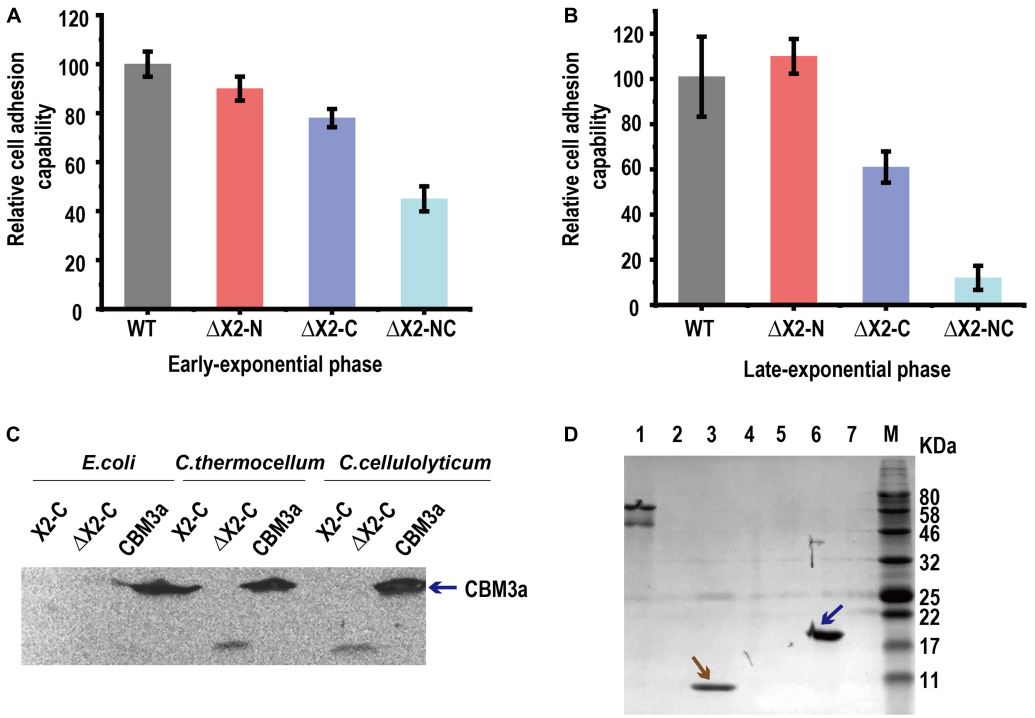


FIGURE 4 | The *in vivo* function of the X2 module was related to the binding affinity between cells and cellulose. Panels **(A,B)**, the relative cell adhesion capability between cells and cellulose for each strain in early exponential **(A)** and late-exponential phase **(B)**. Data are presented as the mean of three biological replicates and error bars represent SD. **(C)** The binding of X2-C, ΔX2-C, and CBM3a proteins to the cell surfaces of *E. coli*, *C. thermocellum,* or *C. cellulolyticum* respectively, determined by western blot. CBM3a was detected in all three strains (as blue arrow indicated), indicating it could bind to the cell surface for both Gram-negative and Gram-positive bacteria. The X2-C could not be detected for any of them, indicating it can not directly bind to the cell surface. The weak band of ΔX2-C was detected in *C. thermocellum* and *C. cellulolyticum*, indicating it had a weak binding affinity with the cell surface of the Gram-positive bacteria. **(D)** Binding of X2-C, CBM3a, and BSA protein to crystalline cellulose, determined by SDS-PAGE. The CBM3a protein and the BSA protein were used as the positive and negative control respectively. The CBM3a was detected in the cellulose pellet (as blue arrow indicated) and X2-C was only detected in the supernatant fraction as same as the BSA negative control (as brown arrow indicated), indicating that X2-C can not directly bind to the cellulose. Lane 1, BSA + Cellulose in supernatant fraction; Lane 2, CBM + Cellulose in supernatant fraction; Lane 3, X2-C + Cellulose in supernatant fraction; Lane 4, blank; Lane 5, BSA + Cellulose in cellulose-containing pellet; Lane 6, CBM + Cellulose in cellulose-containing pellet; Lane 7, X2-C + Cellulose in cellulose-containing pellet.

the conserved short motif) for *in vitro* assays (If not specified, we used X2 to represent X2-C in the following *in vitro* assays). In this binding assay, we separately mixed above-purified module proteins with *C. cellulolyticum* cells or *E. coli* (gram-negative) and *C. thermocellum* (Gram-positive) as controls. Western blotting did not detect the presence of X2 on any of these three strains post incubation with purified X2, indicating that X2 cannot directly bind to cell surfaces (**Figure 4C**). This agrees well with a recent study (Tarraran et al., 2021).

However, purified CBM3a was found in all three strains, indicating it is able to bind bacterial cell walls (**Figure 4C**). More interesting, we found a very weak signal of ΔX2 protein in the group of *C. thermocellum* and *C. cellulolyticum*. It is possible that ΔX2 could enhance the localization of cellulosomes on the surface of these cellulose degrading strains.

X2 Module Cannot Directly Bind to Cellulose in *Clostridium cellulolyticum*

Since the X2 module could not directly bind to cell wall, here we performed *in vitro* X2 module-cellulose binding assays to test if X2 modules mediate the binding affinity between cellulosomes and cellulose. Using BSA as the negative control and purified CBM3a as the positive control for cellulose binding assays, we found that purified X2 only present in the supernatant instead of cellulose pellets, indicating that the X2 protein cannot directly bind to the cellulose (**Figure 4D**). Meanwhile, the ITC assay indicated that there was a weak interaction between the X2 module and the CBM3a (**Supplementary Figure 4**).

In summary, the X2 module in CipC cannot directly bind to cellulose or the bacterial cell surface. However, removing the conserved motif in the X2 modules decreased cellulose utilization and severely reduced cell attachment on cellulose.

DISCUSSION

The function of the X2 module has been studied for many years (Mosbah et al., 2000; Kosugi et al., 2004; Pasari et al., 2017), and almost all previous studies determining the function of the X2 module were based on *in vitro* biochemical assays. However, the *in vivo* function of the X2 module remains elusive.

In *C. cellulolyticum*, two X2 modules are located in the *cipC* gene and the nucleotide sequence identity between them are very similar (65% for the pairwise nucleotide sequence identity). Due to limitations of traditional genome editing, it was very difficult to inactivate both of the X2 modules while maintaining the functional and structural integrity of the CipC scaffoldin protein. Fortunately, with the development and application of CRISPR-Cas9 based genome editing tools, we were able to generate both single and double deletions of two conserved sites in the X2 module. To our knowledge, at the genomic level determination of the functions of the X2 module has never been reported in any strain. A previous study, found that the disruption of the *cipC* gene in *C. cellulolyticum* hardly affected growth on soluble sugar but led to barely any growth on cellulose (Maamar et al., 2004). Compared to their study, we also found that the mutation of X2 modules did not influence growth on cellobiose (**Figure 3A**). Although, the mutation of the X2 modules led to a longer lag phase and decrease in cellulose degradation efficiency, growth on cellulose was not severely affected as was observed for the *cipC* deletion strain. Even the growth rate and maximum cell biomass between WT and ΔX2-NC were similar (**Figure 3B**). All of these data indicated that the deletion of the conserved motif in X2 modules did not affect the basic functions of the CipC and the functional loss of the X2 modules may directly influence certain functions of the cellulosome, as noted by the observed decrease in efficiency of the cellulosome in cellulose degradation.

Our microarray analysis confirmed the disruption of the X2 module in the *cipC* gene did not influence gene expression. On the other hand, the results of the adhesion assay indicated that the function of X2 modules was related to binding affinity between the cells and cellulose (**Figures 4A,B**). It is known that the cell surface cellulosome is the bridge for adhesion between the cells and cellulose (Gal et al., 1997; Schwarz, 2001), although the mechanism of the localization of the cellulosomes on the surface of *C. cellulolyticum* remains unclear (Desvaux, 2005). Therefore, the lower binding affinity between cells and cellulose caused by mutation of X2 modules should be attributed to the functional change of cellulosomes, which was consistent with our assumption.

From the cell adhesion assay (**Figures 4,B**), we hypothesized three possible mechanisms for X2 modules in regulating the binding affinity between cells and cellulose. Although previous studies indicated that the X2 module might directly bind to the cell wall (Kosugi et al., 2004), our *in vitro* protein-cell wall binding assay indicated that the X2 module protein could not directly bind to the cell wall (**Figure 4C**). This is consistent with a recent study that scaffoldins containing X2 domains derived from CpbA were not able to bind to the *L. lactis* surface (Tarraran et al., 2021). Based on structural analysis of the X2 modules (Mosbah et al., 2000), the surface of the X2 module is predominantly covered by hydrophilic amino acids and only contains a hydrophobic shallow groove, which could explain why the X2 module could not directly bind to the cell wall. Notably, the ΔX2 module protein had weak binding with the cell wall. This might promote the binding between the cellulosomes and cell surface, suggesting more numbers of cellulosomes might locate on the cell surface of the mutant strain. When the conserved motif (NGNT) was deleted, the hydrophobic shallow groove became wider as structure modeling indicated (**Figure 2**), which is more similar to typical CBM modules where most hydrophobic residues protrude outside (Luis et al., 2013; Pasari et al., 2017). As a result, this structural change might allow the ΔX2 module to weakly bind to the cell wall. This possible mechanism will be further investigated in our future work.

A previous study indicated that the localization of the cellulosomes (localized on the surface/free living) did not significantly influence the cellulose degradation efficiency of the cellulosomes (Xu et al., 2016). If the function of the X2 modules was only related to the localization of the cellulosomes, we would not expect to observe a significant decrease in cellulose degradation efficiency and release of soluble sugars in the ΔX2-NC strain compared to the WT (**Figures 3C,D**). Meanwhile, the HMMER analysis found that many X2 modules existed in free cellulases, which also indicates that the main function of X2 modules is likely unrelated to localization of cellulosomes. Meanwhile, we did not observe the X2 module from *C. cellulolyticum* could direcly bind to cellulose (**Figure 4D**). In CBM modules, most hydrophobic residues are on the protruded on the surface to promote carbohydrate polymer binding (Luis et al., 2013). In contrast, few hydrophobic residues are on the surface of the X2 module, and all the polar residues are exposed to the solvent (Mosbah et al., 2000), which indicated the reason why the X2 module could not directly bind to the cellulose. Therefore, the only possible reasonable explanation for the cell adhesion assay was that the X2 module cannot directly bind to the cellulose but may promote the binding between the cellulosomes and cellulose.

Given that the CBM3a module of the CipC scaffolding protein is for binding to the cellulose, we speculated that the *in vivo* function of the X2 module was realized by promoting the binding between CBM3a domain and cellulose. Our ITC assay also found there was a weak interaction between the X2 module and the CBM3a (**Supplementary Figure 4**), suggesting that the X2 module might interact with CBM3a and promote its binding function. In fact, previous *in vitro* assay had already found that the CBM3-X2 module had a better cellulose binding affinity to crystalline cellulose compared to CBM3 module alone (Pasari et al., 2017; Zhang et al., 2018). The processivity of *Cc*Cel9A mutant that lacks the CBMX2s was significantly lower compared to that of the wild-type *Cc*Cel9A, indicating the X2 module could indeed promote the binding between the CBM3 module and cellulose (Pasari et al., 2017; Zhang et al., 2018). In addition, the X2 module was thought to be associated with cellulase activity, which rendered the Cel9A cellulase from *L. phytofermentans* to be significantly more efficient on crystalline cellulose than any of the known cellulases from *C. cellulolyticum* (Ravachol et al., 2015). Some cellulase engineering studies also pointed out that the integration of the CBM with the X2 module into some cellulases enhanced avicelase activities, such as Cel48F and Cel9G (Mingardon et al., 2007; Vita et al., 2019). Taken together, these data may explain why we observed that the phenotype of WT grown on cellulose was better than ΔX2-NC and why the X2 modules are always next to the CBM3a modules in free cellulases.

CONCLUSION

In summary, precise deletion of the NGNT conserved sequences of the X module was a useful strategy to carry out functional *in vivo* studies as this approach maintained the structural and functional integrity of the cellulosomes. This strategy can be applied to study the function of X2 modules or other interesting modules within certain proteins in other bacteria with similar cellulases/cellulosome-producing systems. We found that (i) the mutation of the X2 modules in *C. cellulolyticum* could indeed influence the cellulose utilization efficiency, and (ii) the *in vivo* function of the X2 module was determined to be associated with binding affinity between cells and cellulose. Given that the X2 modules are widely distributed in cellulolytic bacteria and play important roles in cellulose degradation, all of these findings provide new perspectives on engineering those potential CBP bacteria to improve their cellulose degradation efficiencies or modifying commercial cellulases to improve their hydrolysis efficiencies.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/**Supplementary Material**, further inquiries can be directed to the corresponding authors.

AUTHOR CONTRIBUTIONS

TX, XT, and JZ designed the experiments. XT, TX, MK, and JL performed all the experiments. XT, JL, MK, and TX wrote the manuscript. MK and JZ edited the manuscript. All authors were given the opportunity to review the results and comment on the manuscript.

FUNDING

This work was supported by the Office of the Vice President for Research at the University of Oklahoma, the National Science Foundation of China (31960469), and the China Scholarship Council (201908360006).

ACKNOWLEDGMENTS

We thank members of the Zhou laboratory for help in the project. We also appreciate the help of Jennifer Byrum and Karla Rodgers from OUHSC for the ITC assay.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmicb.2022.861549/full#supplementary-material

REFERENCES

Balan, V. (2014). Current challenges in commercially producing biofuels from lignocellulosic biomass. *ISRN Biotechnol.* 2014:463074. doi: 10.1155/2014/463074

Boraston, A. B., Bolam, D. N., Gilbert, H. J., and Davies, G. J. (2004). Carbohydrate-binding modules: fine-tuning polysaccharide recognition. *Biochem. J.* 382(Pt 3), 769–781. doi: 10.1042/BJ20040892

Brown, B. L., Lord, D. M., Grigoriu, S., Peti, W., and Page, R. (2013). The *Escherichia coli* toxin MqsR destabilizes the transcriptional repression complex formed between the antitoxin MqsA and the mqsRA operon promoter. *J. Biol. Chem.* 288, 1286–1294. doi: 10.1074/jbc.M112.421008

Chanal, A., Mingardon, F., Bauzan, M., Tardif, C., and Fierobe, H. P. (2011). Scaffoldin modules serving as "cargo" domains to promote the secretion of heterologous cellulosomal cellulases by Clostridium acetobutylicum. *Appl. Environ. Microbiol.* 77, 6277–6280. doi: 10.1128/AEM.00758-11

Desvaux, M. (2005). Clostridium cellulolyticum: model organism of mesophilic cellulolytic clostridia. *FEMS Microbiol. Rev.* 29, 741–764. doi: 10.1016/j.femsre.2004.11.003

Doi, R. H., Goldstein, M., Hashida, S., Park, J. S., and Takagi, M. (1994). The Clostridium cellulovorans cellulosome. *Crit. Rev. Microbiol.* 20, 87–93. doi: 10.3109/10408419409113548

Duff, M. R. Jr., Grubbs, J., and Howell, E. E. (2011). Isothermal titration calorimetry for measuring macromolecule-ligand affinity. *J. Vis. Exp.* 55:e2796. doi: 10.3791/2796

Farrell, A. E., Plevin, R. J., Turner, B. T., Jones, A. D., O'Hare, M., and Kammen, D. M. (2006). Ethanol can contribute to energy and environmental goals. *Science* 311, 506–508. doi: 10.1126/science.1121416

Fierobe, H. P., Pages, S., Belaich, A., Champ, S., Lexa, D., and Belaich, J. P. (1999). Cellulosome from Clostridium cellulolyticum: molecular study of the Dockerin/Cohesin interaction. *Biochemistry* 38, 12822–12832. doi: 10.1021/bi9911740

Gal, L., Pages, S., Gaudin, C., Belaich, A., Reverbel-Leroy, C., Tardif, C., et al. (1997). Characterization of the cellulolytic complex (cellulosome) produced by Clostridium cellulolyticum. *Appl. Environ. Microbiol.* 63, 903–909. doi: 10.1128/AEM.63.3.903-909.1997

Hemme, C. L., Fields, M. W., He, Q., Deng, Y., Lin, L., Tu, Q., et al. (2011). Correlation of genomic and physiological traits of thermoanaerobacter species with biofuel yields. *Appl. Environ. Microbiol.* 77, 7998–8008. doi: 10.1128/AEM.05677-11

Higashide, W., Li, Y., Yang, Y., and Liao, J. C. (2011). Metabolic engineering of Clostridium cellulolyticum for production of isobutanol from cellulose. *Appl. Environ. Microbiol.* 77, 2727–2733. doi: 10.1128/AEM.02454-10

Kosugi, A., Amano, Y., Murashima, K., and Doi, R. H. (2004). Hydrophilic domains of scaffolding protein CbpA promote glycosyl hydrolase activity and localization of cellulosomes to the cell surface of Clostridium cellulovorans. *J. Bacteriol.* 186, 6351–6359. doi: 10.1128/JB.186.19.6351-6359.2004

Lamed, R., and Bayer, E. A. (1988). The cellulosome of Clostridium thermocellum. *Adv. Appl. Microbiol.* 33, 1–46. doi: 10.1016/S0065-2164(08)70203-X

Li, Y., Tschaplinski, T. J., Engle, N. L., Hamilton, C. Y., Rodriguez, M. Jr., Liao, J. C., et al. (2012). Combined inactivation of the Clostridium cellulolyticum lactate and malate dehydrogenase genes substantially increases ethanol yield from cellulose and switchgrass fermentations. *Biotechnol. Biofuels* 5:2. doi: 10.1186/1754-6834-5-2

Liao, J. C., Mi, L., Pontrelli, S., and Luo, S. (2016). Fuelling the future: microbial engineering for the production of sustainable biofuels. *Nat. Rev. Microbiol.* 14, 288–304. doi: 10.1038/nrmicro.2016.32

Liu, Y. J., Li, B., Feng, Y., and Cui, Q. (2020). Consolidated bio-saccharification: leading lignocellulose bioconversion into the real world. *Biotechnol. Adv.* 40:107535. doi: 10.1016/j.biotechadv.2020.107535

Luis, A. S., Venditto, I., Temple, M. J., Rogowski, A., Basle, A., Xue, J., et al. (2013). Understanding how noncatalytic carbohydrate binding modules can display specificity for xyloglucan. *J. Biol. Chem.* 288, 4799–4809. doi: 10.1074/jbc.M112.432781

Lynd, L. R. (2017). The grand challenge of cellulosic biofuels. *Nat. Biotechnol.* 35, 912–915. doi: 10.1038/nbt.3976

Lynd, L. R., van Zyl, W. H., McBride, J. E., and Laser, M. (2005). Consolidated bioprocessing of cellulosic biomass: an update. *Curr. Opin. Biotechnol.* 16, 577–583. doi: 10.1016/j.copbio.2005.08.009

Lynd, L. R., Weimer, P. J., van Zyl, W. H., and Pretorius, I. S. (2002). Microbial cellulose utilization: fundamentals and biotechnology. *Microbiol. Mol. Biol. Rev.* 66, 506–577. doi: 10.1128/mmbr.66.3.506-577.2002

Maamar, H., Valette, O., Fierobe, H. P., Belaich, A., Belaich, J. P., and Tardif, C. (2004). Cellulolysis is severely affected in Clostridium cellulolyticum strain cipCMut1. *Mol. Microbiol.* 51, 589–598. doi: 10.1046/j.1365-2958.2003.03859.x

Mingardon, F., Chanal, A., Tardif, C., Bayer, E. A., and Fierobe, H. P. (2007). Exploration of new geometries in cellulosome-like chimeras. *Appl. Environ. Microbiol.* 73, 7138–7149. doi: 10.1128/AEM.01306-07

Mosbah, A., Belaich, A., Bornet, O., Belaich, J. P., Henrissat, B., and Darbon, H. (2000). Solution structure of the module X2 1 of unknown function of the cellulosomal scaffolding protein CipC of Clostridium cellulolyticum. *J. Mol. Biol.* 304, 201–217. doi: 10.1006/jmbi.2000.4192

Naik, S. N., Goud, V. V., Rout, P. K., and Dalai, A. K. (2010). Production of first and second generation biofuels: a comprehensive review. *Renew. Sustain. Energy Rev.* 14, 578–597. doi: 10.1016/j.rser.2009.10.003

Pasari, N., Adlakha, N., Gupta, M., Bashir, Z., Rajacharya, G. H., Verma, G., et al. (2017). Impact of Module-X2 and Carbohydrate Binding Module-3 on the catalytic activity of associated glycoside hydrolases towards plant biomass. *Sci. Rep.* 7:3700. doi: 10.1038/s41598-017-03927-y

Ravachol, J., Borne, R., Meynial-Salles, I., Soucaille, P., Pages, S., Tardif, C., et al. (2015). Combining free and aggregated cellulolytic systems in the cellulosome-producing bacterium Ruminiclostridium cellulolyticum. *Biotechnol. Biofuels* 8:114. doi: 10.1186/s13068-015-0301-4

Schwarz, W. H. (2001). The cellulosome and cellulose degradation by anaerobic bacteria. *Appl. Microbiol. Biotechnol.* 56, 634–649. doi: 10.1007/s002530100710

Shoham, Y., Lamed, R., and Bayer, E. A. (1999). The cellulosome concept as an efficient microbial strategy for the degradation of insoluble polysaccharides. *Trends Microbiol.* 7, 275–281. doi: 10.1016/s0966-842x(99)01533-4

Tao, X., Xu, T., Kempher, M. L., Liu, J., and Zhou, J. (2020). Precise promoter integration improves cellulose bioconversion and thermotolerance in Clostridium cellulolyticum. *Metab. Eng.* 60, 110–118. doi: 10.1016/j.ymben.2020.03.013

Tarraran, L., Gandini, C., Luganini, A., and Mazzoli, R. (2021). Cell-surface binding domains from Clostridium cellulovorans can be used for surface display of cellulosomal scaffoldins in Lactococcus lactis. *Biotechnol. J.* 16:e2100064. doi: 10.1002/biot.202100064

Vita, N., Borne, R., Perret, S., de Philip, P., and Fierobe, H. P. (2019). Turning a potent family-9 free cellulase into an operational cellulosomal component and vice versa. *FEBS J.* 286, 3359–3373. doi: 10.1111/febs.14858

Xu, Q., Resch, M. G., Podkaminer, K., Yang, S., Baker, J. O., Donohoe, B. S., et al. (2016). Dramatic performance of Clostridium thermocellum explained by its wide range of cellulase modalities. *Sci. Adv.* 2:e1501254. doi: 10.1126/sciadv.1501254

Xu, T., Li, Y., Shi, Z., Hemme, C. L., Li, Y., Zhu, Y., et al. (2015). Efficient Genome Editing in Clostridium cellulolyticum via CRISPR-Cas9 Nickase. *Appl. Environ. Microbiol.* 81, 4423–4431. doi: 10.1128/AEM.00873-15

Zhang, K. D., Li, W., Wang, Y. F., Zheng, Y. L., Tan, F. C., Ma, X. Q., et al. (2018). Processive Degradation of Crystalline Cellulose by a Multimodular Endoglucanase via a Wirewalking Mode. *Biomacromolecules* 19, 1686–1696. doi: 10.1021/acs.biomac.8b00340

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

frontiers | Frontiers in Microbiology

MINI REVIEW
published: 25 April 2022
doi: 10.3389/fmicb.2022.869509

Exploiting Information and Control Theory for Directing Gene Expression in Cell Populations

*Lucas Henrion†, Mathéo Delvenne†, Fatemeh Bajoul Kakahi, Fabian Moreno-Avitia and Frank Delvigne**

Microbial Processes and Interactions (MiPI), Terra Research and Teaching Centre, Gembloux Agro-Bio Tech, University of Liège, Gembloux, Belgium

Microbial populations can adapt to adverse environmental conditions either by appropriately sensing and responding to the changes in their surroundings or by stochastically switching to an alternative phenotypic state. Recent data point out that these two strategies can be exhibited by the same cellular system, depending on the amplitude/frequency of the environmental perturbations and on the architecture of the genetic circuits involved in the adaptation process. Accordingly, several mitigation strategies have been designed for the effective control of microbial populations in different contexts, ranging from biomedicine to bioprocess engineering. Technically, such control strategies have been made possible by the advances made at the level of computational and synthetic biology combined with control theory. However, these control strategies have been applied mostly to synthetic gene circuits, impairing the applicability of the approach to natural circuits. In this review, we argue that it is possible to expand these control strategies to any cellular system and gene circuits based on a metric derived from this information theory, i.e., mutual information (MI). Indeed, based on this metric, it should be possible to characterize the natural frequency of any gene circuits and use it for controlling gene circuits within a population of cells.

Keywords: phenotypic heterogeneity, biological noise, population control, synchronization, cell collective behavior, cell decision-making process

OPEN ACCESS

Edited by:
Ulrike Kappler, The University of Queensland, Australia

Reviewed by:
Peng Xu, Guangdong Technion-Israel Institute of Technology (GTIIT), China

****Correspondence:***
Frank Delvigne
f.delvigne@uliege.be

†These authors have contributed equally to this work and share first authorship

Specialty section:
This article was submitted to Microbial Physiology and Metabolism, a section of the journal Frontiers in Microbiology

Received: *04 February 2022*
Accepted: *06 April 2022*
Published: *25 April 2022*

Citation:
Henrion L, Delvenne M, Bajoul Kakahi F, Moreno-Avitia F and Delvigne F (2022) Exploiting Information and Control Theory for Directing Gene Expression in Cell Populations. Front. Microbiol. 13:869509. doi: 10.3389/fmicb.2022.869509

INTRODUCTION

The parallel advances made at the level of cell culturing procedures [i.e., microfluidics (Grunberger et al., 2014) and cell–machine interfaces (Lugagne and Dunlop, 2019)], as well as the manipulation of gene circuits (Wong and Liao, 2006; Levine et al., 2013; Din et al., 2020), have paved the way for the design of efficient cell population control procedures. It is now possible to act either on cell population (Milias-Argeitis et al., 2016; Sassi et al., 2019; Nguyen et al., 2021) or on individual cells within population (Lugagne et al., 2017; Rullan et al., 2018) for directing gene expression and cellular functions. In this review article, we will focus more precisely on a generic approach that could be used to control gene expression in individual cells among population. A critical aspect that must be taken into account before being able to manipulate gene expression in cell population is related to the inherent noise of cellular systems (Pilpel, 2011).

This noise induces cell-to-cell variability in gene expression, and a potential control procedure must be designed by taking into account the inherent functionality exhibited by noise on the cellular system (Levine et al., 2013; Ackermann, 2015). Indeed, it is known that biological noise is a mechanism exploited by cell population in order to increase its fitness in front of fluctuating environmental conditions (Thattai and van Oudenaarden, 2004; Kussell and Leibler, 2005). As an example, in natural ecosystems, microbial populations are often exposed to unpredictable environmental changes such as nutrient starvation, exposure to antibiotics, temperature variations, and many other sources of stress (Ackermann, 2015) that can fluctuate periodically or randomly. Cellular systems have then evolved accordingly by adapting different cellular components in order to accommodate such fluctuations involving different timescales. If environmental conditions change slowly and regularly, a responsive switching strategy leads to increased fitness for the cell population (Kussell and Leibler, 2005). On the other hand, if environmental conditions are fast and erratic, a random switching mechanism, leading to preadapted cells, is more suited for optimizing population fitness. The study of phenotypic diversification mechanisms involved in antibiotic persistence in bacteria has pointed out that cellular systems can take benefit from both stochastic and responsive switching (Kussell et al., 2005). It is clear that, for designing an efficient population control procedure, stochastic switching must be minimized and responsive switching must be favored. Such responsive mechanisms typically involve gene circuits, able to record environmental changes and to respond accordingly. A spectacular realization of the inference of periodic environmental changes by gene circuits is the implementation of circadian (oscillation with a period of ~24 h; Voigt et al., 2016) or ultradian (oscillation with a period < 24 h; Isomura and Kageyama, 2014) rhythms by cellular systems. Many other gene circuit architectures or motifs are known to be able to infer extracellular signals and trigger appropriate biological responses (Perkins and Swain, 2009; Balazsi et al., 2011). Even if we have now access to a classification of the motifs and their possible dynamics (Shoval and Alon, 2010), it is still a challenge to infer the dynamics when several motifs are combined to each other or when the response interferes with many other cellular components. Indeed, in some case, gene circuit architectures can involve overlaps between different stress response pathways, allowing cells to anticipate environmental changes (Tagkopoulos et al., 2008; Freddolino and Tavazoie, 2012). This anticipatory switching arises in ecosystems where different environmental changes exhibit a strongly correlated time profile. As an example, *Escherichia coli* has evolved in order to be able to grow inside and outside a host (i.e., a mammals; Mitchell et al., 2009). When invading the host, *E. coli* is exposed to heat shock where temperature increases from 20 to 37°C. This heat shock is then followed by oxygen limitation as bacteria are reaching the gastrointestinal tract. The gene circuits involved in heat shock response and oxygen limitation have been found to share common inputs and outputs in *E. coli*, elevation of heat leading to the adaptation to oxygen limitation in order to anticipate correlated environmental changes.

Given all these elements, it is then difficult to infer the mode of switching, i.e., stochastic, responsive, or anticipatory (or a combination of them) based on the gene circuit architecture. Accordingly, we propose in this work, a generalizable approach aiming at stimulating the responsive component of switching for directing gene expression in cell population. Such approach could be made possible through the use of a universal metrics aiming at quantifying the information transfer efficiency in cells and leading to the design of robust cell–machine interfaces.

USING INFORMATION THEORY FOR DETERMINING THE OPTIMAL STIMULATION FREQUENCY LEADING TO COORDINATED GENE EXPRESSION IN CELL POPULATION

Cells are intrinsically programmed in order to react to external stimuli and to adapt appropriately by switching to different phenotypic states (Acar et al., 2008; Schreiber and Ackermann, 2020). It is then unrealistic to try to keep these cells into a specific phenotypic state, even if this would be a nice outcome for several applications, such as the optimization of cell factories for bioprocessing (Binder et al., 2017). Indeed, these phenotypic states are linked to specific environmental states through selection pressure and the resulting fitness advantage, environmental condition being under constant evolution (Thattai and van Oudenaarden, 2004). A more realistic alternative is to control cell switching itself, which is now technically feasible through the use of cell–machine interfaces (Delvigne et al., 2017; Sassi et al., 2019; Nguyen et al., 2021). In order to make this control strategy successful, two specific aspects must be taken into account, i.e., the efficiency in information transmission through the targeted gene circuits and the timing at which cells commit to phenotypic switching. These two aspects will be illustrated through a case study recently addressed, i.e., the synchronization for the activation of the gene circuit responsible for the induction of the arabinose operon in *E. coli* (Nguyen et al., 2021). The relevance of this case study is also justified by the fact that the arabinose operon has been long used as a biological case study for the characterization of the functionality of biological noise in cell population (Megerle et al., 2008) and also by the fact that the genes belonging to the arabinose operon are widely used for synthetic biology applications and notably for the synchronization of cell response (Stricker et al., 2008; Mondragón-Palomino et al., 2011). Finally, the arabinose operon is known to exhibit strong cell-to-cell variability both in the timing for activation (Megerle et al., 2008; Nikolic et al., 2017) and also the level of expression of the corresponding genes (Sagmeister et al., 2014), suggesting that the underlying cell switching mechanisms involves a mix of responsive and stochastic components. This make this system very interesting to be considered for possible coordination at the population level.

The first step in the cell-to-cell coordination for the activation of the arabinose operon is to know the possible effector for the underlying gene circuit. The activation of

the arabinose operon is under the control of a feedforward loop (Mangan et al., 2003; **Figure 1A**) combining the glucose depletion signal (through the accumulation of cAMP inside cells) and the presence of arabinose (through the activation of the transcription factor AraC). Under glucose-limiting conditions, it is then possible to activate or deactivate this gene circuit based on arabinose pulsing (Nguyen et al., 2021; **Figure 1A**). At this stage, the first drawback exhibited by biological noise can be observed. Indeed, upon arabinose pulsing, cells will commit to the activation of the feedforward circuit leading to the synthesis of the different proteins involved in arabinose assimilation. However, due to biological noise, timing in commitment will exhibit cell-to-cell heterogeneity (Yurkovsky and Nachman, 2013; Ghusinga et al., 2017). Timing in cellular commitment to alternative phenotypes depends on the accumulation of regulatory proteins at the single cell level. Transcription and translation processes in individual cells are prone to biological noise (Thattai and van Oudenaarden, 2001; Swain et al., 2002). These processes can be simulated based on the resolution of the chemical master equation or, more practically, based on the Gillespie algorithm (Thattai and van Oudenaarden, 2001). These simulations have been shown to lead to very realistic pictures for mRNA and protein synthesis in individual cells (Balazsi et al., 2011) and pointed out that these processes follow Poisson statistics. Accordingly, the transition of cells between two adjacent phenotypic states (for example, the two states, GFP negative and GFP positive, drawn in **Figure 1A**) can also be represented by a Poisson process. One key property of the Poisson process is that the timing between two consecutive events (e.g., the time between the synthesis of two mRNAs from the same DNA sequence in a single cell) follows an exponential distribution. Based on this statement, the residence time distribution of cells in a given phenotypic state can be represented by an exponential distribution (Norman et al., 2015). It is thus very critical to take into account this residence time distribution for coordinating cell switching at the population level. One way to overcome this use is to rely on the use of a cell–machine interface allowing the on-line monitoring of the switching process at the level of individual cells in the population and to react accordingly. This principle has been notably adopted for developing the segregostat (Sassi et al., 2019; Nguyen et al., 2021). This system is based on the use of on-line flow cytometry for recoding the cell switching rate and to trigger environmental switching accordingly. The fact that the frequency of environmental perturbation must be set based on the phenotypic switching frequency has been previously deduced from numerical simulation (Thattai and van Oudenaarden, 2004). In a similar way, another study has pointed out that the control of gene circuits is dramatically reduced above a critical stimulation frequency (Tan et al., 2007). Under these conditions, the frequency of the extracellular signal is effectively transmitted, leading to a cell population with synchronized gene expression (**Figure 1B**). Effective entrainment of cell population can be assessed based on the oscillatory gene expression profile exhibiting a frequency close to the one of the input stimulations. Such oscillations were experimentally observed during segregostat experiments carried out for controlling the activation of the arabinose operon (Nguyen et al., 2021). It is also important to point out that in this case, square waves are used as stimulatory input. This strategy is also called pulse width modulation (PWM; Davidson et al., 2013; Purvis and Lahav, 2013). We will see in the next section that this strategy has been used several times for controlling different cellular systems (listed in **Table 1**). In the present case, population oscillates according to a frequency corresponding to the one of the input square waves. This is represented in **Figure 1B** based on the period T_{env} ($T_{env} = 1/$ frequency) of the input square wave stimulation, which is transmitted to the population and lead to oscillation in gene expression with the same period $T_{switch} = T_{env}$.

All these observations point out that cells are able to deduce changes in their surroundings based on diverse sensory mechanisms. We do not want here to discuss about the biological diversity of these mechanisms, but rather to quantify the efficiency at which a cell is able to infer extracellular perturbation. A universal way to quantify information transmission through biochemical network can be derived from Shannon theory or information theory (Cheong et al., 2011). In order to be able to understand the importance of information theory in biochemical signal processing, it is important to introduce the concept of input–output (i/o) or dose–response relationship. For many gene circuits, this i/o relationship can be represented by a sigmoidal curve (**Figure 1C**), also called Hill equation (Levchenko and Nemenman, 2014). For example, Hill equation can be used to infer the response of the feedforward loop involved in the regulation of the arabinose operon (Mangan and Alon, 2003). This i/o correlation tells us what will be the output of the gene circuit according to a given input. However, we have seen that cell switching mechanism involves a random component in addition to the responsive one. This random component can be represented by the error bars on the i/o correlation (**Figure 1D**). Accordingly, one input can drive different output trajectories, leading to cell-to-cell heterogeneity. It can be seen that some input leads to a very heterogeneous response, making cell unable to properly infer the state of the environment. This is exactly where information theory can be useful, i.e., by providing a metric for quantifying the amount of information transmitted by the gene network for some specific input environmental conditions. This metric, mutual information (MI), corresponds to the logarithm of the number of distinct, input-dependent, states that can be reached by cells (Levchenko and Nemenman, 2014) and is quantified in bits. For example, a gene network exhibiting a MI of one bit means that only two physiologically distinct states can be resolved by cells based on the input conditions. Generally speaking, most of the gene circuits are corrupted by noise and can carry only a limited amount of information, and most of the studies carried out so far in this area have pointed out that MI equals to only 1–2 bits for different gene networks and model organisms (Mehta et al., 2009; Perkins and Swain, 2009; Cheong et al., 2011; Hansen and O'Shea, 2015; Sarkar et al., 2020). This leads to the conclusion that only these

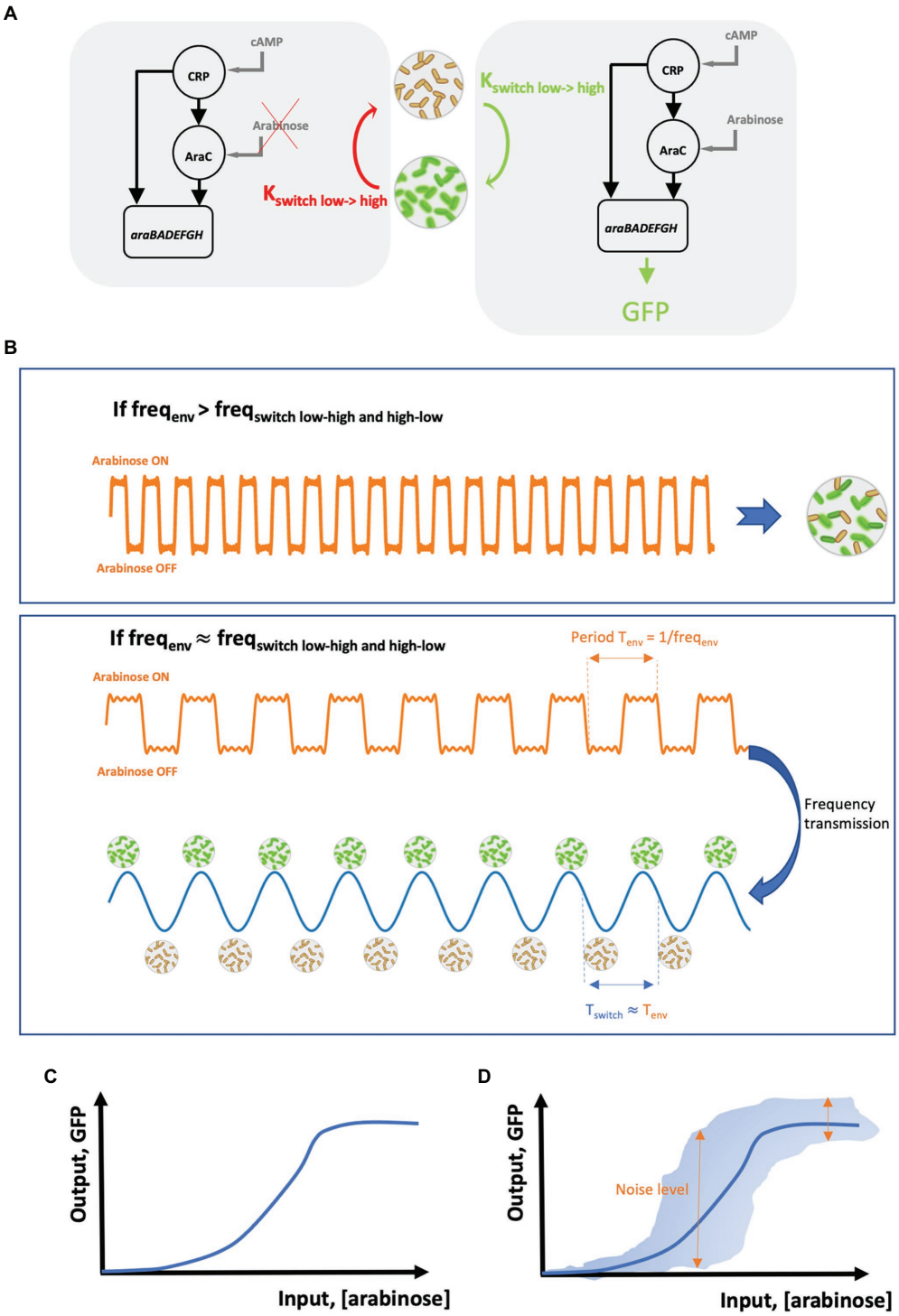


FIGURE 1 | (A) Scheme of the feedforward loop motif involved in the regulation of the arabinose operon. On the left, arabinose is not available and the AraC branch cannot be induced. Accordingly, cell switching does not take place and, eventually, previously induced cells are relaxed back to the uninduced (low) state at a rate $K_{switch\ low->high}$. On the right, arabinose is available and the AraC branch, together with the cAMP-CRP branch, is activated leading to the induction of the genes *araBADEFGH* involved in arabinose metabolism. Under these conditions, cells from the low-state switch actively to the high state at a rate $K_{switch\ high->low}$. **(B)** Proper coordination/synchronization of gene expression can be achieved based on periodic stimulations (or environmental fluctuations) made at a specific frequency $freq_{env}$. If $freq_{env}$ is too high by comparison with the frequency for cell switching $freq_{switch}$, then cells are not coordinated and exhibit strong variability in gene expression. However, when $freq_{env}$ is set close to $freq_{switch}$, coordination in gene expression is possible leading to synchronized gene expression. **(C)** Typical shape of a Hill relationship between an input (here the concentration of arabinose in the medium) and its resulting output (here, detected based on the synthesis of GFP based on a P_{araBAD}::GFP transcriptional reporter). **(D)** Impact of biological noise (represented by double arrows) on the probability for delivering an output based on a given input.

TABLE 1 | Range of environmental stimuli used for controlling gene expression in cell population and range of periodic signal associated with these environmental perturbations.

Approximated period T for the input stimulation[A]	Type of perturbation[B]	Controlled trait	Organism	Culture system	Single-cell analysis tool	References
83 min*	Pulse of phosphate in phosphate-poor medium	Cell cycle	*E. coli*	Bioreactor 240 ml	None	Goodwin, 1969
98 min*	Pulses of methionine to induce (*MET3*pr-*CLN2*)	Cell cycle	*S. cerevisiae*	Microfluidics	Microscopy	Charvin et al., 2009
110 min* 150 min*	Nutrient availability, i.e., poor-rich	Cell cycle	*S. cerevisiae*	Microfluidics	Microscopy	Tian et al., 2012
16 min 45 min	Red-far red pulses of light	*gal1*-responsive genes through optogenetic	*S. cerevisiae*	Microplates with 96 wells	Flow cytometry	Milias-Argeitis et al., 2011, 2016
28 min 88 min	Pulses of sorbitol-enriched medium in normal medium	Osmostress (value of 1,500 a.u.)	*S. cerevisiae*	Microfluidics	Microscopy	Uhlendorf et al., 2012
142 min	Glucose–galactose	p*gal1*-GFP	*S. cerevisiae*	Microfluidics	Microscopy	Fiore et al., 2016
~990 min	Tetracycline	Tetracycline-inducible system p(CMV-TET)-d2EYFP	Mammalian (CHO)	Microfluidics	Microscopy	Fracassi et al., 2016
120 min	Green light intensity (in relation with red light intensity)	CcaS/CcaR gene expression system at a varying level of expression	*E. coli*	Bioreactor 20 ml	Automated flow cytometry	Milias-Argeitis et al., 2016
6 min	Computed open-loop green–red light	ccaSR-based system	*E. coli*	Microfluidics	Microscopy	Chait et al., 2017
150 min 210 min (depending on the amplitude/ concentration)	IPTG—aTc	Maintain a toggle switch at an unstable intermediary level	*E. coli*	Microfluidics	Microscopy	Lugagne et al., 2017
10 min	High and low red intensity blue light	Transcription	*S. cerevisiae*	Microfluidics	Microscopy	Rullan et al., 2018
For GFP: –5 s off/80 s on –8 s off/80 s on –11 s off/80 s on For isobutanol and butanol: –65 s off and 15 s on	Blue light	-GFP expression -production of isobutanol and 2-methyl-1-butanol	*S. cerevisiae*	For GFP: Microplates 24 wells For isobutanol and butanol: Bioreactor 0.5 L	None	Zhao et al., 2018
5–60 min	Pulses of galactose to stabilize synecluine concentration at different values	a-synuclein formation	*S. cerevisiae*	Microfluidics	Microscopy	Perrino et al., 2019
90 min 105 min 180 min	Methionine concentration in cultivation medium	Yeast cell cycle coordination	*S. cerevisiae*	Microfluidics	Microscopy	Perrino et al., 2021
600 min	Arabinose pulses	Arabinose operon induction	*E. coli*	Bioreactor 1 L (continuous mode)	Online flow-cytometry	Nguyen et al., 2021

*[A]This column indicates the range of periodic signal associated with these environmental perturbations. In most of the case, the input signal can be approximated by a square wave of period T and frequency 1/T (see **Figure 1B** for more details).*

[B]This column indicates the range of environmental stimuli used for controlling gene expression.

**These stimulation periods have been determined without monitoring and feedback control (open-loop control).*

states have to be targeted when designing a cell population control strategy.

The next section will be dedicated to the description of some realization in the field of cell population control (also termed cybergenetics), pointing out that the above-mentioned methodology could help at this level by providing a general framework aiming at developing further cell population control procedures.

THE CONTRIBUTION OF CONTROL THEORY AND THE NASCENT FIELD OF CYBERGENETICS

The fact that cell population can be controlled based on pulsatile inputs has been reported a long time ago. Indeed, long before the advent of single-cell technologies, Goodwin (1969) observed that it was possible to synchronize cell

cycle in *E. coli* cells by periodically pulsing phosphate in a phosphate-limited chemostat. This pioneering work has led to the establishment of a robust modeling framework for the understanding of the impact of external conditions on the synchronization of cell cycle for many types of organisms (Ruoff et al., 2001; Gonze and Abou-Jaoudé, 2013; Gonze and Ruoff, 2021). However, these studies have been carried out based on an open-loop control approach and the application of regular pulses with varying frequencies and amplitudes. More recently, the application of control theory to the manipulation of cellular systems, i.e., cybergenetics, has set the ground for a more rational design of cell population control procedures (Milias-Argeitis et al., 2016; Banderas et al., 2020). Cybergenetics is an entirely new and exciting field of research at the interface between control engineering and synthetic biology, and emerged with the recent advances made in genetic engineering combined with the works initially derived from cybernetics (Wiener, 1961). A distinction can be made between "internal cybergenetics" (also called *in vivo* and involving genetic controllers directly embedded in cells) and "external cybergenetics" (also called *in silico* controllers; Lugagne et al., 2017; Lugagne and Dunlop, 2019; Carrasco-López et al., 2020; Pedone et al., 2021). In the context of this review, we will be focused more on the latter technology, since it involves cell–machine interface and pulsatile inputs used as actuators.

Remarkably, although different systems have been used (e.g., various model organisms, type of gene circuits to be controlled, and single cell techniques), all the data accumulated point out that it is possible to effectively control gene circuits at the level of individual cells by applying external periodic signals (**Table 1**). Evidences have been provided suggesting that pulses of inducers tend to decrease noise in biochemical network, leading to synchronized gene expression (Uhlendorf et al., 2012; Benzinger and Khammash, 2018). This effect can be explained based on the dose–response relationship (**Figure 1D**) where input concentrations at the extremities of the dynamic range lead to a homogenous response at the population level. In contrast, input concentrations at the center of the dynamic range produce a heterogeneous population. This strategy, known as PWM, seems to be generalizable for the effective control of diverse gene circuits in diverse cellular systems. Most of the experiments involving the control of gene expression in cellular systems have been performed in microfluidic devices (**Table 1**). This type of cultivation device allows the acquisition of single cell data with a high spatiotemporal resolution, but with a low experimental throughput due to the time and computational power required for image analysis (Dusny and Schmid, 2014) and with possible technical biases by comparison with conventional cultivation devices (Dusny et al., 2015; Westerwalbesloh et al., 2017). Nonetheless, there is a growing interest in using standard cultivation devices (e.g., flasks, bioreactors, etc.) for studying and controlling cell populations (Zhao et al., 2021). In this case, single-cell analyses can be performed based on automated flow cytometry, leading to the rapid accumulation of data at the population level. In this context, the use of cell–machine interface relying on flow cytometry can lead to the automated determination systematic determination of the optimal stimulation frequency for the effective synchronization of gene expression at the population level (Nguyen et al., 2021).

PERSPECTIVE: EXPLOITING INTRINSIC FREQUENCY OF GENE CIRCUITS

Taken altogether, the elements assembled in the previous sections point out that a lot of different gene circuits architectures can exhibit periodic behavior [and not only the motifs reported to behave as natural oscillators, such as the repressilator (Elowitz and Leibler, 2000) or the oscillator motif (Stricker et al., 2008; Mondragón-Palomino et al., 2011)] if stimulated at the appropriate frequencies (Tan et al., 2007). Development made in information theory and in cybergenetics provides the computational framework and the experimental tools in order to generalize this concept to many biological systems. Impressive achievements can be expected from these field of research such as the control of complex cell regulatory program (e.g., control of cell cycle program; Perrino et al., 2021) and the control of microbial communities composition (Fiore et al., 2017; Liao et al., 2019), with applications in various field from bioproduction (Briat and Khammash, 2018; Zhao et al., 2021) to biomedicine (Davidson et al., 2013; Din et al., 2020).

AUTHOR CONTRIBUTIONS

FD drafted the manuscript, designed **Figure 1**, and wrote section "Using information theory for determining the optimal stimulation frequency leading to coordinated gene expression in cell population." FB and FM-A wrote section "Introduction." LH and MD wrote sections "The contribution of control theory and the nascent field of cybergenetics" and "Perspective: exploiting intrinsic frequency of gene circuits" and designed **Table 1**. FB, FM-A, LH, and MD revised the final version of the manuscript. All authors contributed to the article and approved the submitted version.

FUNDING

LH is supported by the FRS-FNRS (Fond National pour la Recherche Scientifique, Belgium) through a FRIA PhD grant. MD is supported by a FNRS PhD grant in the context of an Era-Net Aquatic Pollutant project (ARENA).

ACKNOWLEDGMENTS

The Walloon competitive clusters Wagralim and Greenwin (funded through DGO6 of Wallonia, Belgium) are acknowledged for their support through the Sunup and Elithe projects.

REFERENCES

Acar, M., Mettetal, J. T., and van Oudenaarden, A. (2008). Stochastic switching as a survival strategy in fluctuating environments. *Nat. Genet.* 40, 471–475. doi: 10.1038/ng.110

Ackermann, M. (2015). A functional perspective on phenotypic heterogeneity in microorganisms. *Nat. Rev. Microbiol.* 13, 497–508. doi: 10.1038/nrmicro3491

Balazsi, G., van Oudenaarden, A., and Collins, J. J. (2011). Cellular decision making and biological noise: from microbes to mammals. *Cell* 144, 910–925. doi: 10.1016/j.cell.2011.01.030

Banderas, A., Le Bec, M., Cordier, C., and Hersen, P. (2020). Autonomous and assisted control for synthetic microbiology. *Int. J. Mol. Sci.* 21:9223. doi: 10.3390/ijms21239223

Benzinger, D., and Khammash, M. (2018). Pulsatile inputs achieve tunable attenuation of gene expression variability and graded multi-gene regulation. *Nat. Commun.* 9:3521. doi: 10.1038/s41467-018-05882-2

Binder, D., Drepper, T., Jaeger, K.-E., Delvigne, F., Wiechert, W., Kohlheyer, D., et al. (2017). Homogenizing bacterial cell factories: analysis and engineering of phenotypic heterogeneity. *Metab. Eng.* 42, 145–156. doi: 10.1016/j.ymben.2017.06.009

Briat, C., and Khammash, M. (2018). Perfect adaptation and optimal equilibrium productivity in a simple microbial biofuel metabolic pathway using dynamic integral control. *ACS Synth. Biol.* 7, 419–431. doi: 10.1021/acssynbio.7b00188

Carrasco-López, C., García-Echauri, S. A., Kichuk, T., and Avalos, J. L. (2020). Optogenetics and biosensors set the stage for metabolic cybergenetics. *Curr. Opin. Biotechnol.* 65, 296–309. doi: 10.1016/j.copbio.2020.07.012

Chait, R., Ruess, J., Bergmiller, T., Tkacik, G., and Guet, C. C. (2017). Shaping bacterial population behavior through computer-interfaced control of individual cells. *Nat. Commun.* 8:1535. doi: 10.1038/s41467-017-01683-1

Charvin, G., Cross, F. R., and Siggia, E. D. (2009). Forced periodic expression of G1 cyclins phase-locks the budding yeast cell cycle. *Proc. Natl. Acad. Sci. U. S. A.* 106, 6632–6637. doi: 10.1073/pnas.0809227106

Cheong, R., Rhee, A., Wang, C. J., Nemenman, I., and Levchenko, A. (2011). Information transduction capacity of noisy biochemical signaling networks. *Science* 334, 354–358. doi: 10.1126/science.1204553

Davidson, E. A., Basu, A. S., and Bayer, T. S. (2013). Programming microbes using pulse width modulation of optical signals. *J. Mol. Biol.* 425, 4161–4166. doi: 10.1016/j.jmb.2013.07.036

Delvigne, F., Baert, J., Sassi, H., Fickers, P., Grünberger, A., and Dusny, C. (2017). Taking control over microbial populations: current approaches for exploiting biological noise in bioprocesses. *Biotechnol. J.* 12:549. doi: 10.1002/biot.201600549

Din, M. O., Martin, A., Razinkov, I., Csicsery, N., and Hasty, J. (2020). Interfacing gene circuits with microelectronics through engineered population dynamics. *Sci. Adv.* 6:eaaz8344. doi: 10.1126/sciadv.aaz8344

Dusny, C., Grunberger, A., Probst, C., Wiechert, W., Kohlheyer, D., and Schmid, A. (2015). Technical bias of microcultivation environments on single-cell physiology. *Lab Chip* 15, 1822–1834. doi: 10.1039/C4LC01270D

Dusny, C., and Schmid, A. (2014). Microfluidic single-cell analysis links boundary environments and individual microbial phenotypes. *Environ. Microbiol.* 17, 1839–1856. doi: 10.1111/1462-2920.12667

Elowitz, M. B., and Leibler, S. (2000). A synthetic oscillatory network of transcriptional regulators. *Nature* 403, 335–338. doi: 10.1038/35002125

Fiore, G., Matyjaszkiewicz, A., Annunziata, F., Grierson, C., Savery, N. J., Marucci, L., et al. (2017). In-silico analysis and implementation of a multicellular feedback control strategy in a synthetic bacterial consortium. *ACS Synth. Biol.* 6, 507–517. doi: 10.1021/acssynbio.6b00220

Fiore, G., Perrino, G., di Bernardo, M., and di Bernardo, D. (2016). In vivo real-time control of gene expression: a comparative analysis of feedback control strategies in yeast. *ACS Synth. Biol.* 5, 154–162. doi: 10.1021/acssynbio.5b00135

Fracassi, C., Postiglione, L., Fiore, G., and di Bernardo, D. (2016). Automatic control of gene expression in mammalian cells. *ACS Synth. Biol.* 5, 296–302. doi: 10.1021/acssynbio.5b00141

Freddolino, P. L., and Tavazoie, S. (2012). Beyond homeostasis: a predictive-dynamic framework for understanding cellular behavior. *Annu. Rev. Cell Dev. Biol.* 28, 363–384. doi: 10.1146/annurev-cellbio-092910-154129

Ghusinga, K. R., Dennehy, J. J., and Singh, A. (2017). First-passage time approach to controlling noise in the timing of intracellular events. *Proc. Natl. Acad. Sci. U. S. A.* 114, 693–698. doi: 10.1073/pnas.1609012114

Gonze, D., and Abou-Jaoudé, W. (2013). The goodwin model: behind the hill function. *PLoS One* 8:e69573. doi: 10.1371/journal.pone.0069573

Gonze, D., and Ruoff, P. (2021). The goodwin oscillator and its legacy. *Acta Biotheor.* 69, 857–874. doi: 10.1007/s10441-020-09379-8

Goodwin, B. C. (1969). Synchronization of Escherichia coli in a chemostat by periodic phosphate feeding. *Eur. J. Biochem.* 10, 511–514. doi: 10.1111/j.1432-1033.1969.tb00718.x

Grunberger, A., Wiechert, W., and Kohlheyer, D. (2014). Single-cell microfluidics: opportunity for bioprocess development. *Curr. Opin. Biotechnol.* 29, 15–23. doi: 10.1016/j.copbio.2014.02.008

Hansen, A. S., and O'Shea, E. K. (2015). Limits on information transduction through amplitude and frequency regulation of transcription factor activity. *Elife* 4:e06559. doi: 10.7554/eLife.06559

Isomura, A., and Kageyama, R. (2014). Ultradian oscillations and pulses: coordinating cellular responses and cell fate decisions. *Development* 141, 3627–3636. doi: 10.1242/dev.104497

Kussell, E., Kishony, R., Balaban, N. Q., and Leibler, S. (2005). Bacterial persistence: a model of survival in changing environments. *Genetics* 169, 1807–1814. doi: 10.1534/genetics.104.035352

Kussell, E., and Leibler, S. (2005). Phenotypic diversity, population growth, and information in fluctuating environments. *Science* 309, 2075–2078. doi: 10.1126/science.1114383

Levchenko, A., and Nemenman, I. (2014). Cellular noise and information transmission. *Curr. Opin. Biotechnol.* 28, 156–164. doi: 10.1016/j.copbio.2014.05.002

Levine, J. H., Lin, Y., and Elowitz, M. B. (2013). Functional roles of pulsing in genetic circuits. *Science* 342, 1193–1200. doi: 10.1126/science.1239999

Liao, M. J., Din, M. O., Tsimring, L., and Hasty, J. (2019). Rock-paper-scissors: engineered population dynamics increase genetic stability. *Science* 365, 1045–1049. doi: 10.1126/science.aaw0542

Lugagne, J.-B., Carrillo, S. S., Kirch, M., Kohler, A., Batt, G., and Hersen, P. (2017). Balancing a genetic toggle switch by real-time feedback control and periodic forcing. *Nat. Commun.* 8:1671. doi: 10.1038/s41467-017-01498-0

Lugagne, J.-B., and Dunlop, M. J. (2019). Cell-machine interfaces for characterizing gene regulatory network dynamics. *Curr. Opin. Syst. Biol.* 14, 1–8. doi: 10.1016/j.coisb.2019.01.001

Mangan, S., and Alon, U. (2003). Structure and function of the feed-forward loop network motif. *Proc. Natl. Acad. Sci. U. S. A.* 100, 11980–11985. doi: 10.1073/pnas.2133841100

Mangan, S., Zaslaver, A., and Alon, U. (2003). The coherent feedforward loop serves as a sign-sensitive delay element in transcription networks. *J. Mol. Biol.* 334, 197–204. doi: 10.1016/j.jmb.2003.09.049

Megerle, J. A., Fritz, G., Gerland, U., Jung, K., and Rädler, J. O. (2008). Timing and dynamics of single cell gene expression in the arabinose utilization system. *Biophys. J.* 95, 2103–2115. doi: 10.1529/biophysj.107.127191

Mehta, P., Goyal, S., Long, T., Bassler, B. L., and Wingreen, N. S. (2009). Information processing and signal integration in bacterial quorum sensing. *Mol. Syst. Biol.* 5:325. doi: 10.1038/msb.2009.79

Milias-Argeitis, A., Rullan, M., Aoki, S. K., Buchmann, P., and Khammash, M. (2016). Automated optogenetic feedback control for precise and robust regulation of gene expression and cell growth. *Nat. Commun.* 7:12546. doi: 10.1038/ncomms12546

Milias-Argeitis, A., Summers, S., Stewart-Ornstein, J., Zuleta, I., Pincus, D., El-Samad, H., et al. (2011). In silico feedback for in vivo regulation of a gene expression circuit. *Nat. Biotechnol.* 29, 1114–1116. doi: 10.1038/nbt.2018

Mitchell, A., Romano, G. H., Groisman, B., Yona, A., Dekel, E., Kupiec, M., et al. (2009). Adaptive prediction of environmental changes by microorganisms. *Nature* 460, 220–224. doi: 10.1038/nature08112

Mondragón-Palomino, O., Danino, T., Selimkhanov, J., Tsimring, L., and Hasty, J. (2011). Entrainment of a population of synthetic genetic oscillators. *Science* 333, 1315–1319. doi: 10.1126/science.1205369

Nguyen, T. M., Telek, S., Zicler, A., Martinez, J. A., Zacchetti, B., Kopp, J., et al. (2021). Reducing phenotypic instabilities of a microbial population during continuous cultivation based on cell switching dynamics. *Biotechnol. Bioeng.* 118, 3847–3859. doi: 10.1002/bit.27860

Nikolic, N., Schreiber, F., Dal Co, A., Kiviet, D. J., Bergmiller, T., Littmann, S., et al. (2017). Cell-to-cell variation and specialization in sugar metabolism in clonal bacterial populations. *PLoS Genet.* 13:e1007122. doi: 10.1371/journal.pgen.1007122

Norman, T. M., Lord, N. D., Paulsson, J., and Losick, R. (2015). Stochastic switching of cell fate in microbes. *Annu. Rev. Microbiol.* 69, 381–403. doi: 10.1146/annurev-micro-091213-112852

Pedone, E., de Cesare, I., Zamora-Chimal, C. G., Haener, D., Postiglione, L., La Regina, A., et al. (2021). Cheetah: a computational toolkit for cybergenetic control. *ACS Synth. Biol.* 10, 979–989. doi: 10.1021/acssynbio.0c00463

Perkins, T. J., and Swain, P. S. (2009). Strategies for cellular decision-making. *Mol. Syst. Biol.* 5:326. doi: 10.1038/msb.2009.83

Perrino, G., Napolitano, S., Galdi, F., La Regina, A., Fiore, D., Giuliano, T., et al. (2021). Automatic synchronisation of the cell cycle in budding yeast through closed-loop feedback control. *Nat. Commun.* 12:2452. doi: 10.1038/s41467-021-22689-w

Perrino, G., Wilson, C., Santorelli, M., and di Bernardo, D. (2019). Quantitative characterization of α-synuclein aggregation in living cells through automated microfluidics feedback control. *Cell Rep.* 27, 916.e5–927.e5. doi: 10.1016/j.celrep.2019.03.081

Pilpel, Y. (2011). Noise in biological systems: pros, cons, and mechanisms of control. *Methods Mol. Biol.* 759, 407–425. doi: 10.1007/978-1-61779-173-4_23

Purvis, J. E., and Lahav, G. (2013). Encoding and decoding cellular information through signaling dynamics. *Cell* 152, 945–956. doi: 10.1016/j.cell.2013.02.005

Rullan, M., Benzinger, D., Schmidt, G. W., Milias-Argeitis, A., and Khammash, M. (2018). An optogenetic platform for real-time, single-cell interrogation of stochastic transcriptional regulation. *Mol. Cell* 70, 745.e6–756.e6. doi: 10.1016/j.molcel.2018.04.012

Ruoff, P., Vinsjevik, M., Monnerjahn, C., and Rensing, L. (2001). The Goodwin model: simulating the effect of light pulses on the circadian sporulation rhythm of Neurospora crassa. *J. Theor. Biol.* 209, 29–42. doi: 10.1006/jtbi.2000.2239

Sagmeister, P., Schimek, C., Meitz, A., Herwig, C., and Spadiut, O. (2014). Tunable recombinant protein expression with *E. coli* in a mixed-feed environment. *Appl. Microbiol. Biotechnol.* 98, 2937–2945. doi: 10.1007/s00253-013-5445-1

Sarkar, S., Tack, D., and Ross, D. (2020). Sparse estimation of mutual information landscapes quantifies information transmission through cellular biochemical reaction networks. *Commun. Biol.* 3:203. doi: 10.1038/s42003-020-0901-9

Sassi, H., Nguyen, T. M., Telek, S., Gosset, G., Grünberger, A., and Delvigne, F. (2019). Segregostat: a novel concept to control phenotypic diversification dynamics on the example of gram-negative bacteria. *Microb. Biotechnol.* 12, 1064–1075. doi: 10.1111/1751-7915.13442

Schreiber, F., and Ackermann, M. (2020). Environmental drivers of metabolic heterogeneity in clonal microbial populations. *Curr. Opin. Biotechnol.* 62, 202–211. doi: 10.1016/j.copbio.2019.11.018

Shoval, O., and Alon, U. (2010). Snap shot: network motifs. *Cell* 143, 326–326.e1. doi: 10.1016/j.cell.2010.09.050

Stricker, J., Cookson, S., Bennett, M. R., Mather, W. H., Tsimring, L. S., and Hasty, J. (2008). A fast, robust and tunable synthetic gene oscillator. *Nature* 456, 516–519. doi: 10.1038/nature07389

Swain, P. S., Elowitz, M. B., and Siggia, E. D. (2002). Intrinsic and extrinsic contributions to stochasticity in gene expression. *Proc. Natl. Acad. Sci. U. S. A.* 99, 12795–12800. doi: 10.1073/pnas.162041399

Tagkopoulos, I., Liu, Y.-C., and Tavazoie, S. (2008). Predictive behavior within microbial genetic networks. *Science* 320, 1313–1317. doi: 10.1126/science.1154456

Tan, C., Reza, F., and You, L. (2007). Noise-limited frequency signal transmission in gene circuits. *Biophys. J.* 93, 3753–3761. doi: 10.1529/biophysj.107.110403

Thattai, M., and van Oudenaarden, A. (2001). Intrinsic noise in gene regulatory networks. *Proc. Natl. Acad. Sci. U. S. A.* 98, 8614–8619. doi: 10.1073/pnas.151588598

Thattai, M., and van Oudenaarden, A. (2004). Stochastic gene expression in fluctuating environments. *Genetics* 167, 523–530. doi: 10.1534/genetics.167.1.523

Tian, Y., Luo, C., Lu, Y., Tang, C., and Ouyang, Q. (2012). Cell cycle synchronization by nutrient modulation. *Integr. Biol.* 4, 328–334. doi: 10.1039/c2ib00083k

Uhlendorf, J., Miermont, A., Delaveau, T., Charvin, G., Fages, F., Bottani, S., et al. (2012). Long-term model predictive control of gene expression at the population and single-cell levels. *Proc. Natl. Acad. Sci. U. S. A.* 109, 14271–14276. doi: 10.1073/pnas.1206810109

Voigt, R. M., Forsyth, C. B., Green, S. J., Engen, P. A., and Keshavarzian, A. (2016). Circadian rhythm and the gut microbiome. *Int. Rev. Neurobiol.* 131, 193–205. doi: 10.1016/bs.irn.2016.07.002

Westerwalbesloh, C., Grünberger, A., Wiechert, W., Kohlheyer, D., and von Lieres, E. (2017). Coarse-graining bacteria colonies for modelling critical solute distributions in picolitre bioreactors for bacterial studies on single-cell level. *Microb. Biotechnol.* 10, 845–857. doi: 10.1111/1751-7915.12708

Wiener, N. (1961). *Cybernetics: Or Control and Communication in the Animal and the Machine. 2nd Edn.* Cambridge, MA, USA: MIT Press.

Wong, W. W., and Liao, J. C. (2006). The design of intracellular oscillators that interact with metabolism. *Cell. Mol. Life Sci.* 63, 1215–1220. doi: 10.1007/s00018-005-5611-4

Yurkovsky, E., and Nachman, I. (2013). Event timing at the single-cell level. *Brief Funct. Genomics* 12, 90–98. doi: 10.1093/bfgp/els057

Zhao, E. M., Lalwani, M. A., Chen, J.-M., Orillac, P., Toettcher, J. E., and Avalos, J. L. (2021). Optogenetic amplification circuits for light-induced metabolic control. *ACS Synth. Biol.* 10, 1143–1154. doi: 10.1021/acssynbio.0c00642

Zhao, E. M., Zhang, Y., Mehl, J., Park, H., Lalwani, M. A., and Toettcher, J. E., et al. (2018). Optogenetic regulation of engineered cellular metabolism for microbial chemical production. *Nature* 555, 683–687. doi: 10.1038/nature26141

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

frontiers | Frontiers in Microbiology

REVIEW
published: 25 April 2022
doi: 10.3389/fmicb.2022.867342

Overview of Diverse Methyl/Alkyl-Coenzyme M Reductases and Considerations for Their Potential Heterologous Expression

*Aleksei Gendron and Kylie D. Allen**

Department of Biochemistry, Virginia Polytechnic Institute and State University, Blacksburg, VA, United States

Methyl-coenzyme M reductase (MCR) is an archaeal enzyme that catalyzes the final step of methanogenesis and the first step in the anaerobic oxidation of methane, the energy metabolisms of methanogens and anaerobic methanotrophs (ANME), respectively. Variants of MCR, known as alkyl-coenzyme M reductases, are involved in the anaerobic oxidation of short-chain alkanes including ethane, propane, and butane as well as the catabolism of long-chain alkanes from oil reservoirs. MCR is a dimer of heterotrimers (encoded by *mcrABG*) and requires the nickel-containing tetrapyrrole prosthetic group known as coenzyme F_{430}. MCR houses a series of unusual post-translational modifications within its active site whose identities vary depending on the organism and whose functions remain unclear. Methanogenic MCRs are encoded in a highly conserved *mcrBDCGA* gene cluster, which encodes two accessory proteins, McrD and McrC, that are believed to be involved in the assembly and activation of MCR, respectively. The requirement of a unique and complex coenzyme, various unusual post-translational modifications, and many remaining questions surrounding assembly and activation of MCR largely limit *in vitro* experiments to native enzymes with recombinant methods only recently appearing. Production of MCRs in a heterologous host is an important step toward developing optimized biocatalytic systems for methane production as well as for bioconversion of methane and other alkanes into value-added compounds. This review will first summarize MCR catalysis and structure, followed by a discussion of advances and challenges related to the production of diverse MCRs in a heterologous host.

Keywords: methyl-coenzyme M reductase, MCR, methanogens, anaerobic methanotrophic archaea, ANME

OPEN ACCESS

Edited by:
Cornelia Welte, Radboud University Nijmegen, Netherlands

Reviewed by:
Tristan Wagner, Max Planck Society, Germany
Silvan Scheller, Aalto University, Finland
Anjali Patwardhan, University of Michigan, United States

***Correspondence:**
Kylie D. Allen
kdallen@vt.edu

Specialty section:
This article was submitted to Microbial Physiology and Metabolism, a section of the journal Frontiers in Microbiology

Received: *01 February 2022*
Accepted: *01 April 2022*
Published: *25 April 2022*

Citation:
Gendron A and Allen KD (2022) Overview of Diverse Methyl/Alkyl-Coenzyme M Reductases and Considerations for Their Potential Heterologous Expression. Front. Microbiol. 13:867342. doi: 10.3389/fmicb.2022.867342

INTRODUCTION

Methyl-coenzyme M reductase (MCR) catalyzes the final methane-forming step of methanogenesis in methanogens, and the initial methane activation step in the anaerobic oxidation of methane (AOM) in anaerobic methanotrophic archaea (ANME). MCR is generally highly conserved in sequence and structure among methanogens and ANME, where it consists of three different subunits, α (McrA), β (McrB), and γ (McrG), arranged in a $\alpha_2\beta_2\gamma_2$ configuration harboring two active sites (Ermler et al., 1997; Shima et al., 2012; Wagner et al., 2017). Each active site

contains F_{430}, the nickel hydrocorphin prosthetic group (**Figure 1**). Methanogenic MCR has been extensively studied in the methane formation direction, where it catalyzes the conversion of methyl-coenzyme M (CH_3-S-CoM) and coenzyme B (HS-CoB) to methane and a CoM-S-S-CoB heterodisulfide (Bobik et al., 1987; Ellermann et al., 1987, 1988; **Figure 1**). This reaction is proposed to occur in reverse in ANME that anaerobically oxidize methane to CO_2 *via* reverse methanogenesis (Hallam et al., 2004; Shima and Thauer, 2005; Scheller et al., 2010; Timmers et al., 2017). In addition to methane formation and oxidation, variants of MCR are also involved in the anaerobic oxidation of short- and long-chain alkanes (Borrel et al., 2019; Wang et al., 2021; Zhou et al., 2022).

Given the remarkable chemistry catalyzed by MCR as well as its central importance in the global carbon cycle and potential for bioenergy applications, this enzyme has been of interest to enzymologists since its initial discovery in the 1970s (McBride and Wolfe, 1971; Gunsalus and Wolfe, 1976). Much of what is known about MCR catalysis comes from work by several groups on the natively purified MCR from the methanogen, *Methanothermobacter marburgensis*. This organism grows to high cell densities (3 g dry mass per L) with a doubling time of less than 2 h (Kaster et al., 2011) and, most importantly, effective procedures have been developed for the isolation of an active MCR from this organism.

A major challenge in the MCR field is the heterologous production of recombinant MCRs. This is due to many reasons including, but not limited to, the heterooligomeric structure of MCR that may require chaperones for proper assembly, the requirement of the unique and complex coenzyme F_{430}, the presence of several unusual post-translational modifications that are organism-specific, and the lack of knowledge surrounding proteins required for activation and incorporation of F_{430}. Although undoubtedly a difficult task, successful development of heterologous expression systems for MCRs would transform the field, allowing further investigation into the catalytic properties and mechanistic aspects of different MCRs, as well as facilitate the development of optimized biocatalytic systems for methane production or methane conversion applications.

In this review, we will provide an overview of the most relevant aspects of MCR structure and catalysis, and then will focus on considerations and perspectives related to the production of MCR in a heterologous host. For more detailed reviews on MCR biochemistry, the reader is referred to recent excellent reviews by Thauer (2019) and Ragsdale et al. (2017).

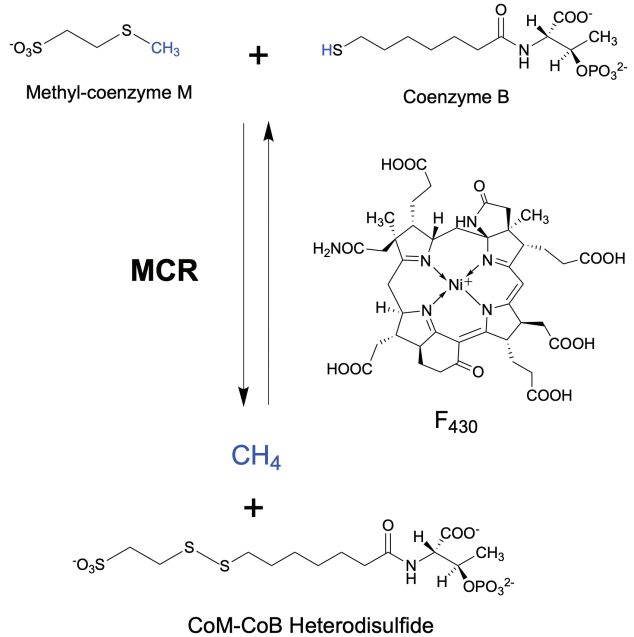


FIGURE 1 | MCR-catalyzed reactions in the final step of methanogenesis in methanogens and the first step of the anaerobic oxidation of methane in anaerobic methanotrophs.

OVERVIEW OF METHANOGENESIS

Methanogenic archaea ("methanogens") are ancient and diverse microorganisms within the archaeal domain of life (Battistuzzi et al., 2004; Adam et al., 2017). They are found in a wide range of anaerobic environments including marine and freshwater habitats, anoxic soils, and as important components of animal microbiomes (Moissl-Eichinger et al., 2018; Lyu et al., 2018a; Borrel et al., 2020). As their sole source of energy, methanogens carry out a form of anaerobic respiration known as methanogenesis, which reduces simple oxidized carbon compounds to generate methane as an end product. There are three main types of methanogenic metabolism depending on the substrate used for methanogenesis (Liu and Whitman, 2008; Costa and Leigh, 2014; Yan and Ferry, 2018). Hydrogenotrophic methanogenesis involves the reduction of CO_2 to CH_4, usually with H_2 as the electron donor. Some hydrogenotrophic methanogens are also able to use other electron donors, such as formate, CO, alcohols, and iron (Dolfing et al., 2008; Ferry, 2010; Kurth et al., 2020). Methylotrophic methanogenesis involves the activation of methylated compounds, such as methanol and trimethylamine *via* substrate-specific corrinoid proteins, which then transfer the methyl group into methanogenesis *via* CH_3-S-CoM. The more recently discovered methoxydotrophic pathway involves methanogenesis from methoxylated aromatic compounds (Mayumi et al., 2016), where the methyl group is transferred to tetrahydromethanopterin instead of coenzyme M (HS-CoM; Kurth et al., 2021a). Finally, acetoclastic methanogenesis utilizes acetate as a methanogenesis substrate, where the carboxyl group is oxidized to CO_2 and the methyl group is reduced to CH_4. Although acetoclastic methanogenesis is the least bioenergetically favorable, 2/3 of biologically derived methane comes from acetate (Lyu et al., 2018a). While there are notable distinctions across the three methanogenic pathways, the key methane-generating step is always catalyzed by MCR (**Figure 1**).

Methanogenesis produces nearly a billion tons of methane each year, which accounts for at least 70% of global methane emissions (Conrad, 2009; Kirschke et al., 2013; Jackson et al., 2020). About half of this methane is consumed by methanotrophic

microorganisms, while the remainder unfortunately escapes to our atmosphere (Kirschke et al., 2013). Although much less abundant compared to CO_2, methane is a more potent greenhouse gas since it has at least a 25-fold higher global warming potential than CO_2 over a 100-year period (Montzka et al., 2011). The rising methane concentration is believed to account for ~20% of the current global warming trend (Lyu et al., 2018a). Thus, the development of strategies to curb methane emissions is essential to mitigate climate change. Indeed, MCR is a highly pursued target for developing inhibitors toward biological methane production (Duin et al., 2016; Yu et al., 2021).

OVERVIEW OF THE ANAEROBIC OXIDATION OF METHANE

Anaerobic methanotrophic archaea (ANME) are related to methanogens and are capable of oxidizing methane in the absence of O_2, consuming substantial amounts of methane in anaerobic environments and thus playing a critical role in the global methane budget (Knittel and Boetius, 2009). Metagenome and gene/protein expression data have revealed that ANME contain and express previously characterized methanogenic genes, indicating that they utilize a reverse methanogenesis pathway to oxidize methane to CO_2 (Hallam et al., 2004; Meyerdierks et al., 2010; Stokke et al., 2012; Timmers et al., 2017). Most commonly, ANME exist with syntrophic sulfate-reducing bacteria that allow AOM to be coupled with sulfate reduction (Nauhaus et al., 2002; Knittel and Boetius, 2009; Holler et al., 2011; Wegener et al., 2016). In these consortia, ANME carry out the oxidation reactions, with MCR presumably catalyzing the initial methane activation step (Scheller et al., 2010; **Figure 1**). The reducing equivalents generated throughout methane oxidation to CO_2 are transferred to the bacteria, likely mediated by multi-heme *c*-type cytochromes, for use in sulfate reduction (McGlynn et al., 2015; Wegener et al., 2015, 2016). Additionally, single archaeal populations have been identified that have the genes necessary for performing AOM as well as sulfite reduction (McKay et al., 2019) or nitrate/nitrite reduction (Haroon et al., 2013), or may transfer electrons directly to metals (He et al., 2018), indicating that there may be exceptions to the paradigm of interspecies redox coupling. On the basis of metagenomic data, ANME are separated into four main clades: ANME-1 (Hinrichs et al., 1999), ANME-2 (Orphan et al., 2001, 2002), ANME-2d (Raghoebarsing et al., 2006; more recently referred to as *Ca.* Methanoperedenaceae; Haroon et al., 2013), and ANME-3 (Knittel et al., 2005; Niemann et al., 2006).

OVERVIEW OF METHYL-COENZYME M REDUCTASE

MCR is a dimer of heterotrimers with a $\alpha_2\beta_2\gamma_2$ configuration (**Figure 2A**), harboring two active sites that are only accessible through a 50 Å channel (Ermler et al., 1997). Each active site contains the nickel hydrocorphin prosthetic group, coenzyme F_{430} (Ellefson et al., 1982, Pfaltz et al., 1982, Livingston et al., 1984, Farber et al., 1991; **Figures 1** and **2**). The active form of MCR contains F_{430} in the Ni(I) oxidation state (Goubeaud et al., 1997). In the methane-forming direction, MCR catalyzes the conversion of CH_3-S-CoM and HS-CoB to methane and the CoM-S-S-CoB heterodisulfide (**Figure 1**). Only one site is activated at any given time ("half-of-the-sites reactivity"), and thus the two active sites are proposed to function similar to a two-stroke engine where binding of the substrates in one active site induces a conformational change that provides energy for the heterodisulfide product to be expelled in the other active site (Goenrich et al., 2005; Scheller et al., 2013). Although any ANME MCR has yet to be enzymatically investigated *in vitro*, MCR from *M. marburgensis* can catalyze the reverse methane oxidation reaction at rates comparable to those measured in AOM consortia *in vivo* (Scheller et al., 2010), supporting the proposal that ANME utilize MCR to oxidize methane.

Recent mechanistic studies have provided evidence that the reaction occurs by "mechanism II," involving a methyl radical intermediate (Wongnate et al., 2016) that was originally proposed on the basis of quantum mechanical modeling studies (Pelmenschikov et al., 2002; Pelmenschikov and Siegbahn, 2003; Chen et al., 2012, 2014). The major alternative mechanism ("mechanism I") involves nucleophilic chemistry with a

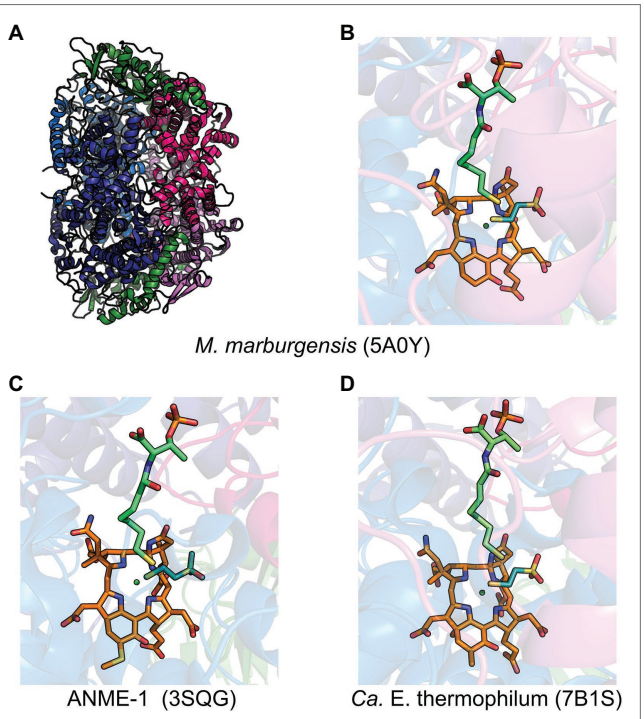


FIGURE 2 | Representative crystal structures of MCRs and ECR. **(A)** MCR from *M. marburgensis* and **(B)** the associated active site with F_{430}, HS-CoB, and HS-CoM. **(C)** Black Sea mat ANME-1 MCR active site with 17^2-methylthio-F_{430}, HS-CoB, and HS-CoM. **(D)** *Ca.* E. thermophilum ECR active site with dimethyl-F_{430}, HS-CoB, and HS-CoM. α subunits are shown in marine and deep blue, *β* subunits in hot pink and violet, γ subunits in forest green, F_{430} in orange, coenzyme M in deep teal, and coenzyme B in lime green.

Ni(III)-methyl intermediate (Yang et al., 2007; Dey et al., 2010). In the proposed radical mechanism, Ni(I) induces homolytic cleavage of the methyl-sulfur bond of CH_3-S-CoM to generate a methyl radical and Ni(II). The methyl radical then reacts with HS-CoB to produce methane and a •S-CoB radical, which reacts with the Ni-bound CoM thiolate to generate a disulfide anion radical. One-electron transfer to Ni(II) then releases the heterodisulfide and regenerates the Ni(I) (Wongnate et al., 2016).

MCR STRUCTURES AND POST-TRANSLATIONAL MODIFICATIONS

The structures of MCR for which crystal structures have been obtained from various methanogens and alkane-oxidizing organisms are all remarkably similar. Based on phylogenetic and structural comparison, MCRs from *Methanobacteriales* and *Methanococcales* were classified into MCR types I, II, and III (Wagner et al., 2017). The different MCR types have representative crystal structures and mainly differ in their electrostatic surface potentials, loop architectures, and the C-terminal end of their γ-subunits that interact with α and β subunits. Based on phylogenetic comparisons, MCRs from other organisms, such as *Methanomicrobiales, Methanosarcinales, Methanocellales,* and ANME-1, are distinct from the defined types I-III present in *Methanobacteriales* and *Methanococcales* (Wagner et al., 2017).

MCR crystal structures with HS-CoB and the substrate analog HS-CoM show the sulfhydryl group of HS-CoM serving as an axial ligand to the Ni(II) of F_{430} (**Figure 2B**). However, all MCR crystal structures are of the enzyme in its inactive Ni(II) state or a chemically modified methyl-Ni(III) state (Cedervall et al., 2011). Thus, the true coordination state of Ni(I) and the binding conformation of CH_3-S-CoM remains unclear since a crystal structure with CH_3-S-CoM has never been obtained. However, recent studies (Patwardhan et al., 2021) have provided new evidence for the possible orientation of CH_3-S-CoM binding to the nickel center of F_{430}. Interestingly, results indicate that there is no nickel-sulfur interaction and thus suggest that the thioether portion of the substrate does not bind to the Ni(I) (Patwardhan et al., 2021). Instead, CH_3-S-CoM appears to bind to Ni(I) through the sulfonate group. This proposed alternate binding scenario puts the reactive portions of the two substrates in close proximity so that the subsequently generated proposed methyl radical is in position to abstract a hydrogen atom from HS-CoB (Patwardhan et al., 2021).

When the first crystal structure of MCR from *M. marburgensis* (**Figures 2A,B**) was solved, five unusual post-translational modifications (PTMs) were revealed in the *α* subunit near the active site (Ermler et al., 1997). Since then, subsequent work has discovered additional MCR PTMs, where the presence of specific PTMs varies depending on the organism (**Table 1**; Grabarse et al., 2000; Kahnt et al., 2007; Wagner et al., 2016; Wagner et al., 2017; Kurth et al., 2021b). PTMs found in methanogens include three strictly conserved modifications—N^1-methylhistidine, 5-(*S*)-methylarginine, and thioglycine—as well as a handful of more variable modifications including *S*-methylcysteine, 2-(*S*)-methylglutamine, didehydroaspartate, and 6-hydroxytryptophan (**Table 1** and **Figure 3**). The impacts of these PTMs remain a major area of research in the field as their precise role in catalysis and/or active site structure remains unclear. MCR PTMs were recently summarized in a mini-review (Chen et al., 2020), but we will outline key aspects here.

Significant progress has been made toward identifying the enzymes involved in installing MCR PTMs, including identification of the radical *S*-adenosylmethionine (SAM) enzyme catalyzing the difficult methylation reaction to produce 5-(*S*)-methylarginine (Deobald et al., 2018; Radle et al., 2019; Lyu et al., 2020) and identification of the enzymes responsible for the thioglycine transformation (Nayak et al., 2017). Both of these PTMs appear to at least be important for the stability of MCR from *Methanosarcina acetivorans*, especially under thermal stress (Nayak et al., 2017, 2020; Deobald et al., 2018). The methylated arginine seems to have a more significant impact on *Methanococcus maripaludis* MCR, where a deletion strain lacking this modification showed a highly impaired growth rate and the rate of methanogenesis was only about half the rate of wild type (Lyu et al., 2020). Most recently, the methyltransferase necessary for the synthesis of *S*-methylcysteine was identified (Nayak et al., 2020). Through the production of a *M. acetivorans* deletion strain lacking the genes involved in 5-(*S*)-methylarginine, thioglycine, and *S*-methylcysteine biosynthesis, the associated MCR variant was produced and its crystal structure was solved, which was surprisingly indistinguishable from the wild-type structure (Nayak et al., 2020). Growth studies with the associated deletion strain suggested that epistatic interactions among MCR PTMs influence the stability and *in vivo* activity of the enzyme (Nayak et al., 2020). However, *in vitro* kinetic studies have not yet been carried out on MCR variants lacking one or more PTMs, which will be required to make any conclusions about the specific functions and importance of the respective PTMs.

ANME-1 is the only ANME clade for which an MCR crystal structure has been obtained. By purifying and crystallizing the enzyme directly from a Black Sea mat sample, the crystal structure of the ANME-1 MCR was solved to 2.1 Å resolution (Shima et al., 2012). This ANME MCR possesses the same overall structure as methanogenic MCRs. In particular, the active site channel seems be strictly conserved in both ANME-1 and methanogens, with HS-CoM and HS-CoB holding virtually the same position and conformation (**Figure 2C**). This result excludes the possibility of ANME MCR using different substrates/products, further supporting that methane oxidation catalyzed by MCR in ANME is the reverse of the methane-generating step in methanogens (**Figure 1**). The ANME-1 MCR structure does possess a few notable differences. First, the active site contains a modified F_{430}, 17^2-methylthio F_{430} (**Figures 2C**, **4**), which was previously structurally characterized by mass spectrometry and NMR (Mayr et al., 2008). This modified F_{430} appears to be accommodated by the replacement of the bulky 2-(*S*)-methylglutamine found in methanogens with Val419 in ANME-1. Second, ANME-1 MCR contains five distinct cysteines between F_{430} and the protein surface, suggesting a potential

TABLE 1 | Summary of MCR crystal structures with associated PTM content.

MCR crystal structure	PDB	N^1-methyl-His	*S*-methyl-Cys	2-(*S*)-methyl-Gln	5-(*S*)-methyl-Arg	Thioglycine	Didehydro-Asp	6-hydroxy-Trp	7-hydroxy-Trp	3-methyl-Ile	N^2-methyl-His	*S*-oxy-Met
Methanothermobacter marburgensis MCR I (Ermler et al., 1997; Wagner et al., 2016)	5A0Y	+	+	+	+	+	+	–	–	–	–	–
Methanothermobacter marburgensis MCR II (Wagner et al., 2016)	5A8R	+	+	+	+	+	+	–	–	–	–	–
Methanosarcina barkeri (Grabarse et al., 2000; Wagner et al., 2016)	1E6Y	+	+	–	+	+	+	–	–	–	–	–
Methanosarcina acetivorans (Nayak et al., 2020)		+	+	–	+	+	+	–	–	–	–	–
Methanopyrus kandleri (Grabarse et al., 2000; Kahnt et al., 2007)	1E6V	+	–	+	+	+	–	–	–	–	–	–
Methanotorris formicicus (Wagner et al., 2017)	5N2A	+	–	+	+	+	–	+	–	–	–	–
Methanothermobacter wolfeii (Wagner et al., 2016)	5A8K	+	+	+	+	+	–	–	–	–	–	–
Methanothermococcus thermolithotrophicus (Wagner et al., 2017)	5N1Q	+	–	+	+	+	–	–	–	–	–	–
Methermicoccus shengliensis (Kurth et al., 2021b)	7NKG	+	–	–	+	+	–	–	–	–	–	–
ANME–1 from Black Sea mats (Shima et al., 2012)	3SQG	+	+/–[a]	–	–	+/–[a]	–	–	+	–	–	+
Ca. Ethanoperedens thermophilum (Hahn et al., 2021)	7B1S	+	+	+	+	+	–	–	–	+	+	–

[a]*Early mass spectrometry data indicated that ANME-1 MCR lacked the S-methylcysteine as well as thioglycine (Kahnt et al., 2007), while the ANME-1 MCR crystal structure showed the thioglycine was present (Shima et al., 2012). However, the sample used for crystallization represented a mixed population where 30% contained thioglycine but not S-methylcysteine, while the majority (70%) contained S-methylcysteine but not thioglycine and did not result in crystal formation (Shima et al., 2012).*

(+) indicates PTM is present and (–) indicates PTM is absent.

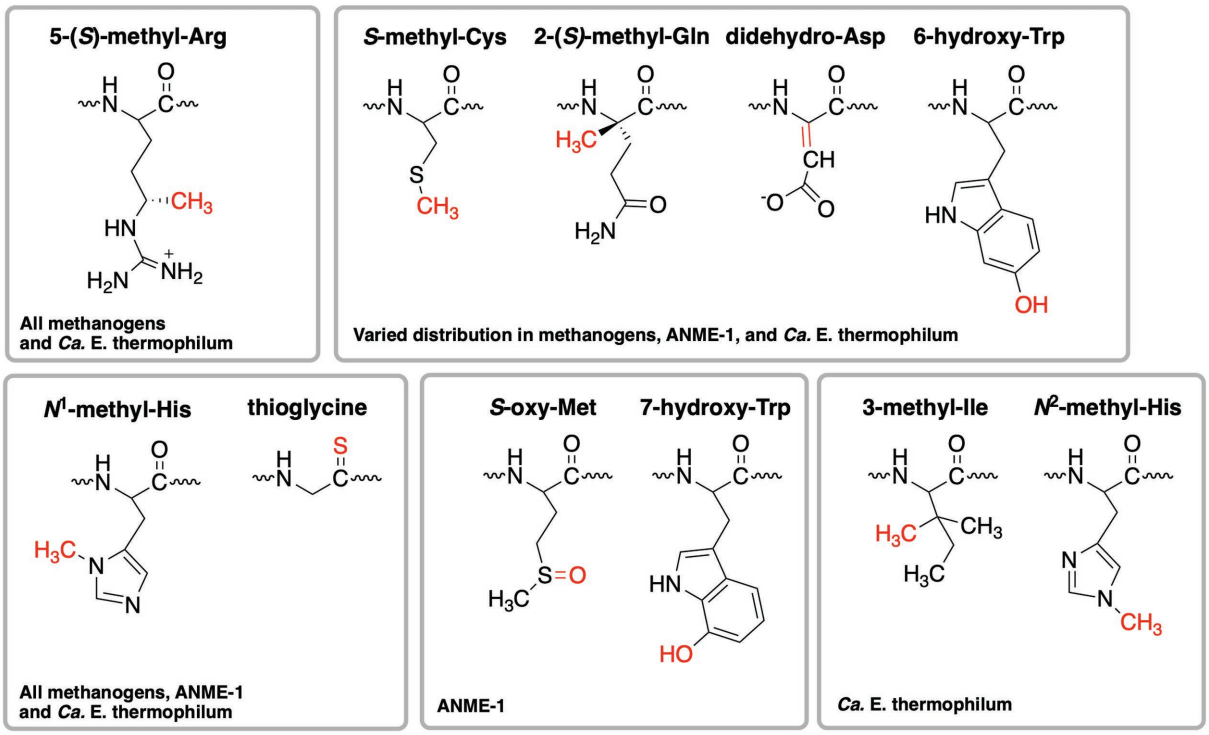


FIGURE 3 | Structures of post-translational modifications identified in the active sites of various MCRs and ECR. The distribution of these PTMs is further summarized in **Table 1**.

redox-relay system that could be used to reduce F_{430} to the active Ni(I) state (Shima et al., 2012). Third, PTM patterns vary between ANME-1 MCR and methanogenic MCR. ANME-1 does not contain the highly conserved arginine methylation seen in methanogens, however, ANME-1 MCR does contain two unique PTMs, 7-hydroxytryptophan and *S*-oxymethionine. Also present in ANME-1 MCR are N^1-methylhistidine and thioglycine (Shima et al., 2012; **Table 1** and **Figure 3**).

Very recently, the crystal structure of the MCR homolog from an anaerobic ethane oxidizing archaeon, *Ca.* Ethanoperedens thermophilum, was determined (Hahn et al., 2021). This enzyme apparently does not take other alkane substrates outside of ethane (Hahn et al., 2020) and thus has been designated as ethyl-coenzyme M reductase (ECR; Hahn et al., 2021). Although ECR follows the overall structural trend of known MCRs, there are notable differences to consider. First, ECR is 20 kDa larger than canonical MCRs, mainly due to three insertions in the α subunit, one insertion in the β subunit, and one insertion in the γ subunit. These insertions impact surface charges and contribute to the unique architecture of the ethane tunnel, which provides a 33 Å hydrophobic path to the active site (Hahn et al., 2021). The tunnel is flanked by post-translationally modified amino acids, including two unique PTMs—N^2-methylhistidine and 3-methylisoleucine (**Table 1** and **Figure 3**). Notably, the ECR structure revealed that the F_{430} nickel is coordinated by methionine rather than the canonical glutamine, and a modified version of F_{430} with two methyl groups is present (dimethyl-F_{430}; **Figures 2D**, **4**). Additionally, an active site loop (α367-374) contains a tryptophan residue (αTrp^{373}) instead of the canonical phenylalanine found in other MCRs. This loop shifts the position of F_{430}, which is stabilized by hydrogen bonds with αAsn^{375} along with a clamping effect from αTyr^{376} and αPhe^{441}, resulting in a 11.4° tilt on the porphinoid ring. Consequently, the active site volume is increased to adequately accommodate ethane. The authors propose that the F_{430} methylations likely serve to maintain the structure and reactivity of the cofactor in the expanded active site (Hahn et al., 2021).

MODIFIED F_{430} COENZYMES

An interesting and underexplored area in the MCR field is the potential functions and importance of F_{430} modifications. F_{430} is only known to function with MCRs and ACRs, indicating that nature has evolved a specialized coenzyme to catalyze the difficult reactions of methane formation and methane/alkane activation. The first modified F_{430} to be discovered was 17^2-methylthio-F_{430} (**Figure 4**), which was originally identified and structurally characterized from Black Sea mat samples enriched with ANME-1 (Mayr et al., 2008). This modified F_{430} is presumed to be the primary physiologically active version of F_{430} in ANME-1 since it was later characterized in the crystal structure of ANME-1 MCR (Shima et al., 2012; **Figure 2C**). ANME-2 organisms apparently do not contain 17^2-methylthio-F_{430} (Mayr et al., 2008; Kaneko et al., 2014), and the potential F_{430} modifications

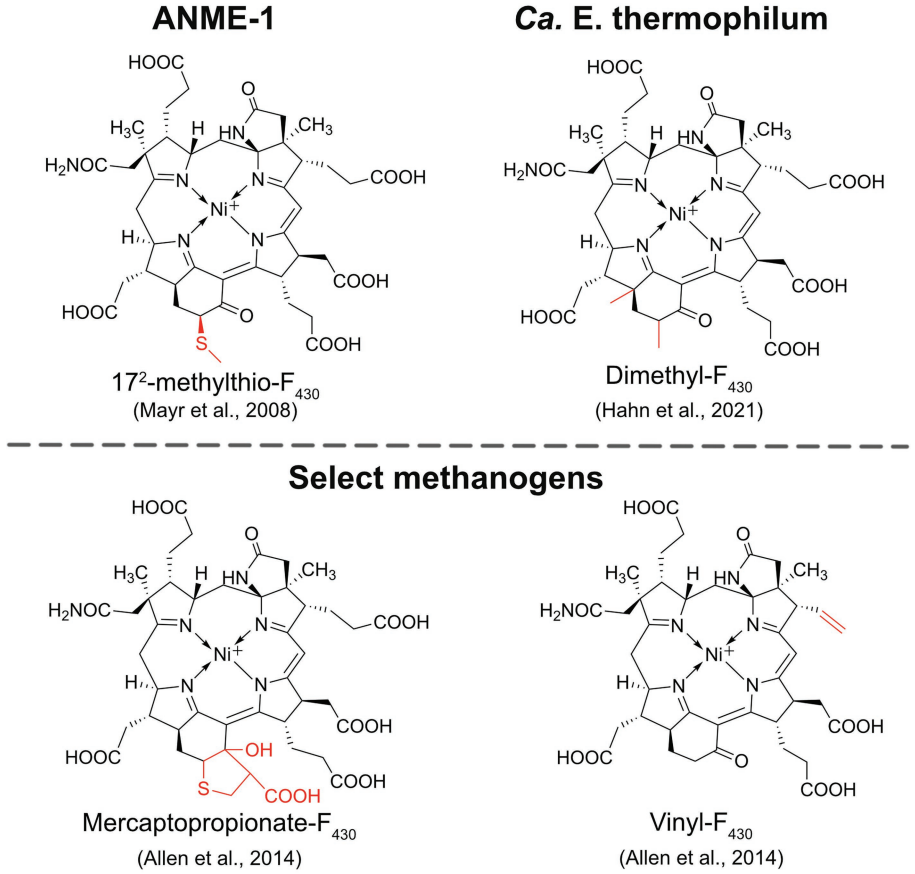


FIGURE 4 | Structures of modified F_{430}s. The structures of the modified F_{430}s in ANME-1 and *Ca. E. thermophilum* are confirmed based on NMR and/or crystal structures while the modifications in select methanogens are proposed based on mass spectrometry data.

present in ANME-3 have not yet been reported to our knowledge. In 2014, additional F_{430} modifications were identified in select *Methanococcales* methanogens (Allen et al., 2014). Mercaptopropionate-F_{430}, containing a cyclized mercaptopropionate moiety bound as a thioether, was identified in *Methanocaldococcus jannaschii* and *Methanococcus maripaludis*, while vinyl-F_{430} was observed in *M. maripaludis* and *Methanococcus vannielii* (**Figure 4**). These structures were proposed based on mass spectrometry data, and thus still need to be confirmed by NMR or crystallography. Finally, most recently, the dimethyl-F_{430} present in ECR was discovered (Hahn et al., 2021; discussed above; **Figure 4**). One important aspect to note is that the F_{430} modifications in methanogens are not always present and the modifications generally exist as a minor component compared to the unmodified F_{430} (Allen et al., 2014). The situation appears to be different in ANME-1, where 17^2-methylthio-F_{430} is the predominant form (Kaneko et al., 2014). Additionally, for both ANME-1 and *Ca.* E. thermophilum, 17^2-methylthio-F_{430}, and dimethyl-F_{430} are confirmed to function with the respective MCR/ECR since they were identified in the crystal structures (Shima et al., 2012; Hahn et al., 2021). In contrast, the modified F_{430}s in methanogens have only so far been identified in small molecule cell extracts and have not been observed in any methanogen MCR crystal structures.

Understanding how modified F_{430}s affect MCR will be important for the design of optimized heterologous systems for recombinant MCR production, especially for anaerobic methane/alkane oxidation applications. It will also be necessary to identify the enzymes involved in their biosynthesis. The complete biosynthetic pathway for F_{430} has been described (Zheng et al., 2016; Moore et al., 2017), but the enzymes required for the installation of F_{430} modifications are currently unknown.

MCR OPERON ORGANIZATION

The three MCR subunits are encoded by *mcrA*, *mcrB*, and *mcrG*, which are usually present in an MCR operon along with two other genes (*mcrD* and *mcrC*). The most common MCR operon across all methanogens is *mcrBDCGA* (**Figure 5**). McrD and McrC are accessory proteins whose functions have yet to be confirmed. McrC was a component of the large complex of proteins identified as being responsible for the reduction of F_{430} to its Ni(I) active state; thus McrC likely plays a role in MCR activation (Prakash et al., 2014). McrD may serve as a chaperone protein that binds F_{430} for subsequent delivery to the MCR active site (Zheng et al., 2016). Early studies demonstrated that

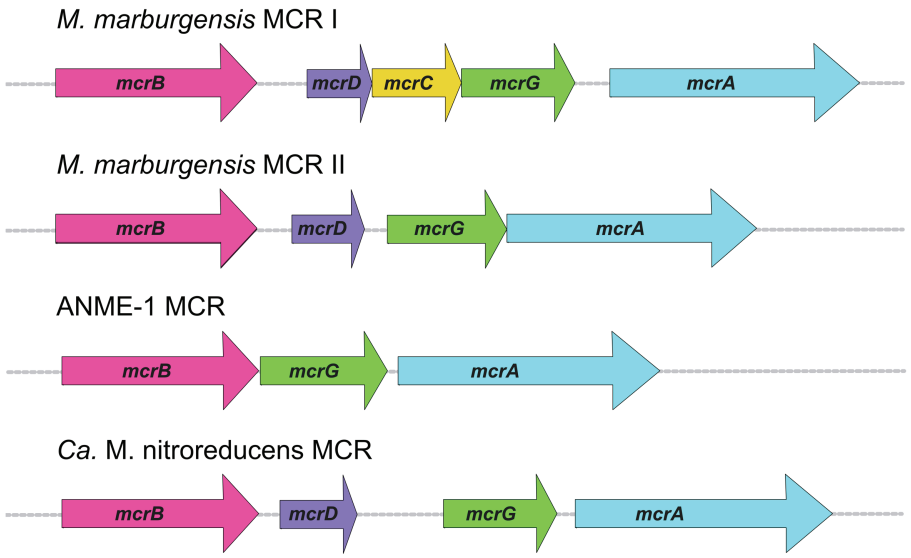


FIGURE 5 | MCR operon organization in selected organisms.

McrD interacts with MCR (Sherf and Reeve, 1990) and it also co-purified with a recombinant MCR expressed in *M. maripaludis* (Lyu et al., 2018b). Finally, McrD was shown to alleviate product inhibition in the final step of F_{430} biosynthesis *in vitro*, presumably due to its ability to bind the newly synthesized F_{430} coenzyme (Zheng et al., 2016).

An interesting aspect of MCR operons is the variability in the presence of accessory proteins, or the presence of additional operons. In some methanogens, including *M. marburgensis*, there is a second MCR operon that encodes for MCR isozyme II (Rospert et al., 1990). Most work on MCR has been performed on MCR isozyme I, whose operon in *M. marburgensis* is the typical *mcrBDCGA* operon. Less studied is MCR II, whose operon lacks *mcrC* (**Figure 5**). The two isozymes are highly similar in sequence as well as overall structure and active site architecture (Wagner et al., 2016), but some notable differences have been reported. MCR I and MCR II show significant differences in electrostatic surface potentials, which allows for convenient separation of the two isozymes *via* anion exchange chromatography (Rospert et al., 1990; Duin et al., 2011). Because MCR II was purified in larger quantities from cells in log phase, and MCR I was purified in larger quantities from cells at the end of growth, it was concluded that MCR II is expressed in non-gas-limiting conditions, while MCR I is expressed during gas-limiting conditions (Rospert et al., 1990; Bonacker et al., 1992). Further studies on both isoenzymes revealed that MCR I and MCR II expression is affected by pH, temperature, and availability of both H_2 and CO_2 (Bonacker et al., 1992), and *in vitro* kinetic studies revealed differing catalytic properties (Bonacker et al., 1993).

In addition to *M. marburgensis*, many other members of *Methanobacteriales*, and some members of *Methanococcales* and *Methanomicrobiales* contain two MCR isozymes. The presence of two MCR isozymes may present a possible evolutionary advantage which could allow methanogens to express different versions of MCR when faced with changes in environmental conditions, such as substrate limitation. The fact that the MCR II gene cluster lacks *mcrC* would indicate that either McrC is not necessary for the function of MCR II or that the McrC encoded in the MCR I gene cluster can be utilized.

The difference between MCR I and MCR II gene clusters is not the only instance in which the MCR operon displays variability. The MCR operon from ANME-1 isolated from a Black Sea mat lacks both *mcrD* and *mcrC* (Meyerdierks et al., 2010; **Figure 5**). A McrC homolog is present outside of the MCR operon in ANME-1 (BSM_08630, 51% identity to McrC from *M. marburgensis*), but McrD appears to be completely absent. The ANME-2d organism, *Ca.* Methanoperedens nitroreducens, contains an MCR operon like that of MCR II in *M. marburgensis*, where only *mcrD* is present without *mcrC* (Haroon et al., 2013; **Figure 5**). Similar to ANME-1, a likely *mcrC* exists outside of the MCR operon (ANME2D_00875, 53% identity to McrC from *M. marburgensis*, 50% identity to McrC from ANME-1 mentioned above). This poses interesting questions regarding the high conservation of the primary MCR I operon *mcrBDCGA* across different methanogenic species, whereas ANME MCR operons lack one or more of the accessory proteins in the operon. Additionally, ANME-1 organisms seem to completely lack *mcrD*. A better understanding of the functions and specificity of these and potentially additional yet to be discovered accessory proteins is crucial for the development of optimized recombinant MCR expression systems.

ALKYL-COENZYME M REDUCTASES

MCR variants known as alkyl-coenzyme M reductases (ACRs) carry out the anaerobic oxidation of various non-methane alkane substrates. The first report of ACR-dependent anaerobic

oxidation of an alkane other than methane was published in 2016 when members of the GoM-Arch87 clade, closely related to *Methanosarcinales*, were proposed to carry out the anaerobic oxidation of butane to CO_2 (Laso-Pérez et al., 2016). Similar to ANME-1, these organisms form consortia with HotSeep-1 sulfate-reducing bacteria and couple butane oxidation to sulfate reduction. Two archaeal genomes were assembled from the butane enrichment cultures that resulted in the proposed names for two new organisms, *Ca.* Syntrophoarchaeum butanivorans and *Ca.* Syntrophoarchaeum caldarius (Laso-Pérez et al., 2016). Intriguingly, these organisms contain four different *mcrBGA* genes clusters—three of the four ACR gene sets are arranged in operons in *Ca.* S. butanivorans, while all four are arranged in operons in *Ca.* S. caldarius. The butane-dependent formation of butyl-S-CoM was confirmed in these cultures, suggesting that butane oxidation involves the use of the MCR homolog(s). The GoM-Arch87 enrichment cultures were also tested with other hydrocarbons and, interestingly, propane enriched cultures resulted in propane dependent sulfate reduction. As with the butane cultures, propyl-S-CoM was detected, indicating the involvement of an ACR in propane oxidation. Other alkane oxidizers include *Ca.* Argoarchaeum ethanivorans (Chen et al., 2019) and *Ca.* E. thermophilum (Hahn et al., 2020), archaeal species that activate ethane using ECR (discussed in MCR structures and post-translational modifications section) to form ethyl-S-CoM, which is subsequently oxidized to CO_2. These ethane oxidizers only utilize ethane and cannot metabolize other alkanes.

In addition to the involvement of ACRs in short-chain alkane oxidation, recently published work suggests the use of ACRs to activate long-chain alkanes in oil reservoirs (Zhou et al., 2022). *Ca.* Methanoliparum couples the degradation of long-chain alkanes to methanogenesis in a process that is independent of syntrophic partners. Metagenomic and transcriptomic data indicated the presence of several different species, which contain and expresses genes encoding putative ACRs as well as MCRs. Additionally, the presence of hexadecyl-S-CoM and other long-chain alkane R-S-CoM derivatives were confirmed by mass spectrometry, thus indicating the involvement of an ACR in alkane activation (Zhou et al., 2022).

OTHER DIVERGENT MCRs

Traditionally, all methanogens were thought to belong to the Euryarchaeota phylum, however, this definition was challenged with the discovery of putative methane metabolism in the Bathyarchaeota phylum (Evans et al., 2015). Metagenome data revealed the presence of *mcrABG* genes and putatively *mcrCD*, as well as several other genes for methylotrophic methanogenesis. The MCR primary sequence and predicted structure analysis showed putative binding sites for HS-CoM, HS-CoB, and F_{430}, suggesting that the enzyme utilizes the same substrates and coenzyme (Evans et al., 2015). In addition to Bathyarchaeota, five metagenomes from a proposed new archaeal phylum termed *Ca.* Vestraetearchaeota were shown to contain divergent *mcrA* sequences (Vanwonterghem et al., 2016). Metabolic reconstructions of the assembled genomes revealed the presence of key genes associated with methylotrophic methanogenesis, including a complete *mcrBDCGA* operon. It is hypothesized that these organisms would perform H_2-dependent methylotrophic methanogenesis. However, it is still unclear if this new phylum is comprised of organisms that perform methanogenesis as a preferential metabolism since the assembled genomes in this study also showed that these organisms likely have the capacity to perform fermentative metabolism (Vanwonterghem et al., 2016).

Archaeoglobi is a class of thermophilic organisms within the Euryarchaeota that were generally believed to be non-methanogenic (Bapteste et al., 2005; Hartzell and Reed, 2006; Boyd et al., 2019). This is because complete MCR-encoding genes and methyl-H_4M(S)PT:coenzyme M methyltransferase (MTR) *MtrABCDEFGH* complex genes had never been identified, even though other characteristic hydrogenotrophic methanogenesis genes as well as archaeal type Wood-Ljungdahl pathway genes have been identified in some Archaeoglobi genomes (Klenk et al., 1997; Bapteste et al., 2005). However, recent metagenome data have demonstrated that some Archaeoglobi contain genes encoding MCR. *Ca.* Polytropus marinifundus contains two divergent *mcrABG* operons similar to Bathyarchaeota and Syntrophoarchaeum as well as other potential genes for alkanotrophic metabolism (Boyd et al., 2019). Another new genus of Archaeoglobi, *Ca.* Methanomixophus, contains MTR complex genes as well as a complete *mcrBDCGA* operon, along with predicted ligand binding sites for HS-CoM, HS-CoB, and F_{430} (Liu et al., 2020). Metatranscriptomic experiments showed active hydrogen-dependent methylotrophic methanogenesis as well as heterotrophic fermentation. Additionally, one of the new proposed organisms, *Ca.* Methanomixophus hydrogenotrophicum, possesses the genes to conserve energy *via* AOM coupled to syntrophic sulfate reduction, while *Ca.* Methanomixophus dulitatem contains its own sulfate reduction genes that would allow for a methane oxidizing lifestyle (Liu et al., 2020).

METHANOGENS AS HOSTS FOR THE HETEROLOGOUS PRODUCTION OF RECOMBINANT MCRs

The recent and continuing discoveries of diverse putative MCRs and ACRs highlight the need to develop effective tools to study the catalytic capabilities of these enzymes with likely very different enzymatic properties. Additionally, MCR is a highly attractive, yet challenging, target for potential use in bioengineering applications for biofuel production, either for methane generation or methane/alkane conversion applications (Conrado and Gonzalez, 2014; Haynes and Gonzalez, 2014). Since ANME utilize MCR in the methane oxidation direction, ANME MCRs are especially appealing biocatalysts for potentially converting abundant methane reserves into more usable liquid fuels and other value-added chemicals (Mueller et al., 2015; Lawton and Rosenzweig,

2016a,b). However, as mentioned previously, no ANME MCR has been studied *in vitro* and thus it is unclear whether ANME MCRs will be better suited for this purpose compared to methanogenic MCRs.

Since AOM consortia grow slowly, with doubling times on the month timescale (Laso-Pérez et al., 2018; Bhattarai et al., 2019), and to low cell densities, obtaining enough cells to purify the native MCR from ANME organisms in sufficient quantities for kinetic and mechanistic studies is not very feasible. Additionally, ANME organisms are not yet genetically tractable and thus are not amenable to bioengineering applications. Thus, the development of heterologous expression systems for ANME MCRs as well as other diverse MCRs from unculturable or difficult-to-culture archaea would be highly advantageous. Traditional hosts, such as *Escherichia coli*, seem to pose currently unsurmountable challenges to achieve this goal since they do not possess the biochemical machinery for F_{430} biosynthesis, post-translational modifications, or MCR assembly and activation. Thus, the current likely best option for a heterologous host is a fast-growing methanogen for which genetic manipulation methods exist. The two highly studied model methanogens for which well-established and robust genetic tools exist are *Methanosarcina acetivorans* and *Methanococcus maripaludis*. The following paragraphs and **Table 2** summarize the major available genetic tools in these organisms as well as recently described genetic tools in *Methanocaldococcus jannaschii* and *Methanothermobacter thermautotrophicus*.

Methanosarcina species are cytochrome-containing methanogens that are capable of using the widest variety of methanogenic substrates compared to other genera. Depending on the growth substrate, *M. acetivorans* exhibits doubling times as low as 8 h at 37°C. Compared to other methanogens with available genetic tools, *Methanosarcina* are phylogenetically most closely related to ANME (Knittel et al., 2005), thus making *Methanosarcina* the most logical choice as potential heterologous hosts for ANME MCRs. *M. acetivorans* C2A is most commonly utilized for genetic experiments, but *Methanosarcina barkeri* Fusaro is also used successfully and the same basic tools have been developed for both species (Buan et al., 2011). Both organisms are mesophilic and require that they are maintained in high-salt medium to prevent cells from growing in clumps, which can pose issues for genetic experiments. Routine genetic experiments normally involve a parental strain containing a deletion of the hypoxanthine phosphoribosyltransferase *hpt* gene that is used as a counterselection marker (Pritchett et al., 2004), and often also contain a ΦC31 *attP* site inserted at the *hpt* locus, allowing for insertion of plasmids containing the complementary *attB* sequence into the host chromosome *via* recombination (Guss et al., 2008). A comprehensive description of the different plasmids and strains of *Methanosarcina* that have been used for various purposes, including gene deletions and recombinant expression, has been reported (Buan et al., 2011). Transforming *Methanosarcina* species involves the well-established liposome-mediated transformation, where the transformation efficiency is as high as 20% in *M. acetivorans*, but is substantially lower in other *Methanosarcina* species (Metcalf et al., 1997). For inducible expression of recombinant proteins in *Methanosarcina*, the tetracycline-inducible system can be used, which allows for the expression of proteins at a desired time point during growth (Guss et al., 2008). The *cdh* operon promoter has

TABLE 2 | Summary of major genetic tools available in methanogens with associated references.

	Methanosarcina acetivorans/barkeri	***Methanosarcina mazei***	***Methanococcus maripaludis* (S2 and JJ)**	***Methanocaldococcus jannaschii***	***Methanothermobacter thermautotrophicus ΔH***
Transformation methods	Liposome mediated (Metcalf et al., 1997; Buan et al., 2011)	Liposome mediated (Metcalf et al., 1997; Ehlers et al., 2005)	Polyethylene glycol mediated (Tumbula et al., 1994; Sarmiento et al., 2011; Natural competence; Fonseca et al., 2020)	Heat shock (Susanti et al., 2019)	Interdomain conjugation (Fink et al., 2021)
Shuttle vectors	Metcalf et al., 1997; Buan et al., 2011	Metcalf et al., 1997	Whitman et al., 1997; Sarmiento et al., 2011	–	Fink et al., 2021
Positive selection marker	Puromycin (Metcalf et al., 1997; Buan et al., 2011)	Puromycin (Metcalf et al., 1997; Ehlers et al., 2005) Neomycin (Mondorf et al., 2012)	Puromycin and neomycin (Whitman et al., 1997; Sarmiento et al., 2011)	Mevinolin and Simvastatin (Susanti et al., 2019)	Neomycin (Fink et al., 2021)
Counterselection marker	*hpt* (8-azahypoxanthine) (Pritchett et al., 2004; Buan et al., 2011)	*hpt* (8-azahypoxanthine) (Ehlers et al., 2011)	*hpt* (8-azahypoxanthine), *upt* (6-azauracil) (Moore and Leigh, 2005; Sarmiento et al., 2011)	–	–
Markerless genetic exchange	Pritchett et al., 2004; Buan et al., 2011	Ehlers et al., 2011	(Moore and Leigh, 2005; Sarmiento et al., 2011	–	–
Inducible promoters	Tetracycline-inducible promoter (Guss et al., 2008) Acetate regulated promoter (Macauley et al., 2009)	Trimethylamine regulated promoter (Mondorf et al., 2012)	*Nif* promoter (Lie and Leigh, 2002; Chaban et al., 2007) Phosphate sensing promoter (Akinyemi et al., 2021)	–	–
CRISPR/Cas System	CRISPR/Cas9 (Nayak and Metcalf, 2017)	–	CRISPR/Cas12 (Bao and Scheller, 2021)	–	–

also been used successfully to drive acetate-dependent overexpression of a carbonic anhydrase in *M. acetivorans* (Macauley et al., 2009). More recent work has generated a new series of suicide plasmids that simplify the cloning process and allow for facile expression and purification of tagged proteins in *Methanosarcina* (Shea et al., 2016). Additionally, an efficient CRISPR/Cas9 system has been developed for *M. acetivorans* (Nayak and Metcalf, 2017). Notably, this system has been used to add a tandem affinity purification tag to the N-terminus of the native McrG gene in *M. acetivorans*, thus greatly facilitating MCR purification (Nayak and Metcalf, 2018). Finally, genetic tools have also been utilized for *Methanosarcina mazei* (Ehlers et al., 2005), including markerless genetic exchange using similar methods developed for *M. acetivorans* and *M. barkeri* (Pritchett et al., 2004; Ehlers et al., 2011). A trimethylamine inducible promoter has also been used to successfully heterologously overexpress a fusion-tagged protein in *M. mazei* (Mondorf et al., 2012).

Methanococcus maripaludis is a hydrogenotrophic mesophilic methanogen with a relatively fast doubling time of 2 h at 37°C (Whitman et al., 1986). The rapid growth of *M. maripaludis* compared to *Methanosarcina* is advantageous for genetic experiments and recombinant protein expression. *M. maripaludis* S2 is the strain for which most reports of genetic manipulation utilize; however, *M. maripaludis* JJ has also been used successfully for this purpose. *Methanococcus maripaludis* can be grown on either H_2/CO_2 or formate, the latter of which provides an easier and safer alternative to dealing with high-pressure gases (Long et al., 2017). Genetic studies often utilize *M. maripaludis* strains lacking the gene for uracil phosphoribosyltransferase, which confers sensitivity to the base analog 6-azauracil and serves as a marker for negative selection (Costa et al., 2010; Sarmiento et al., 2011). Several shuttle vectors have been reported, and well-established methods exist for markerless mutagenesis and recombinant protein expression (Whitman et al., 1997; Sarmiento et al., 2011). *M. maripaludis* plasmids utilize puromycin or neomycin resistance for positive selection. Transformation methods for *M. maripaludis* S2 utilize polyethylene glycol-mediated transformation (Tumbula et al., 1994; Sarmiento et al., 2011). However, natural transformation facilitated by type IV-like pili has been reported for *M. maripaludis* JJ (Fonseca et al., 2020). Notably, a highly efficient CRISPR/Cas system has recently been developed for *M. maripaludis* JJ that utilizes a bacterial Cas12a along with the native homology directed repair machinery (Bao and Scheller, 2021). The capacity for natural transformation as well as the available CRISPR/Cas technology for genetic manipulation makes *M. maripaludis* JJ an especially attractive host for future metabolic engineering applications. Finally, natural transformation *via* type IV-like pili has also been demonstrated in *Methanoculleus thermophilus* (Fonseca et al., 2020). Using an established plasmid employed for generating *M. maripaludis* gene deletions, the authors generated a *M. thermophilus* deletion strain lacking genes for pili to demonstrate that pili are essential for natural transformation (Fonseca et al., 2020). This is the first report of genetic manipulation in a methanogen from the order *Methanomicrobiales*.

Methanocaldococcus jannaschii is a hyperthermophilic methanogen (Jones et al., 1983) that was the first archaeal organism to have its genome sequenced (Bult et al., 1996). *Methanocaldococcus jannaschii* has the fastest doubling time of any methanogen (26 min) and grows optimally at 85°C (Jones et al., 1983). Recently, the first genetic tools for *M. jannaschii* were reported (Susanti et al., 2019). *Methanocaldococcus jannaschii* is resistant to antibiotics commonly used with other archaea, such as previously mentioned puromycin or neomycin, as well as the base analogs used for counter selection in other methanogens (Susanti et al., 2019). However, it was found to be sensitive to mevinolin and simvastatin. These compounds are competitive inhibitors of 3-hydroxy-methylglutaryl (HMG)-CoA reductase and thus overexpression of HMG-CoA reductase can be used as a selection marker (Susanti et al., 2019). Based on this, a suicide vector was developed for generating in-frame gene deletions in *M. jannaschii,* where the gene of interest is replaced with *hmgA*. Further, a similar strategy was used to place a gene of interest under the control of a strong promoter and to add an affinity tag on the chromosome for subsequent purification of the overexpressed protein (Susanti et al., 2019). *M. jannaschii* is transformed using heat shock without the need for chemical treatments, such as PEG in *M. maripaludis* S2 or liposomes in *M. acetivorans* (Susanti et al., 2019). This system is not yet as advanced as for the methanogens described above since only linearized suicide vectors have been used, which does not allow the investigation of potentially essential genes and is not compatible with most strategies for markerless mutagenesis. Additionally, heterologous protein expression in *M. jannaschii* has not yet been reported.

Methanothermobacter marburgensis and *Methanothermobacter thermautotrophicus* ΔH are two highly similar *Methanobacteriales* methanogens that were used as model hydrogenotrophic methanogens during early biochemical investigation of methanogenesis and MCR (Thauer, 1998). *Methanothermobacter* species have also been successfully employed in bioreactors for efficient methane production processes (Thema et al., 2019; Pfeifer et al., 2021). Thus, developing genetic manipulation tools for these organisms has been a major area of interest. Very recently, the first reliable system for *M. thermautotrophicus* ΔH was reported (Fink et al., 2021). *Methanothermobacter thermautotrophicus* ΔH was found to be amenable to plating on solid medium with good efficiencies, and neomycin was shown to be effective for selection. This allowed for the construction of a shuttle vector containing a full array of cloning sites, selections markers for *E. coli* and *M. thermautotrophicus* ΔH, and the β-galactosidase-encoding gene *bgaB,* which can be used as a reporter (Fink et al., 2021). Transformation of *M. thermautotrophicus* ΔH is possible *via* interdomain conjugation with *E. coli* S17-1. As a proof-of-concept, this system was utilized to heterologously express formate dehydrogenase, thus enabling *M. thermautotrophicus* ΔH to grow with formate as a substrate (Fink et al., 2021). This genetic system could potentially be very useful for the future heterologous expression of MCRs since established methods for MCR activation in *M. marburgensis* (see section

below) would likely also be effective for MCRs isolated from *M. thermautotrophicus* ΔH.

CONSIDERATIONS FOR PROMOTERS, OPERON ORGANIZATION, AND PTMs FOR RECOMBINANT MCRs

An important factor to consider when designing an expression construct for recombinant MCR expression is which promoter to use. Promoters are essential for transcription and translation as, like eukarya, archaea possess more complex machinery for translation compared to bacteria (Lyu and Whitman, 2017). Common promoters used for heterologous protein expression in *M. acetivorans* include the MCR promoter (P*mcrB*) and associated tetracycline-inducible forms (Buan et al., 2011), while in *M. maripaludis* the histone promoter A (P*hmvA*) is often used (Sarmiento et al., 2011). An inducible *nif* promoter has also been used for *M. maripaludis* (Lie and Leigh, 2002; Chaban et al., 2007) and, very recently, an inducible expression system based on phosphate limitation was developed (Akinyemi et al., 2021). Notably, a plasmid containing the P*pst* promoter, which becomes activated under phosphate-limiting conditions, was used to express a recombinant *M. maripaludis* MCR in *M. maripaludis*, which represented 6% of the total protein content in a cell-free extract (140% increase compared to when using P*hmva*; Akinyemi et al., 2021). Additionally, the *mmpX* gene, which encodes the radical SAM methylase responsible for the methyl-arginine modification in McrA was successfully expressed. This is a particularly interesting result, as the expression of this protein using the constitutive promoter P*hmva* results in very low protein yields since it is apparently toxic to the cells (Akinyemi et al., 2021). Besides using previously employed constitutive or inducible promoters for recombinant MCR expression, another possibility is to utilize the native promoter for the MCR of interest. This has been done previously for the heterologous expression of ANME-1 MCR in *M. acetivorans* (Soo et al., 2016; discussed more below). If this strategy is chosen, it would be important to consider whether the heterologous host transcription/translation machinery is able to recognize the essential promoter elements present in the foreign promoter.

When designing a plasmid construct for heterologous MCR expression, it is also important to consider operon organization. ANME MCR operons generally lack one or both accessory proteins (e.g., *mcrBGA*—ANME-1 or *mcrBDGA*—*Ca.* M. nitroreducens) in the operon, while methanogens will always contain at least one MCR operon with both accessory proteins within the operon (**Figure 5**). Although it is still unclear how accessory proteins affect MCR assembly and/or activation, generating constructs for heterologous expression with accessory proteins may be necessary, even in the case where they are not present in the MCR operon of interest, especially since the accessory proteins may be organism-specific. Additionally, considering the post-translational machinery present in the heterologous host of choice will be important, as well as whether those enzymes will effectively recognize and correctly modify the recombinant enzyme. Since not all PTMs are consistent across different MCRs, additional and/or replacement PTM genes may need to be incorporated, which may potentially be toxic to the host.

METHODS FOR OBTAINING AN ACTIVATED MCR

Assuming successful expression, assembly, and post-translational modification of a recombinant MCR, the next consideration is obtaining the active form of the enzyme. The three oxidation states of F_{430} are well-described and can be observed *via* electron paramagnetic resonance (EPR) spectroscopy and UV–Vis spectrophotometry. These include the active MCR_{red1} in Ni(I) form, the inactive MCR_{silent} in the Ni(II) form, and the "ready" MCR_{ox1} in the Ni(III) form (Goubeaud et al., 1997; Duin et al., 2011). Many isolation and purification procedures will yield MCR with F_{430} in an inactive Ni(II) state, thus representing a major limitation for the enzymatic investigation of various MCRs.

The first successful MCR activation procedure involved incubating *M. marburgensis* cells with 100% H_2 prior to harvesting and including 10 mM CH_3-S-CoM to stabilize the enzyme during purification (Rospert et al., 1991). This resulted in MCR almost entirely in MCR_{red1} state, which can be used for subsequent enzymatic assays. Subsequent work has shown that HS-CoM can be used instead of CH_3-S-CoM to achieve activation (Duin et al., 2011). Other successful efforts to activate MCR from *M. marburgensis* involved the treatment of cells with 80% N_2/20% CO_2 prior to harvesting, which resulted in the MCR_{ox1} form of the enzyme (Goubeaud et al., 1997). Upon incubation of purified MCR_{ox1} with Ti(III) citrate, EPR spectra revealed the conversion from MCR_{ox1} to MCR_{red1}, and specific activity of the enzyme raised from 2 U/mg protein to 100 U/mg protein. The advantage of isolating MCR in the MCR_{ox1} form is that this strategy minimizes oxidation of MCR_{red1} to the MCR_{silent} Ni(II) state, which cannot be reduced to the active form *in vitro* with chemical reductants. Another protocol for activation of MCR from *M. marburgensis* involved treatment of cells with CO, which was shown to activate MCR at a significantly faster rate than treatment with H_2 (Zhou et al., 2013). CO activation resulted in MCR_{red1} within 1 h of incubation, while H_2 treatment required overnight incubation. Although effective, these described MCR activation protocols have primarily only been used successfully for MCR isolated from *M. marburgensis.*

Toward developing strategies for activating MCR from other organisms, it was reasoned that MCR activation could be achieved by controlling the ligation state of nickel through the addition of different chemical agents (Becker and Ragsdale, 1998). Thus, sodium sulfide was added to *Methanosarcina thermophila* cells prior to harvesting, which successfully elicited the MCR_{ox1} state (Becker and Ragsdale, 1998). Using ^{35}S-labeled sulfide, the authors demonstrated that the sulfide enters the cell, binds to the nickel site in F_{430}, and remains bound during purification. The amount of MCR_{ox1} was correlated to the amount ^{35}S-labeled sulfide. Thus, based on the available protocols described so far, isolation of MCR as MCR_{ox1} using sodium sulfide treatment

will likely be the most widely applicable to different MCRs isolated from various organisms. This is because ligating nickel to control the oxidation state should be independent from the metabolic state of the cell, which would allow for the control of the MCR oxidation state regardless of which organism and/or which methanogenic substrate is used.

In vivo MCR activation remains a poorly understood process. Early work provided initial insights into the cellular components responsible for the reduction of CH_3-S-CoM to methane (see (Thauer, 2019) for a comprehensive discussion of these experiments). In 2014, complete activation of MCR_{ox1} and 65% activation of MCR_{silent} was achieved in the presence of dithiothreitol, ATP, component A2, and component A3a (Prakash et al., 2014). An important discovery for this work was that the heterodisulfide product promotes the inactivation of MCR and thus it is essential to isolate the activation process from the methane formation reaction (Prakash et al., 2014). Further, the authors characterized component A3a as a 700 kDa complex that includes an assortment of redox proteins as well as McrC (Prakash et al., 2014). Any attempts to activate MCR with a smaller version of this complex were unsuccessful. Although much is still unclear about the structural basis of the complex and how it operates to activate MCR, this is a major advancement toward understanding how methanogens are able to supply low potential electrons to reduce F_{430} to the active Ni(I). Additionally, the presence of McrC in this complex finally linked a functional role to McrC, one of two accessory proteins within methanogenic MCR operons. It will be essential to fully elucidate the activation proteins and cofactors required as well as to obtain information about specificity in order to engineer an effective heterologous host for expression of active MCRs.

EXAMPLES OF MCR RECOMBINANT EXPRESSION

The first reported example of the heterologous expression of a recombinant MCR in a methanogen was for the ANME-1 MCR from the Black Sea mat (the same ANME-1 MCR for which the crystal structure is solved), which was expressed in *M. acetivorans* to engineer the methanogen to perform reverse methanogenesis using Fe(III) as an electron acceptor (Soo et al., 2016). The authors found that the strain expressing ANME-1 MCR consumed almost two times more methane compared to *M. acetivorans* with an empty vector. A further engineered air-adapted strain of *M. acetivorans* (Jasso-Chávez et al., 2015) containing ANME-1 MCR was used to generate electricity from methane in a microbial fuel cell containing other engineered microbes (McAnulty et al., 2017). Despite the critical importance of these studies toward the goal of activating methane for a range of biotechnology applications, it is still unclear whether the recombinant ANME MCR was necessary to significantly facilitate methane oxidation or, on a more fundamental level, how much of the recombinant MCR was produced in a complete and active form, especially since the methanogen lacks the 17^2-methylthio-F_{430} utilized by ANME-1 MCR. Notably, another investigation demonstrated that wild-type *M. acetivorans* is also capable of growth on methane using Fe(III) as an electron acceptor (Yan et al., 2018), indicating that the native methanogenic MCR is also capable of methane oxidation.

Another significant study described the heterologous expression of the MCR from *M. okinawensis* in *M. maripaludis* (Lyu et al., 2018b). The recombinant MCR was cloned into the traditional *M. maripaludis* protein expression plasmid under the control of P*hmvA* and with a his-tag on the C-terminus of McrA. This resulted in a highly expressed and uniformly assembled recombinant MCR as determined *via* SDS-PAGE and MALDI-MS. Additionally, expression of a MCR hybrid construct, consisting of *mcrBDCG* from *M. okinawensis* and *mcrA* from *M. maripaludis* resulted in a MCR consisting of the exact gene products from the hybrid construct without any components from the chromosomally encoded MCR. This supports the idea of an ordered MCR assembly, where MCR is simultaneously transcribed and translated (Lyu et al., 2018b). The PTMs were shown to be installed correctly for the heterologously produced *M. okinawensis* MCR and ~20% of the recombinant enzyme contained F_{430}. Interestingly, the portion of the recombinant MCR that did not contain F_{430} was found to be associated with McrD, while the portion that did have F_{430} largely lacked McrD. This further supports a role for McrD in F_{430} delivery. Since only a fraction of the recombinant MCR contained F_{430}, this suggests that potentially the F_{430} biosynthesis machinery cannot keep up with supplying F_{430} to the additional MCR being produced in the cell. Alternatively, the protein(s) potentially required to interact with McrD and/or MCR for F_{430} incorporation may not recognize the non-native McrD/MCR. If F_{430} biosynthesis was the bottleneck, one may expect that the native MCR would also have lower F_{430} incorporation. The authors found that the native MCR expression was not significantly affected in the presence of the recombinant MCR, but they did not report whether the F_{430} incorporation into native MCR was impaired. Finally, the purified heterologously produced MCR was found to exhibit low but detectable methane formation activity, where the authors point out that methods to activate *M. marburgensis* MCR *in vitro* do not appear to be effective for methanococcal MCR (Lyu et al., 2018b). It is important to note that *M okinawensis* and *M. maripaludis* are very closely related organisms, so it is unclear what the threshold of relatedness will be with respect to heterologous expression of MCRs from diverse organisms in model methanogens, such as *M. maripaludis* and *M. acetivorans*.

CONCLUSION

Although the field has seen many significant advancements since the initial discovery of MCR, many questions remain that need to be addressed, especially toward the development of robust heterologous expression systems for diverse MCRs and ACRs. Specifically, the functions of unique PTMs and F_{430} modifications need to be elucidated, which will require

in vitro kinetic studies with mutated MCRs in the presence vs. absence of modified F_{430}s. Further, the activation of MCR remains a poorly understood process, including the roles of specific proteins identified in complex A3a (Prakash et al., 2014) as well as the ATP costs of this process. There is still very little known about the putative accessory proteins necessary for MCR assembly—McrD may serve as an F_{430} chaperone for delivery to the MCR active site, but no other proteins potentially involved in assembly have been discovered. In terms of MCR kinetics and mechanism, it is important to emphasize that the vast majority of what is known results from studies on a single MCR from *M. marburgensis*. Since the substrate specificity and catalytic efficiency of even closely related enzymes can vary, it will be important to develop tools to study the enzymatic capabilities of other MCRs and related ACRs.

AUTHOR CONTRIBUTIONS

AG and KA wrote and edited the manuscript. All authors contributed to the article and approved the submitted version.

FUNDING

MCR research in the Allen lab is funded by the DOE Office of Science (DE-SC0022338).

ACKNOWLEDGMENTS

Thank you to the reviewers for their critical comments and helpful suggestions that greatly improved the manuscript.

REFERENCES

Adam, P. S., Borrel, G., Brochier-Armanet, C., and Gribaldo, S. (2017). The growing tree of Archaea: new perspectives on their diversity, evolution and ecology. *ISME J.* 11, 2407–2425. doi: 10.1038/ismej.2017.122

Akinyemi, T. S., Shao, N., Lyu, Z., Drake, I. J., Liu, Y., and Whitman, W. B. (2021). Tuning gene expression by phosphate in the methanogenic archaeon *Methanococcus maripaludis*. *ACS Synth. Biol.* 10, 3028–3039. doi: 10.1021/acssynbio.1c00322

Allen, K. D., Wegener, G., and White, R. H. (2014). Discovery of multiple modified F_{430} coenzymes in methanogens and anaerobic methanotrophic archaea suggests possible new roles for F_{430} in nature. *Appl. Environ. Microbiol.* 80, 6403–6412. doi: 10.1128/AEM.02202-14

Bao, J., and Scheller, S. (2021). Efficient CRISPR/Cas12a-based genome editing toolbox for metabolic engineering in *Methanococcus maripaludis*. [BioRxiv] doi: 10.1101/2021.12.29.474413

Bapteste, É., Brochier, C., and Boucher, Y. (2005). Higher-level classification of the Archaea: evolution of methanogenesis and methanogens. *Archaea* 1:859728, 353–363. doi: 10.1155/2005/859728

Battistuzzi, F. U., Feijao, A., and Hedges, S. B. (2004). A genomic timescale of prokaryote evolution: insights into the origin of methanogenesis, phototrophy, and the colonization of land. *BMC Evol. Biol.* 4:44. doi: 10.1186/1471-2148-4-44

Becker, D. F., and Ragsdale, S. W. (1998). Activation of methyl-SCoM reductase to high specific activity after treatment of whole cells with sodium sulfide. *Biochemistry* 37, 2639–2647. doi: 10.1021/bi972145x

Bhattarai, S., Cassarini, C., and Lens, P. N. L. (2019). Physiology and distribution of Archaeal Methanotrophs that couple anaerobic oxidation of methane with sulfate reduction. *Microbiol. Mol. Biol. Rev.* 83:e00074-18. doi: 10.1128/MMBR.00074-18

Bobik, T. A., Olson, K. D., Noll, K. M., and Wolfe, R. S. (1987). Evidence that the heterodisulfide of coenzyme M and 7-mercaptoheptanoylthreonine phosphate is a product of the methylreductase reaction in Methanobacterium. *Biochem. Biophys. Res. Commun.* 149, 455–460. doi: 10.1016/0006-291X(87)90389-5

Bonacker, L. G., Baudner, S., Mörschel, E., Bocher, R., and Thauer, R. K. (1993). Properties of the two isoenzymes of methyl-coenzyme M reductase in *Methanobacterium thermoautotrophicum*. *Eur. J. Biochem.* 217, 587–595. doi: 10.1111/j.1432-1033.1993.tb18281.x

Bonacker, L. G., Baudner, S., and Thauer, R. K. (1992). Differential expression of the two methyl-coenzyme M reductases in *Methanobacterium thermoautotrophicum* as determined immunochemically via isoenzyme-specific antisera. *Eur. J. Biochem.* 206, 87–92. doi: 10.1111/j.1432-1033.1992.tb16904.x

Borrel, G., Adam, P. S., McKay, L. J., Chen, L.-X., Sierra-García, I. N., Sieber, C. M. K., et al. (2019). Wide diversity of methane and short-chain alkane metabolisms in uncultured archaea. *Nat. Microbiol.* 4, 603–613. doi: 10.1038/s41564-019-0363-3

Borrel, G., Brugère, J. F., Gribaldo, S., Schmitz, R. A., and Moissl-Eichinger, C. (2020). The host-associated archaeome. *Nat. Rev. Microbiol.* 18, 622–636. doi: 10.1038/s41579-020-0407-y

Boyd, J. A., Jungbluth, S. P., Leu, A. O., Evans, P. N., Woodcroft, B. J., Chadwick, G. L., et al. (2019). Divergent methyl-coenzyme M reductase genes in a deep-subseafloor Archaeoglobi. *ISME J.* 13, 1269–1279. doi: 10.1038/s41396-018-0343-2

Buan, N., Kulkarni, G., and Metcalf, W. (2011). Genetic methods for Methanosarcina species. *Methods Enzymol.* 494, 23–42. doi: 10.1016/B978-0-12-385112-3.00002-0

Bult, C. J., White, O., Olsen, G. J., Zhou, L., Fleischmann, R. D., Sutton, G. G., et al. (1996). Complete genome sequence of the Methanogenic Archaeon, *Methanococcus jannaschii*. *Science* 273, 1058–1073. doi: 10.1126/science.273.5278.1058

Cedervall, P. E., Dey, M., Li, X., Sarangi, R., Hedman, B., Ragsdale, S. W., et al. (2011). Structural analysis of a Ni-methyl species in methyl-coenzyme M reductase from *Methanothermobacter marburgensis*. *J. Am. Chem. Soc.* 133, 5626–5628. doi: 10.1021/ja110492p

Chaban, B., Ng, S. Y., Kanbe, M., Saltzman, I., Nimmo, G., Aizawa, S., et al. (2007). Systematic deletion analyses of the fla genes in the flagella operon identify several genes essential for proper assembly and function of flagella in the archaeon, *Methanococcus maripaludis*. *Mol. Microbiol.* 66, 596–609. doi: 10.1111/j.1365-2958.2007.05913.x

Chen, S. L., Blomberg, M. R., and Siegbahn, P. E. (2012). How is methane formed and oxidized reversibly when catalyzed by Ni-containing methyl-coenzyme M reductase? *Chemistry* 18, 6309–6315. doi: 10.1002/chem.201200274

Chen, S. L., Blomberg, M. R., and Siegbahn, P. E. (2014). An investigation of possible competing mechanisms for Ni-containing methyl-coenzyme M reductase. *Phys. Chem. Chem. Phys.* 16, 14029–14035. doi: 10.1039/c4cp01483a

Chen, H., Gan, Q., and Fan, C. (2020). Methyl-coenzyme M Reductase and its post-translational modifications. *Front. Microbiol.* 11:578356. doi: 10.3389/fmicb.2020.578356

Chen, S. C., Musat, N., Lechtenfeld, O. J., Paschke, H., Schmidt, M., Said, N., et al. (2019). Anaerobic oxidation of ethane by archaea from a marine hydrocarbon seep. *Nature* 568, 108–111. doi: 10.1038/s41586-019-1063-0

Conrad, R. (2009). The global methane cycle: recent advances in understanding the microbial processes involved. *Environ. Microbiol. Rep.* 1, 285–292. doi: 10.1111/j.1758-2229.2009.00038.x

Conrado, R. J., and Gonzalez, R. (2014). Envisioning the bioconversion of methane to liquid fuels. *Science* 343, 621–623. doi: 10.1126/science.1246929

Costa, K. C., and Leigh, J. A. (2014). Metabolic versatility in methanogens. *Curr. Opin. Biotechnol.* 29, 70–75. doi: 10.1016/j.copbio.2014.02.012

Costa, K. C., Wong, P. M., Wang, T., Lie, T. J., Dodsworth, J. A., Swanson, I., et al. (2010). Protein complexing in a methanogen suggests electron bifurcation and electron delivery from formate to heterodisulfide reductase. *Proc. Natl. Acad. Sci. U. S. A.* 107, 11050–11055. doi: 10.1073/pnas.1003653107

Deobald, D., Adrian, L., Schöne, C., Rother, M., and Layer, G. (2018). Identification of a unique radical SAM methyltransferase required for the sp(3)-C-methylation

of an arginine residue of methyl-coenzyme M reductase. *Sci. Rep.* 8:7404. doi: 10.1038/s41598-018-25716-x

Dey, M., Li, X., Zhou, Y., and Ragsdale, S. W. (2010). Evidence for organometallic intermediates in bacterial methane formation involving the nickel coenzyme F_{430}. *Met. Ions Life Sci.* 7, 71–110. doi: 10.1039/BK9781847551771-00071

Dolfing, J., Jiang, B., Henstra, A. M., Stams, A. J., and Plugge, C. M. (2008). Syntrophic growth on formate: a new microbial niche in anoxic environments. *Appl. Environ. Microbiol.* 74, 6126–6131. doi: 10.1128/AEM.01428-08

Duin, E. C., Prakash, D., and Brungess, C. (2011). Methyl-coenzyme M reductase from *Methanothermobacter marburgensis*. *Methods Enzymol.* 494, 159–187. doi: 10.1016/B978-0-12-385112-3.00009-3

Duin, E. C., Wagner, T., Shima, S., Prakash, D., Cronin, B., Yáñez-Ruiz, D. R., et al. (2016). Mode of action uncovered for the specific reduction of methane emissions from ruminants by the small molecule 3-nitrooxypropanol. *Proc. Natl. Acad. Sci. U. S. A.* 113, 6172–6177. doi: 10.1073/pnas.1600298113

Ehlers, C., Jäger, D., and Schmitz, R. A. (2011). Establishing a markerless genetic exchange system for *Methanosarcina mazei* strain Gö1 for constructing chromosomal mutants of small RNA genes. *Archaea* 2011:439608. doi: 10.1155/2011/439608

Ehlers, C., Weidenbach, K., Veit, K., Deppenmeier, U., Metcalf, W. W., and Schmitz, R. A. (2005). Development of genetic methods and construction of a chromosomal glnK1 mutant in *Methanosarcina mazei* strain Gö1. *Mol. Gen. Genomics.* 273, 290–298. doi: 10.1007/s00438-005-1128-7

Ellefson, W. L., Whitman, W. B., and Wolfe, R. S. (1982). Nickel-containing factor F_{430}: chromophore of the methylreductase of Methanobacterium. *Proc. Natl. Acad. Sci. U. S. A.* 79, 3707–3710. doi: 10.1073/pnas.79.12.3707

Ellermann, J., Hedderich, R., Böcher, R., and Thauer, R. K. (1988). The final step in methane formation. Investigations with highly purified methyl-CoM reductase (component C) from *Methanobacterium thermoautotrophicum* (strain Marburg). *Eur. J. Biochem.* 172, 669–677. doi: 10.1111/j.1432-1033.1988.tb13941.x

Ellermann, J., Kobelt, A., Pfaltz, A., and Thauer, R. K. (1987). On the role of N-7-mercaptoheptanoyl-O-phospho-L-threonine (component B) in the enzymatic reduction of methyl-coenzyme M to methane. *FEBS Lett.* 220, 358–362. doi: 10.1016/0014-5793(87)80846-3

Ermler, U., Grabarse, W., Shima, S., Goubeaud, M., and Thauer, R. K. (1997). Crystal structure of methyl-coenzyme M reductase: the key enzyme of biological methane formation. *Science* 278, 1457–1462.

Evans, P. N., Parks, D. H., Chadwick, G. L., Robbins, S. J., Orphan, V. J., Golding, S. D., et al. (2015). Methane metabolism in the archaeal phylum Bathyarchaeota revealed by genome-centric metagenomics. *Science* 350, 434–438. doi: 10.1126/science.aac7745

Farber, G., Keller, W., Kratky, C., Jaun, B., Pfaltz, A., Spinner, C., et al. (1991). Coenzyme F_{430} from methanogenic bacteria–complete assignment of configuration based on an X-ray-analysis of 12,13-Diepi-F430 Pentamethyl Ester and on Nmr-spectroscopy. *Helv. Chim. Acta* 74, 697–716. doi: 10.1002/hlca.19910740404

Ferry, J. G. (2010). CO in methanogenesis. *Ann. Microbiol.* 60, 1–12. doi: 10.1007/s13213-009-0008-5

Fink, C., Beblawy, S., Enkerlin, A. M., Mühling, L., Angenent, L. T., Molitor, B., et al. (2021). A shuttle-vector system allows heterologous gene expression in the thermophilic methanogen *Methanothermobacter thermautotrophicus* ΔH. *mBio* 12:e0276621. doi: 10.1128/mBio.02766-21

Fonseca, D. R., Halim, M. F. A., Holten, M. P., and Costa, K. C. (2020). Type IV-like Pili facilitate transformation in naturally competent Archaea. *J. Bacteriol.* 202:e00355-20. doi: 10.1128/JB.00355-20

Goenrich, M., Duin, E. C., Mahlert, F., and Thauer, R. K. (2005). Temperature dependence of methyl-coenzyme M reductase activity and of the formation of the methyl-coenzyme M reductase red2 state induced by coenzyme B. *J. Biol. Inorg. Chem.* 10, 333–342. doi: 10.1007/s00775-005-0636-6

Goubeaud, M., Schreiner, G., and Thauer, R. K. (1997). Purified methyl-coenzyme-M reductase is activated when the enzyme-bound coenzyme F_{430} is reduced to the nickel(I) oxidation state by titanium(III) citrate. *Eur. J. Biochem.* 243, 110–114. doi: 10.1111/j.1432-1033.1997.00110.x

Grabarse, W., Mahlert, F., Shima, S., Thauer, R. K., and Ermler, U. (2000). Comparison of three methyl-coenzyme M reductases from phylogenetically distant organisms: unusual amino acid modification, conservation and adaptation. *J. Mol. Biol.* 303, 329–344. doi: 10.1006/jmbi.2000.4136.

Gunsalus, R. P., and Wolfe, R. S. (1976). Components and cofactors for enzymatic formation of methane from methyl-coenzyme-M. *Fed. Proc.* 35:1547

Guss, A. M., Rother, M., Zhang, J. K., Kulkkarni, G., and Metcalf, W. W. (2008). New methods for tightly regulated gene expression and highly efficient chromosomal integration of cloned genes for Methanosarcina species. *Archaea* 2, 193–203. doi: 10.1155/2008/534081

Hahn, C. J., Laso-Pérez, R., Vulcano, F., Vaziourakis, K. M., Stokke, R., Steen, I. H., et al. (2020). "Candidatus Ethanoperedens," a Thermophilic genus of Archaea mediating the anaerobic oxidation of ethane. *mBio* 11:e00600-20. doi: 10.1128/mBio.00600-20

Hahn, C. J., Lemaire, O. N., Kahnt, J., Engilberge, S., Wegener, G., and Wagner, T. (2021). Crystal structure of a key enzyme for anaerobic ethane activation. *Science* 373, 118–121. doi: 10.1126/science.abg1765

Hallam, S. J., Putnam, N., Preston, C. M., Detter, J. C., Rokhsar, D., Richardson, P. M., et al. (2004). Reverse methanogenesis: testing the hypothesis with environmental genomics. *Science* 305, 1457–1462. doi: 10.1126/science.1100025

Haroon, M. F., Hu, S., Shi, Y., Imelfort, M., Keller, J., Hugenholtz, P., et al. (2013). Anaerobic oxidation of methane coupled to nitrate reduction in a novel archaeal lineage. *Nature* 500, 567–570. doi: 10.1038/nature12375

Hartzell, P., and Reed, D. W. (2006). "The genus Archaeoglobus," in *The Prokaryotes: Archaea bacteria: Firmicutes, Actinomycetes*. *Vol. 3,* eds. M. Dworkin, S. Falkow, E. Rosenberg, K.-H. Schleifer and E. Stackebrandt (New York, NY: Springer), 82–100.

Haynes, C. A., and Gonzalez, R. (2014). Rethinking biological activation of methane and conversion to liquid fuels. *Nat. Chem. Biol.* 10, 331–339. doi: 10.1038/nchembio.1509

He, Z., Zhang, Q., Feng, Y., Luo, H., Pan, X., and Gadd, G. M. (2018). Microbiological and environmental significance of metal-dependent anaerobic oxidation of methane. *Sci. Total Environ.* 610-611, 759–768. doi: 10.1016/j.scitotenv.2017.08.140

Hinrichs, K.-U., Hayes, J. M., Sylva, S. P., Brewer, P. G., and DeLong, E. F. (1999). Methane-consuming archaebacteria in marine sediments. *Nature* 398, 802–805. doi: 10.1038/19751

Holler, T., Widdel, F., Knittel, K., Amann, R., Kellermann, M. Y., Hinrichs, K. U., et al. (2011). Thermophilic anaerobic oxidation of methane by marine microbial consortia. *ISME J.* 5, 1946–1956. doi: 10.1038/ismej.2011.77

Jackson, R. B., Saunois, M., Bousquet, P., Canadell, J. G., Poulter, B., Stavert, A. R., et al. (2020). Increasing anthropogenic methane emissions arise equally from agricultural and fossil fuel sources. *Environ. Res. Lett.* 15:071002. doi: 10.1088/1748-9326/ab9ed2

Jasso-Chávez, R., Santiago-Martinez, M. G., Lira-Silva, E., Pineda, E., Zepeda-Rodriguez, A., Belmont-Diaz, J., et al. (2015). Air-adapted *Methanosarcina acetivorans* shows high methane production and develops resistance against oxygen stress. *PLoS One* 10:e0117331. doi: 10.1371/journal.pone.0117331

Jones, W. J., Leigh, J. A., Mayer, F., Woese, C. R., and Wolfe, R. S. (1983). Methanococcus jannaschii sp. nov., an extremely thermophilic methanogen from a submarine hydrothermal vent. *Arch. Microbiol.* 136, 254–261. doi: 10.1007/BF00425213

Kahnt, J., Buchenau, B., Mahlert, F., Krüger, M., Shima, S., and Thauer, R. K. (2007). Post-translational modifications in the active site region of methyl-coenzyme M reductase from methanogenic and methanotrophic archaea. *FEBS J.* 274, 4913–4921. doi: 10.1111/j.1742-4658.2007.06016.x

Kaneko, M., Takano, Y., Chikaraishi, Y., Ogawa, N. O., Asakawa, S., Watanabe, T., et al. (2014). Quantitative analysis of coenzyme F_{430} in environmental samples: a new diagnostic tool for methanogenesis and anaerobic methane oxidation. *Anal. Chem.* 86, 3633–3638. doi: 10.1021/ac500305j

Kaster, A. K., Goenrich, M., Seedorf, H., Liesegang, H., Wollherr, A., Gottschalk, G., et al. (2011). More than 200 genes required for methane formation from H_2 and CO_2 and energy conservation are present in *Methanothermobacter marburgensis* and *Methanothermobacter thermautotrophicus*. *Archaea* 2011:973848. doi: 10.1155/2011/973848

Kirschke, S., Bousquet, P., Ciais, P., Saunois, M., Canadell, J. G., Dlugokencky, E. J., et al. (2013). Three decades of global methane sources and sinks. *Nat. Geosci.* 6, 813–823. doi: 10.1038/ngeo1955

Klenk, H.-P., Clayton, R. A., Tomb, J.-F., White, O., Nelson, K. E., Ketchum, K. A., et al. (1997). The complete genome sequence of the hyperthermophilic,

sulphate-reducing archaeon Archaeoglobus fulgidus. *Nature* 390, 364–370. doi: 10.1038/37052

Knittel, K., and Boetius, A. (2009). Anaerobic oxidation of methane: progress with an unknown process. *Annu. Rev. Microbiol.* 63, 311–334. doi: 10.1146/annurev.micro.61.080706.093130

Knittel, K., Lösekann, T., Boetius, A., Kort, R., and Amann, R. (2005). Diversity and distribution of methanotrophic archaea at cold seeps. *Appl. Environ. Microbiol.* 71, 467–479. doi: 10.1128/AEM.71.1.467-479.2005

Kurth, J. M., Muller, M. C., Welte, C. U., and Wagner, T. (2021b). Structural insights into the methane-generating enzyme from a methoxydotrophic methanogen reveal a restrained gallery of post-translational modifications. *Microorganisms*. 9:837. doi: 10.3390/microorganisms9040837

Kurth, J. M., Nobu, M. K., Tamaki, H., de Jonge, N., Berger, S., Jetten, M. S. M., et al. (2021a). Methanogenic archaea use a bacteria-like methyltransferase system to demethoxylate aromatic compounds. *ISME J.* 15, 3549–3565. doi: 10.1038/s41396-021-01025-6

Kurth, J. M., Op den Camp, H. J. M., and Welte, C. U. (2020). Several ways one goal-methanogenesis from unconventional substrates. *Appl. Microbiol. Biotechnol.* 104, 6839–6854. doi: 10.1007/s00253-020-10724-7

Laso-Pérez, R., Krukenberg, V., Musat, F., and Wegener, G. (2018). Establishing anaerobic hydrocarbon-degrading enrichment cultures of microorganisms under strictly anoxic conditions. *Nat. Protoc.* 13, 1310–1330. doi: 10.1038/nprot.2018.030

Laso-Pérez, R., Wegener, G., Knittel, K., Widdel, F., Harding, K. J., Krukenberg, V., et al. (2016). Thermophilic archaea activate butane via alkyl-coenzyme M formation. *Nature* 539, 396–401. doi: 10.1038/nature20152

Lawton, T. J., and Rosenzweig, A. C. (2016a). Biocatalysts for methane conversion: big progress on breaking a small substrate. *Curr. Opin. Chem. Biol.* 35, 142–149. doi: 10.1016/j.cbpa.2016.10.001

Lawton, T. J., and Rosenzweig, A. C. (2016b). Methane-oxidizing enzymes: an upstream problem in biological gas-to-liquids conversion. *J. Am. Chem. Soc.* 138, 9327–9340. doi: 10.1021/jacs.6b04568

Lie, T. J., and Leigh, J. A. (2002). Regulatory response of *Methanococcus maripaludis* to alanine, an intermediate nitrogen source. *J. Bacteriol.* 184, 5301–5306. doi: 10.1128/JB.184.19.5301-5306.2002

Liu, Y. F., Chen, J., Zaramela, L. S., Wang, L. Y., Mbadinga, S. M., Hou, Z. W., et al. (2020). Genomic and Transcriptomic evidence supports methane metabolism in *Archaeoglobi*. *mSystems* 5:e00651-19. doi: 10.1128/mSystems.00651-19

Liu, Y., and Whitman, W. B. (2008). Metabolic, phylogenetic, and ecological diversity of the methanogenic archaea. *Ann. N. Y. Acad. Sci.* 1125, 171–189. doi: 10.1196/annals.1419.019

Livingston, D. A., Pfaltz, A., Schreiber, J., Eschenmoser, A., Ankelfuchs, D., Moll, J., et al. (1984). Factor-F_{430} from Methanogenic bacteria-structure of the protein-free factor. *Helv. Chim. Acta* 67, 334–351. doi: 10.1002/hlca.19840670141

Long, F., Wang, L. L., Lupa, B., and Whitman, W. B. (2017). A flexible system for cultivation of *Methanococcus* and other formate-utilizing methanogens. *Archaea* 2017, 1–12. doi: 10.1155/2017/7046026

Lyu, Z., Chou, C. W., Shi, H., Wang, L., Ghebreab, R., Phillips, D., et al. (2018b). Assembly of methyl coenzyme M Reductase in the Methanogenic Archaeon *Methanococcus maripaludis*. *J. Bacteriol.* 200:e00746-17. doi: 10.1128/JB.00746-17

Lyu, Z., Shao, N., Akinyemi, T., and Whitman, W. B. (2018a). Methanogenesis. *Curr. Biol.* 28, R727–R732. doi: 10.1016/j.cub.2018.05.021

Lyu, Z., Shao, N., Chou, C. W., Shi, H., Patel, R., Duin, E. C., et al. (2020). Posttranslational methylation of arginine in methyl coenzyme M Reductase has a profound impact on both methanogenesis and growth of *Methanococcus maripaludis*. *J. Bacteriol.* 202:e00654-19. doi: 10.1128/JB.00654-19

Lyu, Z., and Whitman, W. B. (2017). Evolution of the archaeal and mammalian information processing systems: towards an archaeal model for human disease. *Cell. Mol. Life Sci.* 74, 183–212. doi: 10.1007/s00018-016-2286-y

Macauley, S. R., Zimmerman, S. A., Apolinario, E. E., Evilia, C., Hou, Y. M., Ferry, J. G., et al. (2009). The archetype gamma-class carbonic anhydrase (cam) contains iron when synthesized in vivo. *Biochemistry* 48, 817–819. doi: 10.1021/bi802246s

Mayr, S., Latkoczy, C., Krüger, M., Günther, D., Shima, S., Thauer, R. K., et al. (2008). Structure of an F430 variant from archaea associated with anaerobic oxidation of methane. *J. Am. Chem. Soc.* 130, 10758–10767. doi: 10.1021/ja802929z

Mayumi, D., Mochimaru, H., Tamaki, H., Yamamoto, K., Yoshioka, H., Suzuki, Y., et al. (2016). Methane production from coal by a single methanogen. *Science* 354, 222–225. doi: 10.1126/science.aaf8821

McAnulty, M. J., Poosarla, V. G., Kim, K. Y., Jasso-Chavez, R., Logan, B. E., and Wood, T. K. (2017). Electricity from methane by reversing methanogenesis. *Nat. Commun.* 8:15419. doi: 10.1038/ncomms15419

McBride, B. C., and Wolfe, R. S. (1971). A new coenzyme of methyl transfer, coenzyme M. *Biochemistry* 10, 2317–2324. doi: 10.1021/bi00788a022

McGlynn, S. E., Chadwick, G. L., Kempes, C. P., and Orphan, V. J. (2015). Single cell activity reveals direct electron transfer in methanotrophic consortia. *Nature* 526, 531–535. doi: 10.1038/nature15512

McKay, L. J., Dlakić, M., Fields, M. W., Delmont, T. O., Eren, A. M., Jay, Z. J., et al. (2019). Co-occurring genomic capacity for anaerobic methane and dissimilatory sulfur metabolisms discovered in the Korarchaeota. *Nat. Microbiol.* 4, 614–622. doi: 10.1038/s41564-019-0362-4

Metcalf, W. W., Zhang, J. K., Apolinario, E., Sowers, K. R., and Wolfe, R. S. (1997). A genetic system for Archaea of the genus *Methanosarcina*: liposome-mediated transformation and construction of shuttle vectors. *Proc. Natl. Acad. Sci. U. S. A.* 94, 2626–2631.

Meyerdierks, A., Kube, M., Kostadinov, I., Teeling, H., Glöckner, F. O., Reinhardt, R., et al. (2010). Metagenome and mRNA expression analyses of anaerobic methanotrophic archaea of the ANME-1 group. *Environ. Microbiol.* 12, 422–439. doi: 10.1111/j.1462-2920.2009.02083.x

Moissl-Eichinger, C., Pausan, M., Taffner, J., Berg, G., Bang, C., and Schmitz, R. A. (2018). Archaea are interactive components of complex microbiomes. *Trends Microbiol.* 26, 70–85. doi: 10.1016/j.tim.2017.07.004

Mondorf, S., Deppenmeier, U., and Welte, C. (2012). A novel inducible protein production system and neomycin resistance as selection marker for *Methanosarcina mazei*. *Archaea* 2012:973743. doi: 10.1155/2012/973743

Montzka, S. A., Dlugokencky, E. J., and Butler, J. H. (2011). Non-CO_2 greenhouse gases and climate change. *Nature* 476, 43–50. doi: 10.1038/nature10322

Moore, B. C., and Leigh, J. A. (2005). Markerless mutagenesis in Methanococcus maripaludis demonstrates roles for alanine dehydrogenase, alanine racemase, and alanine permease. *J. Bacteriol.* 187, 972–979. doi: 10.1128/JB.187.3.972-979.2005

Moore, S. J., Sowa, S. T., Schuchardt, C., Deery, E., Lawrence, A. D., Ramos, J. V., et al. (2017). Elucidation of the biosynthesis of the methane catalyst coenzyme F_{430}. *Nature* 543, 78–82. doi: 10.1038/nature21427

Mueller, T. J., Grisewood, M. J., Nazem-Bokaee, H., Gopalakrishnan, S., Ferry, J. G., Wood, T. K., et al. (2015). Methane oxidation by anaerobic archaea for conversion to liquid fuels. *J. Ind. Microbiol. Biotechnol.* 42, 391–401. doi: 10.1007/s10295-014-1548-7

Nauhaus, K., Boetius, A., Kruger, M., and Widdel, F. (2002). In vitro demonstration of anaerobic oxidation of methane coupled to sulphate reduction in sediment from a marine gas hydrate area. *Environ. Microbiol.* 4, 296–305. doi: 10.1046/j.1462-2920.2002.00299.x

Nayak, D. D., Liu, A., Agrawal, N., Rodriguez-Carerro, R., Dong, S. H., Mitchell, D. A., et al. (2020). Functional interactions between posttranslationally modified amino acids of methyl-coenzyme M reductase in *Methanosarcina acetivorans*. *PLoS Biol.* 18:e3000507. doi: 10.1371/journal.pbio.3000507

Nayak, D. D., Mahanta, N., Mitchell, D. A., and Metcalf, W. W. (2017). Post-translational thioamidation of methyl-coenzyme M reductase, a key enzyme in methanogenic and methanotrophic Archaea. *elife* 6. doi: 10.7554/eLife.29218

Nayak, D. D., and Metcalf, W. W. (2017). Cas9-mediated genome editing in the methanogenic archaeon *Methanosarcina acetivorans*. *Proc. Natl. Acad. Sci. U. S. A.* 114, 2976–2981. doi: 10.1073/pnas.1618596114

Nayak, D. D., and Metcalf, W. W. (2018). Genetic techniques for studies of methyl-coenzyme M reductase from *Methanosarcina acetivorans* C2A. *Methods Enzymol.* 613, 325–347. doi: 10.1016/bs.mie.2018.10.012

Niemann, H., Lösekann, T., de Beer, D., Elvert, M., Nadalig, T., Knittel, K., et al. (2006). Novel microbial communities of the Haakon Mosby mud volcano and their role as a methane sink. *Nature* 443, 854–858. doi: 10.1038/nature05227

Orphan, V. J., Hinrichs, K. U., Ussler, W. 3rd, Paull, C. K., Taylor, L. T., Sylva, S. P., et al. (2001). Comparative analysis of methane-oxidizing archaea and sulfate-reducing bacteria in anoxic marine sediments. *Appl. Environ. Microbiol.* 67, 1922–1934. doi: 10.1128/AEM.67.4.1922-1934.2001

Orphan, V. J., House, C. H., Hinrichs, K. U., McKeegan, K. D., and DeLong, E. F. (2002). Multiple archaeal groups mediate methane oxidation in anoxic cold seep sediments. *Proc. Natl. Acad. Sci. U. S. A.* 99, 7663–7668. doi: 10.1073/pnas.072210299

Patwardhan, A., Sarangi, R., Ginovska, B., Raugei, S., and Ragsdale, S. W. (2021). Nickel-Sulfonate mode of substrate binding for forward and reverse reactions of methyl-SCoM Reductase suggest a radical mechanism involving Long-range electron transfer. *J. Am. Chem. Soc.* 143, 5481–5496. doi: 10.1021/jacs.1c01086

Pelmenschikov, V., Blomberg, M. R., Siegbahn, P. E., and Crabtree, R. H. (2002). A mechanism from quantum chemical studies for methane formation in methanogenesis. *J. Am. Chem. Soc.* 124, 4039–4049. doi: 10.1021/ja011664r

Pelmenschikov, V., and Siegbahn, P. E. M. (2003). Catalysis by methyl-coenzyme M reductase: a theoretical study for heterodisulfide product formation. *J. Biol. Inorg. Chem.* 8, 653–662. doi: 10.1007/s00775-003-0461-8

Pfaltz, A., Jaun, B., Fassler, A., Eschenmoser, A., Jaenchen, R., Gilles, H. H., et al. (1982). Factor-F_{430} from *Methanogenic* bacteria–structure of the porphinoid ligand system. *Helv. Chim. Acta* 65, 828–865. doi: 10.1002/hlca.19820650320

Pfeifer, K., Ergal, İ., Koller, M., Basen, M., Schuster, B., and Simon, K.-M. R. (2021). Archaea biotechnology. *Biotechnol. Adv.* 47:107668. doi: 10.1016/j.biotechadv.2020.107668

Prakash, D., Wu, Y., Suh, S. J., and Duin, E. C. (2014). Elucidating the process of activation of methyl-coenzyme M reductase. *J. Bacteriol.* 196, 2491–2498. doi: 10.1128/JB.01658-14

Pritchett, M. A., Zhang, J. K., and Metcalf, W. W. (2004). Development of a Markerless genetic exchange method for *Methanosarcina acetivorans* C2A and its use in construction of new genetic tools for Methanogenic Archaea. *Appl. Environ. Microbiol.* 70, 1425–1433. doi: 10.1128/AEM.70.3.1425-1433.2004

Radle, M. I., Miller, D. V., Laremore, T. N., and Booker, S. J. (2019). Methanogenesis marker protein 10 (Mmp10) from *Methanosarcina acetivorans* is a radical S-adenosylmethionine methylase that unexpectedly requires cobalamin. *J. Biol. Chem.* 294, 11712–11725. doi: 10.1074/jbc.RA119.007609

Raghoebarsing, A. A., Pol, A., van de Pas-Schoonen, K. T., Smolders, A. J. P., Ettwig, K. F., Rijpstra, W. I. C., et al. (2006). A microbial consortium couples anaerobic methane oxidation to denitrification. *Nature* 440, 918–921. doi: 10.1038/nature04617

Ragsdale, S. W., Raugei, S., Ginovska, B., and Wongnate, T. (2017). Biochemistry of methyl-coenzyme M Reductase. *Rsc. Metallobio. Ser.* 10, 149–169. doi: 10.1039/9781788010580-00149

Rospert, S., Bocher, R., Albracht, S. P., and Thauer, R. K. (1991). Methyl-coenzyme M reductase preparations with high specific activity from H_2-preincubated cells of *Methanobacterium thermoautotrophicum*. *FEBS Lett.* 291, 371–375.

Rospert, S., Linder, D., Ellermann, J., and Thauer, R. K. (1990). Two genetically distinct methyl-coenzyme M reductases in *Methanobacterium thermoautotrophicum* strain Marburg and ΔH. *Eur. J. Biochem.* 194, 871–877. doi: 10.1111/j.1432-1033.1990.tb19481.x

Sarmiento, F., Leigh, J. A., and Whitman, W. B. (2011). Genetic systems for hydrogenotrophic methanogens. *Methods Enzymol.* 494, 43–73. doi: 10.1016/B978-0-12-385112-3.00003-2

Scheller, S., Goenrich, M., Boecher, R., Thauer, R. K., and Jaun, B. (2010). The key nickel enzyme of methanogenesis catalyses the anaerobic oxidation of methane. *Nature* 465, 606–608. doi: 10.1038/nature09015

Scheller, S., Goenrich, M., Thauer, R. K., and Jaun, B. (2013). Methyl-coenzyme M reductase from methanogenic archaea: isotope effects on the formation and anaerobic oxidation of methane. *J. Am. Chem. Soc.* 135, 14975–14984. doi: 10.1021/ja406485z

Shea, M. T., Walter, M. E., Duszenko, N., Ducluzeau, A.-L., Aldridge, J., King, S. K., et al. (2016). pNEB193-derived suicide plasmids for gene deletion and protein expression in the methane-producing archaeon, *Methanosarcina acetivorans*. *Plasmid* 84-85, 27–35. doi: 10.1016/j.plasmid.2016.02.003

Sherf, B. A., and Reeve, J. N. (1990). Identification of the mcrD gene product and its association with component C of methyl coenzyme M reductase in *Methanococcus vannielii*. *J. Bacteriol.* 172, 1828–1833.

Shima, S., Krueger, M., Weinert, T., Demmer, U., Kahnt, J., Thauer, R. K., et al. (2012). Structure of a methyl-coenzyme M reductase from Black Sea mats that oxidize methane anaerobically. *Nature* 481, 98–101. doi: 10.1038/nature10663

Shima, S., and Thauer, R. K. (2005). Methyl-coenzyme M reductase and the anaerobic oxidation of methane in methanotrophic Archaea. *Curr. Opin. Microbiol.* 8, 643–648. doi: 10.1016/j.mib.2005.10.002

Soo, V. W., McAnulty, M. J., Tripathi, A., Zhu, F., Zhang, L., Hatzakis, E., et al. (2016). Reversing methanogenesis to capture methane for liquid biofuel precursors. *Microb. Cell Factories* 15:11. doi: 10.1186/s12934-015-0397-z

Stokke, R., Roalkvam, I., Lanzen, A., Haflidason, H., and Steen, I. H. (2012). Integrated metagenomic and metaproteomic analyses of an ANME-1-dominated community in marine cold seep sediments. *Environ. Microbiol.* 14, 1333–1346. doi: 10.1111/j.1462-2920.2012.02716.x

Susanti, D., Frazier, M. C., and Mukhopadhyay, B. (2019). A genetic system for *Methanocaldococcus jannaschii*: An evolutionary deeply rooted Hyperthermophilic Methanarchaeon. *Front. Microbiol.* 10:1256. doi: 10.3389/fmicb.2019.01256

Thauer, R. K. (1998). Biochemistry of methanogenesis: a tribute to Marjory Stephenson prize lecture. *Microbiology* 144, 2377–2406. doi: 10.1099/00221287-144-9-2377

Thauer, R. K. (2019). Methyl (alkyl)-coenzyme M Reductases: nickel F_{430}-containing enzymes involved in anaerobic methane formation and in anaerobic oxidation of methane or of short chain alkanes. *Biochemistry* 58, 5198–5220. doi: 10.1021/acs.biochem.9b00164

Thema, M., Weidlich, T., Hörl, M., Bellack, A., Mörs, F., Hackl, F., et al. (2019). Biological CO2-methanation: an approach to standardization. *Energies* 12:1670. doi: 10.3390/en12091670

Timmers, P. H., Welte, C. U., Koehorst, J. J., Plugge, C. M., Jetten, M. S., and Stams, A. J. (2017). Reverse Methanogenesis and respiration in Methanotrophic Archaea. *Archaea* 2017:1654237. doi: 10.1155/2017/1654237

Tumbula, D. L., Makula, R. A., and Whitman, W. B. (1994). Transformation of *Methanococcus maripaludis* and identification of a Psti-Like restriction system. *FEMS Microbiol. Lett.* 121, 309–314. doi: 10.1111/j.1574-6968.1994.tb07118.x

Vanwonterghem, I., Evans, P. N., Parks, D. H., Jensen, P. D., Woodcroft, B. J., Hugenholtz, P., et al. (2016). Methylotrophic methanogenesis discovered in the archaeal phylum Verstraetearchaeota. *Nat. Microbiol.* 1:16170. doi: 10.1038/nmicrobiol.2016.170

Wagner, T., Kahnt, J., Ermler, U., and Shima, S. (2016). Didehydroaspartate modification in methyl-coenzyme M Reductase catalyzing methane formation. *Angew. Chem. Int. Ed. Engl.* 55, 10630–10633. doi: 10.1002/anie.201603882

Wagner, T., Wegner, C.-E., Kahnt, J., Ermler, U., and Shima, S. (2017). Phylogenetic and structural comparisons of the three types of methyl coenzyme M Reductase from Methanococcales and Methanobacteriales. *J. Bacteriol.* 199:e00197-17. doi: 10.1128/JB.00197-17

Wang, Y., Wegener, G., Ruff, S. E., and Wang, F. (2021). Methyl/alkyl-coenzyme M reductase-based anaerobic alkane oxidation in archaea. *Environ. Microbiol.* 23, 530–541. doi: 10.1111/1462-2920.15057

Wegener, G., Krukenberg, V., Riedel, D., Tegetmeyer, H. E., and Boetius, A. (2015). Intercellular wiring enables electron transfer between methanotrophic archaea and bacteria. *Nature* 526, 587–590. doi: 10.1038/nature15733

Wegener, G., Krukenberg, V., Ruff, S. E., Kellermann, M. Y., and Knittel, K. (2016). Metabolic capabilities of microorganisms involved in and associated with the anaerobic oxidation of methane. *Front. Microbiol.* 7:46. doi: 10.3389/fmicb.2016.00046

Whitman, W. B., Shieh, J., Sohn, S., Caras, D. S., and Premachandran, U. (1986). Isolation and characterization of 22 Mesophilic Methanococci. *Syst. Appl. Microbiol.* 7, 235–240. doi: 10.1016/S0723-2020(86)80012-1

Whitman, W. B., Tumbula, D. L., Yu, J. P., and Kim, W. (1997). Development of genetic approaches for the methane-producing archaebacterium *Methanococcus maripaludis*. *Biofactors* 6, 37–46.

Wongnate, T., Sliwa, D., Ginovska, B., Smith, D., Wolf, M. W., Lehnert, N., et al. (2016). The radical mechanism of biological methane synthesis by methyl-coenzyme M reductase. *Science* 352, 953–958. doi: 10.1126/science.aaf0616

Yan, Z., and Ferry, J. G. (2018). Electron bifurcation and confurcation in methanogenesis and reverse methanogenesis. *Front. Microbiol.* 9:1322. doi: 10.3389/fmicb.2018.01322

Yan, Z., Joshi, P., Gorski, C. A., and Ferry, J. G. (2018). A biochemical framework for anaerobic oxidation of methane driven by Fe(III)-dependent respiration. *Nat. Commun.* 9:1642. doi: 10.1038/s41467-018-04097-9

Yang, N., Reiher, M., Wang, M., Harmer, J., and Duin, E. C. (2007). Formation of a nickel-methyl species in methyl-coenzyme m reductase, an enzyme catalyzing methane formation. *J. Am. Chem. Soc.* 129, 11028–11029. doi: 10.1021/ja0734501

Yu, G., Beauchemin, K. A., and Dong, R. (2021). A review of 3-Nitrooxypropanol for enteric methane mitigation from ruminant livestock. *Animals* 11:11. doi: 10.3390/ani11123540

Zheng, K., Ngo, P. D., Owens, V. L., Yang, X. P., and Mansoorabadi, S. O. (2016). The biosynthetic pathway of coenzyme F430 in methanogenic and methanotrophic archaea. *Science* 354, 339–342. doi: 10.1126/science.aag2947

Zhou, Y., Dorchak, A. E., and Ragsdale, S. W. (2013). In vivo activation of methyl-coenzyme M reductase by carbon monoxide. *Front. Microbiol.* 4:69. doi: 10.3389/fmicb.2013.00069

Zhou, Z., Zhang, C.-j., Liu, P.-f., Fu, L., Laso-Pérez, R., Yang, L., et al. (2022). Non-syntrophic methanogenic hydrocarbon degradation by an archaeal species. *Nature* 601, 257–262. doi: 10.1038/s41586-021-04235-2

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

frontiers | Frontiers in Microbiology

ORIGINAL RESEARCH
published: 27 April 2022
doi: 10.3389/fmicb.2022.867384

Carbohydrate Metabolism in Bacteria: Alternative Specificities in ADP-Glucose Pyrophosphorylases Open Novel Metabolic Scenarios and Biotechnological Tools

Jaina Bhayani[1], Maria Josefina Iglesias[2], Romina I. Minen[2], Antonela E. Cereijo[2], Miguel A. Ballicora[1], Alberto A. Iglesias[2] and Matias D. Asencion Diez[2]**

[1]*Department of Chemistry and Biochemistry, Loyola University Chicago, Chicago, IL, United States,* [2]*Facultad de Bioquímica y Ciencias Biológicas, Instituto de Agrobiotecnología del Litoral, Universidad Nacional del Litoral, Consejo Nacional de Investigaciones Científicas y Técnicas, Santa Fe, Argentina*

OPEN ACCESS

Edited by:
Ulrike Kappler, The University of Queensland, Australia

Reviewed by:
Adelfo Escalante, Universidad Nacional Autónoma de México, Mexico
Michael Benedik, Texas A&M University, United States

***Correspondence:**
Alberto A. Iglesias
iglesias@fbcb.unl.edu.ar
Matias D. Asencion Diez
masencion@fbcb.unl.edu.ar

Specialty section:
This article was submitted to Microbial Physiology and Metabolism, a section of the journal Frontiers in Microbiology

Received: *01 February 2022*
Accepted: *22 March 2022*
Published: *27 April 2022*

Citation:
Bhayani J, Iglesias MJ, Minen RI, Cereijo AE, Ballicora MA, Iglesias AA and Asencion Diez MD (2022) Carbohydrate Metabolism in Bacteria: Alternative Specificities in ADP-Glucose Pyrophosphorylases Open Novel Metabolic Scenarios and Biotechnological Tools. Front. Microbiol. 13:867384. doi: 10.3389/fmicb.2022.867384

We explored the ability of ADP-glucose pyrophosphorylase (ADP-Glc PPase) from different bacteria to use glucosamine (GlcN) metabolites as a substrate or allosteric effectors. The enzyme from the actinobacteria *Kocuria rhizophila* exhibited marked and distinctive sensitivity to allosteric activation by GlcN-6P when producing ADP-Glc from glucose-1-phosphate (Glc-1P) and ATP. This behavior is also seen in the enzyme from *Rhodococcus* spp., the only one known so far to portray this activation. GlcN-6P had a more modest effect on the enzyme from other Actinobacteria (*Streptomyces coelicolor*), Firmicutes (*Ruminococcus albus*), and Proteobacteria (*Agrobacterium tumefaciens*) groups. In addition, we studied the catalytic capacity of ADP-Glc PPases from the different sources using GlcN-1P as a substrate when assayed in the presence of their respective allosteric activators. In all cases, the catalytic efficiency of Glc-1P was 1–2 orders of magnitude higher than GlcN-1P, except for the unregulated heterotetrameric protein (GlgC/GgD) from *Geobacillus stearothermophilus*. The Glc-1P substrate preference is explained using a model of ADP-Glc PPase from *A. tumefaciens* based on the crystallographic structure of the enzyme from potato tuber. The substrate-binding domain localizes near the N-terminal of an α-helix, which has a partial positive charge, thus favoring the interaction with a hydroxyl rather than a charged primary amine group. Results support the scenario where the ability of ADP-Glc PPases to use GlcN-1P as an alternative occurred during evolution despite the enzyme being selected to use Glc-1P and ATP for α-glucans synthesis. As an associated consequence in such a process, certain bacteria could have improved their ability to metabolize GlcN. The work also provides insights in designing molecular tools for producing oligo and polysaccharides with amino moieties.

Keywords: glucose-1-phosphate, glucosamine-1-phosphate, glucosamine-6-phosphate, allosterism, promiscuity

INTRODUCTION

α-1,4-Glucans are biomolecules produced in several types of bacteria, including Gram-positive, Gram-negative, cyanobacteria, and archaebacteria (Preiss, 1984, 2009, 2014; Preiss and Romeo, 1990). Glycogen, an important α-glucan polysaccharide in bacteria, is synthesized by the classical *glgC-glgA* pathway, where ATP activates glucose-1P (Glc-1P) into ADP-Glc, the glycosyl donor for glucan elongation. These reactions are, respectively, catalyzed by ADP-Glc pyrophosphorylase (ADP-Glc PPase, EC 2.7.7.27) and glycogen synthase (EC 2.4.1.21). The pathway completes the formation of glycogen *via* the branching enzyme (EC 2.4.1.18), which introduces α-1,6-ramifications (Preiss, 2009; Cifuente et al., 2019). In this sequence of reactions, the limiting step is at the level of ADP-Glc PPase, which is regulated by metabolites from the major carbon utilization route in the organism (specified below) (Ballicora et al., 2003, 2004; Figueroa et al., 2021, 2022).

ADP-Glc PPases from numerous bacterial sources have been characterized concerning kinetic, regulatory, and structural properties (Ballicora et al., 2003, 2004; Figueroa et al., 2022). In Proteobacteria and Actinobacteria, the enzyme is a tetramer composed of a single subunit (codified by the *glgC* gene) and is allosterically regulated by critical intermediates of the central metabolism in the respective microorganism. For example, fructose-1,6-bisphosphate or fructose-6-phosphate (Fru-6P) and pyruvate (Pyr) are major activators of the enzyme from *Escherichia coli* (performing glycolysis by the Embden–Meyerhof pathway) and *Agrobacterium tumefaciens* (using the Entner–Doudoroff route), respectively (Ballicora et al., 2003, 2004). Also, the ADP-Glc PPase from Actinobacteria has multiple allosteric activators, notably glucose-6-phosphate (Glc-6P) (Asencion Diez et al., 2012, 2015; Cereijo et al., 2016, 2020), agreeing with the varied metabolic pathways having the hexose-P as a common starting point (Verma et al., 2013; Martín and Liras, 2020). Conversely, in Firmicutes (Gram-positive bacteria having low G + C content), two genes (*glgC* and *glgD*) codify for ADP-Glc PPase, giving rise to tetrameric GlgC or GlgC/GlgD active forms (Takata et al., 1997; Asención Diez et al., 2013a; Cereijo et al., 2018). GlgD is a protein lacking activity and is absent in Actinobacteria and Proteobacteria (Asención Diez et al., 2013a). The enzyme (either GlgC or GlgC/GlgD forms) from *Geobacillus stearothermophilus* (and the Bacillales group) is the only known ADP-Glc PPase exhibiting insensitivity to allosteric regulators to our knowledge (Takata et al., 1997). In other Firmicutes (Lactobacillales and Clostridiales), GlgC and GlgC/GlgD have distinct responses to specific allosteric regulators with the GlgD subunit providing a modulatory function (Asención Diez et al., 2013a; Cereijo et al., 2018). For example, the GlgC/GlgD from *Ruminococcus albus* exhibits higher catalytic efficiency than GlgC and phospho*enol*pyruvate (PEP) is the primary allosteric activator (Cereijo et al., 2018).

The context detailed above supports that ADP-Glc PPase was a target of evolution (Ballicora et al., 2003; Figueroa et al., 2022). This protein acquired sensitivity to varied allosteric effectors (based on the different metabolic pathways in diverse prokaryotes) to orchestrate the regulation of its activity. ADP-Glc PPase distinguishes itself from all other pyrophosphorylases because of its sensitivity to allosteric regulators. This modulatory characteristic is structurally justified by having a C-terminal (absent in other pyrophosphorylases) critical for regulatory function. It has been demonstrated that the tight interaction between the C- and the N-(catalytic) domains (Bejar et al., 2004) exerts allosteric control *via* a mechanism triggered by conformational changes of specific motifs in the enzyme (Figueroa et al., 2011, 2013; Asención Diez et al., 2013b, 2020). A detailed study (Ebrecht et al., 2017) evidenced that the ADP-Glc PPase from *E. coli* exhibits relative promiscuity for substrates (a common trait amongst catalytic proteins) (Jensen, 2003; Dönertaş et al., 2016; Copley and Copley, 2020) and that allosteric activation primarily drives the enzyme to select the production of ADP-Glc from Glc-1P and ATP. In other words, the evolutionary mechanism developed to regulate ADP-Glc PPases operated by adapting the protein to select specific substrates (Ebrecht et al., 2017).

Recent works have pointed out the potential relevance of glucosamine (GlcN, 2-Amino-2-deoxy-glucose) metabolism in actinobacteria of the genus *Rhodococcus* (Cereijo et al., 2016, 2020, 2021). ADP-Glc PPase from rhodococci can utilize GlcN-1P as a substrate and: (*i*) allosteric effectors enhance this activity and (*ii*) GlcN-6P is a primary allosteric activator (Cereijo et al., 2020). These findings were particularly relevant relating evolution and the mechanism for allosteric properties exhibited by this enzyme. These data challenged us to perform studies to explore the sensitivity of ADP-Glc PPases from other bacteria to the GlcN-P metabolites. Herein, we show that GlcN-6P is a common allosteric effector, with significant effect on the protein from Actinobacteria. In addition, GlcN-1P is a suitable substrate, giving rise to different catalytic efficiencies for the enzyme from Actinobacteria, Firmicutes, and Proteobacteria. Using the elucidated crystallographic structure of the enzyme from potato tuber as a starting model, we placed Glc-1P and GlcN-1P in the active site of the *A. tumefaciens* enzyme to discuss the metabolic evolution of GlcN in ADP-Glc PPase from prokaryotes. Finally, based on distinct enzymatic behavior, we discuss some putative applications where ADP-Glc PPases could be incorporated in the design and production of oligo- and polysaccharides with potential biotechnological interests.

MATERIALS AND METHODS

Chemicals, Bacterial Strains, and Plasmids

Chemicals used for enzymatic assays were from Sigma-Aldrich (St. Louis, MO, United States) or the highest quality available. *E. coli* Top 10 (Invitrogen) was used for plasmid maintenance, and proteins were expressed in *E. coli* BL21 (DE3) (Invitrogen) using the pET28b vector (Novagen). DNA manipulations, molecular biology techniques, and *E. coli* cultivation and transformation were performed according to standard protocols (Sambrook and Russell, 2001).

Site-Directed Mutagenesis

Mutagenesis was performed by using a Q5 Site-Directed Mutagenesis Kit from New England BioLabs. Oligonucleotides were synthesized by Integrated DNA Technologies (IDT,

San Diego). The primers listed were used to make E187A in ADP-Glc PPase from *A. tumefaciens*: E187A Forward: CGACTTCATCGCCAAGCCGGC and E187A Reverse: ATGATCTCGTCTTTTTCGTTCACATGC.

Enzyme Production and Purification

Proteins were obtained as previously described (Takata et al., 1997; Asencion Diez et al., 2012, 2020; Cereijo et al., 2018, 2020). Briefly, the general procedure involved a recombinant *E. coli* BL21 (DE3) (Invitrogen) transformed with a recombinant plasmid [in most cases, pET28b vector (Novagen)] harboring the desired *glgC* gene. Transformed cells were cultured at 37°C and 200 rpm in LB medium (10 g/l tryptone; 5 g/l yeast extract; and 10 g/l NaCl) containing the corresponding antibiotics at its proper concentration [kanamycin (50 μg/ml); ampicillin (100 μg/ml)] until optical density at 600 nm reached between ~0.6 and 0.8. Then, protein expression was induced by adding 0.1–0.4 mM isopropyl-β-D-1-thiogalactopyranoside (depending the protein), which continued for 16 h at 18°C–24°C. Then, recombinant cells were harvested with a 10-min centrifugation at room temperature and 5,000 × *g*, and the obtained pellets were stored at −20°C until used for purification purposes.

Immobilized metal affinity chromatography (IMAC) at 4°C was the procedure to purify the His-tagged proteins. Protein purification started with cells resuspension in *Buffer H* [50 mM Tris–HCl pH 8.0, 300 mM NaCl, 10 mM imidazole, and 5% (v/v) glycerol], disruption by sonication (4 s on; 2 s off, for a total time of 10 min in an ice-bath), and then centrifuging twice at 30,000 × *g* for 10 min. The resulting supernatant (crude extract) was loaded on a His-Trap column (1 ml size, GE Healthcare) that had been equilibrated with *Buffer H*. Elution of the recombinant proteins was produced using a linear imidazole gradient (10–300 mM) in the same buffer. Fractions exhibiting the highest ADP-Glc PPase activity were pooled and concentrated. After concentration, the ADP-Glc PPases from *Streptomyces coelicolor*, *Rhodococcus fascians*, and *R. albus* were dialyzed against *Buffer S* [50 mM HEPES-NaOH, 10% (w/v) sucrose, 0.2 mM DTT, and 1 mM EDTA]. The resulting purified enzymes were stored at −80°C until use. Under these conditions, enzymes remained active for 6 months (at least).

Protein Methods

Protein concentration was assayed by the Bradford method (Bradford, 1976), with the use of bovine serum albumin (BSA) as a standard. The purity of the recombinant proteins was assessed by sodium dodecyl sulfate–polyacrylamide gel electrophoresis (SDS-PAGE), as described elsewhere (Laemmli, 1970). After running, the gels (containing between 5 and 50 μg of protein per well) were revealed with Coomassie Brilliant Blue.

Enzyme Activity Assays

The ADP-Glc PPase activity was determined at 37°C in the direction of ADP-Glc synthesis following the formation of P_i after PP_i hydrolysis by inorganic pyrophosphatase, using the colorimetric method described before (Fusari et al., 2006). Reaction mixtures (a total volume of 50 μl) contained (unless otherwise specified) 50 mM MOPS-NaOH pH 8.0, 1–2 mM ATP, 10 mM $MgCl_2$, 0.5 U/ml yeast inorganic pyrophosphatase, 0.2 mg/ml BSA, and a proper enzyme dilution. The addition of 1–2 mM Glc-1P initiated the assays, which were run for 10 min, and terminated by the addition of 400 μl of the Malachite Green reagent. The complex formed with the released P_i was measured at 620 nm in a 96-well microplate reader (Multiskan GO, Thermo).

One unit of activity (U) is defined as the amount of enzyme catalyzing the formation of 1 μmol of product per min under the above-specified conditions.

Saturation curves were constructed by assaying enzyme activity at different concentrations of the variable substrate or effector, while the others remained at saturating levels. Plots of enzyme activity (U/mg) vs. substrate (or effector) concentration (mM) were used to calculate the kinetic constants by fitting the experimental data to a modified Hill equation (Ballicora et al., 2007). Data were fitted using the Levenberg–Marquardt nonlinear least-squares algorithm included in the computer program Origin 8.0 (OriginLab). Accordingly, we calculated the maximal velocity (V_{max}), the Hill coefficient (n_H), and the concentrations of activator ($A_{0.5}$) or substrate (K_m), giving 50% of the maximal activation or velocity, respectively. Kinetic constants were reproducible within a range of ±10%, being the mean of three independent sets of data.

Computational Methods

The *A. tumefaciens* ADP-Glc PPase structure with ADP-Glc was built using the following templates: the crystal structure of the small subunit of the potato (*Solanum tuberosum*) tuber in complex with ADP-Glc (PDB ID: 1YP4) and *A. tumefaciens* ADP-Glc PPase (PDB ID: 5W5R; Jin et al., 2005; Hill et al., 2019). The model was generated using MODELLER (Šali and Blundell, 1993; Eswar et al., 2007). Protein visualization was performed with the program Chimera (Pettersen et al., 2004).

RESULTS

GlcN-6P as a Bacterial ADP-Glc PPase Activator

Recently, it was found that GlcN-6P is a primary activator of the rhodococcal ADP-Glc PPase, improving the catalytic properties of both the canonical and the alternative substrates, Glc-1P and GlcN-1P, respectively (Cereijo et al., 2020). Given that GlcN-6P was reported as a key metabolite in some Gram-positive microorganisms (Jolly et al., 2000; Urem et al., 2016; van der Heul et al., 2018; Manso et al., 2019), we tested whether other ADP-Glc PPases from Actinobacteria, *S. coelicolor* and *Kocuria rhizophila*, are sensitive to the glucosamine derivative. GlcN-6P increased 4-fold the Glc-1P activity of the *S. coelicolor* enzyme (**Figure 1**; $A_{0.5}$ 3.55 ± 0.52 mM), which was a small effect compared to the 27-fold activation exerted by the major effector Glc-6P (Asencion Diez et al., 2012). On the other hand, **Figure 1** shows that the *K. rhizophila* ADP-Glc PPase responded to the GlcN-6P stimulus to the same extent as to Glc-6P (4–5-fold increase), thus behaving similarly to the

rhodococcal enzyme (Cereijo et al., 2020). In these experiments, the *R. fascians* enzyme was activated about 8-fold with either GlcN-6P or Glc-6P. Therefore, GlcN-6P could be ascribed as a common activator in actinobacterial ADP-Glc PPases, with varying intensities, depending on the source. Interestingly, the *S. coelicolor* ADP-Glc PPase is the only enzyme from this group showing a distinct level of activation for various molecules (Asencion Diez et al., 2012), since ADP-Glc PPases from *Mycobacterium tuberculosis* (Asencion Diez et al., 2015), *Rhodococci* (Cereijo et al., 2020), and *K. rhizophila* exhibit similar degrees of activation for each effector.

We then analyzed the GlcN-6P activation of the different quaternary structures (GlgC or GlgC/GlgD) of the ADP-Glc PPase from two Firmicutes species: *R. albus* and *G. stearothermophilus*. Both GlgC and GlgC/GlgD enzymes from *G. stearothermophilus* were insensitive to GlcN-6P (**Figure 1**), even when assayed at 10 mM, in agreement with the observed recalcitrant insensitivity of these proteins to regulation, constituting the exception among all ADP-Glc PPases (Takata et al., 1997; Ballicora et al., 2003; Asención Diez et al., 2013a; Cereijo et al., 2018). In contrast, both *R. albus* GlgC and GlgC/GlgD responded to GlcN-6P activation, although to a lesser extent (about one half) than their main effector PEP (Cereijo et al., 2018), as shown in **Figure 1**. We also assayed the effect of GlcN-6P on the *A. tumefaciens* ADP-Glc PPase and curiously found that it decreased activity about 30%, while allosteric controls with Fru-6P and Pyr increased activity 15–20-fold (Uttaro et al., 1998; Asencion Diez et al., 2020). Taken together, our results show that all regulated ADP-Glc PPase analyzed here responded to GlcN-6P to some extent.

GlcN-1P as a Substrate for ADP-Glc PPase

Recently, two NDP-sugar pyrophosphorylases were characterized with the ability to use GlcN-1P as substrate: one being specific for the amino sugar-1P and UTP (GalU2; Cereijo et al., 2021) and the other using GlcN-1P as an alternative substrate with ATP (Cereijo et al., 2020). These enzymes belong to the *Rhodococcus* genus and are the only known studies regarding GlcN-1P consumption (for further details, see Cereijo et al., 2021). Therefore, we analyzed GlcN-1P as a substrate for ADP-Glc PPases from different bacterial sources in the presence of their main activators as well as GlcN-6P. When possible, parameters were also obtained without effectors, particularly for the allosterically insensitive *G. stearothermophilus* GlgC and

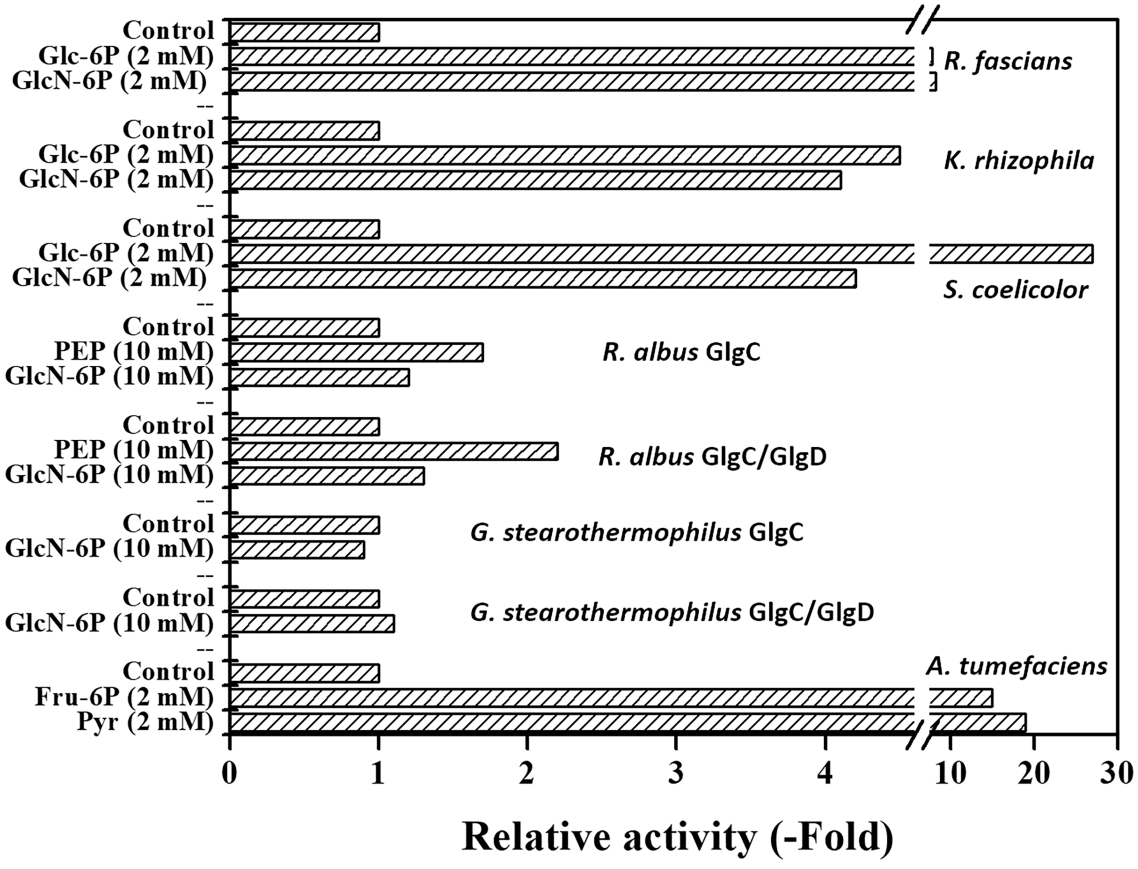


FIGURE 1 | GlcN-6P as effector of bacterial ADP-glucose pyrophosphorylases (ADP-Glc PPases). Activities were measured according to the section "Materials and Methods," using indicated effector concentrations.

GlgC/GlgD enzymes. Selectivity ratios, defined as the catalytic efficiencies of Glc-1P divided by GlcN-1P in the different bacterial ADP-Glc PPases, are shown in **Figure 2**. General kinetic parameters for the characterized enzymes are presented in **Supplementary Table 1**.

First, we studied GlcN-1P utilization by ADP-Glc PPase from the actinobacteria *S. coelicolor* and *K. rhizophila*. Both were active with the alternative substrate, although with values in the mU/mg order of magnitude (9 and 30 mU/mg for the enzymes from *S. coelicolor* and *K. rhizophila*, respectively). The *S. coelicolor* ADP-Glc PPase poorly used GlcN-1P and only showed a low activation degree in the presence of Glc-6P and GlcN-6P. GlcN-6P barely doubled the specific activity, and while Glc-6P increased activity 4-fold, it remained in the mU/mg order (**Supplementary Figure 1**). In contrast, the *K. rhizophila* enzyme behaved similarly to the rhodococcal ADP-Glc PPases (Cereijo et al., 2020), responding to a 2-mM GlcN-6P stimulus for both the sugar phosphate and amino-sugar phosphate as substrates. With Glc-1P, the activator increased the V_{max} 4-fold with an $A_{0.5}$ value of 0.08 mM. Remarkably, GlcN-6P (with an $A_{0.5}$ of 0.84 mM) increased GlcN-1P activity by two orders of magnitude, reaching up to ~20 U/mg. In terms of catalytic efficiency, the *K. rhizophila* ADP-Glc PPase used GlcN-1P with an efficiency that is almost one order of magnitude lower compared to Glc-1P (**Figure 2**). However, the efficiency of the *Kocuria* enzyme for GlcN-1P is 16-fold higher than the one of rhodococcal ADP-Glc PPases, making it a better biocatalyst tool in this respect (**Supplementary Table 1**).

We also explored the ability of the GlgC or GlgC/GlgD structures of the *G. stearothermophilus* ADP-Glc PPase to use GlcN-1P as an alternative substrate. With this substrate (at 1 mM), we observed 10% activity relative to Glc-1P (not shown), similar to the rhodococcal ADP-Glc PPase (Cereijo et al., 2016, 2020).

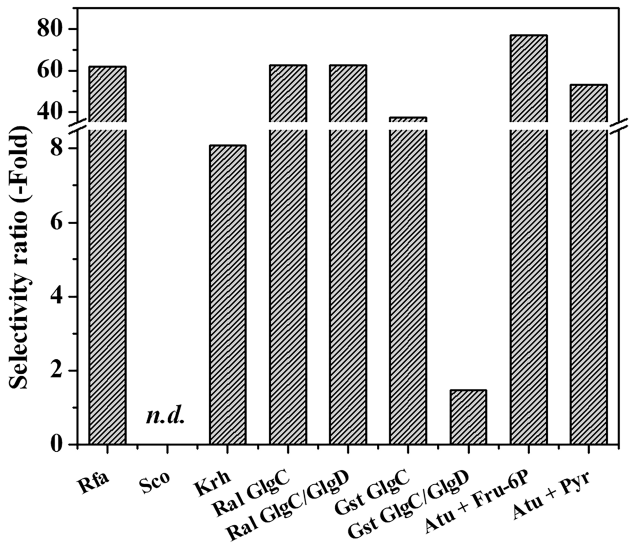


FIGURE 2 | Selectivity ratios for glucose-1-phosphate (Glc-1P) and GlcN-1P (ratio between catalytic efficiency for the use of Glc-1P over that with GlcN-1P).

The GlgC homotetramer had a catalytic efficiency of 0.36 mM^{-1} s^{-1} toward GlcN-1P, a value in the same order of magnitude as the rhodococcal enzymes (between 0.3 and 0.5 mM^{-1} s^{-1}) (Cereijo et al., 2020). It is important to note that the efficiency for these substrates in the *G. stearothermophilus* protein was in the absence of activator and the catalytic efficiency for GlcN-1P is 37-fold lower than Glc-1P (**Figure 2**), sustained by lower values for both maximum activity and apparent affinity. Interestingly, the selectivity ratio decreased to 1.5 for the heterotetramer GlgC/GlgD (**Figure 2**) showing that the presence of both subunits diminishes substrate selectivity. The presence of GlgD subunit increased both the catalytic capacity (around 20-fold) and the apparent affinity (K_m ~ 0.4 mM) for GlcN-1P. Furthermore, in the presence of 1 mM GlcN-1P, ATP reached similar k_{cat} and K_m (0.21 mM) levels (**Supplementary Figure 2**) with a catalytic efficiency for ATP utilization with GlcN-1P at 40.28 $mM^{-1}s^{-1}$, a value 7-fold higher than with Glc-1P, according to data previously reported (Takata et al., 1997).

Therefore, the *G. stearothermophilus* GlgC/GlgD heterotetrameric structure of ADP-Glc PPase has the highest reported catalytic efficiency to use ATP and GlcN-1P allowing us to hypothesize that it may be metabolically meaningful (Bar-Even et al., 2011; Davidia et al., 2016). Also important, at a catalytic level, is the effect of the inactive GlgD subunit on the global behavior of ADP-Glc PPase, which is not observed for the substrate Glc-1P (Takata et al., 1997). The homotetrameric GlgC enzyme from the *R. albus* used GlcN-1P with a lower efficiency than its canonical substrate Glc-1P (selectivity ratio of ~60; **Figure 2**). The heterotetrameric GlgC/GlgD from *R. albus* also utilized GlcN-1P as an alternative substrate, even in the absence of activator. This is like the *G. stearothermophilus* enzymes presented above, but it marks a difference with the actinobacterial ADP-Glc PPases, which need to be activated. The heteromeric GlgC/GlgD from *R. albus* also showed 60-fold lower catalytic efficiency toward GlcN-1P than for Glc-1P (**Figure 2**). Although the enzyme shows a k_{cat} with the amino sugar-1P that is 70% of that with the sugar-1P, the enzyme possesses a low apparent affinity (~2 mM) for GlcN-1P. The catalytic efficiency for GlcN-1P in the heterotetrameric *G. stearothermophilus* enzyme falls within the range of a plausible and meaningful metabolic reaction, according to the hypothesis presented elsewhere (Bar-Even et al., 2011; Davidia et al., 2016).

Complementing the study of GlcN-1P utilization by bacterial ADP-Glc PPase, we analyzed this activity using the model enzyme from *A. tumefaciens*, which has been thoroughly analyzed at the kinetic, regulatory, and structural levels (Uttaro and Ugalde, 1994; Uttaro et al., 1998; Gómez-Casati et al., 2001; Hill et al., 2019; Asencion Diez et al., 2020). The enzyme was active with GlcN-1P as a substrate, but its efficiency was about two orders of magnitude lower than Glc-1P. Both Fru-6P and Pyr activated GlcN-1P utilization (remarkably, the enzyme resulted highly sensitive to activation, with an $A_{0.5}$ between 27 and 33 μM for both effectors) exhibiting selectivity ratios with respect to Glc-1P in the range of 50–80 (**Figure 2**). We observed that the apparent affinity for ATP remains constant whether the other substrate is Glc-1P or GlcN-1P. Considering the availability of the crystallographic structure of the

A. tumefaciens ADP-Glc PPase and that the enzyme used GlcN-1P with a lower preference than the canonical substrate, it was relevant to explore further this kinetic characteristic in the framework of the protein structure.

Structural Approach to GlcN-1P Binding/ Catalysis

The only structural difference between Glc-1P and GlcN-1P is that an amino group is in position 2 rather than a hydroxyl group. Therefore, to understand the selectivity between Glc-1P and GlcN-1P in ADP-Glc PPases, it is important to analyze the structural determinants for the binding of the hydroxyl group in position 2 when Glc-1P is the substrate. Using the elucidated crystallographic structure of ADP-Glc PPase from potato tuber, we modeled the sugar-phosphate binding site in *A. tumefaciens*. Previous studies in crystal structures of proteins have revealed that alcohol binding sites generally consist of a hydrogen bond acceptor in a turn or loop region that is often situated at the N-terminal end of an α-helix (Dwyer and Bradley, 2000). This is certainly the case in the small subunit of the ADP-Glc PPase from potato tuber (**Figure 3**). The hydroxyl group in position 2 of the glucosyl moiety of ADP-Glc, which is the product after Glc-1P reacted with ATP, is in a perfect place to donate hydrogen to the carboxylate of E197 (Jin et al., 2005). In addition, this binding site is near the N-terminal end of the α-helix comprising residues between G255 and L265. No other interactions are apparent. Structural analysis of the *A. tumefaciens* enzyme, where ADP-Glc has been modeled (**Figure 3**), yielded identical results. Therefore, it is expected that these two major structural elements, the H-bond acceptor E187 (E197 in potato tuber, E194 in *E. coli*) and the α-helix, will contribute to selecting a hydroxyl over an amino group, and as such, prefer Glc-1P over GlcN-1P.

A hypothesis was that the difference in strength between an H-bond could explain the difference in substrate specificity. This is because the oxygen from a hydroxyl group is predicted to be a stronger donor than nitrogen from a primary amine (Halder et al., 2019). In both cases, the acceptor is one of the oxygens from the carboxylate (E187). If this hypothesis is true, the elimination of the acceptor by mutagenesis should remove or significantly decrease the differences between Glc-1P and GlcN-1P. Previously, the homologous glutamic acid at position 194 was mutated to alanine in the *E. coli* ADP-Glc PPase and was found to be critical to the binding of Glc-1P (Bejar et al., 2006). We mutated E187 to alanine to determine whether we can eliminate the substrate preference of Glc-1P over GlcN-1P. In the E187A mutant with no effector, Glc-1P had an K_m of 0.4 mM, nearly 4.5-fold worse than the wild-type (WT) enzyme (**Table 1**). In the presence of its major activators, the K_m increased 13-fold for Fru-6P and 12-fold for Pyr for E187A compared to the WT, showing that the binding of the substrate is greatly hindered by removing the negatively charged glutamic acid residue with alanine. When comparing the catalytic efficiencies between Glc-1P and GlcN-1P without an activator, there is still a 79-fold preference for Glc-1P over GlcN-1P in E187A. Even with activators, the mutated enzyme still shows selectivity for Glc-1P over GlcN-1P, meaning that the glutamic acid does not solely control the preference of sugar-phosphate for this enzyme.

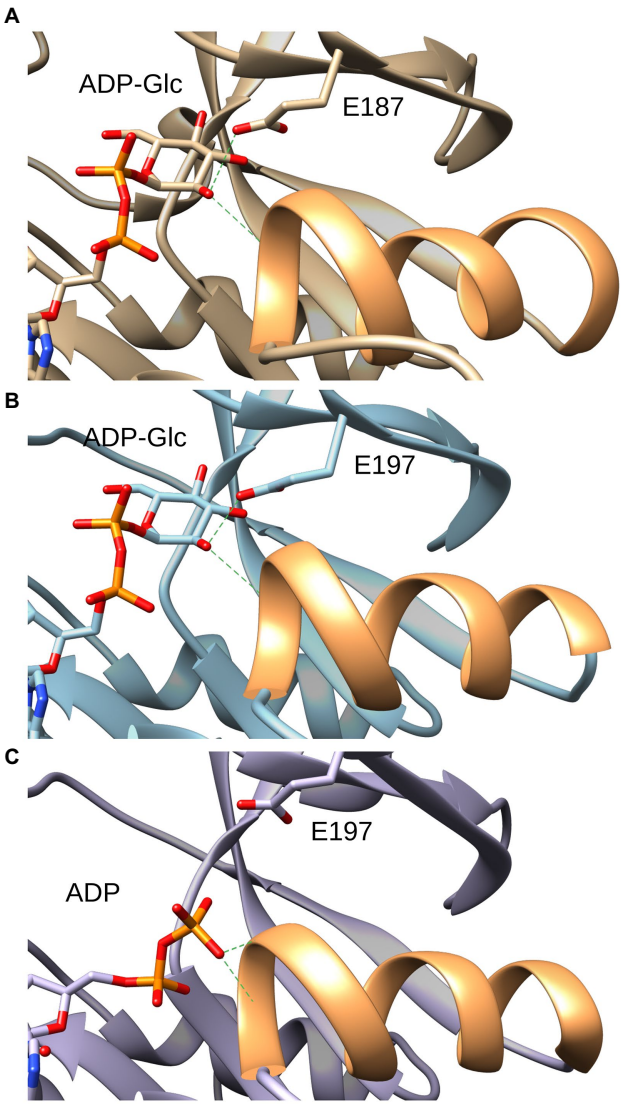


FIGURE 3 | Interaction of the glucose moiety of the product ADP-Glc with the ADP-Glc PPase from *Agrobacterium tumefaciens* and potato tuber. **(A)** Structure of the *A. tumefaciens* ADP-Glc PPase in which ADP-Glc has been modeled as described in the section "Materials and Methods." In orange is the ribbon corresponding to the α-helix from residues F233 to H244. Green dashes indicate the distance from the OH (position 2) to the side chain oxygen of the E187 residue (3.1 Å) and to the backbone N of the residue G234 (4.0 Å), respectively. **(B)**: Structure of the small subunit ADP-Glc PPase from potato tuber (PDB ID: 1YP4, subunit B) with ADP-Glc in the active site. In orange is the ribbon corresponding to the α-helix from residues F254 to S264. Green dashes indicate the distance to the O of side chain of the E197 residue (2.8 Å) and to the N of the backbone of the residue G255 (4.0 Å), respectively. **(C)** Structure of the small subunit ADP-Glc PPase from potato tuber (PDB ID: 1YP4, subunit A) with ADP in the active site. In orange is the ribbon corresponding to the α-helix from residues F254 to S264. Green dashes indicate the distance from the closest O of the β-phosphate of ADP to the backbone nitrogen of the S256 residue (3.3 Å) and to the backbone nitrogen of the residue G255 (3.9 Å), respectively.

TABLE 1 | Kinetic parameters for E187A ADP-Glc PPase from *A. tumefaciens* with Glc-1P and GlcN-1P as substrates in the absence and presence of activators.

	Glc-1P			GlcN-1P			Selectivity ratio
	V_{max}	K_m	V_{max}/K_m	V_{max}	K_m	V_{max}/K_m	$\left(\frac{V_{max}}{K_m}\right)_{Glc1P} / \left(\frac{V_{max}}{K_m}\right)_{GlcN1P}$
	(U·mg^{-1})	(mM)	(U·mg^{-1}·mM^{-1})	(U·mg^{-1})	(mM)	(U·mg^{-1}·mM^{-1})	
No Effector	0.27 ± 0.01	0.427 ± 0.049	0.632	0.012 ± 0.001	1.50 ± 0.23	0.008	79
Fru-6P (1.5 mM)	4.00 ± 0.15	0.56 ± 0.05	7.14	0.193 ± 0.025	1.83 ± 0.37	0.105	68
Pyr (1.5 mM)	3.85 ± 0.13	0.65 ± 0.05	5.92	0.062 ± 0.002	0.93 ± 0.10	0.067	88
GlcN-6P (1.5 mM)	0.208 ± 0.004	1.42 ± 0.04	0.146	0.005 ± 0.0004	0.78 ± 0.15	0.006	24

DISCUSSION

Pyrophosphorylases (PPases) are critical enzymes for carbohydrate metabolism in different organisms, catalyzing the reversible production of nucleoside-diphospho-sugar and their derivatives. It is described that ~70 nucleotide sugars were identified in bacteria (Decker and Kleczkowski, 2019), and this number could be higher if we consider (*i*) bioprospecting of metagenomic data and discovery of new types and families of enzymes or (*ii*) secondary activities and/or promiscuity of enzymes producing molecules at sub-optimal metabolic concentrations (Peracchi, 2018; Rosenberg and Commichau, 2019; Copley and Copley, 2020). The former requires the study of uncharacterized nucleotide-sugar PPases and the latter to revisit the kinetic properties of already known enzymes. To achieve these tasks, we have reported PPases with new activities (Asención Diez et al., 2010, 2017; Ebrecht et al., 2015; Cereijo et al., 2021), and herein, we deepen the study of promiscuity of the extensively characterized bacterial ADP-Glc PPase (Ballicora et al., 2003).

UDP-Glc and ADP-Glc are synthesized from Glc-1P and their respective nucleoside-triphosphate (NTP; Glc-1P + NTP ↔ NDP-Glc + PPi). In bacteria, the formation of NDP-Glc represents a branching point that determines the fate of Glc in the cell: ADP-Glc is used to form glycogen, while UDP-Glc is shuttled to various metabolic routes (Figueroa et al., 2021). UDP-Glc PPase (EC 2.7.7.9) from prokaryotes has been characterized as a non-regulated enzyme, as is the case for other NDP-sugar PPases. Conversely, ADP-Glc PPase from many bacteria is sensitive to allosteric regulation by intermediates of the organism's central metabolic pathway(s). Structurally, ADP-Glc PPases are larger proteins than other NDP-sugar PPases, having a C-terminal domain that confers the regulatory function (Ballicora et al., 2003, 2004; Figueroa et al., 2022).

UDP-Glc PPases from different sources were reported to be specific toward the use of UTP and Glc-1P. The only exception so far is the case of a GalU-type enzyme from *Rhodococcus* spp., which exhibits the ability to use GlcN-1P with high efficiency, being the preferred substrate (Cereijo et al., 2021). On the other hand, the ADP-Glc PPase from *E. coli* was found promiscuous toward different NTPs and hexose-1Ps but, when assayed in the presence of allosteric regulators, only Glc-1P and ATP were major substrates (Ebrecht et al., 2017). Thus, the hypothesis regarding the evolution of regulation in ADP-Glc PPases seems to have selected enzymes, which have a preference toward ADP-Glc synthesis after allosteric activation. Interestingly, a divergence had been reported at that time, with the ADP-Glc PPase from *R. jostii* using GlcN-1P as a substrate and its activity responding to Glc-6P, the primary activator of the actinobacterial ADP-Glc PPase (Cereijo et al., 2016). More recently, it was demonstrated that the catalysis of the *R. jostii* enzyme is efficiently enhanced by Glc-6P, Fru-6P, Man-6P, and PEP. The analysis of the regulation of the GlcN-1P allowed the discovery of a new regulating metabolite, GlcN-6P, which exerted activation mainly on the catalysis of the alternative substrate, analogously to the Glc-6P/Glc-1P pair for its canonical activity (Cereijo et al., 2020).

In the work presented here, we extended the analysis of GlcN phosphates regarding their relationships with prokaryotic ADP-Glc PPases either as a substrate (GlcN-1P) or an effector (GlcN-6P). We identified these molecules are related to all the characterized enzymes with common patterns and characteristics depending on the source. Recently, several crystallographic structures were obtained for ADP-Glc PPases (Cupp-Vickery et al., 2008; Hill et al., 2019; Asencion Diez et al., 2020), and they provide invaluable tools to deeper understand the structure–function relationships in the enzyme. We constructed an *in silico* model to understand the structural determinants for substrate specificity related to GlcN-1P, as described below.

The hydroxyl of Glc-1P (or the product ADP-Glc) in position 2 donates a H to the O of the carboxylate from E187 and stands close to a positive density at the N-terminus end of the α-helix F233-H244. This led to the hypothesis that these two major structural elements contribute to selecting a hydroxyl over an amino group, and as such, preferring Glc-1P over GlcN-1P. However, this was not exactly the case. The mutation E187A decreased the overall activity of the enzyme and the apparent affinity for the sugar phosphates, but the selectivity ratio only decreased minimally. This indicates that the H-bond with E187 is a strong feature for the binding of the sugars, but it is not a major contributor in selecting Glc-1P over GlcN-1P. It is well known that at the N-terminus of an α-helix there is an electrostatic positive density that is important for

binding of anions (Hol, 1985). This electrostatic interaction most likely contributes to differentially attracting a hydroxyl group over a positively charged primary amine. Experimental evidence that this area attracts negatively charged groups is that in a different subunit of the potato tuber crystal structure, the beta-phosphate of ADP binds there (presented in **Figure 3**). This α-helix is a common feature of the GT-A fold of the ADP-Glc PPase family (Figueroa et al., 2022), which would explain the widespread selectivity for Glc-1P over GlcN-1P in the family.

Results with the two PPases from *Rhodococci* using GlcN phosphate metabolites led us to propose a putative metabolic node or bifurcation at the level of GlcN-1P, incorporating the characterization of other enzymes, such as GlmU (Cereijo et al., 2021). This enzymological approach led us to revisit GlcN utilization in *Rhodococcus* spp., using *R. jostii* and *R. fascians* as growing models, and to find new metabolic outputs when the microorganisms use GlcN as the sole carbon source (Cereijo et al., manuscript in preparation). Worthy of mentioning, the literature available concerning GlcN utilization by microorganisms is scarce and limited to a group (Flores and Gancedo, 2018), including some Gram-positive bacteria (Gaugué et al., 2013; Uhde et al., 2013; Álvarez-Añorve et al., 2016). Then, given that the allosteric properties of ADP-Glc PPases are linked to the major carbon assimilation pathways(s) in the organisms they belong to, their kinetic and regulatory characterization may help infer metabolic pathways occurring in the organism. In this regard, we extended the study of GlcN metabolites usage by ADP-Glc PPases beyond the *Rhodococci* case.

Our analysis shows that the *K. rhizophila* ADP-Glc PPase follows the general mechanism we previously described for the rhodococcal enzyme (Cereijo et al., 2021), where the activation (either by GlcN-6P or other effectors, e.g., Glc-6P) switches on GlcN-1P consumption. Particularly, the enzyme from *Kocuria* displays this phenomenon in a more accentuated form. The GlcN-6P presence augments activity by two orders of magnitude with a significant increase in the catalytic efficiency toward the alternative substrate. In addition, compared to the protein from *R. fascians*, ADP-Glc PPase from *Kocuria* depicts a higher degree of promiscuity since the selectivity coefficient is 8-fold lower than the former. On the other hand, the *S. coelicolor* ADP-Glc PPase showed a poor use of GlcN-1P, with activity remaining in milliunit values after a weak response to activators. It is possible that these differences could be attributed to the varied degree of sensitivity toward distinct regulatory effectors by the *S. coelicolor* protein compared to the enzymes from *Rhodococci* and *Kocuria* (Asencion Diez et al., 2012; Cereijo et al., 2016, 2020). These particular characteristics in regulation could be related to the many roles proposed for glycogen in Actinobacteria. In *Rhodococci*, the glucan serves as a temporary carbon allocation molecule, closely interrelated to the metabolism of other storage compounds (fatty acids, triacylglycerols) (Hernández et al., 2008, 2019; Hernandez and Alvarez, 2010), whereas in *Streptomyces* spp. glycogen seems to participate (together with trehalose) in processes regarded to cell differentiation (Hey-Ferguson et al., 1973; Brana et al., 1986; Rueda et al., 2001). According to genomic information, *Kocuria* is metabolically similar to *Rhodococci* (Takarada et al., 2008). In this context, the GlcN-1P switch-on mechanism exerted by the corresponding enzyme activators (specifically GlcN-6P) might be a strategy to cope with temporary GlcN phosphates accumulation. It has been demonstrated that increased intracellular hexose-P concentrations could be toxic and detrimental to cell viability (Boulanger et al., 2021).

The hypothetical GlcN-1P utilization "switch-on mechanism" described above is absent in the ADP-Glc PPases other than from *Kocuria* and *Rhodococcus*. Remarkably, little activation was observed in the *R. albus* GlgC and GlgC/GlgD enzymes, although it was already reported they had an inherent low sensitivity to effectors for the canonical substrates (Cereijo et al., 2018). *R. albus* is a microorganism whose ecological niche is the rumen of animals; where GlcN would not be a major carbon source available (Suen et al., 2011). In *A. tumefaciens*, it is well known that the microorganism has Entner-Doudoroff as its central glucose oxidation pathway (Arthur et al., 1973, 1975; Xu et al., 2021); although a transporter system for the GlcN (Zhao and Binns, 2014) and catabolic genes for GlcN utilization was described (Iwamoto et al., 1998).

In *Bacilli*, two different operons related to transport and catabolism of GlcN were reported (Gaugué et al., 2013). The general implication of these pathways is to channel GlcN-6P to Fru-6P, then feeding the core metabolism. The ADP-Glc PPase from *Bacilli* is the only unregulated enzyme of the family, which implies it is "disconnected" from intracellular metabolic signals to modulate glycogen accumulation (Kiel et al., 1994; Takata et al., 1997; Ballicora et al., 2003, 2004). Exploring the role of GlcN-6P as an effector on the active conformations (GlgC and GlgC/GlgD) of *G. stearothermophilus* ADP-Glc PPase reinforced the precedent of the lack of allosteric regulation. However, the features of this enzyme described in this work unmask its capability to utilize GlcN-1P. Remarkably, the modest activity with GlcN-1P shown by the GlgC homotetramer (displaying a clear preference for Glc-1P) is enhanced in the heterotetramer GlgC/GlgD. In this case, the "allosteric behavior" exerted by the GlgD subunit increases the efficiency 60-times for GlcN-1P which is analogous to the "switch-on" mechanism described in the *Kocuria*/*Rhodococcus* enzymes. For the canonical substrate Glc-1P, the *G. stearothermophilus* GlgD subunit increases the catalytic efficiency ~3-fold. Thus, this GlgD behavior resembles the action of the GlgD subunit from *Streptococcus mutans* on the activity with Glc-1P (Asención Diez et al., 2013a). It has already been observed that an ADP-Glc PPase subunit could change the specificity of a ligand. The large subunit of the unicellular algae *Ostreococcus tauri* was considered a "specifier" because its presence in the heterotetramer only increases the specificity for the activator 3-phosphoglycerate (3-PGA) compared to fructose-1,6-bisphosphate (Kuhn et al., 2013). A classic example of a subunit changing substrate specificity was the β-1,4-galactosyltransferase, which changes its specificity from N-acetylglucosamine to glucose when it interacts with the protein α-lactalbumin (Ramakrishnan et al., 2001). The physiological relevance of the increment in GlcN-1P utilization by the GlgC/GlgD enzyme in *G. stearothermophilus* (or other *Bacilli*) remains to be solved; this case also illustrates how

enzymological data bring new questions regarding the metabolism while opening the door to novel biotechnological tools.

The ability of the ADP-Glc PPases reported here to utilize GlcN phosphates suggests that evolution could harness the capacity to bind the amino sugar in the substrate pocket even with diminished affinity of the amino group at C2. Also, the allosteric properties of these enzymes can increase catalytic efficiencies. This increased substrate promiscuity could also be considered in terms of underground metabolism. The latter comprises the alternative reactions produced by the promiscuity of some enzymes, which are reactions below the metabolic plausibility (Bar-Even et al., 2011; Davidia et al., 2016; Rosenberg and Commichau, 2019). The underground metabolism serves as an evolutionary tool for the appearance of novel pathways in particular circumstances (D'Ari and Casadesús, 1998; Peracchi, 2018; Rosenberg and Commichau, 2019; Glasner et al., 2020). Works performed in the model organism *E. coli* demonstrated its ability to harness the underground metabolism to discover new metabolic functions (Guzmán et al., 2015). Using this organism, specific examples are in recent works where *(i)* new metabolite links ("serine shunt") were identified, allowing to rewire central metabolism, particularly the Embden–Meyerhof–Parnas glycolysis (Iacometti et al., 2021) and *(ii)* elucidated the appearance of new pathways for isoleucine biosynthesis (Cotton et al., 2020). *E. coli* was also demonstrated that GlcN-6P intracellular concentration increases two orders of magnitude (up to 9 mM) when the bacterium grows in GlcN as a sole carbon source (Álvarez-Añorve et al., 2009), where specifically a GlcN-6P utilization enzyme [NagB (EC 3.5.99.6)] is activated to channel the hexosamine-P to the central glycolytic pathway (Álvarez-Añorve et al., 2016). However, little is known regarding other metabolites or enzymatic activities linked to GlcN consumption. The properties reported herein for bacterial ADP-Glc PPases would fit this metabolic scenario. Indeed, since there is no enzyme apart from GlmM (EC 5.4.2.10) to GlmU (EC 2.7.7.23/2.3.1.157) reported to use GlcN-1P as a substrate, the allosteric activation of GlcN-1P (either by metabolites or GlgD in *G. stearothermophilus*) may provide the potential for the evolution of alternative/novel GlcN-related pathway(s), to yet unknown intracellular fates. Reinforcing this hypothesis, we recently have published the discovery of a new type of pyrophosphorylase specific for GlcN-1P (Cereijo et al., 2021). Beyond the structural basis for GlcN-P utilization and its putative metabolic implication, we also stress the importance of producing the recombinant enzymes (particularly PPases) to catalyze GlcN-1P to NDP-GlcN since it would be critical for developing different types of molecular tools. As a leading example, we could consider the TDP-Glc PPase from *Salmonella enterica* where is promiscuous activity toward sugar-1P and NTP was deeply studied and exploited to create a library of different NDP-sugars (using natural and non-natural substrates) (Moretti and Thorson, 2007, 2008). The library was proposed as a glycosyl-transferase substrate to modify diverse aglycon molecules (such as antibiotics), by a process-denominated glycorandomization (Křen and Řezanka, 2008). Among the NDP-sugars, a minor group is constituted with amino sugar moieties. Thus, the product of bacterial ADP-Glc PPases catalyzing GlcN-1P may find a niche to be incorporated. The synthesis, purification, and biophysical characterization of these molecules could allow later *in vivo* experiments to establish their potential physiological relevance. In addition, the availability of purified NDP-GlcN may help to identify novel types of (amino) glycosyltransferases (Zhang et al., 2014; Da Costa et al., 2021) and incorporate them into strategies of cell-free glycobiology for the synthesis of novel glycans (Jaroentomeechai et al., 2020).

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

AUTHOR CONTRIBUTIONS

MA, MB, and AI conceived the study, designed the experiments, and analyzed the results. JB, MI, RM, and AC performed the enzymatic characterizations. JB and MB produced the homology model. JB, MA, MB, and AI wrote the manuscript with contributions from all authors. All authors contributed to the article and approved the submitted version.

FUNDING

This work was supported by grants from ANPCyT to MA (PICT-2018-00698), to AI (PICT-2017-1515; PICT-2018-00929) and from NSF to MB (MCB 1616851). MA and AI are investigators from CONICET. MI is a Fellow (*Cientibeca*) from UNL.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmicb.2022.867384/full#supplementary-material

REFERENCES

Álvarez-Añorve, L. I., Bustos-Jaimes, I., Calcagno, M. L., and Plumbridge, J. (2009). Allosteric regulation of glucosamine-6-phosphate deaminase (NagB) and growth of *Escherichia coli* on glucosamine. *J. Bacteriol.* 191, 6401–6407. doi: 10.1128/JB.00633-09

Álvarez-Añorve, L. I., Gaugué, I., Link, H., Marcos-Viquez, J., Díaz-Jiménez, D. M., Zonszein, S., et al. (2016). Allosteric activation of *Escherichia coli* glucosamine-6-phosphate deaminase (NagB) in vivo justified by intracellular amino sugar metabolite concentrations. *J. Bacteriol.* 198, 1610–1620. doi: 10.1128/JB.00870-15

Arthur, L. O., Bulla, L. A., St. Julian, G., and Nakamura, L. K. (1973). Carbohydrate metabolism in agrobacterium tumefaciens. *J. Bacteriol.* 116, 304–313. doi: 10.1128/jb.116.1.304-313.1973

Arthur, L. O., Nakamura, L. K., Julian, G. S., and Bulla, L. A. (1975). Carbohydrate catabolism of selected strains in the genus *agrobacterium*. *Appl. Microbiol.* 30, 731–737. doi: 10.1128/aem.30.5.731-737.1975

Asención Diez, M. D., Demonte, A., Giacomelli, J., Garay, S., Rodrígues, D., Hofmann, B., et al. (2010). Functional characterization of GDP-mannose pyrophosphorylase from *Leptospira interrogans* serovar Copenhageni. *Arch. Microbiol.* 192, 103–114. doi: 10.1007/s00203-009-0534-3

Asención Diez, M. D., Demonte, A. M., Guerrero, S. A., Ballicora, M. A., and Iglesias, A. A. (2013a). The ADP-glucose pyrophosphorylase from *Streptococcus mutans* provides evidence for the regulation of polysaccharide biosynthesis in Firmicutes. *Mol. Microbiol.* 90, 1011–1027. doi: 10.1111/mmi.12413

Asencion Diez, M. D., Demonte, A. M., Syson, K., Arias, D. G., Gorelik, A., Guerrero, S. A., et al. (2015). Allosteric regulation of the partitioning of glucose-1-phosphate between glycogen and trehalose biosynthesis in *Mycobacterium tuberculosis*. *Biochim. Biophys. Acta Gen. Subj.* 1850, 13–21. doi: 10.1016/j.bbagen.2014.09.023

Asención Diez, M., Ebrecht, A., Martínez, L., Aleanzi, M., Guerrero, S., Ballícora, M., et al. (2013b). A chimeric UDP-glucose pyrophosphorylase produced by protein engineering exhibits sensitivity to allosteric regulators. *Int. J. Mol. Sci.* 14, 9703–9721. doi: 10.3390/ijms14059703

Asencion Diez, M. D., Figueroa, C. M., Esper, M. C., Mascarenhas, R., Aleanzi, M. C., Liu, D., et al. (2020). On the simultaneous activation of *agrobacterium tumefaciens* ADP-glucose pyrophosphorylase by pyruvate and fructose 6-phosphate. *Biochimie* 171-172, 23–30. doi: 10.1016/j.biochi.2020.01.012

Asención Diez, M. D., Miah, F., Stevenson, C. E. M., Lawson, D. M., Iglesias, A. A., and Bornemann, S. (2017). The production and utilization of GDP-glucose in the biosynthesis of Trehalose 6-phosphate by *Streptomyces venezuelae*. *J. Biol. Chem.* 292, 945–954. doi: 10.1074/jbc.M116.758664

Asencion Diez, M. D., Peiru, S., Demonte, A. M., Gramajo, H., and Iglesias, A. A. (2012). Characterization of recombinant UDP- and ADP-glucose pyrophosphorylases and glycogen synthase to elucidate glucose-1-phosphate partitioning into oligo- and polysaccharides in *Streptomyces coelicolor*. *J. Bacteriol.* 194, 1485–1493. doi: 10.1128/JB.06377-11

Ballicora, M., Erben, E., Yazaki, T., et al. (2007). Identification of regions critically affecting kinetics and allosteric regulation of the *Escherichia coli* ADP-glucose pyrophosphorylase by modeling and pentapeptide-scanning mutagenesis. *J. Bacteriol.* 189, 5325–5333. doi: 10.1128/JB.00481-07

Ballicora, M., Iglesias, A., and Preiss, J. (2003). ADP-glucose pyrophosphorylase, a regulatory enzyme for bacterial glycogen synthesis. *Microbiol. Mol. Biol. Rev.* 67, 213–225. doi: 10.1128/MMBR.67.2.213-225.2003

Ballicora, M. A., Iglesias, A. A., and Preiss, J. (2004). ADP-glucose pyrophosphorylase: a regulatory enzyme for plant starch synthesis. *Photosynth. Res.* 79, 1–24. doi: 10.1023/B:PRES.0000011916.67519.58

Bar-Even, A., Noor, E., Savir, Y., Liebermeister, W., Davidi, D., Tawfik, D. S., et al. (2011). The moderately efficient enzyme: evolutionary and physicochemical trends shaping enzyme parameters. *Biochemistry* 50, 4402–4410. doi: 10.1021/bi2002289

Bejar, C. M., Ballicora, M. A., Gómez-Casati, D. F., Iglesias, A. A., and Preiss, J. (2004). The ADP-glucose pyrophosphorylase from *Escherichia coli* comprises two tightly bound distinct domains. *FEBS Lett.* 573, 99–104. doi: 10.1016/j.febslet.2004.07.060

Bejar, C. M., Jin, X., Ballicora, M. A., and Preiss, J. (2006). Molecular architecture of the glucose 1-phosphate site in ADP-glucose pyrophosphorylases. *J. Biol. Chem.* 281, 40473–40484. doi: 10.1074/JBC.M607088200

Boulanger, E. F., Sabag-Daigle, A., Thirugnanasambantham, P., Gopalan, V., and Ahmer, B. M. M. (2021). Sugar-phosphate toxicities. *Microbiol. Mol. Biol. Rev.* 85:e0012321. doi: 10.1128/MMBR.00123-21

Brana, A. F., Mendez, C., Diaz, L. A., Manzanal, M. B., and Hardisson, C. (1986). Glycogen and trehalose accumulation during colony development in *Streptomyces antibioticus*. *J. Gen. Microbiol.* 132, 1319–1326. doi: 10.1099/00221287-132-5-1319/CITE/REFWORKS

Cereijo, A., Alvarez, H., Iglesias, A., and Asencion Diez, M. (2020). Glucosamine-P and rhodococcal ADP-glucose pyrophosphorylases: A hint to (re)discover (actino)bacterial amino sugar metabolism. *Biochimie* 176, 158–161. doi: 10.1016/j.biochi.2020.07.006

Cereijo, A. E., Asencion Diez, M. D., Ballicora, M. A., and Iglesias, A. A. (2018). Regulatory properties of the ADP-glucose pyrophosphorylase from the clostridial Firmicutes member *Ruminococcus albus*. *J. Bacteriol.* 200, e00172–e00218. doi: 10.1128/JB.00172-18

Cereijo, A. E., Asencion Diez, M. D., Davila Costa, J. S., Alvarez, H. M., and Iglesias, A. A. (2016). On the kinetic and allosteric regulatory properties of the ADP-glucose pyrophosphorylase from *Rhodococcus jostii*: An approach to evaluate glycogen metabolism in oleaginous bacteria. *Front. Microbiol.* 7:830. doi: 10.3389/fmicb.2016.00830

Cereijo, A. E., Kuhn, M. L., Hernández, M. A., Ballicora, M. A., Iglesias, A. A., Alvarez, H. M., et al. (2021). Study of duplicated galU genes in *Rhodococcus jostii* and a putative new metabolic node for glucosamine-1P in rhodococci. *Biochim. Biophys. Acta Gen. Subj.* 1865:129727. doi: 10.1016/j.bbagen.2020.129727

Cifuente, J. O., Comino, N., Trastoy, B., D'Angelo, C., and Guerin, M. E. (2019). Structural basis of glycogen metabolism in bacteria. *Biochem. J.* 476, 2059–2092. doi: 10.1042/BCJ20170558

Copley, S. D., and Copley, S. D. (2020). Evolution of new enzymes by gene duplication and divergence. *FEBS J.* 287, 1262–1283. doi: 10.1111/FEBS.15299

Cotton, C. A. R., Bernhardsgrütter, I., He, H., Burgener, S., Schulz, L., Paczia, N., et al. (2020). Underground isoleucine biosynthesis pathways in *E. coli*. *eLife* 9, 1–25. doi: 10.7554/ELIFE.54207

Cupp-Vickery, J. R., Igarashi, R. Y., Perez, M., Poland, M., and Meyer, C. R. (2008). Structural analysis of ADP-glucose pyrophosphorylase from the bacterium *agrobacterium tumefaciens*. *Biochemistry* 47, 4439–4451. doi: 10.1021/BI701933Q

D'Ari, R., and Casadesús, J. (1998). Underground metabolism. *BioEssays* 20, 181–186. doi: 10.1002/(SICI)1521-1878(199802)20:2<181::AID-BIES10>3.0.CO;2-0

Da Costa, M., Gevaert, O., Van Overtveldt, S., Lange, J., Joosten, H. J., Desmet, T., et al. (2021). Structure-function relationships in NDP-sugar active SDR enzymes: fingerprints for functional annotation and enzyme engineering. *Biotechnol. Adv.* 48:107705. doi: 10.1016/J.BIOTECHADV.2021.107705

Davidia, D., Noorb, E., Liebermeisterc, W., Bar-Evend, A., Flamholze, A., Tummlerf, K., et al. (2016). Global characterization of *in vivo* enzyme catalytic rates and their correspondence to *in vitro k*cat measurements. *Proc. Natl. Acad. Sci. U. S. A.* 113, 3401–3406. doi: 10.1073/pnas.1514240113

Decker, D., and Kleczkowski, L. A. (2019). UDP-sugar producing pyrophosphorylases: distinct and essential enzymes with overlapping substrate specificities, providing de novo precursors for glycosylation reactions. *Front. Plant Sci.* 9:1822. doi: 10.3389/fpls.2018.01822

Dönertaş, H. M., Cuesta, S. M., Rahman, S. A., and Thornton, J. M. (2016). Characterising complex enzyme reaction data. *PLoS One* 11:e0147952. doi: 10.1371/JOURNAL.PONE.0147952

Dwyer, D. S., and Bradley, R. J. (2000). Chemical properties of alcohols and their protein binding sites. *Cell. Mol. Life Sci. C.* 57, 265–275. doi: 10.1007/PL00000689

Ebrecht, A. C., Asención Diez, M. D., Piattoni, C. V., Guerrero, S. A., and Iglesias, A. A. (2015). The UDP-glucose pyrophosphorylase from *Giardia lamblia* is redox regulated and exhibits promiscuity to use galactose-1-phosphate. *Biochim. Biophys. Acta* 1850, 88–96. doi: 10.1016/j.bbagen.2014.10.002

Ebrecht, A. C., Solamen, L., Hill, B. L., Iglesias, A. A., Olsen, K. W., and Ballicora, M. A. (2017). Allosteric control of substrate specificity of the *Escherichia coli* ADP-glucose pyrophosphorylase. *Front. Chem.* 5:41. doi: 10.3389/fchem.2017.00041

Eswar, N., Webb, B., Marti-Renom, M. A., Madhusudhan, M. S., Eramian, D., Shen, M., et al. (2007). Comparative protein structure modeling using MODELLER. *Curr. Protoc. Protein Sci.* 50, 2.9.1–2.9.31. doi: 10.1002/0471140864.PS0209S50

Figueroa, C. M., Asencion Diez, M. D., Ballicora, M. A., and Iglesias, A. A. (2022). Structure, function, and evolution of plant ADP-glucose pyrophosphorylase. *Plant Mol. Biol.* 108, 307–323. doi: 10.1007/S11103-021-01235-8

Figueroa, C. M., Esper, M. C., Bertolo, A., Demonte, A. M., Aleanzi, M., Iglesias, A. A., et al. (2011). Understanding the allosteric trigger for the fructose-1,6-bisphosphate regulation of the ADP-glucose pyrophosphorylase from *Escherichia coli*. *Biochimie* 93, 1816–1823. doi: 10.1016/j.biochi.2011.06.029

Figueroa, C. M., Kuhn, M. L., Falaschetti, C. A., Solamen, L., Olsen, K. W., Ballicora, M. A., et al. (2013). Unraveling the activation mechanism of the potato tuber ADP-glucose Pyrophosphorylase. *PLoS One* 8:e66824. doi: 10.1371/JOURNAL.PONE.0066824

Figueroa, C. M., Lunn, J. E., and Iglesias, A. A. (2021). Nucleotide-sugar metabolism in plants: the legacy of Luis F. *Leloir. J. Exp. Bot.* 72, 4053–4067. doi: 10.1093/JXB/ERAB109

Flores, C. L., and Gancedo, C. (2018). Construction and characterization of a Saccharomyces cerevisiae strain able to grow on glucosamine as sole carbon and nitrogen source. *Sci. Report.* 8:16949. doi: 10.1038/s41598-018-35045-8

Fusari, C., Demonte, A. M., Figueroa, C. M., Aleanzi, M., and Iglesias, A. A. (2006). A colorimetric method for the assay of ADP-glucose pyrophosphorylase. *Anal. Biochem.* 352, 145–147. doi: 10.1016/j.ab.2006.01.024

Gaugué, I., Oberto, J., Putzer, H., and Plumbridge, J. (2013). The use of amino sugars by *Bacillus subtilis*: presence of a unique operon for the catabolism of glucosamine. *PLoS One* 8, e63025. doi: 10.1371/JOURNAL.PONE.0063025

Glasner, M. E., Truong, D. P., and Morse, B. C. (2020). How enzyme promiscuity and horizontal gene transfer contribute to metabolic innovation. *FEBS J.* 287, 1323–1342. doi: 10.1111/febs.15185

Gómez-Casati, D. F., Igarashi, R. Y., Berger, C. N., Brandt, M. E., Iglesias, A. A., and Meyer, C. R. (2001). Identification of functionally important amino-terminal arginines of *Agrobacterium tumefaciens* ADP-glucose pyrophosphorylase by alanine scanning mutagenesis. *Biochemistry* 40, 10169–10178. doi: 10.1021/bi002615e

Guzmán, G. I., Utrilla, J., Nurk, S., Brunk, E., Monk, J. M., Ebrahim, A., et al. (2015). Model-driven discovery of underground metabolic functions in *Escherichia coli*. *Proc. Natl. Acad. Sci. U. S. A.* 112, 929–934. doi: 10.1073/PNAS.1414218112/SUPPL_FILE/PNAS.201414218SI.PDF

Halder, A., Data, D., Seelam, P. P., Bhattacharyya, D., and Mitra, A. (2019). Estimating strengths of individual hydrogen bonds in RNA base pairs: Toward a consensus between different computational approaches. *ACS Omega* 4, 7354–7368. doi: 10.1021/ACSOMEGA.8B03689/SUPPL_FILE/AO8B03689_SI_002.ZIP

Hernandez, M. A., and Alvarez, H. M. (2010). Glycogen formation by *Rhodococcus* species and the effect of inhibition of lipid biosynthesis on glycogen accumulation in *Rhodococcus opacus* PD630. *FEMS Microbiol. Lett.* 312, 93–99. doi: 10.1111/j.1574-6968.2010.02108.x

Hernández, M. A., Alvarez, H. M., Lanfranconi, M. P., Silva, R. A., Herrero, O. M., and Villalba, M. S. (2019). "Central metabolism of species of the genus *Rhodococcus*," in *Biology of Rhodococcus*. ed. H. M. Alvarez (Cham: Springer), 61–85.

Hernández, M. A., Mohn, W. W., Martínez, E., Rost, E., Alvarez, A. F., and Alvarez, H. M. (2008). Biosynthesis of storage compounds by *Rhodococcus jostii* RHA1 and global identification of genes involved in their metabolism. *BMC Genomics* 9:600. doi: 10.1186/1471-2164-9-600

Hey-Ferguson, A., Mitchell, M., and Elbein, A. D. (1973). Trehalose metabolism in germinating spores of *Streptomyces hygroscopicus*. *J. Bacteriol.* 116, 1084–1085. doi: 10.1128/jb.116.2.1084-1085.1973

Hill, B. L., Mascarenhas, R., Patel, H. P., Asencion Diez, M. D., Wu, R., Iglesias, A. A., et al. (2019). Structural analysis reveals a pyruvate-binding activator site in the *Agrobacterium tumefaciens* ADP–glucose pyrophosphorylase. *J. Biol. Chem.* 294, 1338–1348. doi: 10.1074/jbc.RA118.004246

Hol, W. G. J. (1985). The role of the α-helix dipole in protein function and structure. *Prog. Biophys. Mol. Biol.* 45, 149–195. doi: 10.1016/0079-6107(85)90001-X

Iacometti, C., Marx, K., Hönick, M., Biletskaia, V., Schulz-Mirbach, H., Satanowski, A., et al. (2021). Activating silent glycolysis bypasses in *Escherichia coli*. bioRxiv [Preprint]. doi: 10.1101/2021.11.18.468982

Iwamoto, R., Tanaka, M., and Ushida, N. (1998). A new pathway of d-glucosamine catabolism in bacteria. *Nippon Nogeikagaku Kaishi* 72, 937–941. doi: 10.1271/NOGEIKAGAKU1924.72.937

Jaroentomeechai, T., Taw, M. N., Li, M., Aquino, A., Agashe, N., Chung, S., et al. (2020). Cell-free synthetic glycobiology: designing and engineering glycomolecules outside of living cells. *Front. Chem.* 8:645. doi: 10.3389/fchem.2020.00645

Jensen, R. A. (2003). Enzyme recruitment in evolution of new function. *Annu. Rev. Microbiol.* 30, 409–425. doi: 10.1146/annurev.mi.30.100176.002205

Jin, X., Ballicora, M. A., Preiss, J., and Geiger, J. H. (2005). Crystal structure of potato tuber ADP-glucose pyrophosphorylase. *EMBO J.* 24, 694–704. doi: 10.1038/SJ.EMBOJ.7600551

Jolly, L., Pompeo, F., Van Heijenoort, J., Fassy, F., and Mengin-Lecreulx, D. (2000). Autophosphorylation of phosphoglucosamine mutase from *Escherichia coli*. *J. Bacteriol.* 182, 1280–1285. doi: 10.1128/JB.182.5.1280-1285.2000

Kiel, J. A. K. W., Boels, J. M., Beldman, G., and Venema, G. (1994). Glycogen in *Bacillus subtilis*: molecular characterization of an operon encoding enzymes involved in glycogen biosynthesis and degradation. *Mol. Microbiol.* 11, 203–218. doi: 10.1111/j.1365-2958.1994.tb00301.x

Křen, V., and Řezanka, T. (2008). Sweet antibiotics—The role of glycosidic residues in antibiotic and antitumor activity and their randomization. *FEMS Microbiol. Rev.* 32, 858–889. doi: 10.1111/j.1574-6976.2008.00124.x

Kuhn, M. L., Figueroa, C. M., Iglesias, A. A., and Ballicora, M. A. (2013). The ancestral activation promiscuity of ADP-glucose pyrophosphorylases from oxygenic photosynthetic organisms. *BMC Evol. Biol.* 13:51. doi: 10.1186/1471-2148-13-51

Laemmli, U. (1970). Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature* 227, 680–685. doi: 10.1038/227680a0

Manso, J. A., Nunes-Costa, D., Macedo-Ribeiro, S., Empadinhas, N., and Pereira, P. J. B. (2019). Molecular fingerprints for a novel enzyme family in Actinobacteria with glucosamine kinase activity. *mBio* 10, e00239–e00319. doi: 10.1128/mBio.00239-19

Martín, J. F., and Liras, P. (2020). The balance metabolism safety net: integration of stress signals by interacting transcriptional factors in *Streptomyces* and related Actinobacteria. *Front. Microbiol.* 10:3120. doi: 10.3389/FMICB.2019.03120

Moretti, R., and Thorson, J. S. (2007). Enhancing the latent nucleotide triphosphate flexibility of the glucose-1-phosphate thymidylyltransferase RmlA. *J. Biol. Chem.* 282, 16942–16947. doi: 10.1074/jbc.M701951200

Moretti, R., and Thorson, J. S. (2008). A comparison of sugar indicators enables a universal high-throughput sugar-1-phosphate nucleotidyltransferase assay. *Anal. Biochem.* 377, 251–258. doi: 10.1016/j.ab.2008.03.018

Peracchi, A. (2018). The limits of enzyme specificity and the evolution of metabolism. *Trends Biochem. Sci.* 43, 984–996. doi: 10.1016/j.tibs.2018.09.015

Pettersen, E. F., Goddard, T. D., Huang, C. C., Couch, G. S., Greenblatt, D. M., Meng, E. C., et al. (2004). UCSF chimera—A visualization system for exploratory research and analysis. *J. Comput. Chem.* 25, 1605–1612. doi: 10.1002/JCC.20084

Preiss, J. (1984). Bacterial glycogen synthesis and its regulation. *Annu. Rev. Microbiol.* 38, 419–458. doi: 10.1146/ANNUREV.MI.38.100184.002223

Preiss, J. (2009). "Glycogen biosynthesis," in *Encyclopedia of Microbiology*. ed. M. Echaech (Cambridge: Academic Press), 145–158.

Preiss, J. (2014). Glycogen: biosynthesis and regulation. *EcoSal Plus* 6, 145–158. doi: 10.1128/ecosalplus.ESP-0015-2014

Preiss, J., and Romeo, T. (1990). Physiology, biochemistry and genetics of bacterial glycogen synthesis. *Adv. Microb. Physiol.* 30, 183–238. doi: 10.1016/S0065-2911(08)60113-7

Ramakrishnan, B., Shah, P. S., and Qasba, P. K. (2001). Alpha-Lactalbumin (LA) stimulates milk beta-1,4-galactosyltransferase I (beta 4Gal-T1) to transfer glucose from UDP-glucose to N-acetylglucosamine. Crystal structure of beta 4Gal-T1 x LA complex with UDP-Glc. *J. Biol. Chem.* 276, 37665–37671. doi: 10.1074/JBC.M102458200

Rosenberg, J., and Commichau, F. M. (2019). Harnessing underground metabolism for pathway development. *Trends Biotechnol.* 37, 29–37. doi: 10.1016/j.tibtech.2018.08.001

Rueda, B., Miguélez, E. M., Hardisson, C., and Manzanal, M. B. (2001). Changes in glycogen and trehalose content of *Streptomyces brasiliensis* hyphae during growth in liquid cultures under sporulating and non-sporulating conditions. *FEMS Microbiol. Lett.* 194, 181–185. doi: 10.1111/J.1574-6968.2001.TB09466.X

Šali, A., and Blundell, T. L. (1993). Comparative protein modelling by satisfaction of spatial restraints. *J. Mol. Biol.* 234, 779–815. doi: 10.1006/JMBI.1993.1626

Sambrook, J., and Russell, D. (2001). *Molecular Cloning: A Laboratory Manual*. New York: Cold Spring Harbor Laboratory Press

Suen, G., Stevenson, D. M., Bruce, D. C., Chertkov, O., Copeland, A., Cheng, J. F., et al. (2011). Complete genome of the cellulolytic ruminal bacterium *Ruminococcus albus* 7. *J. Bacteriol.* 193, 5574–5575. doi: 10.1128/JB.05621-11

Takarada, H., Sekine, M., Kosugi, H., Matsuo, Y., Fujisawa, T., Omata, S., et al. (2008). Complete genome sequence of the soil actinomycete *Kocuria rhizophila*. *J. Bacteriol.* 190, 4139–4146. doi: 10.1128/JB.01853-07

Takata, H., Takaha, T., Okada, S., Takagi, M., and Imanaka, T. (1997). Characterization of a gene cluster for glycogen biosynthesis and a heterotetrameric ADP-glucose pyrophosphorylase from *Bacillus stearothermophilus*. *J. Bacteriol.* 179, 4689–4698. doi: 10.1128/jb.179.15.4689-4698.1997

Uhde, A., Youn, J. W., Maeda, T., Clermont, L., Matano, C., Krämer, R., et al. (2013). Glucosamine as carbon source for amino acid-producing *Corynebacterium glutamicum*. *Appl. Microbiol. Biotechnol.* 97, 1679–1687. doi: 10.1007/S00253-012-4313-8

Urem, M., Świątek-Połatyńska, M. A., Rigali, S., and van Wezel, G. P. (2016). Intertwining nutrient-sensory networks and the control of antibiotic production in *Streptomyces*. *Mol. Microbiol.* 102, 183–195. doi: 10.1111/mmi.13464

Uttaro, A. D., and Ugalde, R. A. (1994). A chromosomal cluster of genes encoding ADP-glucose synthetase, glycogen synthase and phosphoglucomutase in *Agrobacterium tumefaciens*. *Gene* 150, 117–122. doi: 10.1016/0378-1119(94)90869-9

Uttaro, A. D., Ugalde, R. A., Preiss, J., and Iglesias, A. A. (1998). Cloning and expression of the *glgC* gene from *Agrobacterium tumefaciens*: purification and characterization of the ADPglucose synthetase. *Arch. Biochem. Biophys.* 357, 13–21. doi: 10.1006/ABBI.1998.0786

van der Heul, H. U., Bilyk, B. L., McDowall, K. J., Seipke, R. F., and van Wezel, G. P. (2018). Regulation of antibiotic production in Actinobacteria: new perspectives from the post-genomic era. *Nat. Prod. Rep.* 35, 575–604. doi: 10.1039/c8np00012c

Verma, M., Lal, D., Saxena, A., Anand, S., Kaur, J., Kaur, J., et al. (2013). Understanding alternative fluxes/effluxes through comparative metabolic pathway analysis of phylum actinobacteria using a simplified approach. *Gene* 531, 306–317. doi: 10.1016/J.GENE.2013.08.076

Xu, N., Yang, Q., Yang, X., Wang, M., and Guo, M. (2021). Reconstruction and analysis of a genome-scale metabolic model for *Agrobacterium tumefaciens*. *Mol. Plant Pathol.* 22, 348–360. doi: 10.1111/MPP.13032

Zhang, J., Singh, S., Hughes, R. R., Zhou, M., Sunkara, M., Morris, A. J., et al. (2014). A simple strategy for glycosyltransferase-catalyzed aminosugar nucleotide synthesis. *Chembiochem* 15, 647–651. doi: 10.1002/cbic.201300779

Zhao, J., and Binns, A. N. (2014). GxySBA ABC transporter of *Agrobacterium tumefaciens* and its role in sugar utilization and vir gene expression. *J. Bacteriol.* 196, 3150–3159. doi: 10.1128/JB.01648-14/SUPPL_FILE/ZJB999093279SO1.PDF

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

frontiers | Frontiers in Microbiology

REVIEW
published: 09 May 2022
doi: 10.3389/fmicb.2022.865168

Engineering Acetogenic Bacteria for Efficient One-Carbon Utilization

Hyeonsik Lee[1†], Jiyun Bae[1†], Sangrak Jin[1†], Seulgi Kang[1] and Byung-Kwan Cho[1,2]*

[1]*Department of Biological Sciences, Korea Advanced Institute of Science and Technology, Daejeon, South Korea,* [2]*KI for the BioCentury, Korea Advanced Institute of Science and Technology, Daejeon, South Korea*

C1 gases, including carbon dioxide (CO_2) and carbon monoxide (CO), are major contributors to climate crisis. Numerous studies have been conducted to fix and recycle C1 gases in order to solve this problem. Among them, the use of microorganisms as biocatalysts to convert C1 gases to value-added chemicals is a promising solution. Acetogenic bacteria (acetogens) have received attention as high-potential biocatalysts owing to their conserved Wood–Ljungdahl (WL) pathway, which fixes not only CO_2 but also CO. Although some metabolites have been produced *via* C1 gas fermentation on an industrial scale, the conversion of C1 gases to produce various biochemicals by engineering acetogens has been limited. The energy limitation of acetogens is one of the challenges to overcome, as their metabolism operates at a thermodynamic limit, and the low solubility of gaseous substrates results in a limited supply of cellular energy. This review provides strategies for developing efficient platform strains for C1 gas conversion, focusing on engineering the WL pathway. Supplying liquid C1 substrates, which can be obtained from CO_2, or electricity is introduced as a strategy to overcome the energy limitation. Future prospective approaches on engineering acetogens based on systems and synthetic biology approaches are also discussed.

Keywords: acetogenic bacteria, one-carbon utilization, Wood–Ljungdahl pathway, energy metabolism, biocatalyst

OPEN ACCESS

Edited by:
Biswarup Mukhopadhyay, Virginia Tech, United States

Reviewed by:
Volker Müller, Goethe University Frankfurt, Germany
Bastian Molitor, University of Tübingen, Germany

****Correspondence:***
Byung-Kwan Cho
bcho@kaist.ac.kr

[†]These authors have contributed equally to this work

Specialty section:
This article was submitted to Microbial Physiology and Metabolism, a section of the journal Frontiers in Microbiology

Received: *29 January 2022*
Accepted: *19 April 2022*
Published: *09 May 2022*

Citation:
Lee H, Bae J, Jin S, Kang S and Cho B-K (2022) Engineering Acetogenic Bacteria for Efficient One-Carbon Utilization. Front. Microbiol. 13:865168. doi: 10.3389/fmicb.2022.865168

INTRODUCTION

The rapid increase in fossil fuel usage and greenhouse gas emissions has caused one of the biggest problems for humankind today. C1 gases such as carbon dioxide (CO_2) and carbon monoxide (CO), which constitute greenhouse gases, industrial waste gases, and synthesis gases (syngas), are the main culprits of the climate crisis (Anwar et al., 2018; Ritchie and Roser, 2020). To make the earth a sustainable place, reducing emissions is crucial, and urgent solutions for carbon capturing, utilization, and storage are needed.

C1 gas fermentation could be a solution, which utilizes microbes as biocatalysts. This is a preferable approach, because it does not require high pressure, temperature, cost, and energy, unlike chemical catalysts, such as in the Fischer–Tropsch process (Latif et al., 2014; Dürre, 2017; Köpke and Simpson, 2020). C1 gases are utilized by microbes as feedstocks and finally converted to value-added chemicals under mild conditions that are required for the optimal growth of microbes.

Acetogenic bacteria (acetogens) are promising platform microbes for C1 gas fixation. They are facultative autotrophs that fix CO_2 and CO as carbon or energy sources *via* the unique

metabolic pathway, the Wood–Ljungdahl (WL) pathway (Ragsdale and Pierce, 2008; Drake et al., 2013). Of the CO_2-fixing pathways known to date, the WL pathway is considered the most energetically efficient (Fast et al., 2015; Claassens et al., 2019). In addition, it is the only pathway for CO_2 fixation coupled with an energy conservation system that plays a crucial role in generating cellular energy and sustaining life (Schuchmann and Müller, 2014). Numerous studies have utilized acetogens as biocatalysts to convert C1 gases into value-added chemicals (Berzin et al., 2012, 2013; Köpke et al., 2012; Banerjee et al., 2014; Beck et al., 2014; Woolston et al., 2018; Annan et al., 2019; Huang et al., 2019; Jin et al., 2020; Bae et al., 2022; Liew et al., 2022). Among the native metabolites produced from acetogens, acetate, ethanol, and 2,3-butanediol (2,3-BDO) have been produced by C1 gas fermentation on an industrial scale using a non-engineered strain of *Clostridium autoethanogenum* (Marcellin et al., 2016). Although numerous efforts have been made to engineer acetogens to produce various biochemicals from C1 gases, these studies have been limited to a small-scale. The low energy potential of C1 gases compared to that of glucose is one of the limitations that causes slow growth and low productivity under autotrophic conditions. In addition, acetogenesis overall has a small change in free energy that results in thermodynamic constraints and the synthesis of less than one molecule ATP (Schuchmann and Müller, 2014). The energy limitations of acetogens and the low solubility of gaseous substrates are additional hurdles to overcome, as they can ultimately limit the availability of cellular energy, thereby restricting the production of value-added metabolites (Abubackar et al., 2011; Molitor et al., 2017). Therefore, it is necessary to engineer the acetogenic metabolism, including the WL pathway and energy metabolism, and alleviate the solubility issue by supplying liquid C1 substrates (e.g., methanol and formate) or electricity as alternative electron sources. This will enable to fully exploit the potential of acetogens as biocatalysts for C1 utilization.

In this review, the physiology and metabolism of acetogens are addressed, focusing on the WL pathway and the energy conservation system. Strategies to enhance C1 gas fixation efficiency by engineering the WL pathway and overcoming energy limitations under autotrophic conditions are also introduced. Furthermore, future perspectives on engineering acetogens to achieve highly efficient biocatalysts are discussed.

UNDERSTANDING PHYSIOLOGY AND METABOLISM OF ACETOGENS

To date, over 100 acetogens belonging to 23 genera have been isolated that can grow in diverse environments such as under a wide range of temperature and pH. Depending on the species, various biochemicals, including acetate, ethanol, butyrate, or 2,3-BDO, can be produced from C1 feedstocks (**Table 1**). As an essential precursor for synthesizing these products, acetyl-CoA is generated through the WL pathway in acetogens from either CO_2 or CO. As CO_2 can only serve as a carbon source, CO_2 fixation requires an additional source of energy such as H_2 or CO. Although CO can serve as both carbon and energy sources, the addition of H_2 is desirable to refix CO_2 generated from CO oxidation, as two-thirds of the carbon is lost when CO is used as a sole substrate (Jeong et al., 2015; Bertsch and Müller, 2015b; Weghoff et al., 2016). Along with the WL pathway, the energy metabolism in acetogens also generates reducing equivalents from H_2 and CO. Accordingly, the composition of gas mixtures fed in acetogens affects not only the carbon yield of the bioprocess but also the metabolism of acetogens, which points out that understanding the acetogenic metabolism is important to realize the efficient C1 gas fermentation.

Wood–Ljungdahl Pathway

The WL pathway is mainly composed of two linear metabolic branches: a methyl- and a carbonyl-branch (**Figure 1**). The methyl-branch consists of six reactions, starting with the reduction of CO2 to formate using two reducing equivalents, and formate is transferred to tetrahydrofolate (THF) to generate formyl-THF, consuming one molecule of ATP. Water splits from formyl-THF to generate methenyl-THF, which is further reduced *via* methylene-THF to methyl-THF. Finally, the methyl group is transferred to acetyl-CoA synthase (ACS) *via* the corrinoid iron–sulfur protein (CoFeSP). Most genes involved in the methyl-branch are strongly conserved among phylogenetically diverse acetogens. However, some genes have diverse characteristics depending on the species, such as the formation of various protein complexes or the use of different cofactors (Shin et al., 2016). Unlike the methyl-branch, the carbonyl-branch undergoes a one-step reaction by a multi-component enzyme called carbon monoxide dehydrogenase/acetyl-CoA synthase (CODH/ACS), which synthesizes acetyl-CoA by attaching a carbonyl-group from the reduction of CO_2, or CO directly, to methyl-CoFeSP generated from the methyl-branch. Acetyl-CoA can be used as a building block for various chemicals. Consequently, C1 gas fixation using the WL pathway requires only one molecule of ATP, which is the lowest energy requirement among all existing biological CO_2 fixing metabolic pathways (Fast and Papoutsakis, 2012). Thus, acetogens have received attention as biocatalysts that can efficiently fix C1 gases. In this section, we compare and summarize the processes of C1 assimilation by several major acetogens.

Methyl-Branch

The methyl-branch is a linear metabolic branch that converts CO_2 to methyl groups. To compare the methyl-branch between the acetogenic species, we discuss the reactions of the methyl-branch into the following three parts: (i) the first step in which CO_2 is converted to formate, (ii) the second step in which formate is converted to methyl-THF, and (iii) the last step in which methyl-THF is converted to methyl-CoFeSP (**Figure 1**).

Step 1: CO_2 to Formate

The first step in the methyl-branch is the reduction of CO_2 to formate by selenocysteine- or non-selenocysteine-containing formate dehydrogenase (FDH). As the standard redox potential

TABLE 1 | Features of diverse acetogenic species.

Organism	Substrate	Products	Optimal growth temperature (°C)	Optimal pH	Genome complete levels	References
Acetobacterium bakii	H_2/CO_2, CO, methanol	Acetate	20	6.5	Scaffold	Kotsyurbenko et al., 1995; Hwang et al., 2015
Acetobacterium woodii	H_2/CO_2, methanol, formate	Acetate	30	7.0	Complete	Balch et al., 1977; Bache and Pfennig, 1981; Poehlein et al., 2012
Acetohalobium arabaticum	H_2/CO_2, CO	Acetate	38–40	7.6–8.0	Complete	Zhilina and Zavarzin, 1990; Sikorski et al., 2010
Blautia producta	H2/CO2, CO	Acetate	37	7.0	Complete	Lorowitz and Bryant, 1984; Geerligs et al., 1987; Liu et al., 2008; Tourlousse et al., 2020
Clostridium aceticum	H_2/CO_2, CO	Acetate	30	8.3	Complete	Wieringa, 1936; Braun et al., 1981; Lux and Drake, 1992; Poehlein et al., 2015b
Clostridium autoethanogenum	H_2/CO_2, CO	2,3-BDO, acetate, ethanol	37	6.0	Complete	Abrini et al., 1994; Köpke et al., 2011; Brown et al., 2014
Clostridium carboxidivorans	H_2/CO_2, CO	Acetate, ethanol, butyrate, butanol	38	5.0–7.0	Complete	Liou et al., 2005; Li et al., 2015
Clostridium coskatii	H_2/CO_2, CO	Acetate, ethanol	37	6.0	Contig	Zahn and Saxena, 2012; Bengelsdorf et al., 2016
Clostridium drakei	H_2/CO_2, CO	Acetate, ethanol, butyrate	30	5.4–7.5	Complete	Küsel et al., 2000; Liou et al., 2005; Gossner et al., 2008; Jeong et al., 2014
Clostridium formicaceticum	CO, CH_3OH	Acetate, formate	37	8.1	Complete	Andreesen et al., 1970; Lux and Drake, 1992; Karl et al., 2017
Clostridium ljungdahlii	H_2/CO_2, CO, formate	2,3-BDO, acetate, ethanol	37	6.0	Complete	Tanner et al., 1993; Köpke et al., 2010, 2011
Clostridium magnum	H_2/CO_2, methanol	Acetate	30	7.2	Scaffold	Schink, 1984; Bomar et al., 1991; Uhlig et al., 2016
Clostridium ragsdalei	H_2/CO_2, CO	2,3-BDO, acetate, ethanol	37	6.3	Contig	Huhnke et al., 2008; Köpke et al., 2011; Bengelsdorf et al., 2016
Clostridium scatologenes	H_2/CO_2, CO, formate	Acetate, ethanol, butyrate	37	5.4–7.0	Complete	Liou et al., 2005; Zhu et al., 2015
Eubacterium limosum	H_2/CO_2, CO, methanol, formate	Acetate, Butyrate	37	7.0	Complete	Eggerth, 1935; Genthner et al., 1981; Genthner and Bryant, 1982; Genthner and Bryant, 1987; Song and Cho, 2015
Eubacterium callanderi	H_2/CO_2, CO, methanol	Acetate, butyrate	37	7.0	Complete	Chang et al., 1997, 2001; Roh et al., 2011
Sporomusa ovata	H_2/CO_2, methanol, formate	Acetate	34	6.3	Scaffold	Möller et al., 1984; Poehlein et al., 2013
Thermoacetogenium phaeum	H_2/CO_2, methanol, formate	Acetate	60	6.8	Complete	Hattori et al., 2000; Oehler et al., 2012
Thermoanaerobacter kivui	H_2/CO_2, CO, formate	Acetate	66	6.4	Complete	Leigh et al., 1981; Daniel et al., 1990; Hess et al., 2014
Treponema primitia	H_2/CO_2	Acetate	30	7.2	Complete	Graber et al., 2004; Graber and Breznak, 2004; Rosenthal et al., 2011
Moorella thermoacetica	H_2/CO_2, CO, methanol, formate	Acetate	55	7.0	Complete	Fontaine et al., 1942; Kerby and Zeikus, 1983; Daniel et al., 1990; Bengelsdorf et al., 2015; Poehlein et al., 2015a

of this reaction is −432 mV, NADH (−320 mV) is not sufficient as an electron donor to drive this reaction. Instead, ferredoxin (−450 to −500 mV), H_2 (−414 mV), and NADPH (−370 mV) can be used for the reduction of CO_2 to formate (Wang et al., 2013; Schuchmann and Müller, 2014). Depending on the electron delivery system, model acetogens are divided into three types. The first model acetogen, *Acetobacterium woodii*, uses hydrogen-dependent CO_2 reductase (HDCR), an FDH complex linked to hydrogenase, to receive electrons directly from hydrogen and convert CO_2 to formate. In the case of the second model

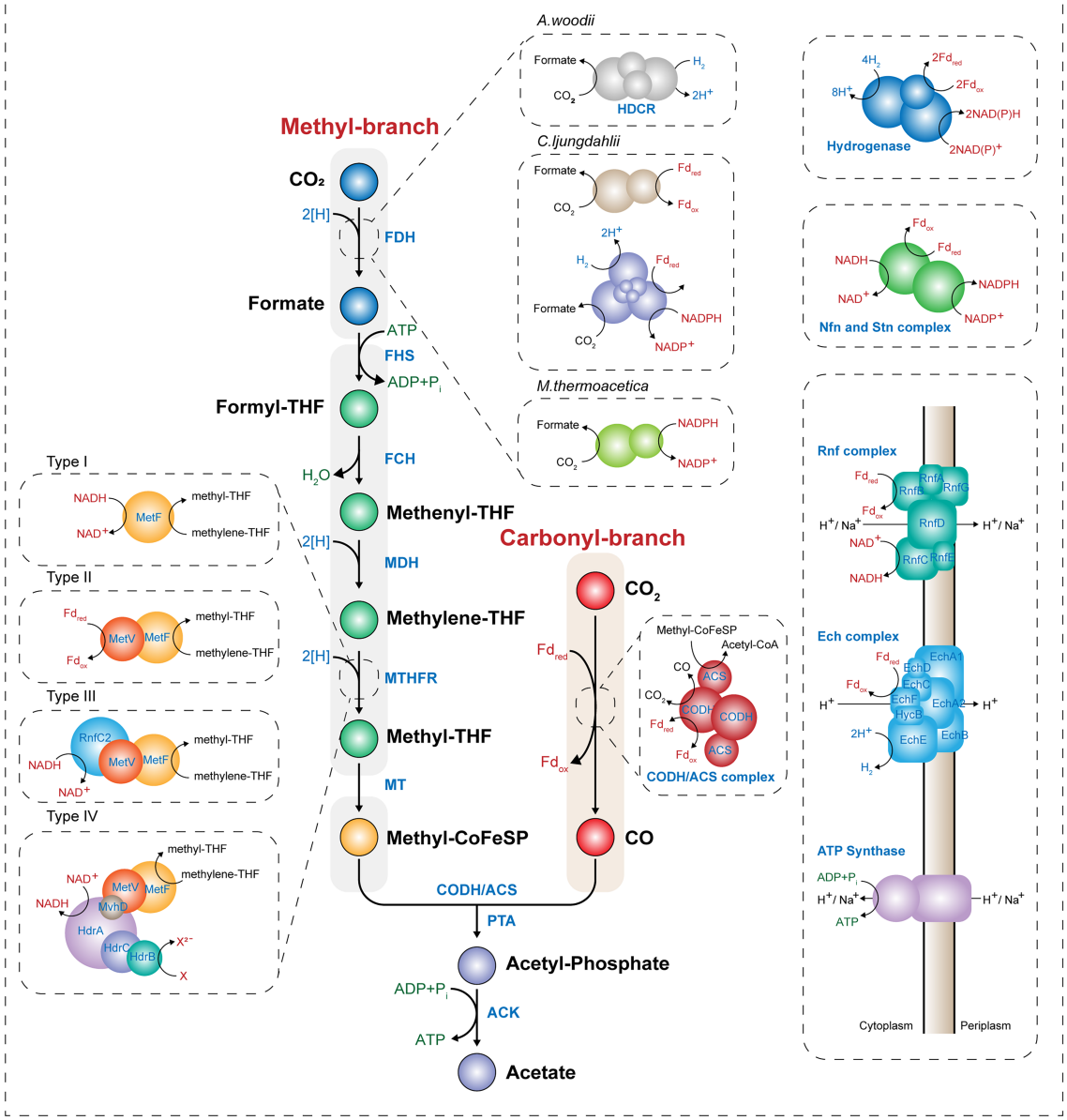


FIGURE 1 | Scheme of the Wood–Ljungdahl pathway and energy conservation. The Wood–Ljungdahl pathway and energy conservation system in acetogens. CO, carbon monoxide; CO_2, carbon dioxide; THF, tetrahydrofolate; FDH, formate dehydrogenase; FHS, formyl-tetrahydrofolate synthase; FCH, formyl-cyclohydrolase; MDH, methylene-tetrahydrofolate dehydrogenase; MTHFR, methylene-tetrahydrofolate reductase; MT, methyltransferase; CoFeSP, corrinoid iron–sulfur protein; CODH, CO dehydrogenase; ACS, acetyl-CoA synthase; PTA, phosphotransacetylase; ACK, acetate kinase; HDCR, hydrogen-dependent CO_2 reductase; Fd_{ox}, oxidized ferredoxin; Fd_{red}, reduced ferredoxin; and P_i, inorganic phosphate.

species, *Clostridium ljungdahlii*, reduced ferredoxin (Fd_{red}) or NADPH is used as cofactor for the two types of FDH. The third model acetogen, *Moorella thermoacetica*, converts CO_2 into formate using only NADPH as a cofactor (**Figure 1**; Wang et al., 2013; Schuchmann and Müller, 2014; Shin et al., 2016).

Step 2: Formate to Methyl-THF

The second step is the conversion of the formyl-group to the methyl group in THF. In this process, one molecule of ATP and four electrons are consumed. First, formyl-THF synthetase (FHS) uses one ATP molecule to bind formate to THF to form formyl-THF. Subsequently, methyl-THF cyclohydrolase (FCH) cyclizes the formyl-group linked to THF to generate a methenyl intermediate, and methylene-THF is finally converted by methylene-THF dehydrogenase (MDH). In *M. thermoacetica*, cyclohydrolase and dehydrogenase function as a single bifunctional protein complex, whereas *Clostridium formicoaceticum* (Clark et al., 1982) and *A. woodii* (Ragsdale and Ljungdahl, 1984) have monofunctional proteins. In particular, when converting methenyl-THF to methylene-THF,

two electrons are used and transferred by either NADH or NADPH depending on the species (Moore et al., 1974; Ragsdale and Ljungdahl, 1984). The last step is the conversion of methylene-THF to methyl-THF by methylene-THF reductase (MTHFR). Interestingly, this enzyme reaction is classified into four types depending on the acetogenic species (Öppinger et al., 2022). In Type I, MetF alone generates methyl-THF using one molecule of NADH in *E. coli* (Sheppard et al., 1999), *Thermus thermophilus* (Igari et al., 2011), and *Blautia producta* (Wohlfarth et al., 1990). The Type II MTHFR system consists of the MetV-MetF complex and is found in *C. formicoaceticum*, *C. ljungdahlii*, and *Thermoanaerobacter kivui*. This Type II MTHFR complex protein obtains electrons from Fd_{red} to convert methylene-THF to methyl-THF (Dietrich et al., 2021; Katsyv et al., 2021; Wiechmann and Müller, 2021). The Type III system used by *A. woodii* has MetV-MetF bound to RnfC2, which has a stoichiometry of 1:1:1, and NADH is used as a cofactor (Bertsch et al., 2015). Finally, in the Type IV MTHFR system, MetV-MetF constitutes a complex in which HdrCBA and MvhD are connected. This complex may use electron bifurcation using two molecules of NADH cofactor to generate methyl-THF and one molecule of a reduced electron carrier (e.g., Fd_{red}). Among acetogens, *M. thermoacetica* and *Sporomusa ovata* are known to have the Type IV MTHFR system (**Figure 1**; Mock et al., 2014). Such differences appear to be due to the utilization of various C1 substrates, including CO and methanol, or difference in optimal growth conditions (e.g., temperature, pH, or metal cofactor), and play a role in the optimal regulation of the intracellular redox balance.

Step 3: Methyl-THF to Methyl-CoFeSP

The final step is to form methyl-CoFeSP by transferring methyl groups from methyl-THF to CoFeSP by methyltransferase (MT; **Figure 1**). THF, from which the methyl group has been removed, is recycled by several metabolic reactions. Methyl-CoFeSP, the final product of the methyl-branch of the WL pathway, is converted into Acetyl-CoA while transferring the methyl group to the CODH/ACS complex of the carbonyl-branch, and the CoFeSP is also recycled.

Carbonyl-Branch

The carbonyl-branch is a pathway for synthesizing acetyl-CoA by combining a carbonyl-group with methyl-CoFeSP obtained from the methyl-branch. When CO is a substrate, CO can be directly incorporated into the carbonyl-branch, while CO_2 can be used after its reduction to CO. These reactions are catalyzed by the CODH/ACS complex. CODH catalyzes either the reduction of CO_2 to CO or the oxidation of CO to CO_2 (**Figure 1**). It is a homodimeric enzyme that contains five Fe-S clusters. Because the CO_2 to CO reduction reaction is the largest thermodynamic barrier in the WL pathway and has a very low standard redox potential ($E_0' = -520$ mV; Thauer et al., 1977), CODH enzymes in most acetogens use only Fd_{red} as an electron donor. Some acetogenic species such as *M. thermoacetica* and *C. ljungdahlii*, which are capable of CO oxidation, have a Ni-Fe-S reaction center in the C-cluster, and Ni insertion accessory proteins are used to construct the Ni insertion C-cluster. Hence, these protein families play an essential role in the growth of the acetogen on CO under autotrophic conditions. When the gene encoding the Ni insertion accessory protein, such as *cooC*, is deleted, the mutant strain needs a high Ni cation concentration for autotrophic growth on CO (Kerby et al., 1997; Jeon et al., 2001).

ACS is a protein containing an A-cluster with a Ni–Ni-Fe_4S_4 active site. This enzyme catalyzes the synthesis of acetyl-CoA, combining CO with methyl-CoFeSP obtained from the carbonyl- and methyl-branch, respectively. Protein crystal analysis of the CODH/ACS complex showed that this complex has two structural forms. In the closed-form case, a gas tunnel is generated from the C-cluster of CODH to the A-cluster of the ACS protein. Along this tunnel, one molecule of CO can be transferred from the C-cluster to the A-cluster, and carboxylation of CO with methyl-CoFeSP in the A-cluster results in acetyl-CoA production. Next, in the open-form, the CO gas tunnel is disconnected by a conformational change in the CODH/ACS complex. In this form, methylation of the A-cluster becomes possible as CoFeSP can access to the active site of A-cluster (Doukov et al., 2008). Acetyl-CoA can be synthesized through this conformational change by attaching a CO molecule and a CoA cofactor to methyl-CoFeSP in the CODH/ACS complex.

Through the cooperation of the methyl- and carbonyl-branch, acetogens synthesize one molecule of acetyl-CoA from two molecules of CO_2 or CO. Acetyl-CoA is an intracellular building block in living organisms and is used to increase cell mass or synthesize one molecule of acetate through SLP *via* catalytic reactions of phosphotransacetylase (PTA) and acetate kinase (ACK). Therefore, the WL pathway consumes one molecule of ATP to convert formate to formyl-THF and produces one molecule of ATP through SLP during acetate synthesis, resulting in a net ATP yield of zero.

Energy Metabolism

As one molecule of ATP is consumed for fixation of C1 gases through the WL pathway, additional cellular energy is required for acetogens to grow C1 gas conditions. Although the oxidation of CO generates Fd_{red}, acetogens usually require an additional energy source, such as H_2, to obtain reducing equivalents and to fix the residual CO_2 produced from the oxidation of CO. Hydrogenase has diverse protein characteristics in different microbial species but the common role is to generate Fd_{red} or NAD(P)H from H_2. Electron bifurcation found in acetogens is a mechanism of biological energy conservation that couples the exergonic oxidation and endergonic reduction reactions. (Peters et al., 2016). Acetogens have electron-bifurcating hydrogenases that couple the exergonic reduction of NAD^+ to endergonic reduction of ferredoxin with exergonic oxidation of H_2 by making the overall reaction exergonic (**Figure 1**; Schuchmann and Müller, 2014). During C1 gas fixation, reducing equivalents are supplied *via* electron carriers to the redox reactions involved in the WL or other metabolic pathways for cell

growth. Interestingly, Fd_{red} produced by CODH or hydrogenase generates an ion gradient between cellular membranes *via* the membrane-bound respiratory enzyme complex of acetogens and yields additional ATP using membrane-bound ATP synthase. Generation of ATP *via* these respiratory enzyme complexes involved in energy conservation systems is the mode used by acetogens to sustain their lives.

There are two different complexes used by acetogens, the Rnf and Ech complexes (**Figure 1**). The Rnf complex is composed of six subunits, as found in *A. woodii*, *C. ljungdahlii*, and *C. autoethanogenum*, and has a ferredoxin:NAD^+ oxidoreductase activity (Biegel and Müller, 2010). It receives electrons from Fd_{red} and transfers them to NAD^+ to generate NADH. It pumps cations such as H^+ or Na^+ from inside the cells. In contrast, the Ech complex, which consists of eight or nine subunits and is found in *M. thermoacetica* and *T. kivui*, has ferredoxin:H^+ oxidoreductase activity that uses protons instead of NAD^+ as the final electron acceptor and pumps out protons during the generation of hydrogen (Hedderich and Forzi, 2005). The generated electrochemical ion gradient results in ATP synthesis by the membrane-bound ATP synthase. The Rnf-containing, Na^+-dependent acetogen *A. woodii* is known to produce one molecule of ATP per 3.3 Na^+, as experimentally determined by solving the crystal structure of *A. woodii* ATP synthase (Matthies et al., 2014). In contrast, a one molecule yield of ATP per 4 H^+ is assumed for other species such as *C. ljungdahlii* and *M. thermoacetica* (Schuchmann and Müller, 2014).

In addition, some acetogens such as *C. autoethanogenum* and *S. ovata* have been reported to have NADH-dependent reduced ferredoxin:$NADP^+$ oxidoreductase (Nfn) and *Sporomusa*-type Nfn (Stn), respectively (**Figure 1**; Wang et al., 2013; Kremp et al., 2020; Mahamkali et al., 2020). The Nfn complex is composed of two subunits of NfnAB, each of which is known to have several Fe-S reaction centers. It is known for its redox balancing potential by producing NADPH from Fd_{red} and NADH through electron bifurcation (Mahamkali et al., 2020). Therefore, both energy conservation and redox balancing systems are highly important for acetogens to generate cellular energy during autotrophic growth under C1 gas conditions.

ENGINEERING THE WL PATHWAY TO ENHANCE THE EFFICIENCY OF C1 GAS FIXATION

Because C1 gases, unlike other substrates such as glucose or glycerol, are gaseous substrates, the gas-to-liquid mass transfer rate is critically affected by the physical properties of gas solubility. Many studies have attempted to increase the fixing efficiency of C1 gases through various gas fermentation techniques, such as increasing the partial pressure of the gas (Phillips et al., 1993; Bredwell and Worden, 1998; Hurst and Lewis, 2010; Orgill et al., 2013; Sathish et al., 2019; Bae et al., 2022). However, this approach also has physical limitations, and thus, genetic attempts should be made to overcome the low productivity, yield, and cell density of acetogens by developing a platform acetogen strain with increased C1 gas fixation efficiency and expanding it to a commercial scale. This chapter summarizes a few approaches that use simple genetic manipulation to increase C1 gas fixation efficiency of acetogens.

Engineering of the Methyl-Branch in the WL Pathway

To increase the C1 gas fixation efficiency *via* the WL pathway, the most straightforward approach is to overexpress genes encoding enzymes of the methyl-branch (**Figure 2A**). When four THF-dependent enzymes (FHS, FCH, MDH, and MTHFR) of *C. ljungdahlii* were overexpressed in *A. woodii*, its growth rate increased approximately 1.1-fold and acetate production increased approximately 1.2-fold compared to the empty vector control under C1 autotrophic batch cultivation. In addition, the engineered strain showed approximately a 1.6-fold increase in the specific activity of ACK, whereas strains overexpressing PTA or ACK showed approximately a 1.2-fold increase in specific activity of ACK (Straub et al., 2014). This result indicated that CO_2 fixation efficiency can be increased by upregulating the genes involved in the methyl-branch.

Another way to increase CO_2 fixation efficiency is to introduce a new metabolic pathway that can fix CO_2 in connection with the WL pathway (**Figure 2B**). An example is the glycine synthase-reductase pathway (GSRP). *Clostridium drakei* has a unique CO_2-fixing metabolic pathway, in which the WL pathway and GSRP are connected. Methylene-THF, an intermediate of the methyl-branch of the WL pathway, is converted to glycine or serine through GSRP or the reductive glycine pathway (RGP), and these amino acids are converted to acetate through acetyl-phosphate. In this process, one ATP molecule can be synthesized through SLP. Heterologous expression of GSRP in *Eubacterium limosum* showed that the CO_2 consumption rate and acetate production rate increased 1.5-fold (Song et al., 2020). These results suggest the possibility of increasing the C1 gas-fixing efficiency by introducing an additional C1 gas-fixing metabolic pathway that can be connected to the WL pathway.

Engineering of the Carbonyl-Branch in the WL Pathway

As mentioned above, the only way to utilize CO as carbon and energy sources in acetogens is using the carbonyl-branch of the WL pathway, and CODH/ACS complex is involved in this process. *C. autoethanogenum* and *T. kivui* can grow well under CO autotrophic condition (Weghoff et al., 2016; Liew et al., 2017), whereas some acetogens such as *A. woodii* and *E. limosum* show low CO oxidation rate and growth retardation at a high concentration of CO (Bertsch and Müller, 2015b; Kang et al., 2020). Hence, engineering the CODH/ACS complex is necessary for efficient utilization of CO and improving growth on CO in these acetogens. To increase the CO oxidation rate, the most straightforward approach is to overexpress all genes encoding proteins of the CODH/ACS complex (**Figure 2C**). The

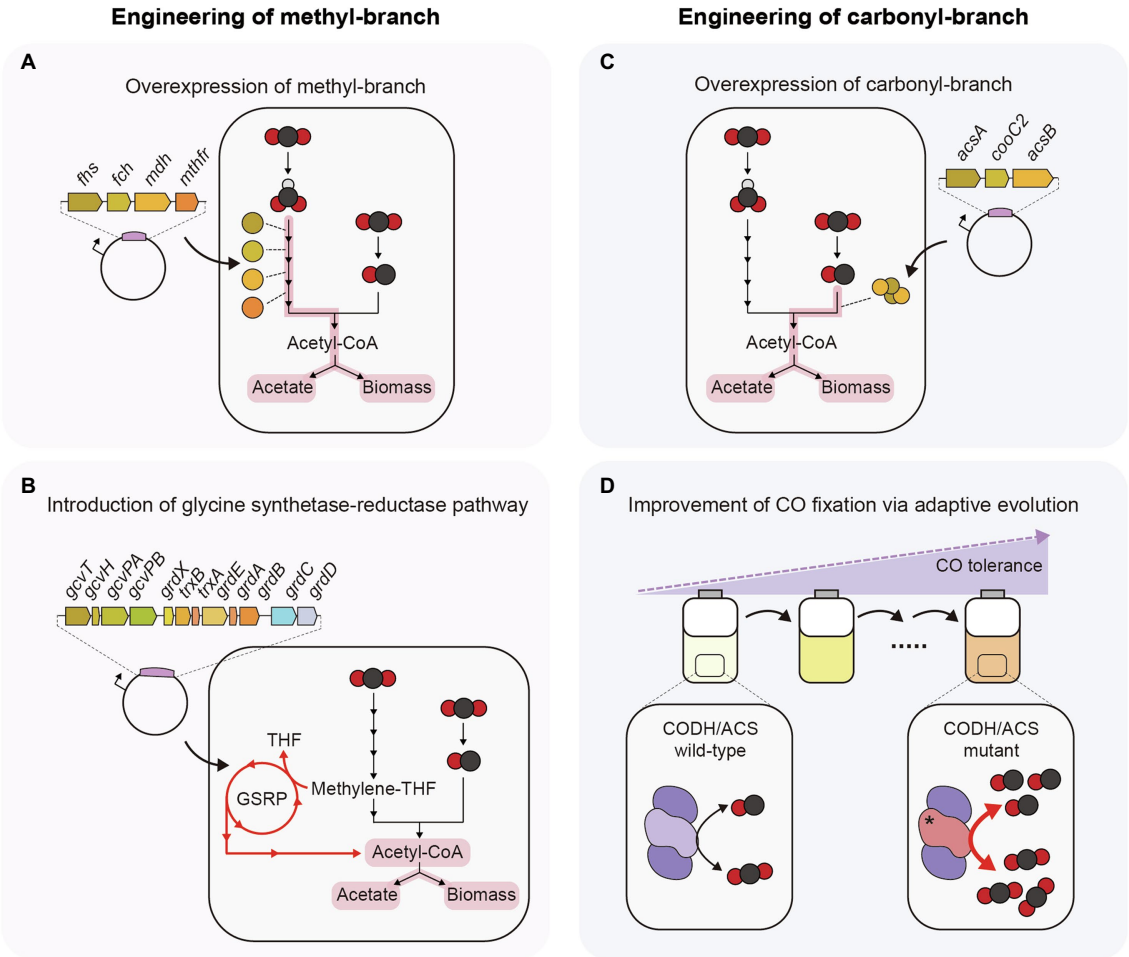


FIGURE 2 | Strategies to enhance efficiency of C1 gas fixation in acetogens. **(A)** Plasmid-based overexpression of THF-dependent enzymes. **(B)** Introduction of glycine synthase-reductase pathway. **(C)** Plasmid-based overexpression of CODH, ACS, and CooC2. **(D)** CODH/ACS mutant obtained from adaptive laboratory evolution under a high concentration of CO. GSRP, Glycine synthase-reductase pathway.

overexpression of *acsA*-encoding CODH, *acsB* encoding ACS, and *cooC2* encoding maturation protein in *Eubacterium callanderi* KIST612 using plasmids led to an increase in the CO oxidation and acetate production rates by 3.1-fold and 1.4-fold, respectively, compared to the control strain, whereas there was no difference when individual genes were overexpressed (Kang et al., 2021).

In one case, the CO oxidation rate was increased by changing the protein sequence of CODH/ACS but not by overexpression of the CODH/ACS complex (**Figure 2D**). In this study, adaptive laboratory evolution was performed on the *E. limosum* ATCC8486 strain under 44% CO conditions to enhance tolerance to CO. In the evolved strain, a C290A single-nucleotide variation (SNV) was found in the *acsA*-encoding CODH catalytic subunit. This SNV caused an A97E amino acid change, and the mutant strain showed a 1.4-fold increase in both the growth rate and the CO consumption rate under autotrophic conditions of 44% CO syngas compared to the wild-type strain (Kang et al., 2020). These results show that C1 gas utilization efficiency can be increased by altering the kinetics of the CODH/ACS protein. Overexpression of such mutant CODH proteins may facilitate higher C1 gas utilization than wild-type CODH/ACS.

STRATEGIES TO OVERCOME ENERGETIC LIMITATIONS IN C1 GAS UTILIZATION

Acetogens suffer from insufficient energy supply in autotrophy, as they are known to live at the thermodynamic edge of life (Schuchmann and Müller, 2014). As mentioned earlier, the energy limitations of acetogens and the low solubility of gaseous substrates restrict the production of energetically high-cost metabolites. Fortunately, besides C1 gases, acetogens can metabolize diverse substrates, including sugars, alcohols, carboxylic acids, and methanol as alternative carbon or energy sources (Perez et al., 2013; Schuchmann and Müller, 2016). The metabolic flexibility of acetogens allows them to overcome energetic limitations by facilitating energy supply.

Liquid C1 Feedstocks: Methanol and Formate

The electrochemical or photochemical reduction of CO_2 generates formate and methanol, which are liquid forms of C1 feedstock with the benefits of mass transfer and energy efficiency (Cotton et al., 2020). Unlike C1 gases, they are easily soluble in water and are transportable, storable, and safe. The most promising aspect is their energy efficiency, as the conversion of methanol or formate is higher than that achieved with H_2/CO_2 or CO (Claassens et al., 2019).

Methanol and formate are directly assimilated into the WL pathway (**Figure 3A**). The utilization of these substrates in acetogens has been shown to improve cell growth and product yields of reduced chemicals such as butyrate. For example, *E. limosum* cultivated on methanol has not only shown higher growth rates than cells cultivated with C1 gases (Genthner and Bryant, 1987; Loubière et al., 1992), but also high product yields with the production of 12 mM acetate and 3.7 mM butyrate from 20 mM methanol (Litty and Müller, 2021). The yield of butyrate under methanol conditions was significantly higher than that obtained under H_2/CO_2 or CO conditions, where acetate was the major product while butyrate was produced in trace amounts. As methanol retains more reducing equivalents than gaseous substrates, a surplus of reducing equivalents can be used to drive biosynthesis and improve the conversion yield of butyrate. The increased butyrate-to-acetate ratio in the presence of methanol was assumed to be due to the role of butyrate production in the NAD(P)H-NAD(P)$^+$ balance during methanol assimilation, as NAD(P)H generation from methanol is catalyzed by 3-hydroxybutyryl-CoA dehydrogenase to regenerate NAD(P)$^+$ (Flaiz et al., 2021).

Similarly, *A. woodii* has been regarded as an interesting organism for anaerobic formate-based bio-production because of the high-energy efficiency achieved through acetate production from formate compared to that from H_2/CO_2 (Cotton et al., 2020). A recent study showed that the cultivation of *A. woodii* with formate as the sole carbon and energy source resulted in conversion of formate to acetate at a higher rate and efficiency than cultivation on gaseous substrates (Moon et al., 2021; Neuendorf et al., 2021). The study also compared the energetic efficiency of different acetogens and other common microbial hosts (e.g., formatotrophs or engineered strains capable of utilizing formate or methanol) during growth and product formation on C1 or sugar substrates, revealing that acetogens show superior energy efficiency on all substrates analyzed, with the highest values for C1 substrates. Formate showed an even higher energy potential than gaseous substrates. Among acetogens, *A. woodii* and *E. limosum* are promising biocatalysts for the conversion of formate into acetate, as they form acetate as a major product during growth on formate (Litty and Müller, 2021).

It should be noted that utilization of formate and methanol has several disadvantages. During formate consumption, pH of the medium can increase slightly (Litty and Müller, 2021; Moon et al., 2021), and the produced alcohols are toxic to some microorganisms at high concentrations, presumably because they damage the cell membrane and cause end-product inhibition of glycolytic enzymes (Dürre et al., 1988; Caldwell, 1989). The increase in pH during formate consumption can be controlled through addition of buffers during fermentation. Regarding the toxicity of methanol, adaptive evolution of acetogens in high concentrations of methanol can increase their tolerance to methanol, as shown by a previous study wherein an adapted strain of *S. ovata* achieved a 5-fold increase in the growth rate of methanol with a higher tolerance to the solvent (Tremblay et al., 2015). More importantly, the reduction of formate to acetate produces two molecules of CO_2 in the absence of excess reducing equivalents to refix the released CO_2, resulting in a carbon efficiency of only 50%. To reduce the carbon loss, addition of H_2 can be a strategy to enable complete fixation

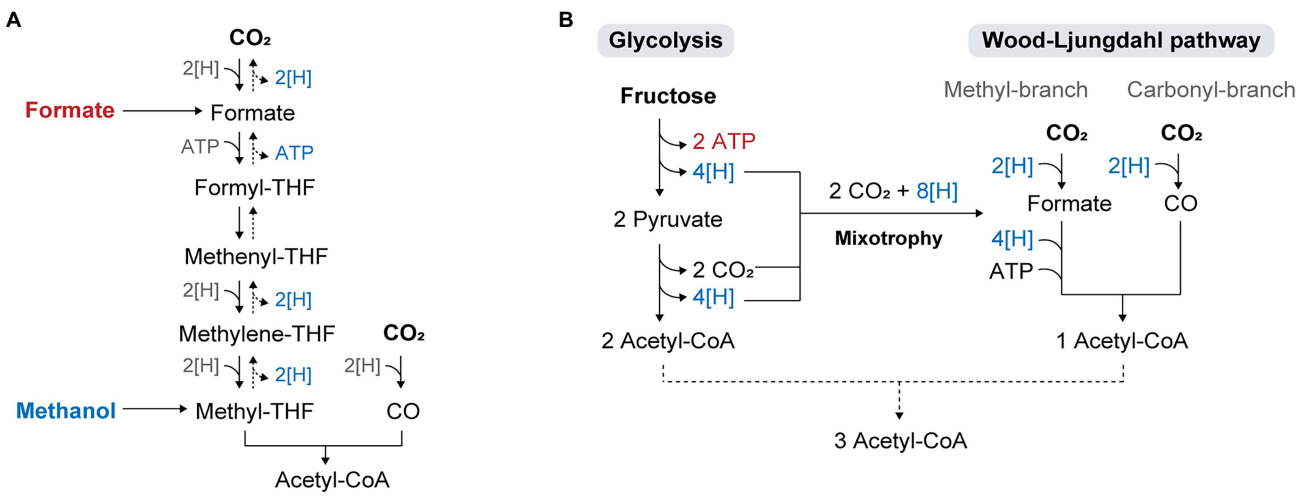


FIGURE 3 | Metabolic flexibility of acetogens utilizing diverse substrates. **(A)** Direct assimilation of formate and methanol into the WL pathway. **(B)** Concurrent utilization of Glycolysis and the WL pathway in acetogenic mixotrophy.

of CO_2, improve the carbon efficiency, and even facilitate net CO_2 uptake (Molitor et al., 2016; Novak et al., 2021).

Renewable Reducing Power: Electricity and Light

H_2 and CO are poor electron donors compared to high-energy substrates, such as methanol and glucose, resulting in insufficient electron supply in acetogens (Emerson and Stephanopoulos, 2019). Instead, the direct supply of such electron sources could be a potential strategy to overcome cellular energy limitations in acetogens. Replacing H_2 or CO with renewable reducing powers for gas fermentation has been demonstrated in acetogens with added advantages of the enhanced efficiency of C1 gas conversion and the metabolic shift toward reduced products (Nevin et al., 2011; Kracke et al., 2016; Im et al., 2018; Cheng et al., 2022).

Microbial electrosynthesis (MES) is a process in which microbes use electrons derived from an electrode to reduce CO_2 to multi-carbon products (Nevin et al., 2010). Several acetogens have been reported to perform MES *via* extracellular electron transfer. For example, direct electron uptake from the electrode has been proposed for *S. ovata, Clostridium aceticum, C. ljungdahlii*, and *M. thermoacetica*, showing reduction of CO_2 into organic acids; however, clear evidence of direct electron transfer is still lacking (Nevin et al., 2010, 2011). In contrast, *C. autoethanogenum* requires an electron mediator such as methyl viologen (MV) or neutral red to transfer electrons from the electrode to cells (Kracke et al., 2016). As some acetogens grow on the electrode by forming a biofilm (Nevin et al., 2010), electron transfer can be spatially limited to cells in contact with the electrode and affected by the biofilm density (Pirbadian et al., 2020), which can result in complex spatial patterns of activity in the bioelectrochemical system (BES).

Extracellular electron supply in the BES induced a significant metabolic shift in *C. autoethanogenum*, resulting in a significant reduction in acetate production, with increased production of lactate and 2,3-BDO by 35-fold and 3-fold, respectively, compared to that under conditions without the electron mediator (Kracke et al., 2016). Recently, *Clostridium carboxidivorans* was found to be capable of MES with MV or biochar as an electron mediator, which facilitated electricity-driven autotrophic CO_2 fixation (Cheng et al., 2022). The standard redox potential of MV is −446 mV, which is sufficiently low to reduce ferredoxin (−450 to −500 mV) and NADH (−320 mV). Hence, it can regulate the redox balance and redistribute the redox flux in acetogens (Michaelis and Hill, 1933; Schuchmann and Müller, 2014). However, increasing the amount of MV supplementation has been reported to have toxic effects on cells (Cheng et al., 2022). Alternatively, neutral red is another electron shuttle with low toxicity and a standard redox potential similar to NADH (Harrington et al., 2015). Addition of neutral red was found to improve the production of volatile fatty acids from CO by increasing the reducing power (Im et al., 2018).

MES powered by solar energy enables artificial photosynthesis in acetogens, with the same net overall reaction as plant-based photosynthesis, in which CO_2 and water are converted to organic compounds and oxygen (Nevin et al., 2010). Light-capturing catalysts, such as cadmium sulfide (CdS) or gold (Au) nanoparticles (NPs), are attached to cells and deliver electrons obtained from light to the cells. For example, *M. thermoacetica* displaying biologically synthesized CdS-NPs or Au-NPs were found to convert CO_2 to acetate (Sakimoto et al., 2016; Zhang et al., 2018). However, the proposed system was limited by the biosynthesis of photo-responsive CdS, which decreases chemical production due to a defense mechanism against heavy metals in bacteria. To circumvent this issue, the chemically synthesized CdS-NP was attached to *C. autoethanogenum*, where the successful conversion of CO_2 into acetate was demonstrated using only light-induced electrons from CdS-NPs (Jin et al., 2021). However, the mechanism of extracellular electron transfer from nanoparticles remains unclear. Transcriptional analysis of *C. autoethanogenum* attached to CdS-NPs suggested that electrons from CdS-NPs are transferred to cells indirectly through mediators such as iron or flavin mononucleotide (FMN). Another study conducted in *M. thermoacetica*-CdS suggested two types of electron transfer pathways: The H_2 generation pathway mediated by membrane-bound hydrogenase and an H_2-independent pathway (Kornienko et al., 2016). Interestingly, a recent study found that neither CdS-NPs nor light was involved in CO_2 fixation, but cysteine present in the medium was used as the carbon source for acetate production in *M. thermoacetica* (Göbbels et al., 2021). These findings indicate that the underlying mechanism of this system is not fully understood; hence, more research is needed to exploit the potential of artificial photosynthesis systems with acetogens.

Mixotrophy

Acetogenic metabolism operates at a thermodynamic limit, which results in an insufficient energy supply during autotrophy (Schuchmann and Müller, 2014). This problem becomes more critical during the cultivation of engineered acetogens, where antibiotics are added to maintain the plasmid, or when the engineered strains express heterologous biosynthesis pathways for energetically high-cost metabolites. In a study, limited autotrophic growth was observed in cells harboring an empty plasmid (Shin et al., 2019), and the non-native chemical mevalonate was only produced under fructose conditions because of the limited availability of reducing equivalents and ATP during C1 gas fermentation (Diner et al., 2018). To overcome the thermodynamic problem and improve autotrophic growth under such conditions, sugar co-feeding is a promising strategy. With a broad range of substrate utilization, acetogens can improve bioenergetics by using sugars along with C1 feedstocks in a process called mixotrophy, where carbon is fixed through both the glycolytic and WL pathways (Fast and Papoutsakis, 2012; Tracy et al., 2012; Fast et al., 2015). Sugar oxidation *via* glycolysis generates eight reducing equivalents and two moles of CO_2, all of which are re-assimilated *via* the WL pathway (**Figure 3B**). Theoretically, it produces a total of three molecules of acetyl-CoA and hence, increases the yield of acetyl-CoA by 50% compared to standard Embden-Meyerhof-Parnas (EMP) glycolysis (Fast et al., 2015). However, in practice,

some of the reducing equivalents are required for biomass generation and maintenance, which indicates that CO_2 evolved in glycolysis cannot be fully converted to acetyl-CoA, resulting in carbon loss as CO_2 and a decrease in the overall carbon yield. Hence, the choice of substrates co-fed with sugars is crucial for achieving complete carbon fixation and high product yields. Supplying exogenous reducing power such as H_2 or syngas during sugar fermentation of acetogens has been demonstrated, which showed high productivity that exceeded the sum of individual substrate productivities and carbon efficiency over 90%, and promoted the generation of more reduced products (Jones et al., 2016; Maru et al., 2018; Cheng et al., 2019).

Another advantage of mixotrophy is the improvement in cellular growth due to the increased availability of acetyl-CoA obtained from additional substrates and improved cellular energy in acetogens. Glucose co-feeding has been shown to improve the methanol uptake rate in *E. limosum* with an increased growth rate and biomass yield due to the improved availability of both carbon and energy for anabolic reactions (Loubière et al., 1992). Recently, *A. woodii* showed mixotrophic growth on formate and fructose, where co-feeding of formate increased the acetate production rate by 50% compared to fructose utilization alone (Neuendorf et al., 2021). Formate utilization improved cell-specific acetate productivity, whereas fructose increased the overall bioenergetics, increasing the amount of ATP wasted. However, co-utilization of formate and fructose resulted in an equal production of CO_2, which was released for growth solely on formate, with a carbon efficiency of only 50%. Although co-feeding of formate and fructose is not preferable to formate utilization alone in terms of carbon efficiency, addition of fructose could be used to improve the bioenergetics of formate-utilizing acetogens and facilitate the redirection of carbon flux from acetate toward desired products, such as ethanol and lactate, in metabolically engineered strains (Neuendorf et al., 2021). Further research is needed to improve carbon yield when formate is co-fed for mixotrophic cultivation.

Although mixotrophy is a promising strategy for increasing the bioenergetics of acetogens, a key concern in its implementation is the possibility of carbon catabolite repression (CCR) of the WL pathway in the presence of a preferred sugar substrate. CCR has been found to be dependent on species and culture conditions (Liu et al., 2015; Maru et al., 2018). For example, CCR in *C. aceticum* (Braun et al., 1981), *M. thermoacetica* (Huang et al., 2012), and *Blautia coccoides* GA-1 (Liu et al., 2015) inhibits autotrophic metabolism and causes poor H_2/CO_2 consumption when glucose or fructose is present at high concentrations. In contrast, *A. woodii, Butyribacterium methylotrophicum, C. autoethanogenum, C. carboxidivorans, C. ljungdahlii*, and *E. limosum* concurrently utilize sugars (glucose or fructose) and C1 gases without CCR (Jones et al., 2016; Maru et al., 2018). However, *E. limosum* was found to be affected by CCR when co-fed with methanol and glucose (Loubière et al., 1992).

Despite the promising aspects of mixotrophy, few studies have elucidated the CCR mechanism in acetogens or the interaction between the WL pathway and glycolysis. Therefore, a thorough understanding of mixotrophy is crucial to fully exploit its potential without CCR. Nevertheless, researchers have recently discovered solutions to overcome CCR in mixotrophy. Controlling the glucose feeding rate with kinetically limiting concentrations during the continuous fermentation of *M. thermoacetica* successfully prevented CCR (Park et al., 2019). This shifted the carbon substrate preferences toward CO_2 and helped achieve mixotrophic growth without CCR. Another recent study similarly employed xylose-limited conditions, in which the simultaneous uptake of xylose and CO for acetate production was observed in *C. autoethanogenum* (Mann et al., 2020).

BIOTECHNOLOGICAL APPLICATIONS OF ACETOGENS

With the enhanced performance of the WL pathway and bioenergetics achieved through the aforementioned strategies, acetogens can increase the yields and titers of native metabolites such as acetate, butyrate, and ethanol (Bertsch and Müller, 2015a; Katsyv and Müller, 2020). In recent years, several efforts have been made to improve the product selectivity of native metabolites by modifying culture conditions (**Table 2**).

Culture Conditions to Shift Metabolite Profiles

Increasing Alcohol Selectivity

Metabolite formation by acetogens is dependent on the culture conditions, including gas composition, substrate, and pH (Arslan et al., 2021; Bae et al., 2022). Most importantly, the availability of a sufficient amount of reducing equivalents is a key factor, as it constrains metabolic flux distribution, energy management, and product formation in acetogens (Hermann et al., 2020). CO was found to be the preferred substrate over H_2/CO_2, glucose, or fructose in *C. ljungdahlii, C. carboxidivorans*, and *C. aceticum* for the production of alcohols (Fernandez-Naveira et al., 2017; Hermann et al., 2020; Arslan et al., 2021). The reduction of CO in the carbonyl-branch generates Fd_{red}, which is then oxidized *via* Rnf to provide NADH. The increased supply of both Fd_{red} and NADH during CO fermentation can facilitate alcohol production by activating aldehyde:ferredoxin oxidoreductase (AOR) or promoting the reaction of bifunctional aldehyde/alcohol dehydrogenase (Hermann et al., 2020; Liu et al., 2020; Arslan et al., 2021). Several studies have found that AOR plays a significant role in ethanol production during autotrophic growth in CO-containing syngas cultures (Richter et al., 2016; Valgepea et al., 2017, 2018). Similarly, medium acidification by pH drop stimulated ethanol production in *C. aceticum* grown on CO, whereas the same effect was not observed for fructose fermentation (Arslan et al., 2021). Similarly, the medium acidification stimulated *C. carboxidivorans* to convert acids to alcohols when grown in a syngas mixture, but this effect was not observed during glucose fermentation (Fernandez-Naveira et al., 2017).

TABLE 2 | Strategies to improve product selectivity of native or non-native biochemicals in acetogens.

Target product	Species	Product type*	Strategy	References
Ethanol	*C. autoethanogenum*	Native	Deletion of *adhE1a* increased ethanol production to 2.46 g/L on CO fermentation	Liew et al., 2017
	C. aceticum	Native	Medium acidification increased ethanol production to 4.4 g/L from CO	Arslan et al., 2021
	C. carboxidivorans	Native	Medium acidification stimulated conversion of acids into alcohols during syngas fermentation	Fernandez-Naveira et al., 2017
Isopropanol	*C. ljungdahlii*	Non-native	Reinforcing acetate reassimilation by overexpressing *aor* and acyl-CoA synthetases (*acs* and *fadKM1/M2*) reduced acetate byproduct and enhanced production of isopropanol and ethanol on syngas fermentation	Jia et al., 2021
3-HB	*C. ljungdahlii*	Non-native	Downregulation of *pta via* CRISPRi increased 3-HB production by reducing acetate production	Woolston et al., 2018
Butyrate	*E. limosum*	Native	Addition of acetate in gas fermentation increased butyrate production with a shift of major product from acetate to butyrate	Loubière and Lindley, 1991; Park et al., 2017
	E. limosum	Native	Feeding methanol promoted butyrate production	Flaiz et al., 2021; Litty and Müller, 2021
	B. methylotrophicum	Native	Varying methanol-to-bicarbonate ratios in the culture media affected butyrate yield and selectivity	Wang et al., 2021
	C. ljungdahlii	Non-native	Deletion of *pta, adhE1* and CoA transferase homolog increased butyrate synthesis to 1.3 g/L, reducing acetate production under H_2/CO_2 fermentation	Ueki et al., 2014
	C. ljungdahlii	Non-native	Downregulation of *adhE1 via* CRISPRi increased butyrate production by reducing ethanol synthesis	Zhao et al., 2019
Butanol	*E. limosum*	Native	High methanol-to-formate ratios induced butanol production with a titer of 38 mg/L	Wood et al., 2021

**Non-native products indicate that the corresponding biosynthesis pathway is introduced into the acetogen to produce the target chemical.*

Interestingly, modifying the co-substrate ratios can induce the production of unexpected metabolites because of the improved bioenergetics in acetogens. Recently, varied ratios of methanol-to-formate were tested in *E. limosum*, where a methanol-to-formate substrate ratio of 7.5:1 achieved a maximum butanol titer of 2.0 ± 1.1 mM (38 mg/L). This is the first evidence of native butanol production in *E. limosum*, as it has not been observed under only syngas or methanol fermentation (Wood et al., 2021). Although the underlying mechanism and butanol production pathway in *E. limosum* remains elusive, butanol production is suspected to be due to overflow metabolism, similar to 2,3-BDO production in *C. ljungdahlii* or *C. autoethanogenum* (Köpke et al., 2011).

Increasing Butyrate Selectivity

E. limosum was reported to produce butyrate only during CO-fed syngas fermentation, and not when using H_2/CO_2, as in the latter case, most of the carbon is directed to acetate synthesis for ATP generation (Litty and Müller, 2021). Butyrate production in acetogens not only provides the ATP needed for cell growth but also balances the redox, indicating that butyrate yield and selectivity depend on NADH availability (Flaiz et al., 2021; Wang et al., 2021). The production of butyrate has been reported to be enhanced by the supplementation of additional reducing equivalents or methanol to provide more NADH (Flaiz et al., 2021; Fu et al., 2021). In addition, the amount of bicarbonate added to the medium can affect the redox levels. When bicarbonate is present in excess in the medium, CO_2 reduction consumes more NADH, resulting in lower NADH availability for reassimilating acetate to produce butyrate. This effect was confirmed in *B. methylotrophicum*, where supplementation with 20 mM bicarbonate and 100 mM methanol significantly improved butyrate production, making it the major product (Wang et al., 2021). In contrast, 40 mM and 60 mM bicarbonate induced the accumulation of a large amount of acetate. Acetate is another co-substrate that influences butyrate selectivity. The addition of acetate into the culture medium during gas fermentation of *E. limosum* increased butyrate production compared to pure gas fermentation (Loubière and Lindley, 1991; Park et al., 2017). Supplementation with acetate elevated energy status and shifted the major product from acetate to butyrate.

Strain Engineering for Non-native Biochemical Production

The development of genetic tools and synthetic biology approaches has opened up possibilities for the metabolic engineering of several acetogens (Köpke et al., 2010; Shin et al., 2019; Zhao et al., 2019; Jin et al., 2020). Using such tools, the heterologous expression of biosynthetic pathways for producing desired chemicals has expanded the product spectrum of acetogens, including isoprene, isopropanol, acetone, and 3-hydroxybutyrate (3-HB; Jones et al., 2016; Diner et al., 2018; Woolston et al., 2018; Karim et al., 2020; Jia et al., 2021). Although they were proof-of-concept production, as their titers were considerably low due to concurrent production of native

metabolites as byproducts, significant improvements can be achieved with strain engineering that redirects carbon fluxes toward desired chemicals.

Blocking competing pathways is an effective method for redirecting the carbon flux. Deletion of three genes encoding PTA, AdhE1, and CoA transferase in *C. ljungdahlii* increased butyrate synthesis and reduced acetate and ethanol byproducts (Ueki et al., 2014). CRISPR-mediated downregulation of *adhE1* also redirected the carbon from ethanol to butyrate in engineered butyrate-producing *C. ljungdahlii* (Zhao et al., 2019).

Considering that acetate synthesis is coupled with ATP formation and is essential for energy supply in acetogens living in energy-limited autotrophic conditions, blocking acetate-producing pathways can result in poor cell growth (Huang et al., 2016). A feasible strategy to overcome this issue is to reassimilate the acetate to form the target product. The AOR present in acetogens converts acetate into ethanol with acetaldehyde as an intermediate. Its role in acetate reassimilation has been found in a study in which knockout of AOR enzymes significantly impaired the conversion of acetate into alcohols in *C. autoethanogenum* (Liew et al., 2017). The positive effect of this strategy was confirmed by the overexpression of native AOR in *C. carboxidivorans,* which resulted in higher ethanol production (Cheng et al., 2019). Recently, a novel acetate reassimilation pathway containing two acyl-CoA synthetases (ACS from *C. ljungdahlii* and FadKM1/2 from *E. coli*) that catalyze acetate conversion to form acetyl-CoA was tested in *C. ljungdahlii* along with overexpression of native AOR enzymes (Jia et al., 2021). Introduction of the pathway into an engineered strain to produce isopropanol successfully reinforced acetate assimilation, achieving significantly reduced acetate formation and increased production of ethanol and isopropanol through syngas fermentation.

CONCLUSION AND FUTURE PERSPECTIVES

Acetogens are attractive microorganisms that can convert C1 gases into value-added biochemicals. Recently, as carbon-neutral technologies have received much attention to mitigate climate change, acetogens have been considered as promising biocatalysts capable of fixing C1 gases. To date, many studies on acetogens have been conducted to understand the enzymes involved in the WL pathway and to convert acetyl-CoA into value-added chemicals. As various omics analyses based on systems biology approaches have been applied for studying acetogens, systems-level understanding of metabolic pathways, including the WL pathway and energy conservation system, has accumulated and expanded our knowledge on acetogenic metabolism. An example is the transcriptional or translational regulation of genes encoding enzymes of the WL pathway, based on the transcriptome or translatome analyses (Marcellin et al., 2016; Al-Bassam et al., 2018; Kremp et al., 2018; Shin et al., 2018, 2021; Song et al., 2018; Neuendorf et al., 2021). In addition, several genome-scale metabolic models have been constructed for several acetogens. They have provided information to develop a chassis strain and tools for *in silico* simulations, such as intracellular carbon flow or cellular energy prediction in acetogenic metabolism (Liu et al., 2019; Mahamkali et al., 2020; Song et al., 2020; Zhu et al., 2020). Recently, protein structure prediction tools, such as AlphaFold2 or RoseTTAFold, have been launched (Baek et al., 2021; Jumper et al., 2021). They are expected to play an important role in revealing the functions and mutations of many proteins in the WL pathway and energy conservation systems. Furthermore, proteins involved in the WL pathway may be engineered or a novel C1 fixation pathway could be designed and constructed using deep learning tools. In addition, genome manipulation techniques, whose application in studying acetogens has been limited, are also being rapidly developed. In particular, the genome-editing technologies based on CRISPR/Cas have been applied in engineering acetogens to identify genes essential for cell growth and metabolite production (Huang et al., 2016; Shin et al., 2019; Zhao et al., 2019; Jeong et al., 2020; Xia et al., 2020). Furthermore, the development of various genetic parts, modules, and circuits based on the synthetic biology approach is expected to enable the transcriptional or translational regulation of genes involved in the acetogenesis under specific culture conditions. Therefore, we expect that the application of various studies based on systems and synthetic biology approaches to acetogen engineering will further improve the C1 gas conversion efficiency and ultimately lead to the development of highly efficient biocatalysts for C1 gas fixation.

AUTHOR CONTRIBUTIONS

B-KC conceptualized and supervised the project. HL, JB, SJ, SK, and B-KC wrote the manuscript. All authors contributed to the article and approved the submitted version.

FUNDING

This work was supported by the C1 Gas Refinery Program (2018M3D3A1A01055733 to B-KC) through the National Research Foundation of Korea (NRF) funded by the Ministry of Science and ICT (MSIT).

REFERENCES

Abrini, J., Naveau, H., and Nyns, E.-J. (1994). *Clostridium autoethanogenum*, sp. nov., an anaerobic bacterium that produces ethanol from carbon monoxide. *Arch. Microbiol.* 161, 345–351. doi: 10.1007/BF00303591

Abubackar, H. N., Veiga, M. C., and Kennes, C. (2011). Biological conversion of carbon monoxide: rich syngas or waste gases to bioethanol. *Biofuels Bioprod. Biorefin.* 5, 93–114. doi: 10.1002/bbb.256

Al-Bassam, M. M., Kim, J. N., Zaramela, L. S., Kellman, B. P., Zuniga, C., Wozniak, J. M., et al. (2018). Optimization of carbon and energy utilization

through differential translational efficiency. *Nat. Commun.* 9:4474. doi: 10.1038/s41467-018-06993-6

Andreesen, J. R., Gottschalk, G., and Schlegel, H. G. (1970). *Clostridium formicoaceticum* nov. spec. isolation, description and distinction from *C. aceticum* and *C. thermoaceticum*. *Arch. Microbiol.* 72, 154–174. doi: 10.1007/BF00409521

Annan, F. J., Al-Sinawi, B., Humphreys, C. M., Norman, R., Winzer, K., Köpke, M., et al. (2019). Engineering of vitamin prototrophy in *Clostridium ljungdahlii* and *Clostridium autoethanogenum*. *Appl. Microbiol. Biotechnol.* 103, 4633–4648. doi: 10.1007/s00253-019-09763-6

Anwar, M. N., Fayyaz, A., Sohail, N. F., Khokhar, M. F., Baqar, M., Khan, W. D., et al. (2018). CO_2 capture and storage: A way forward for sustainable environment. *J. Environ. Econ. Manage.* 226, 131–144. doi: 10.1016/j.jenvman.2018.08.009

Arslan, K., Veiga, M. C., and Kennes, C. (2021). Autotrophic (C1-gas) versus heterotrophic (fructose) accumulation of acetic acid and ethanol in *Clostridium aceticum*. *Bioresour. Technol.* 337:125485. doi: 10.1016/j.biortech.2021.125485

Bache, R., and Pfennig, N. (1981). Selective isolation of *Acetobacterium woodii* on methoxylated aromatic acids and determination of growth yields. *Arch. Microbiol.* 130, 255–261. doi: 10.1007/BF00459530

Bae, J., Song, Y., Lee, H., Shin, J., Jin, S., Kang, S., et al. (2022). Valorization of C1 gases to value-added chemicals using acetogenic biocatalysts. *Chem. Eng. J.* 428:131325. doi: 10.1016/j.cej.2021.131325

Baek, M., DiMaio, F., Anishchenko, I., Dauparas, J., Ovchinnikov, S., Lee, G. R., et al. (2021). Accurate prediction of protein structures and interactions using a three-track neural network. *Science* 373, 871–876. doi: 10.1126/science.abj8754

Balch, W. E., Schoberth, S., Tanner, R. S., and Wolfe, R. S. (1977). *Acetobacterium*, a New Genus of Hydrogen-Oxidizing, Carbon Dioxide-Reducing, Anaerobic Bacteria. *Int. J. Syst. Evol. Microbiol.* 27, 355–361. doi: 10.1099/00207713-27-4-355

Banerjee, A., Leang, C., Ueki, T., Nevin, K. P., and Lovley, D. R. (2014). Lactose-inducible system for metabolic engineering of *Clostridium ljungdahlii*. *Appl. Environ. Microbiol.* 80, 2410–2416. doi: 10.1128/AEM.03666-13

Beck, Z. Q., Cervin, M. A., Chotani, G. K., Diner, B. A., and Fan, J. (2014). *Recombinant anaerobc acetogenic bacteria for production of isoprene and/or industrial bio-products using synthesis gas*. *U.S. Patent No. 20140234926, A1*. Washington, DC: U.S. Patent and Trademark Office.

Bengelsdorf, F. R., Poehlein, A., Esser, C., Schiel-Bengelsdorf, B., Daniel, R., and Dürre, P. (2015). Complete genome sequence of the acetogenic bacterium *Moorella thermoacetica* DSM 2955^T. *Genome Announc.* 3:15. doi: 10.1128/genomeA.01157-15

Bengelsdorf, F. R., Poehlein, A., Linder, S., Erz, C., Hummel, T., Hoffmeister, S., et al. (2016). Industrial acetogenic biocatalysts: a comparative metabolic and genomic analysis. *Front. Microbiol.* 7:1036. doi: 10.3389/fmicb.2016.01036

Bertsch, J., and Müller, V. (2015a). Bioenergetic constraints for conversion of syngas to biofuels in acetogenic bacteria. *Biotechnol. Biofuels* 8:210. doi: 10.1186/s13068-015-0393-x

Bertsch, J., and Müller, V. (2015b). CO metabolism in the acetogen *Acetobacterium woodii*. *Appl. Environ. Microbiol.* 81, 5949–5956. doi: 10.1128/AEM.01772-15

Bertsch, J., Öppinger, C., Hess, V., Langer, J. D., and Müller, V. (2015). Heterotrimeric NADH-oxidizing methylenetetrahydrofolate reductase from the acetogenic bacterium *Acetobacterium woodii*. *J. Bacteriol.* 197, 1681–1689. doi: 10.1128/JB.00048-15

Berzin, V., Kiriukhin, M., and Tyurin, M. (2012). Selective production of acetone during continuous synthesis gas fermentation by engineered biocatalyst *Clostridium* sp. MAceT113. *Lett. Appl. Microbiol.* 55, 149–154. doi: 10.1111/j.1472-765X.2012.03272.x

Berzin, V., Tyurin, M., and Kiriukhin, M. (2013). Selective n-butanol production by *Clostridium* sp. MTButOH1365 during continuous synthesis gas fermentation due to expression of synthetic thiolase, 3-hydroxy butyryl-CoA dehydrogenase, crotonase, butyryl-CoA dehydrogenase, butyraldehyde dehydrogenase, and NAD-dependent butanol dehydrogenase. *Appl. Biochem. Biotechnol.* 169, 950–959. doi: 10.1007/s12010-012-0060-7

Biegel, E., and Müller, V. (2010). Bacterial Na^+-translocating ferredoxin:NAD^+ oxidoreductase. *Proc. Natl. Acad. Sci. U. S. A.* 107, 18138–18142. doi: 10.1073/pnas.1010318107

Bomar, M., Hippe, H., and Schink, B. (1991). Lithotrophic growth and hydrogen metabolism by *Clostridium magnum*. *FEMS Microbiol. Lett.* 83, 347–349. doi: 10.1111/j.1574-6968.1991.tb04488.x

Braun, M., Mayer, F., and Gottschalk, G. (1981). *Clostridium aceticum* (Wieringa), a microorganism producing acetic acid from molecular hydrogen and carbon dioxide. *Arch. Microbiol.* 128, 288–293. doi: 10.1007/BF00422532

Bredwell, M. D., and Worden, R. M. (1998). Mass-transfer properties of microbubbles. 1. *Exp. Stud. Biotechnol. Prog.* 14, 31–38. doi: 10.1021/bp970133x

Brown, S. D., Nagaraju, S., Utturkar, S., De Tissera, S., Segovia, S., Mitchell, W., et al. (2014). Comparison of single-molecule sequencing and hybrid approaches for finishing the genome of *Clostridium autoethanogenum* and analysis of CRISPR systems in industrial relevant Clostridia. *Biotechnol. Biofuels* 7:40. doi: 10.1186/1754-6834-7-40

Caldwell, D. R. (1989). Effects of methanol on the growth of gastrointestinal anaerobes. *Can. J. Microbiol.* 35, 313–317. doi: 10.1139/m89-047

Chang, I.-S., Kim, D. H., Kim, B. H., Shin, P. K., Yoon, J. H., Lee, J. S., et al. (1997). Isolation and identificaltion of carbon monoxide utilizing anaerobe, *Eubacterium limosum* KIST612. *Kor. J. Appl. Microbiol. Biotechnol.* 25, 1–8.

Chang, I. S., Kim, B. H., Lovitt, R. W., and Bang, J. S. (2001). Effect of CO partial pressure on cell-recycled continuous CO fermentation by *Eubacterium limosum* KIST612. *Process Biochem.* 37, 411–421. doi: 10.1016/S0032-9592(01)00227-8

Cheng, C., Li, W., Lin, M., and Yang, S. T. (2019). Metabolic engineering of *Clostridium carboxidivorans* for enhanced ethanol and butanol production from syngas and glucose. *Bioresour. Technol.* 284, 415–423. doi: 10.1016/j.biortech.2019.03.145

Cheng, C., Shao, Y., Li, W., Liu, J., Liu, X., Zhao, Y., et al. (2022). Electricity-enhanced anaerobic, non-photosynthetic mixotrophy by *Clostridium carboxidivorans* with increased carbon efficiency and alcohol production. *Energy Convers. Manag.* 252:118. doi: 10.1016/j.enconman.2021.115118

Claassens, N. J., Cotton, C. A. R., Kopljar, D., and Bar-Even, A. (2019). Making quantitative sense of electromicrobial production. *Nat. Catal.* 2, 437–447. doi: 10.1038/s41929-019-0272-0

Clark, J. E., Ragsdale, S. W., Ljungdahl, L. G., and Wiegel, J. (1982). Levels of enzymes involved in the synthesis of acetate from CO_2 in *Clostridium thermoautotrophicum*. *J. Bacteriol.* 151, 507–509. doi: 10.1128/jb.151.1.507-509.1982

Cotton, C. A., Claassens, N. J., Benito-Vaquerizo, S., and Bar-Even, A. (2020). Renewable methanol and formate as microbial feedstocks. *Curr. Opin. Biotechnol.* 62, 168–180. doi: 10.1016/j.copbio.2019.10.002

Daniel, S. L., Hsu, T., Dean, S. I., and Drake, H. L. (1990). Characterization of the H_2- and CO-dependent chemolithotrophic potentials of the acetogens *Clostridium thermoaceticum* and *Acetogenium kivui*. *J. Bacteriol.* 172, 4464–4471. doi: 10.1128/jb.172.8.4464-4471.1990

Dietrich, H. M., Kremp, F., Öppinger, C., Ribaric, L., and Müller, V. (2021). Biochemistry of methanol-dependent acetogenesis in *Eubacterium callanderi* KIST612. *Environ. Microbiol.* 23, 4505–4517. doi: 10.1111/1462-2920.15643

Diner, B. A., Fan, J., Scotcher, M. C., Wells, D. H., Whited, G. M., and Drake, H. L. (2018). Synthesis of Heterologous Mevalonic Acid Pathway Enzymes in *Clostridium ljungdahlii* for the Conversion of Fructose and of Syngas to Mevalonate and Isoprene. *Appl. Environ. Microbiol.* 84, e01723–e01717. doi: 10.1128/AEM.01723-17

Doukov, T. I., Blasiak, L. C., Seravalli, J., Ragsdale, S. W., and Drennan, C. L. (2008). Xenon in and at the end of the tunnel of bifunctional carbon monoxide dehydrogenase/acetyl-CoA synthase. *Biochemistry* 47, 3474–3483. doi: 10.1021/bi702386t

Drake, H. L., Küsel, K., and Matthies, C. (2013). "Acetogenic Prokaryotes," in *The Prokaryotes: Prokaryotic Physiology and Biochemistry*. eds. E. Rosenberg, E. F. DeLong, S. Lory, E. Stackebrandt and F. Thompson (Berlin, Heidelberg: Springer Berlin Heidelberg), 3–60.

Dürre, P. (2017). Gas fermentation - a biotechnological solution for today's challenges. *Microb. Biotechnol.* 10, 14–16. doi: 10.1111/1751-7915.12431

Dürre, P., Bahl, H., and Gottschalk, G. (1988). "Membrane processes and product formation in anaerobes," in *Handbook of anaerobic fermentation*. eds. L. E. Erickson and D. Y-C. Fung (New York, NY: Marcel Dekker), 187–206.

Eggerth, A. H. (1935). The Gram-positive non-spore-bearing anaerobic bacilli of human feces. *J. Bacteriol.* 30, 277–299. doi: 10.1128/jb.30.3.277-299.1935

Emerson, D. F., and Stephanopoulos, G. (2019). Limitations in converting waste gases to fuels and chemicals. *Curr. Opin. Biotechnol.* 59, 39–45. doi: 10.1016/j.copbio.2019.02.004

Fast, A. G., and Papoutsakis, E. T. (2012). Stoichiometric and energetic analyses of non-photosynthetic CO_2-fixation pathways to support synthetic biology

strategies for production of fuels and chemicals. *Curr. Opin. Chem. Eng.* 1, 380–395. doi: 10.1016/j.coche.2012.07.005

Fast, A. G., Schmidt, E. D., Jones, S. W., and Tracy, B. P. (2015). Acetogenic mixotrophy: novel options for yield improvement in biofuels and biochemicals production. *Curr. Opin. Biotechnol.* 33, 60–72. doi: 10.1016/j.copbio.2014.11.014

Fernandez-Naveira, A., Abubackar, H. N., Veiga, M. C., and Kennes, C. (2017). Production of chemicals from C1 gases (CO, CO_2) by *Clostridium carboxidivorans*. *World J. Microbiol. Biotechnol.* 33:43. doi: 10.1007/s11274-016-2188-z

Flaiz, M., Ludwig, G., Bengelsdorf, F. R., and Dürre, P. (2021). Production of the biocommodities butanol and acetone from methanol with fluorescent FAST-tagged proteins using metabolically engineered strains of *Eubacterium limosum*. *Biotechnol. Biofuels* 14:117. doi: 10.1186/s13068-021-01966-2

Fontaine, F., Peterson, W., McCoy, E., Johnson, M. J., and Ritter, G. J. (1942). A new type of glucose fermentation by *Clostridium thermoaceticum*. *J. Bacteriol.* 43, 701–715. doi: 10.1128/jb.43.6.701-715.1942

Fu, H., Lin, M., Tang, I. C., Wang, J., and Yang, S. T. (2021). Effects of benzyl viologen on increasing NADH availability, acetate assimilation, and butyric acid production by *Clostridium tyrobutyricum*. *Biotechnol. Bioeng.* 118, 770–783. doi: 10.1002/bit.27602

Geerligs, G., Aldrich, H. C., Harder, W., and Diekert, G. (1987). Isolation and characterization of a carbon monoxide utilizing strain of the acetogen *Peptostreptococcus productus*. *Arch. Microbiol.* 148, 305–313. doi: 10.1007/BF00456709

Genthner, B. R. S., and Bryant, M. P. (1982). Growth of *Eubacterium limosum* with Carbon Monoxide as the Energy Source. *Appl. Environ. Microbiol.* 43, 70–74. doi: 10.1128/aem.43.1.70-74.1982

Genthner, B. R. S., and Bryant, M. P. (1987). Additional characteristics of one-carbon-compound utilization by *Eubacterium limosum* and *Acetobacterium woodii*. *Appl. Environ. Microbiol.* 53, 471–476. doi: 10.1128/aem.53.3.471-476.1987

Genthner, B. R., Davis, C. L., and Bryant, M. P. (1981). Features of rumen and sewage sludge strains of *Eubacterium limosum*, a methanol- and H_2-CO_2-utilizing species. *Appl. Environ. Microbiol.* 42, 12–19. doi: 10.1128/aem.42.1.12-19.1981

Göbbels, L., Poehlein, A., Dumnitch, A., Egelkamp, R., Kröger, C., Haerdter, J., et al. (2021). Cysteine: an overlooked energy and carbon source. *Sci. Rep.* 11, 1–15. doi: 10.1038/s41598-021-81103-z

Gossner, A. S., Picardal, F., Tanner, R. S., and Drake, H. L. (2008). Carbon metabolism of the moderately acid-tolerant acetogen *Clostridium drakei* isolated from peat. *FEMS Microbiol. Lett.* 287, 236–242. doi: 10.1111/j.1574-6968.2008.01313.x

Graber, J. R., and Breznak, J. A. (2004). Physiology and nutrition of *Treponema primitia*, an H_2/CO_2-acetogenic spirochete from termite hindguts. *Appl. Environ. Microbiol.* 70, 1307–1314. doi: 10.1128/AEM.70.3.1307-1314.2004

Graber, J. R., Leadbetter, J. R., and Breznak, J. A. (2004). Description of *Treponema azotonutricium* sp. nov. and *Treponema primitia* sp. nov., the first spirochetes isolated from termite guts. *Appl. Environ. Microbiol.* 70, 1315–1320. doi: 10.1128/AEM.70.3.1315-1320.2004

Harrington, T. D., Tran, V. N., Mohamed, A., Renslow, R., Biria, S., Orfe, L., et al. (2015). The mechanism of neutral red-mediated microbial electrosynthesis in *Escherichia coli*: menaquinone reduction. *Bioresour. Technol.* 192, 689–695. doi: 10.1016/j.biortech.2015.06.037

Hattori, S., Kamagata, Y., Hanada, S., and Shoun, H. (2000). *Thermacetogenium phaeum* gen. nov., sp. nov., a strictly anaerobic, thermophilic, syntrophic acetate-oxidizing bacterium. *Int. J. Syst. Evol. Microbiol.* 50, 1601–1609. doi: 10.1099/00207713-50-4-1601

Hedderich, R., and Forzi, L. (2005). Energy-converting [NiFe] hydrogenases: More than just H_2 activation. *J. Mol. Microbiol. Biotechnol.* 10, 92–104. doi: 10.1159/000091557

Hermann, M., Teleki, A., Weitz, S., Niess, A., Freund, A., Bengelsdorf, F. R., et al. (2020). Electron availability in CO_2, CO and H_2 mixtures constrains flux distribution, energy management and product formation in *Clostridium ljungdahlii*. *Microb. Biotechnol.* 13, 1831–1846. doi: 10.1111/1751-7915.13625

Hess, V., Poehlein, A., Weghoff, M. C., Daniel, R., and Müller, V. (2014). A genome-guided analysis of energy conservation in the thermophilic, cytochrome-free acetogenic bacterium *Thermoanaerobacter kivui*. *BMC Genomics* 15:1139. doi: 10.1186/1471-2164-15-1139

Huang, H., Chai, C., Li, N., Rowe, P., Minton, N. P., Yang, S., et al. (2016). CRISPR/Cas9-based efficient genome editing in *Clostridium ljungdahlii*, an autotrophic gas-fermenting bacterium. *ACS Synth. Biol.* 5, 1355–1361. doi: 10.1021/acssynbio.6b00044

Huang, H., Chai, C., Yang, S., Jiang, W., and Gu, Y. (2019). Phage serine integrase-mediated genome engineering for efficient expression of chemical biosynthetic pathway in gas-fermenting *Clostridium ljungdahlii*. *Metab. Eng.* 52, 293–302. doi: 10.1016/j.ymben.2019.01.005

Huang, H., Wang, S., Moll, J., and Thauer, R. K. (2012). Electron bifurcation involved in the energy metabolism of the acetogenic bacterium *Moorella thermoacetica* growing on glucose or H_2 plus CO_2. *J. Bacteriol.* 194, 3689–3699. doi: 10.1128/JB.00385-12

Huhnke, R. L., Lewis, R. S., and Tanner, R. S. (2008). *Isolation and characterization of novel clostridial species*. U.S. Patent No. 20080057554. Washington, DC: U.S. Patent and Trademark Office.

Hurst, K. M., and Lewis, R. S. (2010). Carbon monoxide partial pressure effects on the metabolic process of syngas fermentation. *Biochem. Eng. J.* 48, 159–165. doi: 10.1016/j.bej.2009.09.004

Hwang, S., Song, Y., and Cho, B. K. (2015). Draft genome sequence of *Acetobacterium bakii* DSM 8239, a potential psychrophilic chemical producer through syngas fermentation. *Genome Announc.* 3:15. doi: 10.1128/genomeA.01070-15

Igari, S., Ohtaki, A., Yamanaka, Y., Sato, Y., Yohda, M., Odaka, M., et al. (2011). Properties and crystal structure of methylenetetrahydrofolate reductase from *Thermus thermophilus* HB8. *PLoS One* 6:e23716. doi: 10.1371/journal.pone.0023716

Im, C. H., Kim, C., Song, Y. E., Oh, S. E., Jeon, B. H., and Kim, J. R. (2018). Electrochemically enhanced microbial CO conversion to volatile fatty acids using neutral red as an electron mediator. *Chemosphere* 191, 166–173. doi: 10.1016/j.chemosphere.2017.10.004

Jeon, W. B., Cheng, J., and Ludden, P. W. (2001). Purification and characterization of membrane-associated CooC protein and its functional role in the insertion of nickel into carbon monoxide dehydrogenase from *Rhodospirillum rubrum*. *J. Biol. Chem.* 276, 38602–38609. doi: 10.1074/jbc.M104945200

Jeong, J., Bertsch, J., Hess, V., Choi, S., Choi, I.-G., Chang, I. S., et al. (2015). Energy conservation model based on genomic and experimental analyses of a carbon monoxide-utilizing, butyrate-forming acetogen, *Eubacterium limosum* KIST612. *Appl. Environ. Microbiol.* 81, 4782–4790. doi: 10.1128/AEM.00675-15

Jeong, J., Kim, J.-Y., Park, B., Choi, I.-G., and Chang, I. S. (2020). Genetic engineering system for syngas-utilizing acetogen, *Eubacterium limosum* KIST612. *Bioresour. Technol. Rep.* 11:452. doi: 10.1016/j.biteb.2020.100452

Jeong, Y., Song, Y., Shin, H. S., and Cho, B. K. (2014). Draft genome sequence of acid-tolerant *Clostridium drakei* SL1^T, a potential chemical producer through syngas fermentation. *Genome Announc.* 2:14. doi: 10.1128/genomeA.00387-14

Jia, D., He, M., Tian, Y., Shen, S., Zhu, X., Wang, Y., et al. (2021). Metabolic engineering of gas-fermenting *Clostridium ljungdahlii* for efficient co-production of isopropanol, 3-ydroxybutyrate, and ethanol. *ACS Synth. Biol.* 10, 2628–2638. doi: 10.1021/acssynbio.1c00235

Jin, S., Bae, J., Song, Y., Pearcy, N., Shin, J., Kang, S., et al. (2020). Synthetic biology on acetogenic bacteria for highly efficient conversion of C1 gases to biochemicals. *Int. J. Mol. Sci.* 21:639. doi: 10.3390/ijms21207639

Jin, S., Jeon, Y., Jeon, M. S., Shin, J., Song, Y., Kang, S., et al. (2021). Acetogenic bacteria utilize light-driven electrons as an energy source for autotrophic growth. *Proc. Natl. Acad. Sci. U. S. A.* 118:2118. doi: 10.1073/pnas.2020552118

Jones, S. W., Fast, A. G., Carlson, E. D., Wiedel, C. A., Au, J., Antoniewicz, M. R., et al. (2016). CO_2 fixation by anaerobic non-photosynthetic mixotrophy for improved carbon conversion. *Nat. Commun.* 7:12800. doi: 10.1038/ncomms12800

Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., et al. (2021). Highly accurate protein structure prediction with AlphaFold. *Nature* 596, 583–589. doi: 10.1038/s41586-021-03819-2

Kang, H., Park, B., Oh, S., Pathiraja, D., Kim, J. Y., Jung, S., et al. (2021). Metabolism perturbation Caused by the overexpression of carbon monoxide dehydrogenase/Acetyl-CoA synthase gene complex accelerated gas to acetate

conversion rate of *Eubacterium limosum* KIST612. *Bioresour. Technol.* 341:125879. doi: 10.1016/j.biortech.2021.125879

Kang, S., Song, Y., Jin, S., Shin, J., Bae, J., Kim, D. R., et al. (2020). Adaptive laboratory evolution of *Eubacterium limosum* ATCC 8486 on carbon monoxide. *Front. Microbiol.* 11:402. doi: 10.3389/fmicb.2020.00402

Karim, A. S., Dudley, Q. M., Juminaga, A., Yuan, Y., Crowe, S. A., Heggestad, J. T., et al. (2020). In vitro prototyping and rapid optimization of biosynthetic enzymes for cell design. *Nat. Chem. Biol.* 16, 912–919. doi: 10.1038/s41589-020-0559-0

Karl, M. M., Poehlein, A., Bengelsdorf, F. R., Daniel, R., and Dürre, P. (2017). Complete genome sequence of the autotrophic acetogen *Clostridium formicaceticum* DSM 92^{T} using nanopore and illumina sequencing data. *Genome Announc.* 5:17. doi: 10.1128/genomeA.00423-17

Katsyv, A., Jain, S., Basen, M., and Müller, V. (2021). Electron carriers involved in autotrophic and heterotrophic acetogenesis in the thermophilic bacterium *Thermoanaerobacter kivui*. *Extremophiles* 25, 513–526. doi: 10.1007/s00792-021-01247-8

Katsyv, A., and Müller, V. (2020). Overcoming energetic barriers in acetogenic C1 conversion. *Front. Bioeng. Biotechnol.* 8:621166. doi: 10.3389/fbioe.2020.621166

Kerby, R. L., Ludden, P. W., and Roberts, G. P. (1997). In vivo nickel insertion into the carbon monoxide dehydrogenase of *Rhodospirillum rubrum*: Molecular and physiological characterization of *cooCTJ*. *J. Bacteriol.* 179, 2259–2266. doi: 10.1128/jb.179.7.2259-2266.1997

Kerby, R., and Zeikus, J. G. (1983). Growth of *Clostridium thermoaceticum* on H_2/CO_2 or CO as energy source. *Curr. Microbiol.* 8, 27–30. doi: 10.1007/BF01567310

Köpke, M., Held, C., Hujer, S., Liesegang, H., Wiezer, A., Wollherr, A., et al. (2010). *Clostridium ljungdahlii* represents a microbial production platform based on syngas. *Proc. Natl. Acad. Sci. U. S. A.* 107, 13087–13092. doi: 10.1073/pnas.1004716107

Köpke, M., Mihalcea, C., Liew, F., Tizard, J. H., Ali, M. S., Conolly, J. J., et al. (2011). 2,3-butanediol production by acetogenic bacteria, an alternative route to chemical synthesis, using industrial waste gas. *Appl. Environ. Microbiol.* 77, 5467–5475. doi: 10.1128/AEM.00355-11

Köpke, M., and Simpson, S. D. (2020). Pollution to products: recycling of 'above ground' carbon by gas fermentation. *Curr. Opin. Biotechnol.* 65, 180–189. doi: 10.1016/j.copbio.2020.02.017

Köpke, M., Simpson, S., Leiw, F., and Chen, W. (2012). *Fermentation process for producing isopropanol using a recombinant microorganism. U.S. Patent No. 20120252083, A1*. Washington, DC: U.S. Patent and Trademark Office.

Kornienko, N., Sakimoto, K. K., Herlihy, D. M., Nguyen, S. C., Alivisatos, A. P., Harris, C. B., et al. (2016). Spectroscopic elucidation of energy transfer in hybrid inorganic–biological organisms for solar-to-chemical production. *Proc. Natl. Acad. Sci. U. S. A.* 113, 11750–11755. doi: 10.1073/pnas.1610554113

Kotsyurbenko, O. R., Simankova, M. V., Nozhevnikova, A. N., Zhilina, T. N., Bolotina, N. P., Lysenko, A. M., et al. (1995). New species of psychrophilic acetogens: *Acetobacterium bakii* sp. nov., *A. paludosum* sp. nov., *A. fimetarium* sp. nov. *Arch. Microbiol.* 163, 29–34. doi: 10.1007/BF00262200

Kracke, F., Virdis, B., Bernhardt, P. V., Rabaey, K., and Kromer, J. O. (2016). Redox dependent metabolic shift in *Clostridium autoethanogenum* by extracellular electron supply. *Biotechnol. Biofuels* 9:249. doi: 10.1186/s13068-016-0663-2

Kremp, F., Poehlein, A., Daniel, R., and Müller, V. (2018). Methanol metabolism in the acetogenic bacterium *Acetobacterium woodii*. *Environ. Microbiol.* 20, 4369–4384. doi: 10.1111/1462-2920.14356

Kremp, F., Roth, J., and Müller, V. (2020). The *Sporomusa* type Nfn is a novel type of electron-bifurcating transhydrogenase that links the redox pools in acetogenic bacteria. *Sci. Rep.* 10:14872. doi: 10.1038/s41598-020-71038-2

Küsel, K., Dorsch, T., Acker, G., Stackebrandt, E., and Drake, H. L. (2000). *Clostridium scatologenes* strain SL1 isolated as an acetogenic bacterium from acidic sediments. *Int. J. Syst. Evol. Microbiol.* 50, 537–546. doi: 10.1099/00207713-50-2-537

Latif, H., Zeidan, A. A., Nielsen, A. T., and Zengler, K. (2014). Trash to treasure: production of biofuels and commodity chemicals *via* syngas fermenting microorganisms. *Curr. Opin. Biotechnol.* 27, 79–87. doi: 10.1016/j.copbio.2013.12.001

Leigh, J. A., Mayer, F., and Wolfe, R. S. (1981). *Acetogenium kivui*, a new thermophilic hydrogen-oxidizing acetogenic bacterium. *Arch. Microbiol.* 129, 275–280. doi: 10.1007/BF00414697

Li, N., Yang, J., Chai, C., Yang, S., Jiang, W., and Gu, Y. (2015). Complete genome sequence of *Clostridium carboxidivorans* P7^{T}, a syngas-fermenting bacterium capable of producing long-chain alcohols. *J. Biotechnol.* 211, 44–45. doi: 10.1016/j.jbiotec.2015.06.430

Liew, F., Henstra, A. M., Kpke, M., Winzer, K., Simpson, S. D., and Minton, N. P. (2017). Metabolic engineering of *Clostridium autoethanogenum* for selective alcohol production. *Metab. Eng.* 40, 104–114. doi: 10.1016/j.ymben.2017.01.007

Liew, F. E., Nogle, R., Abdalla, T., Rasor, B. J., Canter, C., Jensen, R. O., et al. (2022). Carbon-negative production of acetone and isopropanol by gas fermentation at industrial pilot scale. *Nat. Biotechnol.* 40, 335–344. doi: 10.1038/s41587-021-01195-w

Liou, J. S., Balkwill, D. L., Drake, G. R., and Tanner, R. S. (2005). *Clostridium carboxidivorans* sp. nov., a solvent-producing *clostridium* isolated from an agricultural settling lagoon, and reclassification of the acetogen *Clostridium scatologenes* strain SL1 as *Clostridium drakei* sp. nov. *Int. J. Syst. Evol. Microbiol.* 55, 2085–2091. doi: 10.1099/ijs.0.63482-0

Litty, D., and Müller, V. (2021). Butyrate production in the acetogen *Eubacterium limosum* is dependent on the carbon and energy source. *Microb. Biotechnol.* 14, 2686–2692. doi: 10.1111/1751-7915.13779

Liu, C., Finegold, S. M., Song, Y., and Lawson, P. A. (2008). Reclassification of *Clostridium coccoides*, *Ruminococcus hansenii*, *Ruminococcus hydrogenotrophicus*, *Ruminococcus luti*, *Ruminococcus productus* and *Ruminococcus schinkii* as *Blautia coccoides* gen. nov., comb. nov., *Blautia hansenii* comb. nov., *Blautia hydrogenotrophica* comb. nov., *Blautia luti* comb. nov., *Blautia producta* comb. nov., *Blautia schinkii* comb. nov. and description of *Blautia wexlerae* sp. nov., isolated from human faeces. *Int. J. Syst. Evol. Microbiol.* 58, 1896–1902. doi: 10.1099/ijs.0.65208-0

Liu, Z.-Y., Jia, D.-C., Zhang, K.-D., Zhu, H.-F., Zhang, Q., Jiang, W.-H., et al. (2020). Ethanol metabolism dynamics in *Clostridium ljungdahlii* grown on carbon monoxide. *Appl. Environ. Microbiol.* 86, e00730–e00720. doi: 10.1128/AEM.00730-20

Liu, C., Li, J., Zhang, Y., Philip, A., Shi, E., Chi, X., et al. (2015). Influence of glucose fermentation on CO_2 assimilation to acetate in homoacetogen *Blautia coccoides* GA-1. *J. Ind. Microbiol. Biotechnol.* 42, 1217–1224. doi: 10.1007/s10295-015-1646-1

Liu, J. K., Lloyd, C., Al-Bassam, M. M., Ebrahim, A., Kim, J. N., Olson, C., et al. (2019). Predicting proteome allocation, overflow metabolism, and metal requirements in a model acetogen. *PLoS Comp. Biol.* 15:e1006848. doi: 10.1371/journal.pcbi.1006848

Lorowitz, W. H., and Bryant, M. P. (1984). *Peptostreptococcus productus* strain that grows rapidly with CO as the energy source. *Appl. Environ. Microbiol.* 47, 961–964. doi: 10.1128/aem.47.5.961-964.1984

Loubière, P., Gros, E., Paquet, V., and Lindley, N. D. (1992). Kinetics and physiological implications of the growth behaviour of *Eubacterium limosum* on glucose/methanol mixtures. *Microbiology* 138, 979–985. doi: 10.1099/00221287-138-5-979

Loubière, P., and Lindley, N. D. (1991). The use of acetate as an additional co-substrate improves methylotrophi growth of the acetogenic anaerobe *Eubacterium limosum* when CO_2 fixation is rate-limiting. *Microbiology* 137, 2247–2251. doi: 10.1099/00221287-137-9-2247

Lux, M. F., and Drake, H. L. (1992). Re-examination of the metabolic potentials of the acetogens *Clostridium aceticum* and *Clostridium formicoaceticum*: chemolithoautotrophic and aromatic-dependent growth. *FEMS Microbiol. Lett.* 95, 49–56. doi: 10.1111/j.1574-6968.1992.tb05341.x

Mahamkali, V., Valgepea, K., de Souza Pinto Lemgruber, R., Plan, M., Tappel, R., Köpke, M., et al. (2020). Redox controls metabolic robustness in the gas-fermenting acetogen *Clostridium autoethanogenum*. *Proc. Natl. Acad. Sci. U. S. A.* 117, 13168–13175. doi: 10.1073/pnas.1919531117

Mann, M., Munch, G., Regestein, L., and Rehmann, L. (2020). Cultivation strategies of *Clostridium autoethanogenum* on xylose and carbon monoxide combination. *ACS Sustain. Chem. Eng.* 8, 2632–2639. doi: 10.1021/acssuschemeng.9b05439

Marcellin, E., Behrendorff, J. B., Nagaraju, S., DeTissera, S., Segovia, S., Palfreyman, R. W., et al. (2016). Low carbon fuels and commodity chemicals from waste gases – systematic approach to understand energy metabolism in a model acetogen. *Green Chem.* 18, 3020–3028. doi: 10.1039/C5GC02708J

Maru, B. T., Munasinghe, P. C., Gilary, H., Jones, S. W., and Tracy, B. P. (2018). Fixation of CO_2 and CO on a diverse range of carbohydrates using

anaerobic, non-photosynthetic mixotrophy. *FEMS Microbiol. Lett.* 365:39. doi: 10.1093/femsle/fny039

Matthies, D., Zhou, W., Klyszejko, A. L., Anselmi, C., Yildiz, Ö., Brandt, K., et al. (2014). High-resolution structure and mechanism of an F/V-hybrid rotor ring in a Na^+-coupled ATP synthase. *Nat. Commun.* 5:5286. doi: 10.1038/ncomms6286

Michaelis, L., and Hill, E. S. (1933). The viologen indicators. *J. Gen. Physiol.* 16, 859–873. doi: 10.1085/jgp.16.6.859

Mock, J., Wang, S., Huang, H., Kahnt, J., and Thauer, R. K. (2014). Evidence for a hexaheteromeric methylenetetrahydrofolate reductase in *Moorella thermoacetica*. *J. Bacteriol.* 196, 3303–3314. doi: 10.1128/JB.01839-14

Molitor, B., Marcellin, E., and Angenent, L. T. (2017). Overcoming the energetic limitations of syngas fermentation. *Curr. Opin. Chem. Biol.* 41, 84–92. doi: 10.1016/j.cbpa.2017.10.003

Molitor, B., Richter, H., Martin, M. E., Jensen, R. O., Juminaga, A., Mihalcea, C., et al. (2016). Carbon recovery by fermentation of CO-rich off gases–turning steel mills into biorefineries. *Bioresour. Technol.* 215, 386–396. doi: 10.1016/j.biortech.2016.03.094

Möller, B., Oßmer, R., Howard, B. H., Gottschalk, G., and Hippe, H. (1984). *Sporomusa*, a new genus of gram-negative anaerobic bacteria including *Sporomusa sphaeroides* spec. nov. and *Sporomusa ovata* spec. nov. *Arch. Microbiol.* 139, 388–396. doi: 10.1007/BF00408385

Moon, J., Donig, J., Kramer, S., Poehlein, A., Daniel, R., and Müller, V. (2021). Formate metabolism in the acetogenic bacterium *Acetobacterium woodii*. *Environ. Microbiol.* 23, 4214–4227. doi: 10.1111/1462-2920.15598

Moore, M. R., O'Brien, W. E., and Ljungdahl, L. G. (1974). Purification and Characterization of Nicotinamide Adenine Dinucleotide-dependent Methylenetetrahydrofolate Dehydrogenase from *Clostridium formicoaceticum*. *J. Biol. Chem.* 249, 5250–5253. doi: 10.1016/S0021-9258(19)42355-7

Neuendorf, C. S., Vignolle, G. A., Derntl, C., Tomin, T., Novak, K., Mach, R. L., et al. (2021). A quantitative metabolic analysis reveals *Acetobacterium woodii* as a flexible and robust host for formate-based bioproduction. *Metab. Eng.* 68, 68–85. doi: 10.1016/j.ymben.2021.09.004

Nevin, K. P., Hensley, S. A., Franks, A. E., Summers, Z. M., Ou, J., Woodard, T. L., et al. (2011). Electrosynthesis of organic compounds from carbon dioxide is catalyzed by a diversity of acetogenic microorganisms. *Appl. Environ. Microbiol.* 77, 2882–2886. doi: 10.1128/AEM.02642-10

Nevin, K. P., Woodard, T. L., Franks, A. E., Summers, Z. M., and Lovley, D. R. (2010). Microbial electrosynthesis: Feeding microbes electricity to convert carbon dioxide and water to multicarbon extracellular organic compounds. *mBio* 1:10. doi: 10.1128/mBio.00103-10

Novak, K., Neuendorf, C. S., Kofler, I., Kieberger, N., Klamt, S., and Pflügl, S. (2021). Blending industrial blast furnace gas with H_2 enables *Acetobacterium woodii* to efficiently co-utilize CO, CO_2 and H_2. *Bioresour. Technol.* 323:124573. doi: 10.1016/j.biortech.2020.124573

Oehler, D., Poehlein, A., Leimbach, A., Müller, N., Daniel, R., Gottschalk, G., et al. (2012). Genome-guided analysis of physiological and morphological traits of the fermentative acetate oxidizer *Thermacetogenium phaeum*. *BMC Genomics* 13:723. doi: 10.1186/1471-2164-13-723

Öppinger, C., Kremp, F., and Müller, V. (2022). Is reduced ferredoxin the physiological electron donor for MetVF-type methylenetetrahydrofolate reductases in acetogenesis? A hypothesis. *Int. Microbiol.* 25, 75–88. doi: 10.1007/s10123-021-00190-0

Orgill, J. J., Atiyeh, H. K., Devarapalli, M., Phillips, J. R., Lewis, R. S., and Huhnke, R. L. (2013). A comparison of mass transfer coefficients between trickle-bed, hollow fiber membrane and stirred tank reactors. *Bioresour. Technol.* 133, 340–346. doi: 10.1016/j.biortech.2013.01.124

Park, J. O., Liu, N., Holinski, K. M., Emerson, D. F., Qiao, K., Woolston, B. M., et al. (2019). Synergistic substrate cofeeding stimulates reductive metabolism. *Nat. Metab.* 1, 643–651. doi: 10.1038/s42255-019-0077-0

Park, S., Yasin, M., Jeong, J., Cha, M., Kang, H., Jang, N., et al. (2017). Acetate-assisted increase of butyrate production by Eubacterium limosum KIST612 during carbon monoxide fermentation. *Bioresour. Technol.* 245, 560–566. doi: 10.1016/j.biortech.2017.08.132

Perez, J. M., Richter, H., Loftus, S. E., and Angenent, L. T. (2013). Biocatalytic reduction of short-chain carboxylic acids into their corresponding alcohols with syngas fermentation. *Biotechnol. Bioeng.* 110, 1066–1077. doi: 10.1002/bit.24786

Peters, J. W., Miller, A. F., Jones, A. K., King, P. W., and Adams, M. W. (2016). Electron bifurcation. *Curr. Opin. Chem. Biol.* 31, 146–152. doi: 10.1016/j.cbpa.2016.03.007

Phillips, J. R., Klasson, K. T., Clausen, E. C., and Gaddy, J. L. (1993). Biological production of ethanol from coal synthesis gas. *Appl. Biochem. Biotechnol.* 39, 559–571. doi: 10.1007/BF02919018

Pirbadian, S., Chavez, M. S., and El-Naggar, M. Y. (2020). Spatiotemporal mapping of bacterial membrane potential responses to extracellular electron transfer. *Proc. Natl. Acad. Sci. U. S. A.* 117, 20171–20179. doi: 10.1073/pnas.2000802117

Poehlein, A., Bengelsdorf, F. R., Esser, C., Schiel-Bengelsdorf, B., Daniel, R., and Dürre, P. (2015a). Complete genome sequence of the type strain of the acetogenic bacterium *Moorella thermoacetica* DSM 521^T. *Genome Announc.* 3:15. doi: 10.1128/genomeA.01159-15

Poehlein, A., Cebulla, M., Ilg, M. M., Bengelsdorf, F. R., Schiel-Bengelsdorf, B., Whited, G., et al. (2015b). The complete genome sequence of *Clostridium aceticum*: A missing link between Rnf- and cytochrome-containing autotrophic acetogens. *mBio* 6:e01168-15. doi: 10.1128/mBio.01168-15

Poehlein, A., Gottschalk, G., and Daniel, R. (2013). First insights into the genome of the gram-negative, endospore-forming organism *Sporomusa ovata* strain H1 DSM 2662. *Genome Announc.* 1:13. doi: 10.1128/genomeA.00734-13

Poehlein, A., Schmidt, S., Kaster, A. K., Goenrich, M., Vollmers, J., Thurmer, A., et al. (2012). An ancient pathway combining carbon dioxide fixation with the generation and utilization of a sodium ion gradient for ATP synthesis. *PLoS One* 7:e33439. doi: 10.1371/journal.pone.0033439

Ragsdale, S. W., and Ljungdahl, L. G. (1984). Purification and properties of NAD-dependent 5,10-methylenetetrahydrofolate dehydrogenase from *Acetobacterium woodii*. *J. Biol. Chem.* 259, 3499–3503. doi: 10.1016/S0021-9258(17)43122-X

Ragsdale, S. W., and Pierce, E. (2008). Acetogenesis and the Wood-Ljungdahl pathway of CO_2 fixation. *Biochim. Biophys. Acta* 1784, 1873–1898. doi: 10.1016/j.bbapap.2008.08.012

Richter, H., Molitor, B., Wei, H., Chen, W., Aristilde, L., and Angenent, L. T. (2016). Ethanol production in syngas-fermenting *Clostridium ljungdahlii* is controlled by thermodynamics rather than by enzyme expression. *Energy Environ. Sci.* 9, 2392–2399. doi: 10.1039/C6EE01108J

Ritchie, H., and Roser, M. (2020). CO_2 and Greenhouse Gas Emission. Our World in Data.

Roh, H., Ko, H. J., Kim, D., Choi, D. G., Park, S., Kim, S., et al. (2011). Complete genome sequence of a carbon monoxide-utilizing acetogen, *Eubacterium limosum* KIST612. *J. Bacteriol.* 193, 307–308. doi: 10.1128/JB.01217-10

Rosenthal, A. Z., Matson, E. G., Eldar, A., and Leadbetter, J. R. (2011). RNA-seq reveals cooperative metabolic interactions between two termite-gut spirochete species in co-culture. *ISME J.* 5, 1133–1142. doi: 10.1038/ismej.2011.3

Sakimoto, K. K., Wong, A. B., and Yang, P. (2016). Self-photosensitization of nonphotosynthetic bacteria for solar-to-chemical production. *Science* 351, 74–77. doi: 10.1126/science.aad3317

Sathish, A., Sharma, A., Gable, P., Skiadas, I., Brown, R., and Wen, Z. (2019). A novel bulk-gas-to-atomized-liquid reactor for enhanced mass transfer efficiency and its application to syngas fermentation. *Chem. Eng. J.* 370, 60–70. doi: 10.1016/j.cej.2019.03.183

Schink, B. (1984). *Clostridium magnum* sp. nov., a non-autotrophic homoacetogenic bacterium. *Arch. Microbiol.* 137, 250–255. doi: 10.1007/BF00414553

Schuchmann, K., and Müller, V. (2014). Autotrophy at the thermodynamic limit of life: a model for energy conservation in acetogenic bacteria. *Nat. Rev. Microbiol.* 12, 809–821. doi: 10.1038/nrmicro3365

Schuchmann, K., and Müller, V. (2016). Energetics and application of heterotrophy in acetogenic bacteria. *Appl. Environ. Microbiol.* 82, 4056–4069. doi: 10.1128/AEM.00882-16

Sheppard, C. A., Trimmer, E. E., and Matthews, R. G. (1999). Purification and Properties of NADH-Dependent 5,10-Methylenetetrahydrofolate Reductase (MetF) from *Escherichia coli*. *J. Bacteriol.* 181, 718–725. doi: 10.1128/JB.181.3.718-725.1999

Shin, J., Kang, S., Song, Y., Jin, S., Lee, J. S., Lee, J. K., et al. (2019). Genome engineering of *Eubacterium limosum* using expanded genetic tools and the CRISPR-Cas9 system. *ACS Synth. Biol.* 8, 2059–2068. doi: 10.1021/acssynbio.9b00150

Shin, J., Song, Y., Jeong, Y., and Cho, B. K. (2016). Analysis of the core genome and pan-genome of autotrophic acetogenic bacteria. *Front. Microbiol.* 7:1531. doi: 10.3389/fmicb.2016.01531

Shin, J., Song, Y., Jin, S., Lee, J. K., Kim, D. R., Kim, S. C., et al. (2018). Genome-scale analysis of *Acetobacterium bakii* reveals the cold adaptation of psychrotolerant acetogens by post-transcriptional regulation. *RNA* 24, 1839–1855. doi: 10.1261/rna.068239.118

Shin, J., Song, Y., Kang, S., Jin, S., Lee, J. K., Kim, D. R., et al. (2021). Genome-Scale Analysis of *Acetobacterium woodii* Identifies Translational Regulation of Acetogenesis. *mSystems* 6:e0069621. doi: 10.1128/mSystems.00696-21

Sikorski, J., Lapidus, A., Chertkov, O., Lucas, S., Copeland, A., Glavina Del Rio, T., et al. (2010). Complete genome sequence of *Acetohalobium arabaticum* type strain (Z-7288). *Stand. Genom. Sci.* 3, 57–65. doi: 10.4056/sigs.1062906

Song, Y., and Cho, B. K. (2015). Draft genome sequence of chemolithoautotrophic acetogenic butanol-producing *Eubacterium limosum* ATCC 8486. *Genome Announc.* 3:14. doi: 10.1128/genomeA.01564-14

Song, Y., Lee, J. S., Shin, J., Lee, G. M., Jin, S., Kang, S., et al. (2020). Functional cooperation of the glycine synthase-reductase and Wood-Ljungdahl pathways for autotrophic growth of *Clostridium drakei*. *Proc. Natl. Acad. Sci. U. S. A.* 117, 7516–7523. doi: 10.1073/pnas.1912289117

Song, Y., Shin, J., Jin, S., Lee, J. K., Kim, D. R., Kim, S. C., et al. (2018). Genome-scale analysis of syngas fermenting acetogenic bacteria reveals the translational regulation for its autotrophic growth. *BMC Genomics* 19:837. doi: 10.1186/s12864-018-5238-0

Straub, M., Demler, M., Weuster-Botz, D., and Dürre, P. (2014). Selective enhancement of autotrophic acetate production with genetically modified *Acetobacterium woodii*. *J. Biotechnol.* 178, 67–72. doi: 10.1016/j.jbiotec.2014.03.005

Tanner, R. S., Miller, L. M., and Yang, D. (1993). *Clostridium ljungdahlii* sp. nov., an Acetogenic Species in Clostridial rRNA Homology Group I. *Int. J. Syst. Evol. Microbiol.* 43, 232–236. doi: 10.1099/00207713-43-2-232

Thauer, R. K., Jungermann, K., and Decker, K. (1977). Energy conservation in chemotrophic anaerobic bacteria. *Bacteriol. Rev.* 41, 100–180. doi: 10.1128/br.41.1.100-180.1977

Tourlousse, D. M., Sakamoto, M., Miura, T., Narita, K., Ohashi, A., Uchino, Y., et al. (2020). Complete Genome Sequence of *Blautia producta* JCM 1471^{T}. *Microbiol. Resour. Announc.* 9:20. doi: 10.1128/MRA.00141-20

Tracy, B. P., Jones, S. W., Fast, A. G., Indurthi, D. C., and Papoutsakis, E. T. (2012). *Clostridia*: the importance of their exceptional substrate and metabolite diversity for biofuel and biorefinery applications. *Curr. Opin. Biotechnol.* 23, 364–381. doi: 10.1016/j.copbio.2011.10.008

Tremblay, P. L., Hoglund, D., Koza, A., Bonde, I., and Zhang, T. (2015). Adaptation of the autotrophic acetogen *Sporomusa ovata* to methanol accelerates the conversion of CO_2 to organic products. *Sci. Rep.* 5:16168. doi: 10.1038/srep16168

Ueki, T., Nevin, K. P., Woodard, T. L., and Lovley, D. R. (2014). Converting carbon dioxide to butyrate with an engineered strain of Clostridium ljungdahlii. *mBio* 5, e01636–e01614. doi: 10.1128/mBio.01636-14

Uhlig, R., Poehlein, A., Fischer, R. J., Daniel, R., and Bahl, H. (2016). Genome Sequence of the Autotrophic Acetogen *Clostridium magnum* DSM 2767. *Genome Announc.* 4:16. doi: 10.1128/genomeA.00464-16

Valgepea, K., de Souza Pinto Lemgruber, R., Meaghan, K., Palfreyman, R. W., Abdalla, T., Heijstra, B. D., et al. (2017). Maintenance of ATP homeostasis triggers metabolic shifts in gas-fermenting acetogens. *Cell Syst.* 4:e505, 505–515.e5. doi: 10.1016/j.cels.2017.04.008

Valgepea, K., de Souza Pinto Lemgruber, R., Abdalla, T., Binos, S., Takemori, N., Takemori, A., et al. (2018). H_2 drives metabolic rearrangements in gas-fermenting *Clostridium autoethanogenum*. *Biotechnol. Biofuels* 11:55. doi: 10.1186/s13068-018-1052-9

Wang, S., Huang, H., Kahnt, J., Mueller, A. P., Köpke, M., and Thauer, R. K. (2013). NADP-specific electron-bifurcating [FeFe]-hydrogenase in a functional complex with formate dehydrogenase in *Clostridium autoethanogenum* grown on CO. *J. Bacteriol.* 195, 4373–4386. doi: 10.1128/JB.00678-13

Wang, X., Qin, J., Ma, C., Wang, J., Wang, X., Xu, S., et al. (2021). Methanol assimilation with CO_2 reduction in *Butyribacterium methylotrophicum* and development of genetic toolkits for its engineering. *ACS Sustain. Chem. Eng.* 9, 12079–12090. doi: 10.1021/acssuschemeng.1c02365

Weghoff, M. C., Müller, V., and Liu, S.-J. (2016). CO metabolism in the thermophilic acetogen *Thermoanaerobacter kivui*. *Appl. Environ. Microbiol.* 82, 2312–2319. doi: 10.1128/AEM.00122-16

Wiechmann, A., and Müller, V. (2021). Energy conservation in the acetogenic bacterium *Clostridium aceticum*. *Microorganisms* 9:258. doi: 10.3390/microorganisms9020258

Wieringa, K. (1936). Over het verdwijnen van waterstof en koolzuur onder anaerobe voorwaarden. *Antonie Van Leeuwenhoek* 3, 263–273. doi: 10.1007/BF02059556

Wohlfarth, G., Geerligs, G., and Diekert, G. (1990). Purification and properties of a NADH-dependent 5,10-methylenetetrahydrofolate reductase from *Peptostreptococcus productus*. *Eur. J. Biochem.* 192, 411–417. doi: 10.1111/j.1432-1033.1990.tb19242.x

Wood, J. C., Marcellin, E., Plan, M. R., and Virdis, B. (2021). High methanol-to-formate ratios induce butanol production in *Eubacterium limosum*. *Microb. Biotechnol.* 1–8. doi: 10.1111/1751-7915.13963

Woolston, B. M., Emerson, D. F., Currie, D. H., and Stephanopoulos, G. (2018). Rediverting carbon flux in *Clostridium ljungdahlii* using CRISPR interference (CRISPRi). *Metab. Eng.* 48, 243–253. doi: 10.1016/j.ymben.2018.06.006

Xia, P. F., Casini, I., Schulz, S., Klask, C. M., Angenent, L. T., and Molitor, B. (2020). Reprogramming acetogenic bacteria with CRISPR-targeted base editing *via* deamination. *ACS Synth. Biol.* 9, 2162–2171. doi: 10.1021/acssynbio.0c00226

Zahn, J. A., and Saxena, J. (2012). *Ethanologenic Clostridium species, Clostridium coskatii*. U.S. Patent No. 8802405, B2. Washington, DC: U.S. Patent and Trademark Office.

Zhang, H., Liu, H., Tian, Z., Lu, D., Yu, Y., Cestellos-Blanco, S., et al. (2018). Bacteria photosensitized by intracellular gold nanoclusters for solar fuel production. *Nat. Nanotechnol.* 13, 900–905. doi: 10.1038/s41565-018-0267-z

Zhao, R., Liu, Y., Zhang, H., Chai, C., Wang, J., Jiang, W., et al. (2019). CRISPR-Cas12a-mediated gene deletion and regulation in *Clostridium ljungdahlii* and its application in carbon flux redirection in synthesis gas fermentation. *ACS Synth. Biol.* 8, 2270–2279. doi: 10.1021/acssynbio.9b00033

Zhilina, T. N., and Zavarzin, G. A. (1990). Extremely halophilic, methylotrophic, anaerobic bacteria. *FEMS Microbiol. Rev.* 7, 315–321. doi: 10.1111/j.1574-6968.1990.tb04930.x

Zhu, Z., Guo, T., Zheng, H., Song, T., Ouyang, P., and Xie, J. (2015). Complete genome sequence of a malodorant-producing acetogen, Clostridium scatologenes ATCC 25775^{T}. *J. Biotechnol.* 212, 19–20. doi: 10.1016/j.jbiotec.2015.07.013

Zhu, H. F., Liu, Z. Y., Zhou, X., Yi, J. H., Lun, Z. M., Wang, S. N., et al. (2020). Energy conservation and carbon flux distribution during fermentation of CO or H_2/CO_2 by *Clostridium ljungdahlii*. *Front. Microbiol.* 11:416. doi: 10.3389/fmicb.2020.00416

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

frontiers | Frontiers in Microbiology

REVIEW
published: 20 May 2022
doi: 10.3389/fmicb.2022.872708

Atypical Carboxysome Loci: JEEPs or Junk?

*USF Genomics Class 2020[1], USF Genomics Class 2021[1], Markus Sutter[2,3], Cheryl A. Kerfeld[2,3] and Kathleen M. Scott[1]**

[1]Integrative Biology Department, University of South Florida, Tampa, FL, United States, [2]MSU-DOE Plant Research Laboratory, Michigan State University, East Lansing, MI, United States, [3]Environmental Genomics and Systems Biology and Molecular Biophysics and Integrated Bioimaging Divisions, Lawrence Berkeley National Laboratory, Berkeley, CA, United States

Carboxysomes, responsible for a substantial fraction of CO_2 fixation on Earth, are proteinaceous microcompartments found in many autotrophic members of domain *Bacteria*, primarily from the phyla *Proteobacteria* and *Cyanobacteria*. Carboxysomes facilitate CO_2 fixation by the Calvin-Benson-Bassham (CBB) cycle, particularly under conditions where the CO_2 concentration is variable or low, or O_2 is abundant. These microcompartments are composed of an icosahedral shell containing the enzymes ribulose 1,5-carboxylase/oxygenase (RubisCO) and carbonic anhydrase. They function as part of a CO_2 concentrating mechanism, in which cells accumulate HCO_3^- in the cytoplasm *via* active transport, HCO_3^- enters the carboxysomes through pores in the carboxysomal shell proteins, and carboxysomal carbonic anhydrase facilitates the conversion of HCO_3^- to CO_2, which RubisCO fixes. Two forms of carboxysomes have been described: α-carboxysomes and β-carboxysomes, which arose independently from ancestral microcompartments. The α-carboxysomes present in *Proteobacteria* and some *Cyanobacteria* have shells comprised of four types of proteins [CsoS1 hexamers, CsoS4 pentamers, CsoS2 assembly proteins, and α-carboxysomal carbonic anhydrase (CsoSCA)], and contain form IA RubisCO (CbbL and CbbS). In the majority of cases, these components are encoded in the genome near each other in a gene locus, and transcribed together as an operon. Interestingly, genome sequencing has revealed some α-carboxysome loci that are missing genes encoding one or more of these components. Some loci lack the genes encoding RubisCO, others lack a gene encoding carbonic anhydrase, some loci are missing shell protein genes, and in some organisms, genes homologous to those encoding the carboxysome-associated carbonic anhydrase are the only carboxysome-related genes present in the genome. Given that RubisCO, assembly factors, carbonic anhydrase, and shell proteins are all essential for carboxysome function, these absences are quite intriguing. In this review, we provide an overview of the most recent studies of the structural components of carboxysomes, describe the genomic context and taxonomic distribution of atypical carboxysome loci, and propose functions for these variants. We suggest that these atypical loci are JEEPs, which have modified functions based on the presence of Just Enough Essential Parts.

Keywords: carboxysome, microcompartment, carbonic anhydrase, carbon dioxide fixation, autotroph

OPEN ACCESS

Edited by:
Biswarup Mukhopadhyay, Virginia Tech, United States

Reviewed by:
James Ferry, The Pennsylvania State University (PSU), United States
Iris Maldener, University of Tübingen, Germany
John Albert Raven, University of Dundee, United Kingdom

***Correspondence:**
Kathleen M. Scott
kmscott@usf.edu

Specialty section:
This article was submitted to Microbial Physiology and Metabolism, a section of the journal Frontiers in Microbiology

Received: *09 February 2022*
Accepted: *10 March 2022*
Published: *20 May 2022*

Citation:
USF Genomics Class 2020, USF Genomics Class 2021, Sutter M, Kerfeld CA and Scott KM (2022) Atypical Carboxysome Loci: JEEPs or Junk? Front. Microbiol. 13:872708. doi: 10.3389/fmicb.2022.872708

INTRODUCTION

Autotrophic organisms that use the Calvin-Benson-Bassham cycle (CBB) for carbon dioxide fixation must grapple with the catalytic constraints of ribulose 1,5-bisphosphate carboxylase/oxygenase (RubisCO). This enzyme has poor substrate specificity; it catalyzes both the carboxylase reaction of the CBB, as well as a wasteful oxygenase reaction, which results in added energetic expense to regenerate the ribulose 1,5 bisphosphate (RuBP) necessary for the CBB (Tabita, 1999). In addition, RubisCO enzymes have relatively low affinities for CO_2 (5–250 μM; Tabita, 1999). RubisCO affinities for CO_2 are particularly low for autotrophic bacteria (25–250 μM; tabulated in Horken and Tabita, 1999). Furthermore, RubisCO is not able to use HCO_3^- (Cooper and Filmer, 1969), the predominant form in the equilibrium between CO_2 and HCO_3^- at the circumneutral pH typical for cytoplasm.

In order to grow while using CO_2 as a major carbon source, many autotrophic bacteria using the CBB cycle have CO_2-concentrating mechanisms (CCMs). CCMs consist of two components: (1) membrane transporters for dissolved inorganic carbon (DIC; = $CO_2 + HCO_3^- + CO_3^{2-}$), which generate high concentrations of cytoplasmic HCO_3^-, and (2) carboxysomes, which are present in the cytoplasm and facilitate high rates of CO_2 fixation by RubisCO (reviewed in Price et al., 2009; Long et al., 2016). Carboxysomes are a type of bacterial microcompartment, and consist of a protein shell filled with RubisCO and a trace of carbonic anhydrase activity (reviewed in Kerfeld et al., 2018). Cytoplasmic HCO_3^- enters carboxysomes, where carbonic anhydrase converts some of it to CO_2, which is then fixed by RubisCO. CO_2 is prevented from escaping from the carboxysome before fixation because the shell is impermeable to this gas (Dou et al., 2008; Cai et al., 2009). The components of CCMs, including carboxysomes, are often upregulated when autotrophic microorganisms are cultivated under low DIC conditions (Dobrinski et al., 2012; Esparza et al., 2019; Scott et al., 2019).

Two types of carboxysomes (α and β) are currently recognized (reviewed in Cannon et al., 2010; Kerfeld and Melnicki, 2016). Members of *Proteobacteria* and certain marine members of *Cyanobacteria* have α-carboxysomes, while the remaining members of *Cyanobacteria* have β-carboxysomes (Price et al., 2009; Scott et al., 2019). These types can be distinguished by the form of RubisCO they carry (α-carboxysomes carry form IA RubisCO; β-carboxysomes carry form IB RubisCO), as well as differences in carbonic anhydrases, scaffolding proteins, and carboxysome shell components (Kerfeld and Melnicki, 2016).

The composition of α-carboxysomes from members of phyla *Proteobacteria* and *Cyanobacteria* is mostly conserved (Kinney et al., 2011; Roberts et al., 2012; Sutter et al., 2021). The icosahedral shells of carboxysomes are comprised of (1) hexagonal units, consisting of hexamers of CsoS1 proteins that assemble into single-layers (Tsai et al., 2007), as well as trimers of CsoS1D proteins that assemble into single and double layers (Klein et al., 2009; Roberts et al., 2012), and (2) pentamers of CsoS4 proteins which assemble into pentagonal truncated pyramids and cap the vertices of the icosahedral shells (Tanaka et al., 2008; Cai et al., 2009; Zhao et al., 2019). Hexamers, trimers, and pentamers typically have central pores, which in some cases open and close. The size and charge of these pores are likely to dictate the selective permeability of carboxysome shells (Tsai et al., 2007; Kinney et al., 2011), which are impermeable to CO_2 (Dou et al., 2008; Cai et al., 2009), and permeable to protons (Menon et al., 2010). α-carboxysomes contain RubisCO and carbonic anhydrase, as described above. Based on amino acid sequence, α-carboxysomal carbonic anhydrase (CsoSCA) was initially believed to be a new form of this enzyme, but its structure clarified that it is a deeply divergent β-carbonic anhydrase (So et al., 2004; Sawaya et al., 2006). α-carboxysomes also contain CsoS2, which facilitates the assembly of these microcompartments by binding to RubisCO and CsoS1 (Cai et al., 2015; Oltrogge et al., 2020). The conserved nature of α-carboxysome shell proteins and contents is reflected in gene synteny apparent in the loci encoding them; typical gene order in α-carboxysome loci is *cbbL*, *cbbS*, *csoS2*, *csoSCA*, *csoS4AB*, and *csoS1ABC*, with *csoS1D* genes, when present, often encoded a few genes downstream or elsewhere (**Figure 1**; Cannon et al., 2002; Cai et al., 2008; Roberts et al., 2012; Axen et al., 2014; Sutter et al., 2021).

Atypical carboxysome loci are scattered among several phyla of *Bacteria* (**Table 1**). Most are present in genomes from members of *Proteobacteria*, as expected, given the abundance of organisms from this phylum with typical α-carboxysome loci (Cannon et al., 2002; Axen et al., 2014; Sutter et al., 2021). The atypical loci described here fall into four categories: (1) *csoSCA* is present without any of the other carboxysome-associated genes; (2) *cbbL* and *cbbS* and *csoSCA* are present, without genes encoding shell proteins; (3) genes encoding RubisCO are missing from the locus, with *cbbL* and *cbbS* encoded in a location distinct from *csoS1*, *csoS2*, *csoSCA*, and *csoS4*; and (4) *csoSCA* is absent, though the other carboxysome-associated genes are present (**Figure 1**). It seems likely that these atypical loci originated from typical loci, and were selected for in some lineages. The objective of this review is to assess the likelihood that the genes of these atypical loci are functional, predict the function of the loci, and describe how they may have originated.

DO THE GENES FROM ATYPICAL CARBOXYSOME LOCI ENCODE FUNCTIONAL PROTEINS?

The majority of genes from atypical carboxysome loci appear to encode proteins that could function similarly to their homologs from typical carboxysome loci, based on the presence of conserved amino acids predicted from their sequences (**Table 2**). For CbbS sequences from members of genus *Nitrobacter*, conserved residue Y25 (tyrosine) is replaced with histidine; given that both are large polar amino acids, this substitution may not disrupt the function of CbbS in these organisms. *Pseudonocardia* sp. N23 has two *cbbS* genes, with one (IMG gene ID 2868417193) immediately upstream of

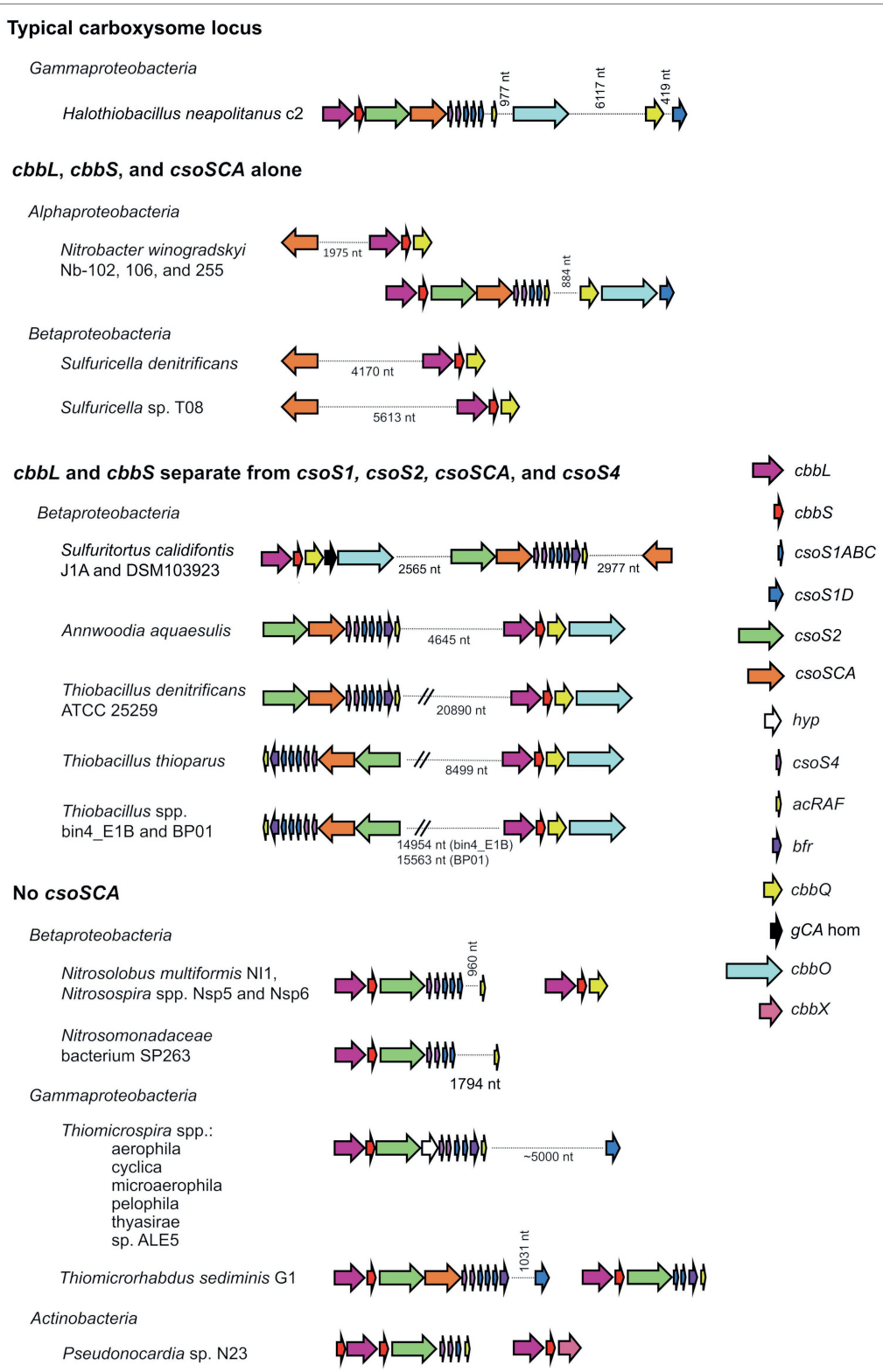


FIGURE 1 | Atypical carboxysome loci. Arrows connected by dotted lines are collocated on the genome, and the distance between them is indicated in nucleotides (nt). *cbbL*, ribulose 1,5-carboxylase/oxygenase (RubisCO) large subunit; *cbbS*, RubisCO small subunit; *csoS1ABC*, hexamer shell proteins; *csoS1D*, pseudohexamer shell protein; *csoS2*, carboxysome assembly protein; *csoSCA*, carboxysomal carbonic anhydrase; *hyp*, hypothetical protein; *csoS4*, pentameric shell protein; *acRAF*, RubisCO assembly factor; *bfr*, bacterioferritin-like protein; *cbbQ*, RubisCO activase; *gCA* hom: gamma carbonic anhydrase homolog; cbbO: adaptor for CbbQ protein; and *cbbX*, RubisCO activase.

TABLE 1 | Number of genomes with atypical carboxysome loci.

Type of locus[a]	*Proteobacteria*			Other phyla
	Alpha	*Beta*	*Gamma*	
Just *csoSCA*	2	27	63	28 from eight phyla[b]
cbbLS and *csoSCA*	3	2	–	–
No *CsoSCA*	–	4	7	One from *Actinobacteria*
cbbLS and *csoS1-4* separate	–	4	–	–

[a]*Atypical carboxysome loci were gathered from the Integrated Microbial Genomes & Microbiomes database (IMG; https://img.jgi.doe.gov; Chen et al., 2019). To find these atypical carboxysome loci, two lists of genomes from isolates were compared: (1) the list of all genomes containing genes encoding members of Pfam12288 (csoS2) or Pfam08936 (csoSCA), believed to be exclusive to carboxysomes (collected using the "find functions" feature at IMG), and (2) the list of all genomes containing typical α-carboxysome loci, with 10 kb regions of genome sequence encoding members of Pfam00016 (cbbL), Pfam00101 (cbbS), Pfam00936 (csoS1), Pfam12288 (csoS2), Pfam08936 (csoSCA), and Pfam03319 (csoS4) (collected using the "cassette search" feature at IMG). Genomes absent from list (2) were examined more closely to determine whether they had atypical carboxysome loci, or whether genes were absent due to sequencing gaps. To remove carboxysome loci likely to be incomplete due to sequencing gaps, draft genomes, and genomes from this list with >100 scaffolds were removed. For the remainder of the genomes on the list, the gene neighborhoods of the csoS2 and csoSCA genes were examined, and those in which these genes were located at the end of a scaffold were removed. The remaining loci were manually reviewed to verify the presence and absence of cbbL, cbbS, csoS1, csoS2, csoSCA, and csoS4.*

[b]*Phyla in which CsoSCA homologs are present outside of carboxysome loci include Candidatus Falkowbacteria (eight genomes), Candidatus Magasanikbacteria (two genomes), Candidatus Moranbacteria (nine genomes), Candidatus Pacebacteria (one genome), Candidatus Staskawiczbacteria (one genome), Candidatus Uhrbacteria (three genomes), Chrysogenetes (two genomes), and Nitrospira (two genomes).*

cbbL, and the other (IMG gene ID 2868417191) immediately downstream (**Figure 1**). The protein encoded by the upstream *cbbS* is only 63 amino acids long, shorter than is typical for CbbS (~90 amino acids), and is missing several conserved amino acids (L53, P54, and F56) in the portion that is present. The protein encoded by the *cbbS* gene downstream *cbbL*, as annotated in IMG, has a truncated amino terminus, but selecting an alternative start codon results in a predicted amino acid sequence including S2, L11, and P12. Based on these observations, the *cbbS* gene downstream of *cbbL* in *Pseudonocardia* sp. N23 is likely to be functional, while the *cbbS* upstream is not.

Genes encoding CsoS1A-C from atypical carboxysome loci have some amino acid substitutions at conserved positions. In many cases the substitutions are biochemically similar; e.g., V36 (valine) is replaced with an isoleucine, R70 (arginine) is replaced with a lysine, I80 (isoleucine) is replaced with a valine, which suggests similar functionality. However, there are some instances, e.g., for *Pseudonocardia* sp. N23, where the amino acids are not biochemically similar [V36 (valine) is replaced with glutamate; G37 (glycine) is replaced with aspartate]; given that these are core residues of CsoS1 monomers, folding may be problematic, suggesting that these carboxysomes may not be able to assemble.

TABLE 2 | Conserved residues in carboxysome-associated proteins.

Protein	Model organism	Conserved amino acids*	Function	Reference
CbbL	*Rhodospirillum rubrum* (CbbM)†	K166, K191, D193, E194, H287, G393, and G395	Active site residues	Watson et al., 1999
CbbS	*Cupriavidus necator*	S9, L11, P12, Y25, E36, W48, L53, P54, F56, and E67	Nearly universally conserved in CbbS	Spreitzer, 2003
CsoS1A,B,C	*Halothiobacillus neapolitanus*	D25, K29, V36, G37, R51, G52, V57, A63, G64, A67, R70, I80, I81, R83, H85, L92, and P93	Conserved in both α- and β-carboxysomes	Tsai et al., 2007; Kinney et al., 2011
CsoS2	*Halothiobacillus neapolitanus*	N region repeats; M region repeats; and Conserved C terminus		Cai et al., 2015; Oltrogge et al., 2020
CsoSCA (CsoS3)	*Halothiobacillus neapolitanus*	C173, D175, R177, H242, C253, H397, and E399	Active site residues	Sawaya et al., 2006
CsoS4	*Halothiobacillus neapolitanus*	V6, D40, G43, V50, S56, A58, D70, and D/E78	conserved	Zhao et al., 2019

**Conserved amino acids are numbered relative to their position in the amino acid sequence from the model organism.*

†*CbbM (form II RubisCO) is homologous to CbbL (large subunit of form I RubisCO); both CbbM and CbbL catalyze the carboxylation of ribulose 1,5-bisphosphate (Tabita et al., 2008).*

Genes encoding CsoS2 from atypical carboxysome loci have features that have been found to be conserved among sequences from typical carboxysome loci. All have at least one N-terminal [RK]XXXXX[HKR]R motif, which binds RubisCO (Cai et al., 2015; Oltrogge et al., 2020). Of the six repetitive motifs from the M (middle) region of CsoS2 (Cai et al., 2015), M1–M4 and M6 are present, while M5 is less conserved. All share a conserved carboxy terminus as described in (Cai et al., 2015).

α-Carboxysomal carbonic anhydrase encoded by atypical carboxysome loci, including those from loci consisting solely of *csoSCA* homologs, have all of the active site residues. In typical carboxysome loci, *csoSCA* follows *csoS2*. Members of

genus *Thiomicrospira* have a gene following *csoS2* which in some cases matches weakly with Pfam08936 (see section "No *csoSCA*" below), but lack all of the active site residues necessary for carbonic anhydrase activity.

Amino acid sequences predicted from genes encoding CsoS4A and B from all of the atypical carboxysome loci include all of the conserved residues, though in some cases S56 (serine) is replaced with threonine; given that a hydroxyl moiety is present in both of these amino acids, this substitution is less likely to be disruptive to the function of these proteins.

TAXONOMIC DISTRIBUTION, ORIGIN, AND POTENTIAL FUNCTION OF THE FOUR TYPES OF ATYPICAL CARBOXYSOME LOCI

As described above, based on predicted amino acid sequences, most of the individual genes of atypical carboxysome loci appear to encode proteins sufficiently conserved to be capable of the same function as their homologs from typical carboxysome loci. Below are detailed descriptions of the taxonomic distribution of atypical loci, possible mechanisms for their origins, and predictions of how the proteins encoded by atypical carboxysome loci could function together.

csoSCA Alone

Genes homologous to those encoding CsoSCA are quite widespread beyond carboxysome loci, and are present in genomes from autotrophic (e.g., *Sulfuritortus caldifontis*, *Nitrospia marina*) and heterotrophic (e.g., *Cand*. Accumulibacter phosphatis; *Chrysiogenes arsenatis*) bacteria. Given their widespread distribution, it is surprising that they have yet to be studied (**Table 1**; **Figure 2**; referred to as *csoSCA2* to distinguish them from those present in carboxysome loci). The amino acid sequences predicted from *csoSCA2* genes share many features with CsoSCA proteins; they include both an active and defunct domain (Sawaya et al., 2006), and the active domain includes all of the residues necessary for catalytic activity as carbonic anhydrase (**Table 2**; **Figure 3**).

There are two variants of CsoSCA2. The first variant closely resembles carboxysomal CsoSCA (found in *Nitrobacter vulgaris*, *Nitrobacter winogradskyi*, *Nitrosomonas nitrosa*, and *Nitrosomonas* sp. 51), clustering with carboxysomal CsoSCA in phylogenetic analyses, but missing the N-terminal 40 residues (**Figure 2**; **Figure 3**). The second variant is further truncated at the N-terminus, is missing short stretches of sequence throughout, and does not cluster with carboxysomal CsoSCA sequences (**Figure 2**; **Figure 3**). The more substantially truncated version has an N-terminal domain of only 40 amino acids (instead of 144 for the CsoSCA from *Halothiobacillus neapolitanus*) that does not align with the CsoSCA equivalent on a sequence level but is also predicted to form two short alpha helices in an AlphaFold2 model (Jumper et al., 2021; **Figure 3**). Further truncations include shorter loops connecting secondary structure elements (**Figure 3**). Some of those extra elements are involved in dimer contacts (Sawaya et al., 2006), so it is possible that this homolog has lost the ability to form dimers, which would be unusual for a β-carbonic anhydrase (Cannon et al., 2010); however, this would need to be verified experimentally. The N-terminus of CsoSCA from *Htb. neapolitanus* facilitates interaction between CsoSCA and RubisCO (Blikstad et al., 2021). Presumably, since CsoSCA2 proteins do not interact with RubisCO, this N-terminal region is not necessary for CsoSCA2 to function outside of carboxysomes. Altogether, this form of CsoSCA2 seems to be a more compact version, possibly due to the fact that it is not necessary to encapsulate this protein in a carboxysome. This is particularly interesting for genomes that include both *csoSCA* and *csoSCA2* genes (e.g., members of *Nitrobacter*, *Nitrosomonas*, *Ectothiorhodospira*, and *Halorhodospira*). Presumably, the CsoSCA2 proteins cannot assemble within the carboxysomes present in these organisms.

The presence of CsoSCA2 sequences in numerous phyla, and the more restricted distribution of CsoSCA, suggest that CsoSCA may have originated from CsoSCA2. However, in some cases, the reverse appears to be the case. *Nitrobacter vulgaris*, *Nb. winogradskyi* (Nb-102, 106, and 255), *Ns. nitrosa*, and *Nitrosomonas* sp. 51 have genes encoding both a CsoSCA (encoded in a typical carboyxsome locus), and a CsoSCA2 encoded elsewhere. The two copies cluster together within the larger clade of carboxysomal CsoSCA sequences, despite having the truncated N-termini seen in other CsoSCA2 sequences (**Figure 2**; **Figure 3**). Sequence similarities between CsoSCA and CsoSCA2 proteins in these organisms suggest that these CsoSCA2 sequences duplicated and diverged from CsoSCA.

One wonders why these deeply divergent β-carbonic anhydrases are so widespread, and what the features of these proteins might be that make them particularly useful to their host organisms. Though CsoSCA2 proteins lack the residues needed to associate with RubisCO, they may have residues that facilitate the formation of other types of enzyme complexes. Alternatively, CsoSCA2 proteins may not require aggregation with other proteins for activity. Indeed, even carboxysomal CsoSCA is active when expressed in the absence of other carboxysomal proteins (Heinhorst et al., 2006), which suggests that "free" CsoSCA2 could also be active in the cytoplasm of its host organisms.

cbbL, *cbbS*, and *csoSCA* Alone

In genomes from some members of *Alpha*- and *Betaproteobacteria*, *csoSCA* homologs are present near genes encoding RubisCO (**Figure 1**; **Figure 4**). In *Alphaproteobacteria*, three strains of *Nb. winogradskyi* share this arrangement of genes (though the average nucleotide identities of strains Nb-102 and Nb-106 versus Nb-255 are 94.6%, suggesting they may be a different species; Richter and Rosselló-Móra, 2009). Among *Betaproteobacteria*, two species of *Sulfuricella* have *cbbL* and *cbbS* genes near *csoSCA* homologs (**Figure 4**).

For both the *Alphaproteobacteria* and *Betaproteobacteria*, if these genomically juxtaposed *cbbL*, *cbbS*, and *csoSCA* genes are the fragments of a single degraded carboxysome locus, one would anticipate that phylogenetic analysis would place

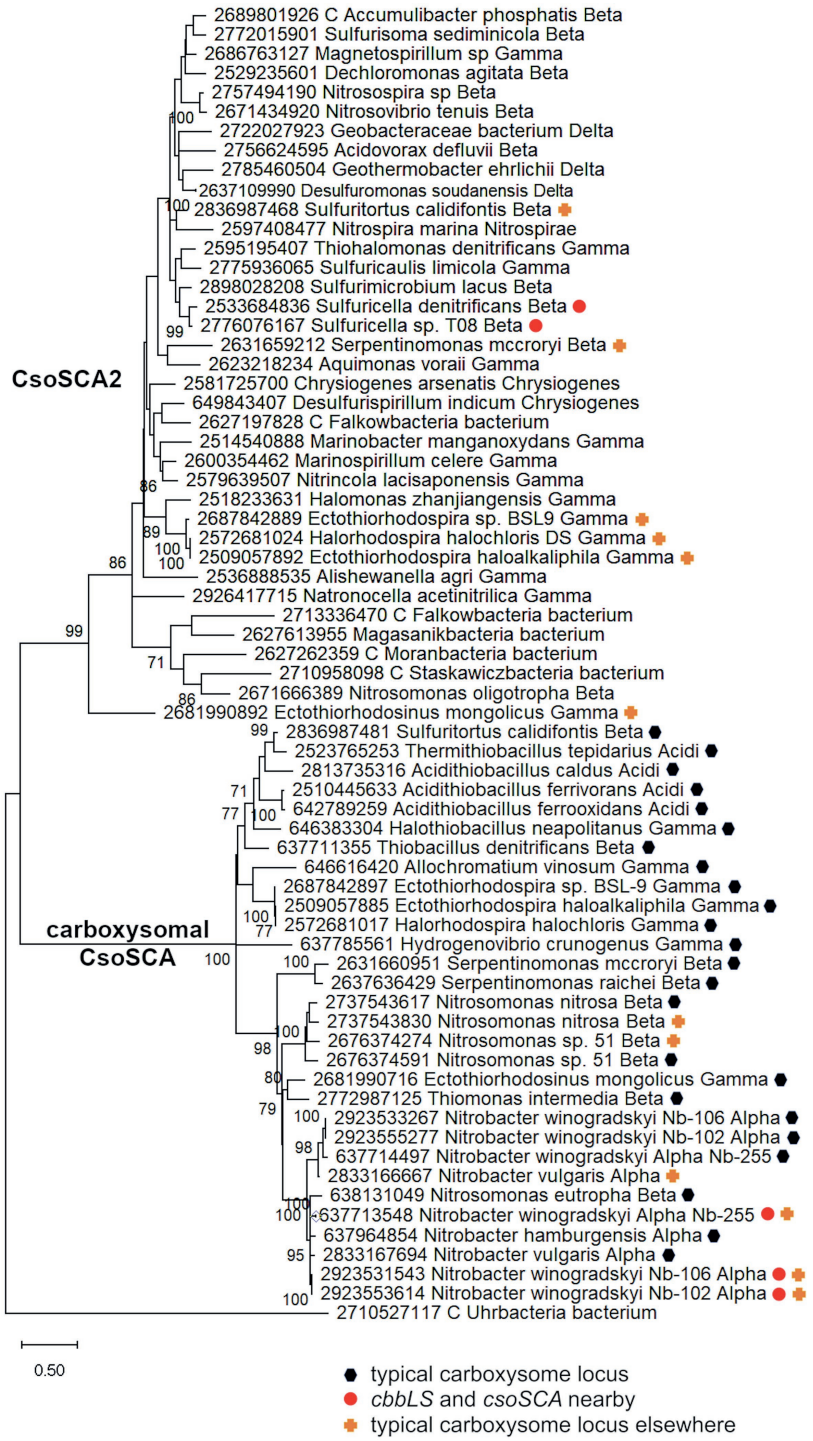


FIGURE 2 | Maximum likelihood analysis of CsoSCA homologs from carboxysome loci and elsewhere (CsoSCA2). Amino acid sequences were gathered from the IMG database, aligned by MUSCLE in MEGA 11, and trimmed *via* GBLOCKS to a final length of 278 aa (Edgar, 2004; Talavera and Castresana, 2007; Tamura et al., 2021). The maximum likelihood tree was constructed with partial deletion of gaps (95% cut-off) and the JTT model (Jones et al., 1992; discrete Gamma distribution with five categories, gamma parameter = 1.9314, 3.55% of sites evolutionarily invariant; this model had the lowest AIC calculated *via* the Find Best DNA/Protein Models feature in MEGA 11; Hurvich and Tsai, 1989; Akaike, 1998). Branch lengths are proportional to the number of substitutions (scale bar = substitutions per site). Bootstrap values are based on 500 resamplings of the alignment, with values <70% omitted. Taxon labels include abbreviated names of classes of *Proteobacteria* (*Alpha*, *Beta*, *Gamma*, and *Delta*; *Acidi* = *Acidithiobacillia*), and full names of phyla beyond *Proteobacteria*. "C" indicates candidate status of species or phylum names. Taxon names also include symbols indicating the position of CsoSCA homologs relative to carboxysome-related genes, if present in the genomes. "Typical carboxysome locus" indicates that the CsoSCA homolog is part of a typical carboxysome locus, "*cbbLS* and *csoSCA* nearby" indicates that genes encoding RubisCO and a CsoSCA homolog are juxtaposed on the genome, and "typical carboxysome locus elsewhere" indicates that a typical carboxysome locus is present elsewhere on the genome.

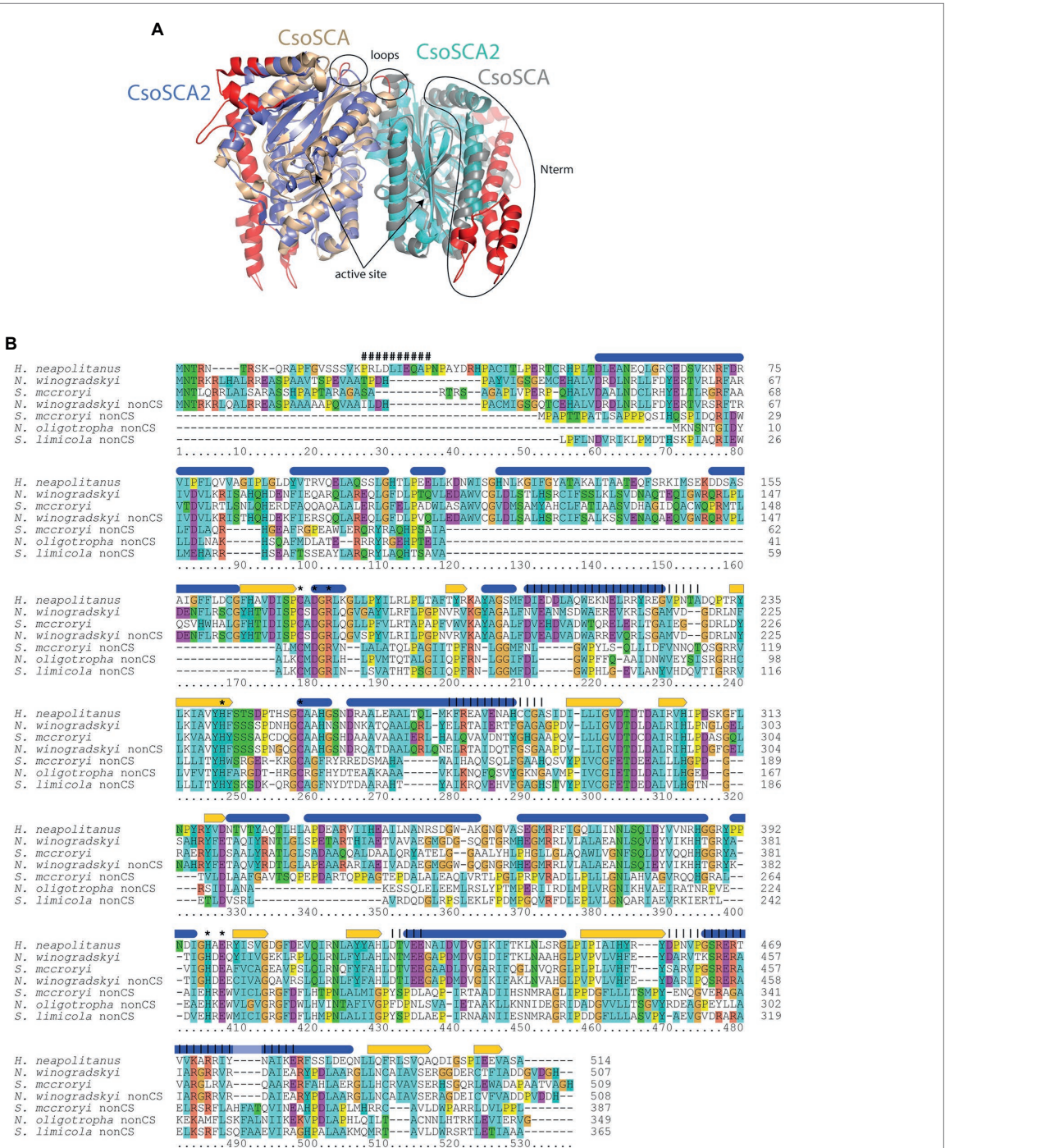


FIGURE 3 | Comparison of carboxysomal CsoSCA and CsoSCA2. **(A)** Cartoon representation (www.pymol.org) of a dimer of CsoSCA subunits (pdb id 2FGY) in gray and wheat, and AlphaFold2 model of CsoSCA2 in blue and cyan. The active site zinc is labeled and shown as a sphere. Differences due to truncations are shown on the CsoSCA in red. The two N-terminal helices of CsoSCA2 are shown as slightly darker blue/cyan. Loop regions with truncations at the dimer interface are also labeled. **(B)** Alignment of carboxysomal CsoSCA and CsoSCA2 sequences. Structural and functional information from *Halothiobacillus neapolitanus* CsoSCA is indicated above the alignment: Blue ovals = alpha-helices, yellow arrows = beta strands, "#" = RubisCO binding site, "*" = active site residue, vertical lines = regions involved in dimerization. Coloring of conserved residues is according to chemical properties. Sequences from typical carboxysome loci included in the alignment are: *Halothiobacillus neapolitanus*, *H. neapolitanus* from *Gammaproteobacteria*, IMG gene object ID 646383304; *Nitrobacter winogradskyi*, *N. winogradskyi* from *Alphaproteobacteria*, IMG gene object ID 2923555277; *Serpentimonas mccroryi*, *S. mccroryi* from *Betaproteobacteria*, IMG gene object ID 2631660951. CsoSCA2 sequences included in the alignment are: *N. winogradskyi* nonCS, *N. winogradskyi* from *Alphaproteobacteria*, IMG gene object ID 2923553614; *Serpentimonas mccroryi* nonCS, *S. mccroryi* from *Betaproteobacteria*, IMG gene object ID 2631659212; *Nitrosomonas oligotropha* nonCS, *N. oligotropha* from *Betaproteobacteria*, IMG gene object ID 2671666389; *Sulfuricaulis limicola* nonCS, *S. limicola* from *Gammaproteobacteria*, IMG gene object ID 2775936065.

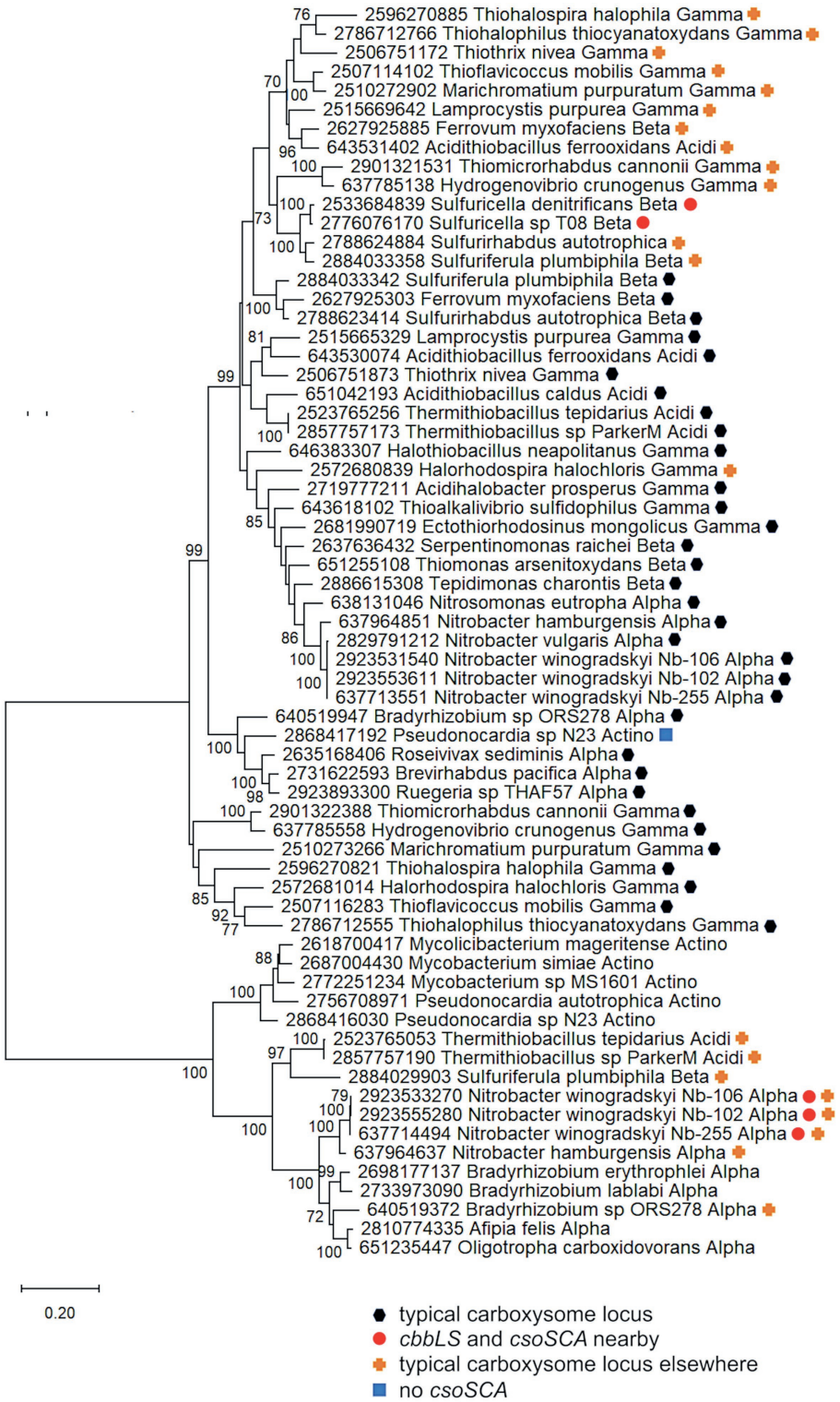


FIGURE 4 | Ribulose 1,5-carboxylase/oxygenase subunits (CbbL and CbbS) encoded by genes collocated with *csoSCA* homologs. Maximum likelihood analysis of CbbL and CbbS sequences was undertaken on sequences gathered from the IMG database, aligned by MUSCLE in MEGA 11, and trimmed *via* GBLOCKS (Edgar, 2004; Talavera and Castresana, 2007; Tamura et al., 2021). CbbL and CbbS alignments were then concatenated using the FASTA alignment joiner feature at FABOX (https://birc.au.dk/~palle/php/fabox/index.php), resulting in an alignment of 527 residues. The tree was constructed with partial deletion of gaps (95% cut-off) and the Le_Gascuel model (Le and Gascuel, 2008; discrete Gamma distribution with five categories, gamma parameter = 0.5775, 17.46% of sites evolutionarily invariant; this model had the lowest AIC calculated *via* the Find Best DNA/Protein Models feature in MEGA 11; Hurvich and Tsai, 1989; Akaike, 1998). Branch

(Continued)

FIGURE 4 | lengths are proportional to the number of substitutions (scale bar = substitutions per site). Bootstrap values are based on 500 resamplings of the alignment, with values <70% omitted. Taxon labels include abbreviated names of classes of *Proteobacteria* (*Alpha*, *Beta*, *Gamma*, and *Delta*; *Acidi* = *Acidithiobacillia*), and members of *Actinobacteria* (*Actino*). Taxon names also include symbols indicating the position of *cbbL* and *cbbS* genes relative to carboxysome-related genes, if present in the genomes. "Typical carboxysome locus" indicates that the *cbbL* and *cbbS* genes are part of a typical carboxysome locus, "*cbbLS* and *csoSCA* nearby" indicates that genes encoding RubisCO and a CsoSCA homolog are juxtaposed on the genome, "typical carboxysome locus elsewhere" indicates that a typical carboxysome locus is present elsewhere on the genome, and "No CsoSCA" indicates that the carboxysome locus lacks *csoSCA*.

them among genes encoding their carboxysomal cognates from taxonomically affiliated organisms. For the *Nitrobacter* spp. and *Sulfuricella* spp., the *cbbLS* genes cluster with noncarboxysomal RubisCO genes (**Figure 4**). The situation is more complicated for the *csoSCA* homologs (**Figure 2**). For the *Nitrobacter* spp., the *csoSCA* homologs appear to be recent duplicates of those present in the typical carboxysome loci in their genomes. For the *Sulfuricella* spp., the *csoSCA* homologs fall within the *csoSCA2* clade and are unlikely to have arisen from carboxysome loci. Given the noncarboxysomal origin of the *cbbLS* genes in both classes, and *csoSCA* gene in the *Sulfuricella* spp., these are not fragments of a single degraded carboxysome locus.

Despite the likelihood that they do not share evolutionary history cohabitating carboxysomes, it is still possible that these two enzymes might function together in the cytoplasm to facilitate CO_2 fixation in their host organisms, all of which are capable of autotrophic growth (Winogradsky, 1892; Kojima and Fukui, 2010). Coregulation is possible for both, but seems more likely for the members of *Nitrobacter*, since their genes are <2 kb apart (**Figure 1**). The juxtaposition of noncarboxysomal RubisCO genes to those encoding typical β-carbonic anhydrase has been noted for two members of *Hydrogenovibrio* (Yoshizawa et al., 2004; Scott et al., 2006), and is apparent in genome data from in many other members of *Hydrogenovibrio* and *Thiomicrorhabdus*,[1] suggesting such juxtaposition may be selected for in some organisms. While the expression of cytoplasmic carbonic anhydrase results in a high CO_2-requiring phenotype in organisms with CCMs (Price and Badger, 1989), there is evidence for carbonic anhydrase activity in the chloroplasts of certain algae and plants (reviewed in Moroney et al., 2001). If these enzymes do function together in *Nitrobacter* and *Sulfuricella*, perhaps the carbonic anhydrase facilitates RubisCO-mediated CO_2 fixation by maintaining intracellular HCO_3^- and CO_2 near chemical equilibrium, preventing RubisCO from diminishing the concentration of intracellular CO_2 under conditions where CCMs are not induced (e.g., moderate environmental CO_2 concentrations; Yoshizawa et al., 2004).

cbbL and *cbbS* Separate From *csoS1*, *csoS2*, *csoSCA*, and *csoS4*

Many members of family *Thiobacillaceae* (Boden et al., 2017; Boden, 2019) have *csoS1–S4* genes in a separate genomic locus from *cbbL* and *cbbS* genes (**Figure 1**), as has previously been described for *Thiobacillus denitrificans* (Cannon et al., 2003; Beller et al., 2006a). Of the eight genome sequences from cultivated members of this family, all of which grow autotrophically (Boden et al., 2017; Boden, 2019), five include a homolog to *csoS2* (Pfam012288; *Annwoodia aquaesulis*, *Sulfuritortus calidifontis* DSM103923 and J1A, *Thiobacillus denitrificans* ATCC25259, and *Thiobacillus thioparus*). In all cases, these *csoS2* genes do not have *cbbL* and *cbbS* genes immediately upstream. Instead, *cbbL* and *cbbS* are located 2.6–21 kb away from *csoS1-4* (**Figure 1**). The other three genome sequences lack *csoS2* homologs; since these three sequences are incomplete (38–98 scaffolds), it is not possible to know whether *csoS2* is truly absent from these organisms. Nine genomes (15–407 scaffolds) inferred to belong to members of *Thiobacillaceae* have been gathered from metagenomes, and five of these include *csoS2* homologs. Three of these genes are present at the ends of scaffolds, making it impossible to determine whether *cbbL* and *cbbS* genes are nearby. For the two remaining (*Thiobacillus* spp. Bin4_E1B and BP01), *cbbL* and *cbbS* are encoded separately from *csoS1–4*. Based on these observations, it seems likely that having *csoS1–4* genes apart from *cbbL* and *cbbS* genes may be a trait shared by all members of this family.

There are two mechanisms that could have resulted in the separate *cbbLS* and *csoS1-4* loci found in members of *Thiobacillaceae*. In the first scenario, a typical ancestral carboxysome locus containing all of these genes was severed by genome re-arrangement. In the second scenario, *cbbLS* and *csoS1-4* did not share an ancestral locus. Instead, carboxysomal *cbbLS* genes could have been lost from the genome entirely, and the *cbbLS* genes currently located 2–20 kb away are noncarboxysomal in origin. An additional possibility is that either *cbbLS* or *csoS1-4* were acquired *via* horizontal gene transfer.

To provide evidence for different mechanisms for formation of *cbbLS* and *csoS1-4* loci, phylogenetic analyses were conducted using concatenated alignments of *cbbL* and *cbbS* genes (*cbbLS*), and *csoS2*, *S3*, *S4a*, and *S4b* (*csoS2-4*). Genes encoding CsoS1A-C were omitted from these analyses, due to difficulties distinguishing the three types of *csoS1A-C* genes. The results of these analyses raise the possibility that *cbbLS* and *csoS1-4* loci did not originate from a single ancestral typical carboxysome locus in these organisms (**Figure 5**). In *Sf. calidifontis* (here collapsed to strain J1A, since sequences for the two strains are identical), *cbbLS* genes fall in a small well-supported clade with *Tb. denitrificans*, and with other members of *Thiobacillaceae* in larger clades, though these larger clades are not as well-supported (**Figure 5**). The *csoS2-4* genes from *S. calidifontis* fall among completely different taxa than its *cbbLS* genes, suggesting independent origins for its two loci. For all four isolates from *Thiobacillaceae*, *cbbLS* genes do not fall among those from typical carboxysome loci, though it should be noted that carboxysomal and non-carboxysomal *cbbLS* genes are not distinguished by two distinct, well-supported clades (**Figure 5**). Together, these observations suggest independent origins for *cbbLS* and *csoS2-4* loci in *Thiobacillaceae*, but low bootstrap values for these phylogenetic analyses compromise the

[1]https://img.jgi.doe.gov/

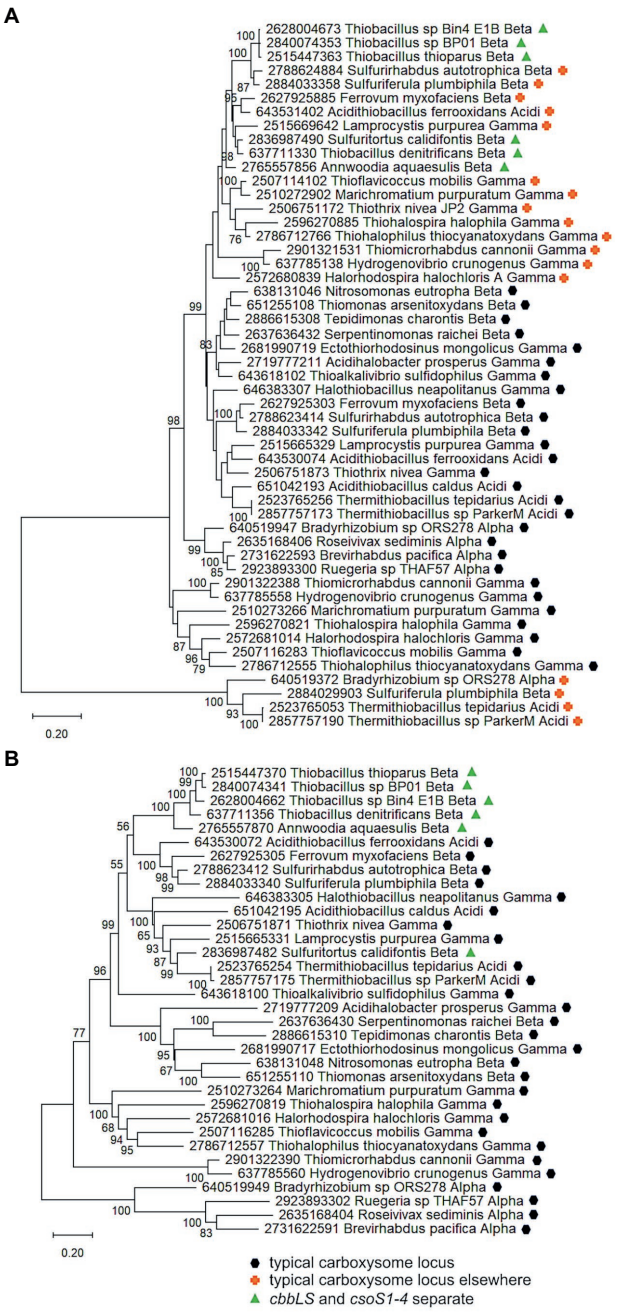


FIGURE 5 | Analysis of RubisCO subunits (CbbL and CbbS) and carboxysome shell proteins (CsoS2, SCA, S4A, and S4B) that are encoded at two separate loci. Maximum likelihood analysis of amino acid sequences of **(A)**. RubisCO subunits and **(B)**. carboxysome shell proteins was undertaken on sequences that were gathered from the IMG database, aligned by MUSCLE in MEGA 11, and trimmed *via* GBLOCKS (Edgar, 2004; Talavera and Castresana, 2007; Tamura et al., 2021). CbbL and CbbS alignments were then concatenated using the FASTA alignment joiner feature at FABOX (https://birc.au.dk/~palle/php/fabox/index.php), as were CsoS2, SCA, S4A, and S4B, resulting in alignments of 550 (CbbLS) and 997 (CsoS2–4) residues. The trees were constructed with partial deletion of gaps (95% cut-off) and the Le_Gascuel model [Le and Gascuel, 2008; discrete Gamma distribution with five categories, gamma parameter = 0.5298, 17.64% of sites evolutionarily invariant (CbbLS); gamma parameter = 1.2921, 7.86% of sites evolutionarily invariant (CsoS2-4); this model had the lowest AIC calculated *via* the Find Best DNA/Protein Models feature in MEGA 11; Hurvich and Tsai, 1989; Akaike, 1998]. Branch lengths are proportional to the number of substitutions (scale bar = substitutions per site). Bootstrap values are based on 500 resamplings of the alignment, with values <70% omitted. Taxon labels include abbreviated names of classes of *Proteobacteria* (*Alpha*, *Beta*, *Gamma*, and *Delta*; *Acidi* = *Acidithiobacillia*). Taxon names also include symbols indicating the position of genes relative to carboxysome-related genes, if present in the genomes. "Typical carboxysome locus" indicates that the genes are part of a typical carboxysome locus, "*cbbLS* and *csoS1-4* separate" indicates that genes encoding RubisCO and CsoS1–4 are encoded by separate loci, and "typical carboxysome locus elsewhere" indicates that a typical carboxysome locus is present elsewhere on the genome.

confidence of this assertion. If the *cbbL* and *cbbS* genes in members of *Thiobacillaceae* did not originate from a typical carboxysome locus, it would be very interesting to verify that they were capable of being packed into carboxysomes, as thus far it seems that noncarboxysomal RubisCO from other organisms cannot be packed into carboxysomes (Menon et al., 2008).

Currently, evidence for the presence of carboxysomes in members of *Thiobacillaceae* is limited to *Tb. thioparus*. Transmission electron micrographs have only been published for *Tb. thioparus* and *Tb. denitrificans*; polyhedral inclusions are apparent in *Tb. thioparus* cells, but not in *Tb. denitrificans* (Shively et al., 1970). Given the synteny of the carboxysome loci among *Thiobacillus* sp. Bin4 E1B, *Thiobacillus* sp. BP01, and *Tb. thioparus*, as well as the placement of their *cbbLS* and *csoS2-4* genes together in clades (**Figure 4**), it seems likely that all three of these organisms are capable of synthesizing carboxysomes. For *Tb. denitrificans*, the absence of carboxysomes despite the presence of *csoS1-4* is puzzling, and cannot be attributed to strain-level differences, since both ultrastructure and genome sequence were obtained from the same strain (ATCC25259). Perhaps their synthesis could be induced under growth conditions different from those used to cultivate the cells for ultrastructural study.

If carboxysomes are indeed synthesized by these organisms, one possible advantage of having separate loci would be independent regulation of *cbbLS* and *csoS1-4* loci. Transcriptome analysis of *Tb. denitrificans* is consistent with this possibility. Based on hybridization with microarrays, transcripts of the *Tb. denitrificans cbbL* and *cbbS* genes are particularly abundant under aerobic conditions, but no such changes are apparent for *csoS1-4* (Beller et al., 2006b). Other organisms have two sets of *cbbLS*, with one in a typical carboxysome locus and the other located elsewhere on the genome. These organisms synthesize noncarboxysomal RubisCO when CO_2 concentrations are moderate, and selectively synthesize carboxysomal RubisCO and shell proteins when CO_2 concentrations are very low (Yoshizawa et al., 2004). Perhaps some members of *Thiobacillaceae* upregulate the *cbbLS* locus when CO_2 concentrations are low to moderate, and reserve upregulation of the *csoS1-4* locus for low CO_2 conditions, or other circumstances where carboxysomes facilitate growth.

No *csoSCA*

Carboxysome loci lacking *csoSCA* genes arose multiple times; they are present in some autotrophic organisms from *Beta-*

and *Gammaproteobacteria*. A carboxysome locus lacking *csoSCA* is also present in *Pseudonocardia* sp. N23, a member *Actinobacteria* (**Figure 1**); though it has not been determined whether this organism could grow autotrophically, other members of its genus can (e.g., *Pseudonocardia autotrophica*; Takamiya and Tubaki, 1956). None of these organisms have *csoSCA* homologs elsewhere in their genomes (aside from *Thiomicrorhabdus sediminis*, which has a copy in its "typical" carboxysome operon). In *Betaproteobacteria*, they are present in *Nitrosospira muliformis* and also *Nitrosospira* spp. Nsp5 and Nsp6, which may be strains of *Nsp. multiformis*, based on average nucleotide identities >99% (Richter and Rosselló-Móra, 2009). Within *Gammaproteobacteria*, they appear to have arisen independently three times. All members of *Thiomicrospira* have carboxysome loci lacking *csoSCA*. Within *Thiomicrorhabdus*, such loci seem to have arisen twice. In *Thiomicrorhabdus sediminis*, two carboxysome loci are present; one is typical, while the second appears to be a recent duplicate of the typical locus. The amino acid sequences predicted from both *cbbL* and *cbbS* genes are 100% identical. Both copies of CsoS2 are 100% identical at amino termini; however, at residue 330, they diverge, and this continues to the carboxy termini. CsoS1 sequences also are identical at the amino termini and have small differences at their carboxy termini. The carboxysome locus from *Thiomicrorhabdus aquaedulcis* does not fall within a clade with those from other members of its genus (**Figure 6**), suggesting that it may have been acquired *via* horizontal gene transfer.

The carboxysome locus from *Pseudonocardia* sp. N23 includes *cbbL* and *cbbS* genes distinct from those present in other members of phylum *Actinobacteria* (**Figure 4**). Other members of this phylum carry *cbbL* and *cbbS*, and *Pseudonocardia* sp. N23 does include a copy that falls within a clade of these sequences (**Figure 4**). However, the *cbbL* and *cbbS* genes present in the carboxysome locus, as well as *csoS2*, *csoS4A*, and *csoS4B*, fall among genes from carboxysome loci from members of *Alpha-* and *Betaproteobacteria* (**Figures 4**, **6**), suggesting this locus was acquired *via* horizontal gene transfer.

It is apparent that these carboxysome loci originated from typical carboxysome loci, given that they cluster with others that contain *csoSCA* genes (**Figure 6**). Indeed, in members of *Thiomicrospira*, a gene is present between *csoS2* and *csoS4A* (**Figure 1**), which is likely to be a degraded form of *csoSCA*. In *Thiomicrospiras pelophila*, *thyasirae*, and *microaerophila*, these genes do match Pfam08936 (*csoSCA*), but e-values range from 0.006 to 4.5e−05, and none of the residues necessary for carbonic anhydrase activity are present. However, the amino termini of the proteins predicted from these genes align well with those from CsoSCA proteins. Given that the amino termini of CsoSCA proteins may facilitate interactions among carboxysome proteins (Blikstad et al., 2021), perhaps these degraded genes may still encode proteins that facilitate packing of RubisCO molecules into carboxysomes.

There is evidence that these carboxysome loci are transcribed and translated. Carboxysome locus genes are transcribed in *Tms. pelophila* (Scott et al., 2019), carboxysomes are visible in transmission electron micrographs of members of *Thiomicrospira* (Sorokin et al., 2001, 2002a,b; Scott et al., 2019), and have been purified from *Tms. thyasirae* (Lanaras et al., 1991). Electron dark inclusions are abundant in *Nsp. muliformis*, but staining patterns suggest these consist of glycogen (Watson et al., 1971). Ultrastructural studies of *Pseudonocardia* sp. N23, as well as *Tmr. aquadulcis* and *sediminis*, have not been published.

Given the presence of carboxysomes in at least some of these taxa, the conservation of residues necessary for the function of the CbbL, CbbS, CsoS2, and CsoS4A and B proteins, and the convergent evolution of this sort of carboxysome locus in multiple lineages of microorganisms, they are likely to be functional in their host organisms (however, see the comments on CsoS1 sequences from *Pseudonocardia* sp. N23 in Section "Do the Genes From Atypical Carboxysome Loci Encode Functional Proteins" above). The current understanding of carboxysome function requires the presence of carbonic anhydrase activity within these microcompartments in order for them to facilitate CO_2 fixation by RubisCO (see above). One possibility is that these modified carboxysomes have shells that are permeable to CO_2, allowing this gas to enter from the cytoplasm. CsoS4 proteins are necessary for carboxysome shell impermeability to CO_2; the absence of CsoS4 to seal the vertices of their shells renders the microcompartments CO_2-permeable (Cai et al., 2009). Their critical function perhaps accounts for their strong sequence conservation, hence redundancy, which is unusual for bacterial microcompartments that have multiple pentamer-forming paralogs (Melnicki et al., 2021). Interestingly, the carboxysome locus from *Tmr. aquaedulcis* lacks genes encoding CsoS4A, and the *Tmr. sediminis* locus lacking *csoSCA* lacks both *csoS4A* and *csoS4B*. Perhaps carboxysomes from these organisms operate without CsoS4 proteins, and are permeable to CO_2. Given that carboxysome shells are assumed to require only 12 pentamers, and their pores are small (~4 Å in diameter), they are assumed to not play a significant role in metabolite conductance. However, a recent study of the protein stoichiometry of β-carboxysomes showed varying occupation of the vertices by the CcmL, the lone pentamer-forming gene product in beta carboxysome loci (Sun et al., 2019). The occupancy was correlated with environmental conditions, suggesting that pentamer association with shells is dynamic and perhaps serves as one way to alter permeability. Because *Tmr. sediminis* has two carboxysome loci (one typical, one lacking *csoSCA*, *csoS4A*, and *csoS4B*), determining the conditions under which it expresses typical, vs. atypical, carboxysomes could provide useful information about how its atypical carboxysomes might function, including whether pentamers and carbonic anhydrase are provided by the other locus. If these carboxysomes are permeable to CO_2, cytoplasmic CO_2 concentrations would need to be elevated in order to enhance RubisCO activity, running the risk of high rates of CO_2 loss from the cells *via* diffusion, unless this loss is counterbalanced by living in a high CO_2 habitat. These organisms have been cultivated in growth media supplemented with HCO_3^- (10–30 mM; Kojima and Fukui, 2019) or CO_2 (20% headspace; Liu et al., 2021). For *Tmr. sediminis*, the lack of *csoS4* genes suggests that this

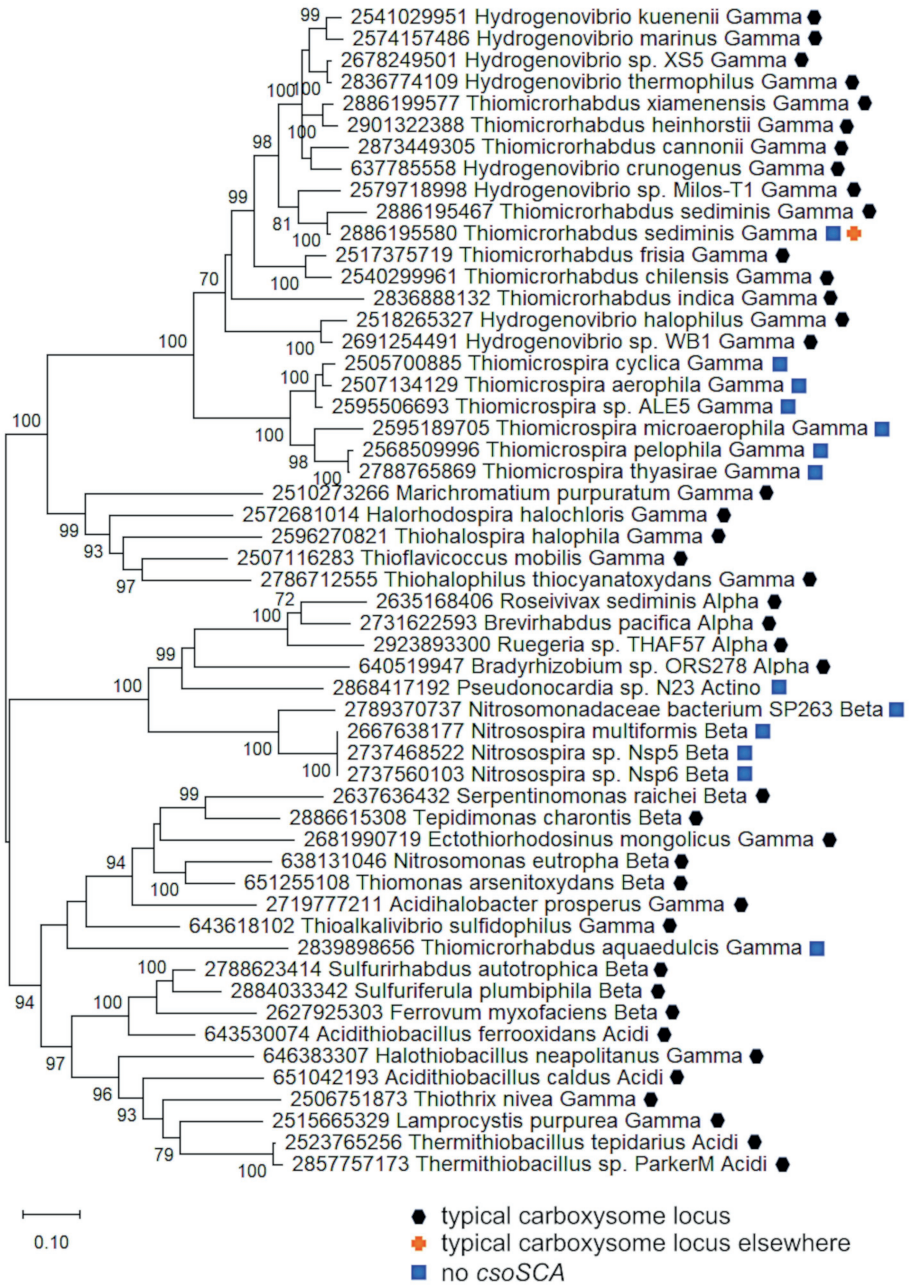


FIGURE 6 | Carboxysome loci lacking *csoSCA* genes. Maximum likelihood analysis was undertaken on amino acid sequences of RubisCO subunits (CbbL and CbbS) and carboxysome shell proteins (CsoS2, S4A, and S4B) gathered from the IMG database, aligned by MUSCLE in MEGA 11, and trimmed *via* GBLOCKS (Edgar, 2004; Talavera and Castresana, 2007; Tamura et al., 2021). CbbL, CbbS, CsoS2, CsoS4A, and CsoS4B alignments were then concatenated using the FASTA alignment joiner feature at FABOX (https://birc.au.dk/~palle/php/fabox/index.php), resulting in an alignment of 940 residues. The trees were constructed with partial deletion of gaps (95% cut-off) and the Le_Gascuel model (Le and Gascuel, 2008; discrete Gamma distribution with five categories, gamma parameter = 0.7037, 16.40% of sites evolutionarily invariant; this model had the lowest AIC calculated *via* the Find Best DNA/Protein Models feature in MEGA 11; Hurvich and Tsai, 1989; Akaike, 1998). Branch lengths are proportional to the number of substitutions (scale bar = substitutions per site). Bootstrap values are based on 500 resamplings of the alignment, with values <70% omitted. Taxon labels include abbreviated names of classes of *Proteobacteria* (*Alpha*, *Beta*, *Gamma*, and *Delta*; *Acidi* = *Acidithiobacillia*) and members of *Actinobacteria* (*Actino*). Taxon names also include symbols indicating the position of genes relative to carboxysome-related genes, if present in the genomes. "Typical carboxysome locus" indicates that the genes are part of a typical carboxysome locus, "typical carboxysome locus elsewhere" indicates that a typical carboxysome locus is present elsewhere on the genome, and "No CsoSCA" indicates that the carboxysome locus lacks *csoSCA*.

organism may not be capable of growth under low CO_2 conditions, and it would be interesting to determine whether this is the case.

The other organisms lacking CsoSCA have loci including genes encoding CsoS1 and CsoS4; perhaps their shells are permeable to CO_2 based on modifications to these two types

of shell proteins. However, such differences are not detected when shell proteins from typical carboxysomes are compared to those from carboxysomes lacking carbonic anhydrase. For CsoS1ABC proteins, the sequence FVGGGY, corresponding to residues 40–45 from *Htb. neapolitanus*, comprises the narrowest part of the pore and the residues surrounding it (Tsai et al., 2007). In all of the CsoS1ABC sequences from atypical carboxysome loci lacking *csoSCA*, these residues are conserved, suggesting the pores have characteristics similar to those in typical carboxysomes. To determine whether there are other residues that vary systematically for these atypical carboxysome loci, and to detect changes in the sequence that are more likely due to the presence/absence of CsoSCA rather than evolutionary distance, CsoS1ABC sequences within *Piscirickettsiaceae* were compared, since genomes from this family include both typical (all 10 species of *Hydrogenovibrio*, 8/10 species from *Thiomicrorhabdus*) and atypical (2/10 species from *Thiomicrorhabdus*, all six members of *Thiomicrospira*) loci. Among all of these organisms, CsoS1ABC sequences are highly conserved throughout the sequences. Likewise, CsoS1D sequences from these organisms have small differences throughout, and mapping those differences on a homology model does not reveal significant patches of variability. CsoS4A and B sequences are also very similar across all three genera and there are no distinguishable differences between them. If these shell proteins actually are permeable to CO_2, the mechanism mediating this change is not apparent from their sequences.

Another mechanism for preserving the activities of these carboxysomes would be their recruitment of a carbonic anhydrase encoded elsewhere on the genome, as may be the case for some β-carboxysomes from *Cyanobacteria*. β-carboxysomes carry homologs to γ-carbonic anhydrase (Dearaujo et al., 2014). In some cases, these homologs are enzymatically active as carbonic anhydrases, while in others, these homologs have apparently lost enzymatic activity (Cot et al., 2008), although the active site residues are intact. In these cases, the carboxysomes also carry a functional β-carbonic anhydrase (deeply divergent to CsoSCA; So et al., 2002; Cot et al., 2008; Rae et al., 2013), and the gene encoding this β-carbonic anhydrase is not present in or near the operon encoding the essential components of the carboxysome (Rae et al., 2013). Evaluation of these possibilities awaits purification of carboxysomes from organisms with carboxysome loci lacking *csoSCA* genes, to test the permeabilities of their shells and the potential presence of carbonic anhydrase activity within them.

CONCLUSION

The unusual carboxysome-related loci described here are common enough to suggest relevance. Genes encoding CsoSCA2 are extremely widespread. Colocalization of *csoSCA* homologs and *cbbLS* is present in genomes from two classes of *Proteobacteria*. "Split" carboxysome loci (*cbbLS* and *csoS1-4*) are likely present in all members of family *Thiobacillaceae*. Carboxysome loci lacking *csoSCA* homologs (or homologs unlikely to be active) are present in at least two classes of *Proteobacteria* and have been horizontally transferred to phylum *Actinobacteria*. Together, all of this indicates that modified carboxysome loci have been evolutionarily selected for in some lineages, and are not the tattered remnants of typical carboxysome loci, captured on their journey to degradation and loss. Understanding how the proteins encoded by these atypical carboxysome loci function could help us understand better how typical carboxysomes function (the exceptions that prove—or disprove—the rule), as well as the selective pressures driving their origins from the assembly of their components over time.

The nature of the selective advantage provided by these atypical loci is not apparent at this point. All of the organisms carrying these atypical carboxysome loci (except for *csoSCA2*) are chemolithoautotrophs, so these atypical loci are likely to play a role in CO_2 fixation. The habitats from which they were isolated are very diverse, with CO_2 concentrations ranging from extremely low (alkaline soda lakes; Sorokin et al., 2001, 2002a,b), to high (e.g., soils, marine sediments; Bock and Wagner, 2006; Kelly and Wood, 2006). Particularly for those organisms from low CO_2 habitats, these atypical carboxysome loci are likely to play a role in CCMs. Consistent with this possibility, most of these organisms have genes for likely DIC transporters either associated with their atypical loci or elsewhere in their genomes (Scott et al., 2020). However, based on the current understanding of CCMs in bacteria, which requires that both RubisCO and carbonic anhydrase are present in carboxysomes, it is difficult to understand how organisms lacking carboxysomal RubisCO (as in section "*cbbL* and *cbbS* Separate From *csoS1*, *csoS2*, *csoSCA*, and *csoS4*" above) or carbonic anhydrase (as in section "No *csoSCA*" above) could have functioning CCMs. This conceptual gap may result from the relative paucity of studies on CCMs in organisms besides *Cyanobacteria*, in which CCMs have been well-studied (reviewed in Price et al., 2009). Though carboxysomes from chemolithoautotrophs have been well-studied (Kerfeld et al., 2010, 2018; Sutter et al., 2021), their integration with the other components of CCMs in these organisms (e.g., DIC transporters) has not. CCM function (carboxysome presence and elevated intracellular DIC concentration) has been demonstrated for only one bacterium beyond *Cyanobacteria* (*Hydrogenovibrio crunogenus*; Dobrinski et al., 2005). Upregulation of genes encoding both DIC transporters and carboxysomes under low DIC conditions has only been demonstrated for a handful of chemolithoautotrophic *Gammaproteobacteria* (Mangiapia et al., 2017; Desmarais et al., 2019; Scott et al., 2019). Despite this undersampling, it is already apparent that CCMs in *Proteobacteria* are more diverse than those from *Cyanobacteria*, in their reliance on a different arsenal of DIC transporters and multiple types of RubisCO (Dobrinski et al., 2012; Scott et al., 2019, 2020). Atypical carboxysomes could represent yet another layer of diversity in these CCMs; evaluating this possibility awaits further study of CCMs in these organisms as well as those in other members of *Proteobacteria*.

AUTHOR CONTRIBUTIONS

KS contributed to conception and design of the manuscript. KS, USF GC 2020, USF GC 2021, MS, and CK ran the analyses. All authors contributed to manuscript writing and revision, and have read and approved the submitted version.

FUNDING

This work was supported by the National Science Foundation (NSF-MCB-1952676 to KS), and the Office of Science of the United States Department of Energy (DE-FG02-91ER20021 to MS and CK), and MSU AgBio Research (MS and CK).

ACKNOWLEDGMENTS

The authors are grateful to the University of South Florida for the resources to make it possible for the students in Genomics Class 2020 and 2021 to contribute to this manuscript. Genomics Class 2020 and 2021 are Salvador Ayala-Preciado, Alizain K. Bangash, Eleanor Brodrick, Deanna Byram, Madilyn Carney, Michael Colosi, Holly David, Gabriela Dolard, Anelys Figueroa, Christyn A. Geniesse, Lucy B. Gravitt, M. Andrew Glass, Rhiannon A. Hendrick, Arianna Henry, Thomas Henry, Alexis Hunt, Nicole Juern, Jason A. Kalathia, Taylor Kelsay, Jacob Klingman, Sydney T. Marshall, Ian M. Matosdeaza, Aaron S. Mohammed, Jeremias Munoz, Nicole Nauman, Dhanvi Patel, Alexandra Rodier, Beatriz Sanchez Uribazo, Logan Suits, Angel D. Valladares, Daniela Varon-Garcia, Alex Wagner, Lauren Wallace, and Jana Wieschollek.

REFERENCES

Akaike, H. (1998). "Information theory and an extension of the maximum likelihood principle," in *Selected Papers of Hirotugu Akaike*. eds. E. Parzen, K. Tanabe and G. Kitagawa (New York, NY: Springer New York), 199–213.

Axen, S. D., Erbilgin, O., and Kerfeld, C. A. (2014). A taxonomy of bacterial microcompartment loci constructed by a novel scoring method. *PLoS Comput. Biol.* 10:e1003898. doi: 10.1371/journal.pcbi.1003898

Beller, H. R., Chain, P. S., Letain, T. E., Chakicherla, A., Larimer, F. W., Richardson, P. M., et al. (2006a). The genome sequence of the obligately chemolithoautotrophic, facultatively anaerobic bacterium *Thiobacillus denitrificans*. *J. Bacteriol.* 188, 1473–1488. doi: 10.1128/JB.188.4.1473-1488.2006

Beller, H. R., Letain, T. E., Chakicherla, A., Kane, S. R., Legler, T. C., and Coleman, M. A. (2006b). Whole-genome transcriptional analysis of chemolithoautotrophic thiosulfate oxidation by *Thiobacillus denitrificans* under aerobic versus denitrifying conditions. *J. Bacteriol.* 188, 7005–7015. doi: 10.1128/JB.00568-06

Blikstad, C., Dugan, E.J., Laughlin, T.G., Liu, M.D., Shoemaker, S.R., Remis, J.P., et al. (2021). Discovery of a carbonic anhydrase-Rubisco supercomplex within the alpha-carboxysome. bioRxiv [Preprint]. doi: 10.1101/2021.11.05.467472

Bock, E., and Wagner, M. (2006). "Oxidation of inorganic nitrogen compounds as an energy source" in *The Prokaryotes*. ed. M. Dworkin. *3rd* ed (New York: Springer), 457–495.

Boden, R. (2019). "Sulfuritortus" in *Bergey's Manual of Systematics of Archaea and Bacteria*. ed. W. B. Whitman (John Wiley & Sons, Inc.), 1–3.

Boden, R., Hutt, L. P., and Rae, A. W. (2017). Reclassification of *Thiobacillus aquaesulis* (Wood & Kelly, 1995) as *Annwoodia aquaesulis* gen. nov., comb. nov., transfer of *Thiobacillus* (Beijerinck, 1904) from the Hydrogenophilales to the Nitrosomonadales, proposal of Hydrogenophilalia class. nov. within the 'Proteobacteria', and four new families within the orders Nitrosomonadales and Rhodocyclales. *Int. J. Syst. Evol. Microbiol.* 67, 1191–1205. doi: 10.1099/ijsem.0.001927

Cai, F., Dou, Z., Bernstein, S. L., Leverenz, R., Williams, E. B., Heinhorst, S., et al. (2015). Advances in understanding carboxysome assembly in *Prochlorococcus* and *Synechococcus* implicate CsoS2 as a critical component. *Life* 5, 1141–1171. doi: 10.3390/life5021141

Cai, F., Heinhorst, S., Shively, J. M., and Cannon, G. C. (2008). Transcript analysis of the *Halothiobacillus neapolitanus* cso operon. *Arch. Microbiol.* 189, 141–150. doi: 10.1007/s00203-007-0305-y

Cai, F., Menon, B. B., Cannon, G. C., Curry, K. J., Shively, J. M., and Heinhorst, S. (2009). The pentameric vertex proteins are necessary for the icosahedral carboxysome shell to function as a CO_2 leakage barrier. *PLoS One* 4:e7521. doi: 10.1371/journal.pone.0007521

Cannon, G. C., Baker, S. H., Soyer, F., Johnson, D. R., Bradburne, C. E., Mehlman, J. L., et al. (2003). Organization of carboxysome genes in the thiobacilli. *Curr. Microbiol.* 46, 115–119. doi: 10.1007/s00284-002-3825-3

Cannon, G. C., Heinhorst, S., Bradburne, C. E., and Shively, J. M. (2002). Carboxysome genomics: a status report. *Funct. Plant Biol.* 29, 175–182. doi: 10.1071/PP01200

Cannon, G. C., Heinhorst, S., and Kerfeld, C. A. (2010). Carboxysomal carbonic anhydrases: structure and role in microbial CO_2 fixation. *Biochim. Biophys. Acta* 1804, 382–392. doi: 10.1016/j.bbapap.2009.09.026

Chen, I. A., Chu, K., Palaniappan, K., Pillay, M., Ratner, A., Huang, J., et al. (2019). IMG/M v.5.0: an integrated data management and comparative analysis system for microbial genomes and microbiomes. *Nucleic Acids Res.* 47, D666–D677. doi: 10.1093/nar/gky901

Cooper, T. G., and Filmer, D. (1969). The active species of "CO_2" utilized by ribulose diphosphate carboxylase. *J. Biol. Chem.* 244, 1081–1083. doi: 10.1016/S0021-9258(18)91899-5

Cot, S. S. W., So, A. K. C., and Espie, G. S. (2008). A multiprotein bicarbonate dehydration complex essential to carboxysome function in cyanobacteria. *J. Bacteriol.* 190, 936–945. doi: 10.1128/JB.01283-07

Dearaujo, C., Arefeen, D., Tadesse, Y., Long, B. M., Price, G. D., Rowlett, R. S., et al. (2014). Identification and characterization of a carboxysomal g-carbonic anhydrase from the cyanobacerium *Nostoc* sp. PCC 7120. *Photosynth. Res.* 121, 135–150. doi: 10.1007/s11120-014-0018-4

Desmarais, J. J., Flamholz, A. I., Blikstad, C., Dugan, E. J., Laughlin, T. G., Oltrogge, L. M., et al. (2019). DABs are inorganic carbon pumps found throughout prokaryotic phyla. *Nat. Microbiol.* 4, 2204–2215. doi: 10.1038/s41564-019-0520-8

Dobrinski, K. P., Enkemann, S. A., Yoder, S. J., Haller, E., and Scott, K. M. (2012). Transcription response of the sulfur chemolithoautotroph *Thiomicrospira crunogena* to dissolved inorganic carbon limitation. *J. Bacteriol.* 194, 2074–2081. doi: 10.1128/JB.06504-11

Dobrinski, K. P., Longo, D. L., and Scott, K. M. (2005). A hydrothermal vent chemolithoautotroph with a carbon concentrating mechanism. *J. Bacteriol.* 187, 5761–5766. doi: 10.1128/JB.187.16.5761-5766.2005

Dou, Z., Heinhorst, S., Williams, E., Murin, E., Shively, J., and Cannon, G. (2008). CO_2 fixation kinetics of *Halothiobacillus neapolitanus* mutant carboxysomes lacking carbonic anhydrase suggest the shell acts as a diffusional barrier for CO_2. *J. Biol. Chem.* 283, 10377–10384. doi: 10.1074/jbc.M709285200

Edgar, R. C. (2004). MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* 32, 1792–1797. doi: 10.1093/nar/gkh340

Esparza, M., Jedlicki, E., Gonzalez, C., Dopson, M., and Holmes, D. S. (2019). Effect of CO_2 concentration on uptake and assimilation of inorganic carbon in the extreme acidophile *Acidithiobacillus ferrooxidans*. *Front. Microbiol.* 10:603. doi: 10.3389/fmicb.2019.00603

Heinhorst, S., Williams, E., Cai, F., Murin, D., Shively, J., and Cannon, G. (2006). Characterization of the carboxysomal carbonic anhydrase CsoCSA from *Halothiobacillus neapolitanus*. *J. Bacteriol.* 188, 8087–8094. doi: 10.1128/JB.00990-06

Horken, K., and Tabita, F. R. (1999). Closely related form I ribulose bisphosphate carboxylase/oxygenase molecules that possess different CO_2/O_2 substrate specificities. *Arch. Biochem. Biophys.* 361, 183–194. doi: 10.1006/abbi.1998.0979

Hurvich, C. M., and Tsai, C. L. (1989). Regression and time series model selection in small samples. *Biometrika* 76, 297–307. doi: 10.1093/biomet/76.2.297

Jones, D. T., Taylor, W. R., and Thornton, J. M. (1992). The rapid generation of mutation data matrices from protein sequences. *Comput. Appl. Biosci.* 8, 275–282. doi: 10.1093/bioinformatics/8.3.275

Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., et al. (2021). Highly accurate protein structure prediction with AlphaFold. *Nature* 596, 583–589. doi: 10.1038/s41586-021-03819-2

Kelly, D. P., and Wood, A. P. (2006). "The chemolithotrophic prokaryotes" in *The Prokaryotes*. ed. M. Dworkin. *3rd ed.* (New York: Springer), 441–456.

Kerfeld, C. A., Aussignargues, C., Zarzycki, J., Cai, F., and Sutter, M. (2018). Bacterial microcompartments. *Nat. Rev. Microbiol.* 16:277. doi: 10.1038/nrmicro.2018.10

Kerfeld, C. A., Heinhorst, S., and Cannon, G. C. (2010). Bacterial microcompartments. *Annu. Rev. Microbiol.* 64, 391–408. doi: 10.1146/annurev.micro.112408.134211

Kerfeld, C. A., and Melnicki, M. R. (2016). Assembly, function and evolution of cyanobacterial carboxysomes. *Curr. Opin. Plant Biol.* 31, 66–75. doi: 10.1016/j.pbi.2016.03.009

Kinney, J. N., Axen, S. D., and Kerfeld, C. A. (2011). Comparative analysis of carboxysome shell proteins. *Photosynth. Res.* 109, 21–32. doi: 10.1007/s11120-011-9624-6

Klein, M. G., Zwart, P., Bagby, S. C., Cai, F., Chisholm, S. W., Heinhorst, S., et al. (2009). Identification and structural analysis of a novel carboxysome shell protein with implications for metabolite transport. *J. Mol. Biol.* 392, 319–333. doi: 10.1016/j.jmb.2009.03.056

Kojima, H., and Fukui, M. (2010). *Sulfuricella denitrificans* gen. nov., sp. nov., a sulfur-oxidizing autotroph isolated from a freshwater lake. *Int. J. Syst. Evol. Microbiol.* 60, 2862–2866. doi: 10.1099/ijs.0.016980-0

Kojima, H., and Fukui, M. (2019). *Thiomicrorhabdus aquaedulcis* sp. nov., a sulfur-oxidizing bacterium isolated from lake water. *Int. J. Syst. Evol. Microbiol.* 69, 2849–2853. doi: 10.1099/ijsem.0.003567

Lanaras, T., Cook, C. M., Wood, A. P., Kelly, D. P., and Codd, G. A. (1991). Purification of ribulose 1,5-bisphosphate carboxylase/oxygenase from *Thiobacillus thyasiris* the putative symbiont of *Thyasira flexuosa* (Montagu). *Arch. Microbiol.* 156, 338–343. doi: 10.1007/BF00248707

Le, S. Q., and Gascuel, O. (2008). An improved general amino acid replacement matrix. *Mol. Biol. Evol.* 25, 1307–1320. doi: 10.1093/molbev/msn067

Liu, X., Chen, B., Lai, Q., Shao, Z., and Jiang, L. (2021). *Thiomicrorhabdus sediminis* sp. nov. and *Thiomicrorhabdus xiamenensis* sp. nov., novel sulfur-oxidizing bacteria isolated from coastal sediments and an emended description of the genus *Thiomicrorhabdus*. *Int. J. Syst. Evol. Microbiol.* 71:4660. doi: 10.1099/ijsem.0.004660

Long, B. M., Rae, B. D., Rolland, V., Förster, B., and Price, G. D. (2016). Cyanobacterial CO_2-concentrating mechanism components: function and prospects for plant metabolic engineering. *Curr. Opin. Plant Biol.* 31, 1–8. doi: 10.1016/j.pbi.2016.03.002

Mangiapia, M., Microbialphysiology, U., Brown, T.-R. W., Chaput, D., Haller, E., Harmer, T. L., et al. (2017). Proteomic and mutant analysis of the CO_2 concentrating mechanism of hydrothermal vent chemolithoautotroph *Thiomicrospira crunogena*. *J. Bacteriol.* 199, e00871–e00916. doi: 10.1128/JB.00871-16

Melnicki, M. R., Sutter, M., and Kerfeld, C. A. (2021). Evolutionary relationships among shell proteins of carboxysomes and metabolosomes. *Curr. Opin. Microbiol.* 63, 1–9. doi: 10.1016/j.mib.2021.05.011

Menon, B. B., Dou, Z., Heinhorst, S., Shively, J. M., and Cannon, G. C. (2008). *Halothiobacillus neapolitanus* carboxysomes sequester heterologous and chimeric RubisCO species. *PLoS One* 3:e3570. doi: 10.1371/journal.pone.0003570

Menon, B. B., Heinhorst, S., Shively, J. M., and Cannon, G. C. (2010). The carboxysome shell is permeable to protons. *J. Bacteriol.* 192, 5881–5886. doi: 10.1128/JB.00903-10

Moroney, J., Bartlett, S., and Samuelsson, G. (2001). Carbonic anhydrases in plants and algae. *Plant Cell Environ.* 24, 141–153. doi: 10.1111/j.1365-3040.2001.00669.x

Oltrogge, L., Chaijarasphong, T., Chen, A., Bolin, E., Marqusee, S., and Savage, D. (2020). Multivalent interactions between CsoS2 and Rubisco mediate α-carboxysome formation. *Nat. Struct. Mol. Biol.* 27, 281–287. doi: 10.1038/s41594-020-0387-7

Price, G. D., and Badger, M. R. (1989). Expression of human carbonic anhydrase in the cyanobacterium *Synechococcus* PCC7942 creates a high CO_2-requiring phenotype. *Plant Physiol.* 91, 505–513. doi: 10.1104/pp.91.2.505

Price, G. D., Badger, M. R., Woodger, F. J., and Long, B. M. (2009). Advances in understanding the cyanobacterial CO_2-concentrating-mechanism (CCM): functional components, Ci transporters, diversity, genetic regulation and prospects for engineering into plants. *J. Exp. Bot.* 59, 1441–1461. doi: 10.1093/jxb/erm112

Rae, B. D., Long, B. M., Badger, M. R., and Price, G. D. (2013). Functions, compositions, and evolution of the two types of carboxysomes: polyhedral microcompartments that facilitate CO_2 fixation in cyanobacteria and some proteobacteria. *Microbiol. Mol. Biol. Rev.* 77, 357–379. doi: 10.1128/MMBR.00061-12

Richter, M., and Rosselló-Móra, R. (2009). Shifting the genomic gold standard for the prokaryotic species definition. *Proc. Natl. Acad. Sci. U. S. A.* 106, 19126–19131. doi: 10.1073/pnas.0906412106

Roberts, E. W., Cai, F., Kerfeld, C. A., Cannon, G. C., and Heinhorst, S. (2012). Isolation and characterization of the *Prochlorococcus* carboxysome reveal the presence of the novel shell protein CsoS1D. *J. Bacteriol.* 194, 787–795. doi: 10.1128/JB.06444-11

Sawaya, M. R., Cannon, G. C., Heinhorst, S., Tanaka, S., Williams, E. B., Yeates, T. O., et al. (2006). The structure of beta-carbonic anhydrase from the carboxysomal shell reveals a distinct subclass with one active site for the price of two. *J. Biol. Chem.* 281, 7546–7555. doi: 10.1074/jbc.M510464200

Scott, K. M., Harmer, T. L., Gemmell, B. J., Kramer, A. M., Sutter, M., Kerfeld, C. A., et al. (2020). Ubiquity and functional uniformity in CO_2 concentrating mechanisms in multiple phyla of bacteria is suggested by a diversity and prevalence of genes encoding candidate dissolved inorganic carbon transporters. *FEMS Microbiol. Lett.* 367:fnaa106. doi: 10.1093/femsle/fnaa106

Scott, K. M., Leonard, J., Boden, R., Chaput, D., Dennison, C., Haller, E., et al. (2019). Diversity in CO_2 concentrating mechanisms among chemolithoautotrophs from the genera *Hydrogenovibrio*, *Thiomicrorhabdus*, and *Thiomicrospira*, ubiquitous in sulfidic habitats worldwide. *Appl. Environ. Microbiol.* 85, e02096–e03018. doi: 10.1128/AEM.02096-18

Scott, K. M., Sievert, S. M., Abril, F. N., Ball, L. A., Barrett, C. J., Blake, R. A., et al. (2006). The genome of deep-sea vent chemolithoautotroph *Thiomicrospira crunogena*. *PLoS Biol.* 4:e383. doi: 10.1371/journal.pbio.0040383

Shively, J. M., Decker, G. L., and Greenawalt, J. W. (1970). Comparative ultrastructure of the thiobacilli. *J. Bacteriol.* 101, 618–627. doi: 10.1128/jb.101.2.618-627.1970

So, A. K., Espie, G. S., Williams, E. B., Shively, J. M., Heinhorst, S., and Cannon, G. C. (2004). A novel evolutionary lineage of carbonic anhydrase (epsilon class) is a component of the carboxysome shell. *J. Bacteriol.* 186, 623–630. doi: 10.1128/JB.186.3.623-630.2004

So, A. K., Johnmckay, M., and Espie, G. S. (2002). Characterization of a mutant lackin carboxysomal carbonic anhydrase from the cyanobacterium *Synechocystis* PCC6803. *Planta* 213, 456–467. doi: 10.1007/s004250100638

Sorokin, D. Y., Gorlenko, V. M., Tourova, T. P., Tsapin, A. I., Nealson, K. H., and Kuenen, G. J. (2002a). *Thioalkalimicrobium cyclicum* sp. nov. and *Thioalkalivibrio jannaschii* sp. nov., novel species of haloalkaliphilic, obligately chemolithoautotrophic sulfur-oxidizing bacteria from hypersaline alkaline mono Lake (California). *Int. J. Syst. Evol. Microbiol.* 52, 913–920. doi: 10.1099/00207713-52-3-913

Sorokin, D. Y., Lysenko, A. M., Mityushina, L. L., Tourova, T. P., Jones, B. E., Rainey, F. A., et al. (2001). *Thioalkalimicrobium aerophilum* gen. nov., sp. nov. and *Thioalkalimicrobium sibericum* sp. nov., and *Thioalkalivibrio versutus* gen. nov., sp. nov., *Thioalkalivibrio nitratis* sp. nov., novel and *Thioalkalivibrio denitrificans* sp. nov., novel obligately alkaliphilic and obligately chemolithoautotrophic sulfur-oxidizing bacteria from soda lakes. *Int. J. Syst. Evol. Microbiol.* 51, 565–580. doi: 10.1099/00207713-51-2-565

Sorokin, D. Y., Tourova, T. P., Kolganova, T. V., Sjollema, K. A., and Kuenen, J. G. (2002b). *Thioalkalispira microaerophila* gen. nov., sp nov., a novel lithoautotrophic, sulfur-oxidizing bacterium from a soda lake. *Int. J. Syst. Evol. Microbiol.* 52, 2175–2182. doi: 10.1099/00207713-52-6-2175

Spreitzer, R. J. (2003). Role of the small subunit in ribulose-1,5-bisphosphate carboxylase/oxygenase. *Arch. Biochem. Biophys.* 414, 141–149. doi: 10.1016/S0003-9861(03)00171-1

Sun, Y., Wollman, A. J. M., Huang, F., Leake, M. C., and Liu, L. N. (2019). Single-organelle quantification reveals stoichiometric and structural variability of carboxysomes dependent on the environment. *Plant Cell* 31, 1648–1664. doi: 10.1105/tpc.18.00787

Sutter, M., Melnicki, M. R., Schulz, F., Woyke, T., and Kerfeld, C. A. (2021). A catalog of the diversity and ubiquity of bacterial microcompartments. *Nat. Commun.* 12:3809. doi: 10.1038/s41467-021-24126-4

Tabita, F. R. (1999). Microbial ribulose 1,5-bisphosphate carboxylase/oxygenase: a different perspective. *Photosynth. Res.* 60, 1–28. doi: 10.1023/A:1006211417981

Tabita, F. R., Hanson, T. E., Satagopan, S., Witte, B. H., and Kreel, N. E. (2008). Phylogenetic and evolutionary relationships of RubisCO and the RubisCO-like proteins and the functional lessons provided by diverse molecular forms. *Philos. Trans. R. Soc. Lond. Ser. B Biol. Sci.* 363, 2629–2640. doi: 10.1098/rstb.2008.0023

Takamiya, A., and Tubaki, K. (1956). A new form of Streptomyces capable of growing autotrophically. *Arch. Mikrobiol.* 25, 58–64. doi: 10.1007/BF00424890

Talavera, G., and Castresana, J. (2007). Improvement of phylogenies after removing divergent and ambiguously aligned blocks from protein sequence alignments. *Syst. Biol.* 56, 564–577. doi: 10.1080/10635150701472164

Tamura, K., Stecher, G., and Kumar, S. (2021). MEGA11: molecular evolutionary genetics analysis version 11. *Mol. Biol. Evol.* 38, 3022–3027. doi: 10.1093/molbev/msab120

Tanaka, S., Kerfeld, C. A., Sawaya, M. R., Cai, F., Heinhorst, S., Cannon, G. C., et al. (2008). Atomic-level models of the bacterial carboxysome shell. *Science* 319, 1083–1086. doi: 10.1126/science.1151458

Tsai, Y., Sawaya, M. R., Cannon, G. C., Cai, F., Williams, E. B., Heinhorst, S., et al. (2007). Structural analysis of CsoS1A and the protein shell of the *Halothiobacillus neapolitanus* carboxysome. *PLoS Biol.* 5:e144. doi: 10.1371/journal.pbio.0050144

Watson, S. W., Graham, L. B., Remsen, C. C., and Valois, F. W. (1971). Lobular, ammonia-oxidizing bacterium, *Nitrosolobus multiformis* nov. gen. nov. sp. *Arch. Mikrobiol.* 76, 183–203. doi: 10.1007/BF00409115

Watson, G. M. F., Yu, J.-P., and Tabita, F. R. (1999). Unusual ribulose 1,5-bisphosphate carboxylase/oxygenase of anoxic *Archaea*. *J. Bacteriol.* 181, 1569–1575. doi: 10.1128/JB.181.5.1569-1575.1999

Winogradsky, S. (1892). *Contributions à la morphologie des organismes de la nitrification*. St. Petersbourg: L'Institut Imperial de Medecine Experimentale.

Yoshizawa, Y., Toyoda, K., Arai, H., Ishii, M., and Igarashi, Y. (2004). CO_2-responsive expression and gene organization of three ribulose-1,5-bisphosphate carboxylase/oxygenase enzymes and carboxysomes in *Hydrogenovibrio marinus* strain MH-110. *J. Bacteriol.* 186, 5685–5691. doi: 10.1128/JB.186.17.5685-5691.2004

Zhao, Y.-Y., Jiang, Y.-L., Chen, Y., Zhou, C.-Z., and Li, Q. (2019). Crystal structure of pentameric shell protein CsoS4B of *Halothiobacillus neapolitanus* α-carboxysome. *Biochem. Biophys. Res. Commun.* 515, 510–515. doi: 10.1016/j.bbrc.2019.05.047

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

frontiers | Frontiers in Microbiology

REVIEW
published: 26 May 2022
doi: 10.3389/fmicb.2022.856547

Catching a Walker in the Act—DNA Partitioning by ParA Family of Proteins

Dipika Mishra[1,2]* *and Ramanujam Srinivasan*[1,2]*

[1] School of Biological Sciences, National Institute of Science Education and Research, Bhubaneswar, India, [2] Homi Bhabha National Institutes, Mumbai, India

Partitioning the replicated genetic material is a crucial process in the cell cycle program of any life form. In bacteria, many plasmids utilize cytoskeletal proteins that include ParM and TubZ, the ancestors of the eukaryotic actin and tubulin, respectively, to segregate the plasmids into the daughter cells. Another distinct class of cytoskeletal proteins, known as the Walker A type Cytoskeletal ATPases (WACA), is unique to Bacteria and Archaea. ParA, a WACA family protein, is involved in DNA partitioning and is more widespread. A centromere-like sequence *parS*, in the DNA is bound by ParB, an adaptor protein with CTPase activity to form the segregation complex. The ParA ATPase, interacts with the segregation complex and partitions the DNA into the daughter cells. Furthermore, the Walker A motif-containing ParA superfamily of proteins is associated with a diverse set of functions ranging from DNA segregation to cell division, cell polarity, chemotaxis cluster assembly, cellulose biosynthesis and carboxysome maintenance. Unifying principles underlying the varied range of cellular roles in which the ParA superfamily of proteins function are outlined. Here, we provide an overview of the recent findings on the structure and function of the ParB adaptor protein and review the current models and mechanisms by which the ParA family of proteins function in the partitioning of the replicated DNA into the newly born daughter cells.

Keywords: DNA segregation, plasmid, ParA, Walker A motif, ParB

OPEN ACCESS

Edited by:
Ulrike Kappler, The University of Queensland, Australia

Reviewed by:
Dagmara Jakimowicz, University of Wrocław, Poland
Ling Chin Hwang, Anglia Ruskin University, United Kingdom

***Correspondence:**
Dipika Mishra
dipika.mishra@niser.ac.in
Ramanujam Srinivasan
rsrini@niser.ac.in

Specialty section:
This article was submitted to Microbial Physiology and Metabolism, a section of the journal Frontiers in Microbiology

Received: *17 January 2022*
Accepted: *28 April 2022*
Published: *26 May 2022*

Citation:
Mishra D and Srinivasan R (2022) Catching a Walker in the Act—DNA Partitioning by ParA Family of Proteins. Front. Microbiol. 13:856547. doi: 10.3389/fmicb.2022.856547

INTRODUCTION

The genetic material in all organisms needs to be equipartitioned during each round of cell division. This mechanism has been very well-studied in eukaryotes. The initial study on eukaryotic chromosome segregation dates back to the later part of the nineteenth century with the discovery of thread-like structures within the nucleus of the stained newt cells. These thread-like structures, observed with a light microscope, were named chromatin (Flemming, 1882). Subsequently, the entire mechanism of eukaryotic chromosome segregation was characterized. It is now known to be carried out by spindle fibers, composed of microtubules that pull apart the chromosomes and assist in segregation (Scholey et al., 2003; Kline-Smith and Walczak, 2004). This mechanism is very well-coordinated and is programmed into four different phases of the cell cycle- S, G1, M, and G2 (Cooper, 2000; Walczak et al., 2010).

Unlike the eukaryotic genetic material, the prokaryotic DNA is not encased within a nuclear membrane. Instead, it spreads over the entire cytosol of bacteria and is called the nucleoid. The term "nucleoid" was first coined by Piekarski (1937). With the progression of the cell cycle, the nucleoid

changes its shape to a bilobed one and soon segregates into two daughter cells (Zimmerman, 2003; Yamaichi and Niki, 2004; Fisher et al., 2013). The genetic material in the eukaryotes is held together by histone and cohesion proteins (Losada and Hirano, 2005; Nasmyth and Haering, 2005). However, in the case of prokaryotes, the chromosomes are held together by DNA binding proteins called Nucleoid Associated Proteins (NAP) (Kar et al., 2005; Badrinarayanan et al., 2015) that help in chromosomal compaction and organization of domains known as the high-density regions (HDRs). These NAPs include HU, HNS, Fis, and IHF (Ali Azam et al., 1999; Verma et al., 2019). The nucleoid occupies a major proportion of the bacterial cytosol and plays an integral and decisive role in positioning the cytokinetic Z-ring (Harry et al., 1999; Yu and Margolin, 1999; Harry, 2001; Sun and Margolin, 2001; Bernhardt and de Boer, 2005; Rothfield et al., 2005) as well as driving the ParA mediated DNA partitioning (Castaing et al., 2008; Le Gall et al., 2016; McLeod et al., 2017).

DNA segregation in bacteria has been extensively studied using plasmids as a model system. Plasmids are extrachromosomal self-replicating species of DNA that usually encode genes for antibiotic resistance, production of bacteriocins, resistance to heavy metals, ultraviolet light, pathogen virulence factors and many other metabolic functions (Birge, 2006). Plasmids generally vary in size from a few kilobases to hundreds of kilobases, and their geometry is commonly circular or linear. Plasmids have been traditionally classified into different types based on their replication and copy numbers (Million-Weaver and Camps, 2014).

High copy number plasmids are generally small and replicate randomly during the cell cycle. These plasmids can reach up to a 100 copies per cell, and thus random assortment and segregation during cytokinesis can ensure sufficient distribution of these plasmids into two daughter cells (Birge, 2006; Million-Weaver and Camps, 2014). On the contrary, low copy number plasmids with <15 copies per cell cannot solely rely on random distribution for maintenance. Instead, they depend upon dedicated partitioning proteins to distribute them into daughter cells. Further, this is also true for single-copy number plasmids and includes many bacterial genomes. In contrast to the universal nature of the eukaryotic DNA segregation, microbes employ diverse mechanisms to partition their DNA. Active DNA partitioning systems have been broadly classified into Type-I, Type-II, and Type-III according to the protein family of the NTPase in the partitioning system (reviewed in Hayes and Barillà, 2006; Gerdes et al., 2010; Lutkenhaus, 2012). The DNA partitioning functions usually rely upon a member of the actin-like ATPase, tubulin-like GTPase, or Walker A Cytoskeletal ATPase (WACA) family of proteins. The WACA family of proteins is part of a larger group of P-loop ATPases with a deviant Walker A motif and is often referred to as the ParA superfamily (Walker et al., 1982; Motallebi-Veshareh et al., 1990; Koonin, 1993). The ParA superfamily of proteins is also involved in a wide range of diverse cellular functions in bacteria, including DNA segregation. Type-I systems utilize the ParA superfamily of proteins for DNA partitioning. They are found in plasmids such as the F plasmid (Fertility Plasmid that encodes the Fertility factor or the F factor) and in the chromosomes of many bacteria (Abeles et al., 1985; Davis et al., 1992, 1996; Koonin, 1993; Lutkenhaus, 2012). The *Salmonella paratyphi* R1 plasmid carries a Type-II system and contains an actin-like ATPase, ParM, that undergoes insertional polymerisation to push apart the plasmids to the two opposite ends of the cell (Gerdes et al., 2000; Møller-Jensen et al., 2003). The Type-III system is exemplified by the pBToxis plasmid, wherein the tubulin homolog TubZ segregates the two plasmids to the poles (Larsen et al., 2007). The Type-I system is the most widespread, with chromosomes in organisms ranging from archaea to bacterial pathogens and low copy number plasmids carrying virulence factors or multi-antibiotic resistance determinants relying upon the ParA family of proteins for partitioning functions (Motallebi-Veshareh et al., 1990; Koonin, 1993; Lutkenhaus, 2012; Stephens et al., 2020). The low copy number plasmids have been studied for a long time, and together with the recent findings in many bacterial species on chromosomal DNA partitioning, they provide us insights into the mechanism of DNA segregation in prokaryotes.

In this review, we provide a historical overview of the evolution of our understanding of DNA segregation in bacteria, with particular emphasis on the ParA protein from the F plasmid of *E. coli* ($ParA_F$) and highlight some of the recent findings. We briefly discuss the various types of plasmid segregation systems (Type-I–Type-III), the functions of ParA and the mechanism by which ParA orchestrates DNA partitioning. The recent findings on the CTPase activity of the centromere binding protein, ParB and its bearings on DNA segregation functions are discussed. We list a few examples of the ParA family of proteins involved in bacterial and archaeal chromosome segregation. Finally, we touch upon the functions of the ParA superfamily of proteins in other diverse cellular processes in bacteria.

TYPES OF PLASMID SEGREGATION SYSTEMS

Type-I—Walker A-Type ATPases in DNA Partitioning

Most bacterial chromosomes and many low-copy number plasmids depend on the Type-I system for the segregation of DNA. The Type-I system is characterized by the presence of a P-loop ATPase superfamily of protein, known as the ParA/MinD family of ATPases, that have a deviant Walker A box motif KGGXXK[S/T] (Koonin, 1993; Lutkenhaus, 2012). The genetic loci important for partitioning also carry a centromeric repeat sequence, *parS*, and encode a second protein, the centromere binding protein (CBP) or adaptor protein ParB that binds *parS* sites and interacts with ParA as well. The Walker A motif in ParA is directly involved in interactions with the bound ATP molecule. Walker A motifs in the Type-I ParA family differ from the classical Walker A motif in having this additional signature lysine residue (**Table 1**), and thus the term deviant Walker A motif follows (Motallebi-Veshareh et al., 1990; Hayes, 2000; Lutkenhaus and Sundaramoorthy, 2003; Wendler et al., 2012). A second motif, the B box, characterized by conserved negatively charged residue (D/E), plays an important role in Mg^{2+} coordination and ATP hydrolysis (Fung et al., 2001;

TABLE 1 | The deviant Walker A motif in different members of P loop ATPase.

Protein	Deviant Walker A motif (KGGXXK[S/T])
$ParA_F$ (SopA)	KGGVYKT
$ParA_{Bsu}$ & $ParA_{Hpy}$ [Soj (*B. subtilis* & *H. pylori*)]	KGGVGKT
MinD	KGGVGKT
FleN	KGGVGKT
PpfA	KGGVGKT
FlhG	KGGVGKS
MipZ	KGGAGKS
ParI	KGGVGKS
McdA	SGGQGKT
BcsQ	RGGVGTT

The signature lysine is marked in red and the catalytic lysine is marked in blue.

Schumacher, 2008). Mutations in the conserved Walker A box lead to the loss of plasmids and point to a key role for ATP hydrolysis in mediating segregation (Ebersbach and Gerdes, 2001; Fung et al., 2001; Libante et al., 2001; Barillà et al., 2005; Pratto et al., 2008). The Type-I system is further sub-divided into Type-Ia and Type-Ib based on the structure of ParA (Hayes, 2000; Schumacher, 2008).

Type-Ia/large ParA proteins (∼300–450 amino acids) have an extended N-terminal helix-turn-helix (HTH) motif and can act as repressors of their gene expression (**Figures 1A,B**) (Abeles et al., 1985; Hirano et al., 1998; Libante et al., 2001). The HTH domains help in this sequence-specific DNA binding to operator regions near their promoters (Mori et al., 1989; Davis et al., 1992; Hayes et al., 1994; Ravin et al., 2003). Type-Ia is one of the first identified DNA segregation systems and is typically encoded by partitioning loci carried in plasmids (Austin and Abeles, 1983; Ogura and Hiraga, 1983; Mori et al., 1986). Specific examples of Type-Ia include the *parABS* and the *sopABC* (*parABS*$_F$) system of P1 bacteriophage and F plasmid, respectively. In the case of F plasmid (the one encoding for Fertility factor), the genes appear in the order of *sopA*, *sopB*, and *sopC*, hereafter referred to as *parA*$_F$, *parB*$_F$, and *parS*$_F$, respectively. The expression of $ParA_F$ and $ParB_F$ are driven by a single promoter located upstream of *parA*$_F$ (Mori et al., 1986, 1989; Hirano et al., 1998). The *parABS*$_F$ cluster serves as the partitioning system, wherein *parS*$_F$ serves as the centromeric sequence and comprises twelve 43-bp base pair repeats (**Figure 1A**). Each 43-bp sequence contains a short 16-bp inverted repeat to which $ParB_F$ binds as a dimer (Hayakawa et al., 1985; Lane et al., 1987; Mori et al., 1989; Hanai et al., 1996). This $ParB_F$-*parS*$_F$ complex is then recruited to $ParA_F$, which is the NTPase and functions as the motor protein (Ogura and Hiraga, 1983).

On the contrary, members of the Type-Ib subfamily are smaller (∼200–250 amino acids) and lack the N-terminal HTH domain. They are thus referred to as smaller ParAs and are found in the chromosomal loci of many bacteria. However, they are also found in a few plasmids, like in TP228 of *Salmonella newport*. The ParA ATPases in the Type-Ib systems do not have any repressor functions due to the lack of the N-terminal HTH domain required for site-specific DNA binding at their promoter regions (**Figure 1A**). Examples of the Type-Ib system on plasmids include *parFGH* loci on the *S. newport* plasmid TP228 and the *parABS* loci (δ/*parA*, ω/*parB* and *parS*) in the plasmid pSM19035 carried by *Streptococcus pyrogenes* (de la Hoz et al., 2004; Fothergill et al., 2005; Pratto et al., 2008). Type-Ib also includes ParA from *Caulobacter crescentus* (Mohl and Gober, 1997) and Soj from *Bacillus subtilis* ($ParA_{Bsu}$/BsSoj) (Leonard et al., 2005a; Lee and Grossman, 2006) and *Helicobacter pylori* ($ParA_{Hpy}$/HpSoj) (Lee et al., 2006) encoded by bacterial genomes. ParB members of the Type-Ib family carry an N-terminal protein-protein interaction domain, central HTH domain and a self-dimerization domain at the C-terminus. Further, residues near the N-terminus of ParB specify their interactions with the cognate ParA (Funnell, 2016; Kawalek et al., 2020). Similar to the Type-1a system, ParB binds to *parS* and interacts with ParA, thus linking the plasmid to the motor protein ParA.

The ParA proteins, both the larger Type-1a and the smaller Type-Ib, also bind DNA in a non-sequence specific manner and is critical for their DNA segregation functions (Ebersbach and Gerdes, 2005; Leonard et al., 2005a; Hatano et al., 2007; Hester and Lutkenhaus, 2007; Castaing et al., 2008; Vecchiarelli et al., 2010; Roberts et al., 2012; Le Gall et al., 2016). The currently accepted models for DNA segregation by ParA, i.e., the diffusion-ratchet, DNA-relay and DNA hitch-hiking mechanisms (described in detail below), emphasize the importance of bacterial nucleoids. Briefly, the DNA hitch-hiking model proposes that the ParA-ATP dimers localize on the HDRs of the nucleoid to which the ParB-*parS* complex binds and stimulates the ATPase activity of ParA. The hydrolysis event converts the nucleoid bound ParA-ATP into ParA-ADP, resulting in the release of ParA from the nucleoid. This creates a zone of depletion of DNA bound ParA, and the released ParB-*parS* complex undergoes diffusive motion until captured by ParA-ATP dimers within the HDRs, resulting in a directional motion toward the highest concentration of DNA bound ParA-ATP in the cell (Vecchiarelli et al., 2012; Le Gall et al., 2016; McLeod et al., 2017). The bacterial nucleoid thus forms a key substrate for the ParA in its function as a motor protein in the equipartitioning of plasmids and chromosomes into daughter cells (Vecchiarelli et al., 2012).

Type-II/Actin-Like ATPases

Type-II partitioning system was first discovered in the resistance plasmid R1 (Jensen and Gerdes, 1997). The partitioning system in this group contains an actin-like ATPase called ParM, an adaptor protein, ParR and a centromere site, *parC*. Monomeric ParM has an actin-like fold, with its crystal structures bearing a close resemblance to the eukaryotic actin (Bork et al., 1992; van den Ent et al., 2002). A search of the microbial genomes revealed as many as 40 different families of actin-like proteins, mostly found on plasmids (Derman et al., 2009). A few have been shown to assemble into polymeric structures and filaments with varying architectures (Becker et al., 2006; Derman et al., 2009; Popp et al., 2012; Jiang et al., 2016; Koh et al., 2019). However, the ParMRC system is the most extensively studied among the DNA partitioning systems. ParM, like actin, assembles

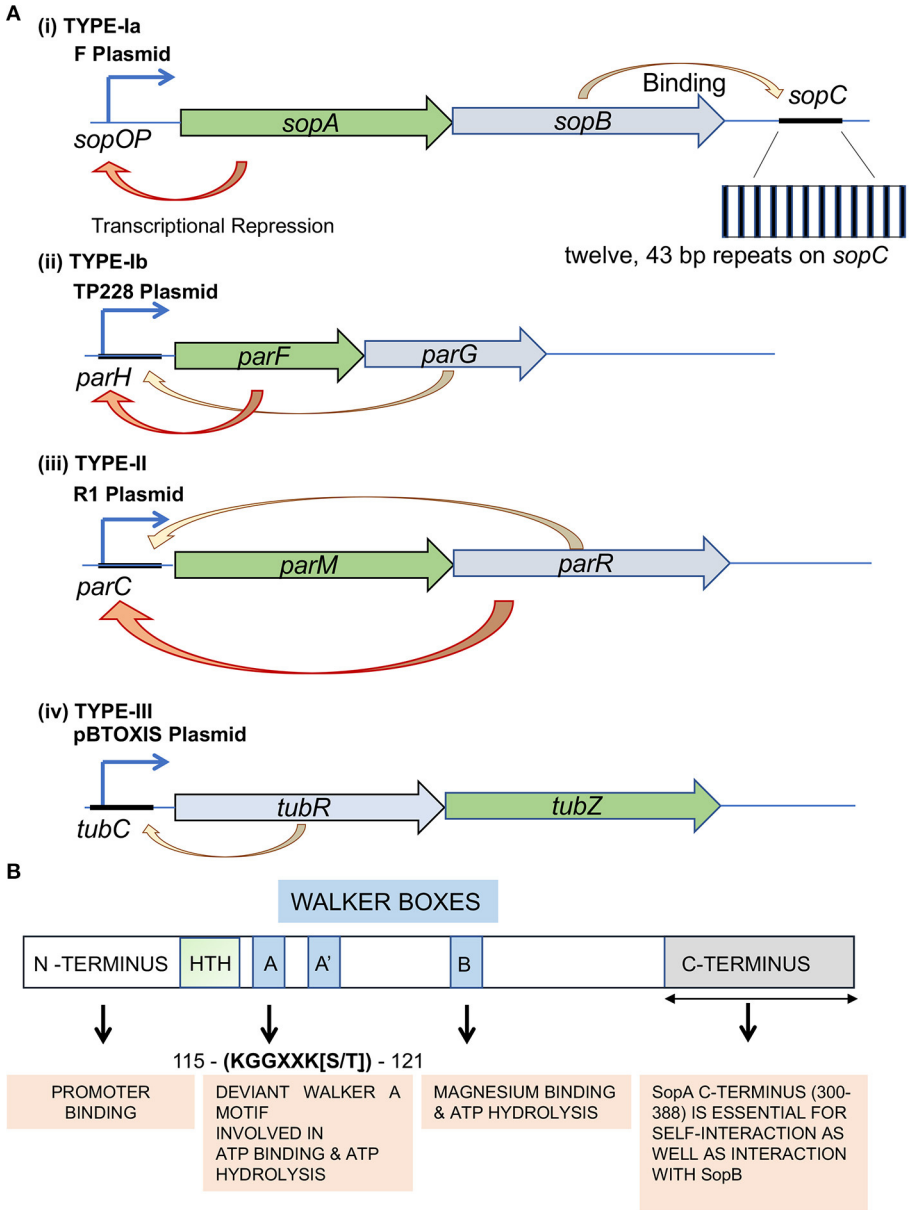


FIGURE 1 | Types of plasmid partitioning systems and domain organization in the ParA family of proteins Type-I system. **(A)** Genetic organization of Type-Ia, Type-Ib, Type-II, and Type-III partition loci. Genes encoding the ATPase (green) and adaptor protein (blue) are indicated. The auto-repression activity is represented by an orange arrow. F plasmid, TP228 plasmid, R plasmid and pBToxis plasmid are used as representatives of (i) Type-Ia, (ii) Type-Ib, (iii) Type-II and (iv) Type-III segregation mechanisms, respectively. The *sopA* (green box) and *sopB* (blue box) genes are expressed from the P*sop* promoter. The partition site *sopC* is located downstream of the *sopAB* genes. The *sopC* gene sequence is represented by a black line. The inset shows the repetitive stretch of twelve 43 bp repeats. The respective NTPases in Type-Ib (ParF), Type-II (ParM) and Type-III (TubZ) are shaded in green, and their corresponding centromere binding proteins (CBPs; ParG, ParR and TubR) are shaded in blue. The orange arrows represent transcriptional repression, whereas the yellow arrows point to centromere binding by the CBPs. **(B)** The conserved domains of F plasmid partitioning $ParA_F$ protein. The N-terminal (shaded pale green), C-terminal (shaded gray) and the Walker A motifs of $ParA_F$ have been represented. The Walker motifs have been highlighted in blue, and the function of each domain has been represented in orange boxes below. The amino acid residue numbers correspond to $ParA_F$.

into a two-stranded helix but with an opposite twist. While actin forms a right-handed filament, ParM helices are mainly left-handed, with 12 subunits per turn as opposed to 13 monomers in actin (Orlova et al., 2007; Popp et al., 2008; Gayathri et al., 2012). Interestingly, in the ParR (C-terminal 17 residue peptide) bound form, domains IA and IIB show a rotation of 9.1° and 10.7° toward the nucleotide, reminiscent of the shift in actin structure from G-actin to F-actin. Further, the ParR peptide binding site overlaps with the ParM polymerisation interface and thus allows ParR to bind only to the barbed end or the

polymerizing end (Gayathri et al., 2012). Also, ParM grows bidirectionally with similar monomer addition rates at both ends and exhibits dynamic instability, a characteristic feature of microtubules (Gerdes et al., 2000; Garner et al., 2004).

Moreover, in the presence of a non-hydrolysable analog of ATP, these *in vitro* filaments were longer suggesting that the growth and retraction of filaments were driven by ATP hydrolysis (Garner et al., 2007). Although ParM can bind both ATP and GTP, the most preferred substrate is ATP, for which ParM has a 10-fold higher affinity than GTP (Popp et al., 2008; Galkin et al., 2009). Cryo-electron microscopy of cells overexpressing ParM has shown that the protein assembles into closely packed filament bundles with an average of 3–5 ParM filaments per bundle (Salje et al., 2009). Further studies suggest that the filaments grow by insertional polymerisation, wherein new ParM subunits are inserted at the interface of the filaments with the ParR-*parC* complex (Møller-Jensen et al., 2003; Garner et al., 2004, 2007; Salje et al., 2010; Gayathri et al., 2012). *In vitro* reconstitution of the ParMRC segregation machinery using polystyrene beads for immobilizing *parC*, purified ParR and ParM proteins, and ATP resulted in the assembly of dynamic filaments of ParM. Surprisingly, the ParM filaments were seen to grow and shrink in a manner that was reminiscent of the dynamic instability of microtubules (Garner et al., 2004, 2007). However, in the presence of another *parC* coated bead in proximity, the filaments appeared to bundle and push the beads further away, consistent with the bipolar antiparallel filament structure (Gayathri et al., 2012, 2013). Thus, ParMRC is a minimalistic tripartite system that is sufficient for plasmid segregation and does not require any additional host factors (Garner et al., 2007). Based on cryo-electron microscopy, *in vitro* reconstitution experiments, and fluorescence imaging, a "search, capture, and push" model for plasmid segregation by ParM and other actin-like proteins has been proposed (Campbell and Mullins, 2007; Garner et al., 2007; Salje et al., 2009, 2010; Gerdes et al., 2010; Gayathri et al., 2012).

Type-III/Tubulin-Like GTPases

Partitioning machinery of this kind has been described in the pXO1 plasmid of *B. anthracis* and *B. cereus* (Tinsley and Khan, 2006; Hoshino and Hayashi, 2012), pBToxis plasmid of *B. thuringiensis* (Tang et al., 2006; Larsen et al., 2007) and more recently in bacteriophages c-st of *Clostridium botulinum* (Oliva et al., 2012) and 201φ2-1 of Pseudomonas chlororaphis (Kraemer et al., 2012). The segregation machinery comprises three components (*tubZRC*) similar to those found in Type-I and Type-II systems. However, unlike the others, the motor NTPase TubZ or PhuZ (for Phage Tubulin/FtsZ) belongs to the tubulin/FtsZ superfamily of cytoskeletal proteins (Tinsley and Khan, 2006; Larsen et al., 2007; Anand et al., 2008; Oliva et al., 2012). While, in general, the sequence similarity with eukaryotic tubulin amounts to even <15%, TubZ exhibits striking structural similarity with that of the bacterial cell division protein FtsZ and tubulin (Löwe and Amos, 1998; Nogales et al., 1998; Larsen et al., 2007; Aylett et al., 2010). TubZ undergoes polymerisation upon binding GTP (Anand et al., 2008; Chen and Erickson, 2008; Hoshino and Hayashi, 2012) and assembles into two or four-stranded filaments (Aylett et al., 2010; Montabana and Agard, 2014). Interestingly, the subunit interactions upon polymerisation leading to the coupling of GTP hydrolysis are closer to that of α/β-tubulin. Further, the protofilaments in the presence of GTPγS show a right-handed twist with 14 subunits over one complete turn (360°) and assemble into a parallel double-helical filament when expressed in *E. coli* as well (Aylett et al., 2010). The filaments undergo treadmilling, i.e., it undergoes directional polymerisation, where one end exhibits growth and the other shrinkage, which assists in DNA partitioning (Larsen et al., 2007; Aylett et al., 2010). The Type-III system is thus an example of the involvement of tubulin-like proteins in the DNA segregation in bacteria. TubR constitutes the CBP in this Type-III system that binds to the centromeric sequence *tubC* and recruits TubZ (Tang et al., 2007; Ni et al., 2010). The structure of the TubR-*tubC* complex shows that TubR binds *tubC* through an N-terminal winged helix-turn-helix motif and assembles into a DNA-nucleoprotein complex *in vitro,* forming a slight right-handed superhelix (Ni et al., 2010; Aylett and Löwe, 2012). Further, the TubR-*tubC* complex stabilizes the TubZ polymers and possibly exerts its effect by preventing the depolymerisation of TubZ filaments (Aylett and Löwe, 2012; Oliva et al., 2012). Reconstitution experiments reveal that the TubR-*tubC* complex tracks the depolymerizing minus-end of the TubZ filaments (Fink and Löwe, 2015). The bacteriophage tubulin-like protein PhuZ ($TubZ_{\Phi KZ}$), on the other hand, resembles the eukaryotic microtubule in many aspects. It was the first prokaryotic tubulin that was shown to exhibit dynamic instability, a property dependent upon the energy derived from GTP hydrolysis (Kraemer et al., 2012; Aylett et al., 2013; Erb et al., 2014). However, unlike tubulin, it assembles into a triple-stranded helical filament (Zehr et al., 2014). One end of the PhuZ filaments is anchored to the cell poles, and they further assemble into bipolar spindle-like structures. These bipolar spindles of PhuZ resemble eukaryotic microtubules during mitosis and help place the viral nuclei at mid-cell and control viral reproduction (Erb et al., 2014).

THE *sopABC* OR *parABS* LOCUS IN TYPE-I PLASMID SEGREGATION SYSTEMS

ParA—A Walker a Cytoskeletal ATPase

ParA protein function in Type-I partitioning systems is essential for DNA segregation of the plasmids and the bacterial chromosome, as described above. ParA belongs to the Walker A type Cytoskeletal ATPases (WACA) family of proteins and contains a Walker A motif, a Walker A' motif, a Walker B motif and a ParA specific sequence (Walker et al., 1982; Motallebi-Veshareh et al., 1990; Koonin, 1993; Lutkenhaus, 2012) (**Figure 1B**). As described above, the larger ParA found in Type-Ia systems has an auto-regulatory function and thus directly regulates the transcription of the *parAB* operon from its promoter (P_{par}) and control levels of both ParA and ParB proteins (Mori et al., 1989; Davis et al., 1992; Hayes et al., 1994; Davey and Funnell, 1997; Hirano et al., 1998; Ravin et al., 2003; Komai et al., 2011). ParA exists in a monomer-dimer equilibrium in the cell wherein only the ATP-bound ParA dimers associate

with the nucleoid, and ADP-bound dimers (and monomers) are free (Vecchiarelli et al., 2010, 2013a; Havey et al., 2012). The conformational change involves nucleotide-binding wherein the ATP bound form is dimeric (Schumacher et al., 2012; Vecchiarelli et al., 2013a). ParA exhibits a weak ATPase activity similar to the other members of the WACA superfamily (Davis et al., 1992; Watanabe et al., 1992; Libante et al., 2001; Leonard et al., 2005a; Barillà et al., 2007; Havey et al., 2012) and ATP hydrolysis activity can be stimulated 3-fold by ParB, and 1.5 fold by DNA. However, the DNA-ParB complex can exert a much stronger effect and stimulate the ATPase activity of ParA by 10–15-fold (Davis et al., 1992; Watanabe et al., 1992; Bouet et al., 2007; Ah-Seng et al., 2009). Moreover, mutants defective in ATPase activity, like $ParA_{P1}$ K122Q and $ParF_{TP228}$ D111A, or the mutants displaying hyper-ATPase activity, $ParF_{TP228}$ P104A, R169A, and G179A, are all impaired in plasmid stability. These studies highlight the crucial role played by ATP hydrolysis of ParA in plasmid segregation (Fung et al., 2001; Libante et al., 2001; Dobruk-Serkowska et al., 2012; Vecchiarelli et al., 2013a; McLeod et al., 2017; Caccamo et al., 2020). Further, ParA-ATP dimers themselves have been postulated to exist in two different states/conformations, an active conformation (ParA-ATP*) that binds to non-specific DNA and an inactive state (ParA-ATP) that cannot bind DNA. Thus, it is the active state ATP-bound dimer $(ParA\text{-}ATP^*)_2$ that associates with the bacterial nucleoid (Vecchiarelli et al., 2010, 2013a). The ParA-ADP dimers produced soon after ATP hydrolysis can no longer associate with the bacterial nucleoid. ParA can rebind DNA only upon nucleotide exchange with ATP or reassociation with ATP and the conformational change to the ParA-ATP* state. However, ParA-ADP can act as a transcriptional repressor of the ParA promoter (Davey and Funnell, 1997; Bouet and Funnell, 1999; Libante et al., 2001; Hao and Yarmolinsky, 2002; Baxter et al., 2020).

ParA has a weak auto-repression activity, and its full repressor function depends on the co-repressor ParB, together with which it strongly represses transcription of its promoter P_{par} (Friedman and Austin, 1988; Hayes et al., 1994; Libante et al., 2001). Using surface-plasmon-resonance, Bouet and group have revealed that three $ParA_F$ dimers, i.e., a trimer of dimers, bind to the promoter region to suppress gene expression (Boudsocq et al., 2021). The ParB-*parS* (SopA-*sopC*) complex further enhances this auto-regulatory function. It was initially thought that only ParA-ATP and ParA-ADP, but not ParA-ATP* states, were competent in binding to the operator sites *parOP* at the promoter (Davey and Funnell, 1997; Fung et al., 2001; Ravin et al., 2003; Vecchiarelli et al., 2013a). However, recent studies using the non-specific DNA binding mutant $ParA_{P1}$ R351A suggest that ParA-ATP* state is also proficient in binding *parOP*. The abrogated ns-DNA binding results in a free pool of excess ParA-ATP* dimers, resulting in repression of transcription from *parOP*. This auto-repression activity of ParA thus can solely be attributed to its specific DNA binding activity mediated by the HTH domain of the protein (Baxter et al., 2020).

The active state ATP-bound dimeric conformation $(ParA\text{-}ATP^*)_2$, enables the binding of ParA molecules to the nucleoid (Vecchiarelli et al., 2010, 2013a). ParA has an affinity for DNA in a non-sequence-specific manner in this conformation. Since bacterial cells have a lot of nsDNA in the form of the nucleoid, ParA molecules are found localized to the nucleoid. Work from laboratories around the world on several ParA family proteins has shown that ParA is predominantly nucleoid bound, and its nucleoid binding function is essential for plasmid maintenance (Leonard et al., 2005a; Hayes and Barillà, 2006; Hatano et al., 2007; Hester and Lutkenhaus, 2007; Castaing et al., 2008; Vecchiarelli et al., 2010; Roberts et al., 2012; Lim et al., 2014; Volante and Alonso, 2015; Le Gall et al., 2016; McLeod et al., 2017; Caccamo et al., 2020). The interaction of ParA with nsDNA has been probed *in vitro* by several studies and visualized *in vivo* by using sophisticated fluorescence microscopy techniques (Lim et al., 2005, 2014; Hatano et al., 2007; Castaing et al., 2008; Hatano and Niki, 2010; Roberts et al., 2012; Le Gall et al., 2016; McLeod et al., 2017). Further, nsDNA binding defective mutants, $ParA_F$ K340A and $ParA_{P1}$ R351A have segregation defects suggestive of the critical role of ns-DNA binding in the process of plasmid maintenance (Castaing et al., 2008; Baxter et al., 2020). In addition, fluorescence microscopy and time-lapse imaging of $ParA_F$ have suggested that the protein assembles into helical polymeric structures that undergo oscillatory behavior on the nucleoid. These observations had earlier led to suggestions that the polymerisation dynamics of ParA drove plasmid segregation (Barillà et al., 2005; Lim et al., 2005; Hatano et al., 2007; Ringgaard et al., 2009). However, as described in detail below, more recent studies on ParA dynamics *in vivo* and *in vitro* biochemical assays have favored a polymerisation-independent mechanism of DNA partitioning by ParA.

ParB—Centromere Binding Protein or the Adaptor Protein

ParB is a DNA binding protein that binds to specific repeat sequences found in plasmids or genomes (*parS*) and forms an active component of the bacterial DNA segregation machinery. The role of chromosomally encoded ParB in various cellular functions and DNA partitioning has been recently reviewed in detail by Kawalek et al. (2020). The crystal structure of the full-length ParB protein from *B. subtilis* (21–218 aa), also known as SpoOJ, provides details about its domain organization (Soh et al., 2019). ParB comprises of different domains: the N-terminal domain, a central DNA binding HTH motif and a C-terminal domain, all connected by flexible linkers (Funnell, 1991, 2016; Schumacher and Funnell, 2005; Badrinarayanan et al., 2015; Soh et al., 2019) (**Figure 2A**). The C-terminal domain plays a pivotal role in the homo-dimerization of ParB (Khare et al., 2004; Leonard et al., 2005a). Mutation R149G within the HTH motif affects *parS* binding (Autret et al., 2001; Gruber and Errington, 2009; Fisher et al., 2017) and the central DNA binding domain thus is critical for the recognition and binding of ParB to *parS* sites (Leonard et al., 2005a; Schumacher and Funnell, 2005). The DNA binding domain or HTH motif plays a vital role in specific DNA interaction with *parS* sites and spreading to the adjacent DNA after binding to the *parS* sites. Spreading is a

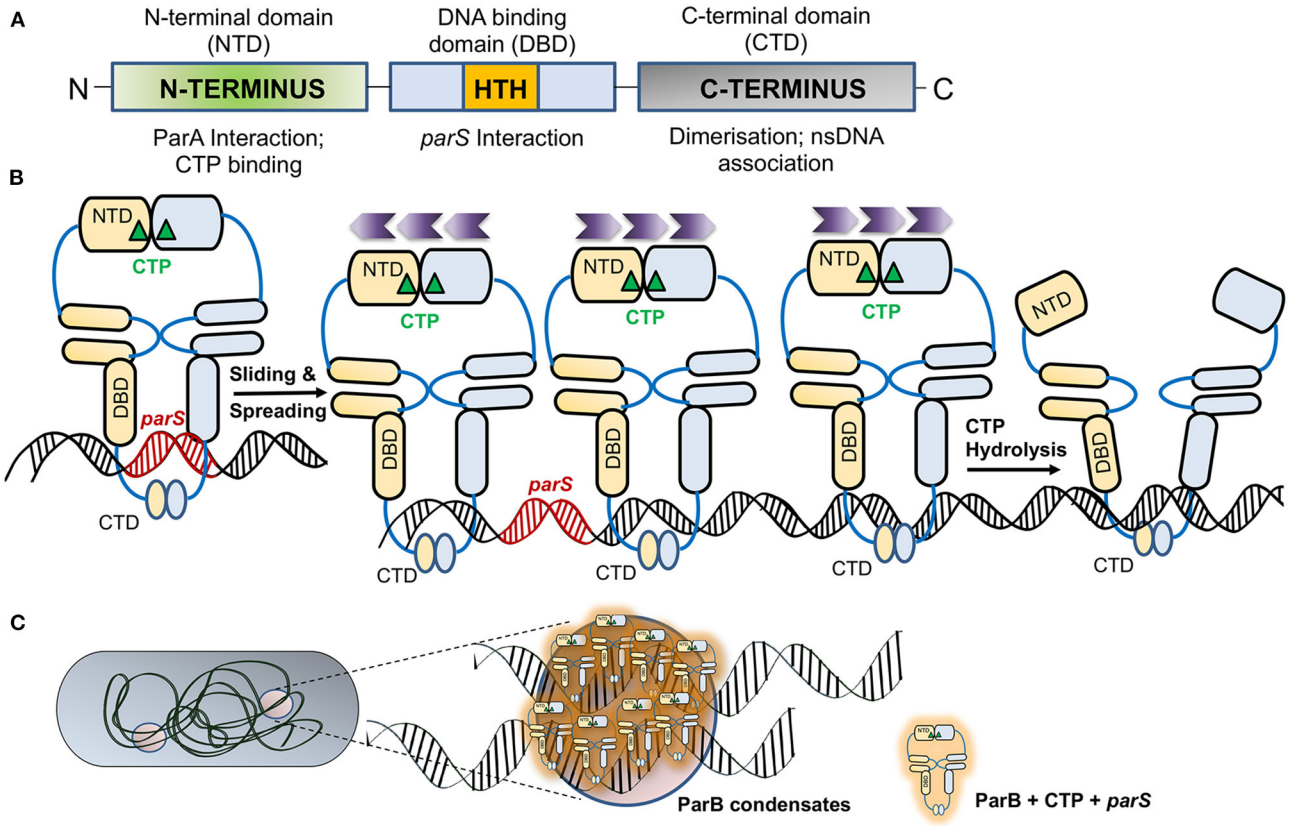


FIGURE 2 | ParB—an adaptor protein and regulation of its activity by CTP. **(A)** The domain organization of ParB. The schematic representation of the ParB domain organization indicates the NTD (N-terminal domain) marked by a green box, the central DNA binding domain (DBD; marked by a blue box) contains a helix-turn-helix (HTH) motif that allows ParB nucleation on *parS* sites and is represented by an orange box. The C-terminal domain (CTD) responsible for dimerization is shaded gray. The function of each domain is shown below. **(B)** Association of ParB with CTP. Upon association of apo-ParB NTD with CTP (marked by a green triangle), the ParB molecules form homodimers and associate with *parS* (indicated in red). The ParB-CTP dimers clamped to the DNA further undergo sliding and spread across the *parS* proximal sites. Sliding and spreading of ParB across DNA may not require CTP hydrolysis. The direction of spreading is indicated by the purple chevrons ( ). However, upon CTP hydrolysis, the NTDs of two ParB are disassociated, leading to a switch from dimer to apo-form, leading to the release of ParB from the DNA. **(C)** Liquid-Liquid Phase Separated (LLPS) condensates formed by ParB. The schematic depicts a bacterial cell with ParB condensates, and the zoomed inset shows several ParB-CTP dimers that associate *parS* sites containing DNA to form the condensates.

crucial feature of ParB, and it involves the formation of higher-ordered complexes. ParB is known to initiate binding to DNA at *parS* sites and spread over *parS* flanking regions, often covering a large span of the nsDNA (reviewed in Jalal and Le, 2020; Kawalek et al., 2020). The N-terminal stretch is necessary for protein oligomerisation as well as interaction with ParA. It is defined by a conserved stretch of Arginine residues GERRxRA, referred to as the arginine patch. Mutations within the arginine patch impair ParB spreading as well as nucleoid segregation functions (Yamaichi and Niki, 2000; Ah-Seng et al., 2013; Chen et al., 2015). This patch is thus essential for the spreading of ParB, foci formation and DNA partitioning (Rodionov et al., 1999; Autret et al., 2001; Bartosik et al., 2004; Breier and Grossman, 2007; Kusiak et al., 2011; Graham et al., 2014; Funnell, 2016). Further, ParB is thought to stimulate the ATPase activity of ParA *via* an arginine finger (R-finger) motif contained within the N-terminal region (Ah-Seng et al., 2009; Zhang and Schumacher, 2017).

ParB Activity Is Regulated by CTP Binding and Hydrolysis

Recent structural and biochemical studies on ParB have revealed that ParB is a CTPase with the arginine patch forming the catalytic center (Soh et al., 2019; Jalal et al., 2020; Osorio-Valeriano et al., 2021). This surprising finding was revealed by crystal structures of ParB of *B. subtilis* and PadC, a ParB member from *M. xanthus,* which were found to be associated with CDP and CTP, respectively (Soh et al., 2019; Osorio-Valeriano et al., 2021). The revelation of these ParB structures with CTP as a cofactor has opened up new avenues of research and questions on the role of CTP in regulating the process of ParB spreading and DNA partitioning (Jalal et al., 2020, 2021). CTP binding results in dimerization of the N-terminal domains of two ParBs in the presence of Mg^{2+}, which acts as a cofactor in the process. Such self-dimerization is sufficient to cause the formation of a clamp that entraps the DNA (Soh et al., 2019; Osorio-Valeriano et al.,

2021). Recent studies using optical tweezers and microscopy have shown that binding of CTP or the non-hydrolysable analog CTPγS enhances ParB binding to *parS* sites and results in DNA condensation (Balaguer et al., 2021). A conformational change mediates ParB dimerization upon binding to CTP and *parS*. Upon CTP hydrolysis, CDP has a low affinity for ParB, and thus, it dissociates from the NBD of ParB and results in a switch to its apo-form (**Figure 2B**). Moreover, studies have shown that ParB networks show a continuous exchange and turnover of ParB at both protein-protein and protein-DNA interfaces (Madariaga-Marcos et al., 2019). Consistently, CTP hydrolysis defective mutants, Q52A and E93A in *M. xanthus* ParB, exhibit abnormal chromosome segregation (Osorio-Valeriano et al., 2021). Further, ParB spreading to distal regions from *parS* sites could be prevented by the presence of protein roadblocks, suggesting that ParB spreading on nsDNA was one-dimensional. CTP binding thus plays an essential role in ParB spreading and DNA condensation (Balaguer et al., 2021).

ParB has been recently shown to undergo Liquid-liquid phase separation (LLPS), favoring condensate formation in the presence of CTP (Babl et al., 2022). Liquid-liquid phase separation is a thermodynamic process wherein a homogeneous mixture of two or more molecules de-mixes into distinct phases and thus helps compartmentalize membrane-less organelles (Hyman et al., 2014; Guilhas et al., 2020; Babl et al., 2022). In most eukaryotic subcellular structures, like germline P-bodies (Brangwynne et al., 2009), stress granules (Protter and Parker, 2016) etc., compartmentalization depends upon LLPS. In bacteria, LLPS has been implicated in the assembly of carboxysomes and divisome protein FtsZ (Monterroso et al., 2019; Wang et al., 2019; MacCready et al., 2020). Recently, ParB has also been shown to exist in two different phases, a gas phase and a liquid phase. Further, in a process similar to the LLPS in eukaryotic proteins, ParB from *Corynebacterium glutamicum* also forms spherical assemblies in the presence of crowding agents, is affected by alterations in ionic strength and is stabilized by potassium glutamate (Babl et al., 2022). The condensation of ParB is further favored by the presence of *parS* containing DNA and the motor protein ParA helps in the separation of these condensates (**Figure 2C**). Moreover, in the case of the $ParA_F$ K120Q mutant, the condensates merge and do not segregate suggesting that the ParB stimulated ATPase activity of ParA is required to separate these condensates (Guilhas et al., 2020; Babl et al., 2022).

THE MECHANISM AND MODELS OF DNA PARTITIONING BY ParA

François Jacob put forth the first DNA segregation and separation model in bacteria. The bacterial inner membrane and cell growth were proposed to play a central role in pulling the chromosome apart during cell division. This model was primarily derived from electron microscopy of bacterial cells showing tethering of the genetic material to the bacterial inner membrane. As per this model, replicated DNA becomes tethered to the cytoplasmic membrane. As the cell elongates, the chromosomes are pulled apart to the two opposite poles of the cells, following which cell division ensues and separates the replicated DNA (Jacob et al., 1963). This mode of DNA segregation was also assumed to be true for F plasmids. A timeline depicting the evolution of our understanding of DNA segregation in bacteria is shown in **Figure 3**. The Jacob model was further supported by the findings that the plasmid partitioning proteins $ParA_F$ of F plasmid (FSopA) and that of Q plasmid ($ParA_Q$/QSopA) of *Coxiella burnetii* associate with the cell membranes (Lin and Mallavia, 1998). The study involved biochemical membrane fractionation, floatation assays and immunoelectron microscopy, which suggested that a fraction of the respective ParA proteins were localized to the bacterial inner membranes. Further, phosphatase assays using the periplasmic PhoA protein indicated that the N-terminal residues in $ParA_F$ and $ParA_Q$ might specify membrane association (Lin and Mallavia, 1998). Consistent with the ideas around the time, these studies resulted in a model being proposed for plasmid partitioning, wherein the plasmid-ParB complex became associated with the membrane *via* ParA, and DNA partitioning was driven by cell elongation (**Figure 4A**). Plasmid partitioning *via* membrane association of ParA was further supported by beautiful genetic and plasmid localization studies showing the abundance of F plasmids in anucleate cells (Ezaki et al., 1991), although early studies using *mukB* indicated a general role for the nucleoid as well (Niki et al., 1991). More recently, we have identified a potential amphipathic helix in the C-terminus of $ParA_F$ (Mishra et al., 2021). However, the sufficiency of the C-terminal helix to bind bacterial membranes has not been examined.

However, further studies that directly visualized F plasmid and other partitioning proteins in bacteria using fluorescence microscopy and imaging techniques never revealed any membrane localization for ParA. On the contrary, ParA was predominantly associated with chromosomal DNA and localized on the bacterial nucleoid. At the turn of the millennium, as the concept of bacterial cytoskeleton had just emerged, a cytoskeletal filament model in the year 2005 was proposed (**Figure 4B**). The cytoskeletal model was based on the observations that $ParA_F$ and ParF ($ParA_{TP228}$) assembled into polymeric structures (Barillà et al., 2005; Lim et al., 2005). Pogliano and colleagues, using Nile red staining and light microscopy, found that $ParA_F$ assembled into filaments *in vitro* in the presence of ATP and grew at a rate of $\sim 0.18 \pm 0.05\,\mu m$ per minute (Lim et al., 2005). The polymerisation of $ParA_F$ was further ascertained by *in vivo* fluorescence imaging of functional C-terminal GFP fusions to $ParA_F$ in *E. coli* (Lim et al., 2005; Hatano et al., 2007) and *in vitro* by transmission electron microscopy (Bouet et al., 2007). ParA of pB171 plasmid also exhibits similar polymeric structures in the presence of DNA (Ringgaard et al., 2009). The cytoskeletal filament model was later supported by super-resolution imaging studies of *C. crescentus* ParA (Ptacin et al., 2010). A chromosome pulling model similar to the eukaryotic burnt-bridge model for tubulin mediated chromosome segregation was proposed (Ptacin et al., 2010). As per this model, ParA undergoes polymerisation forming a filament structure that pulls the plasmid to the opposite ends of the cell *via* the ParB-*parS* complex (Ptacin et al., 2010) and contrasted with the pushing mechanism employed in

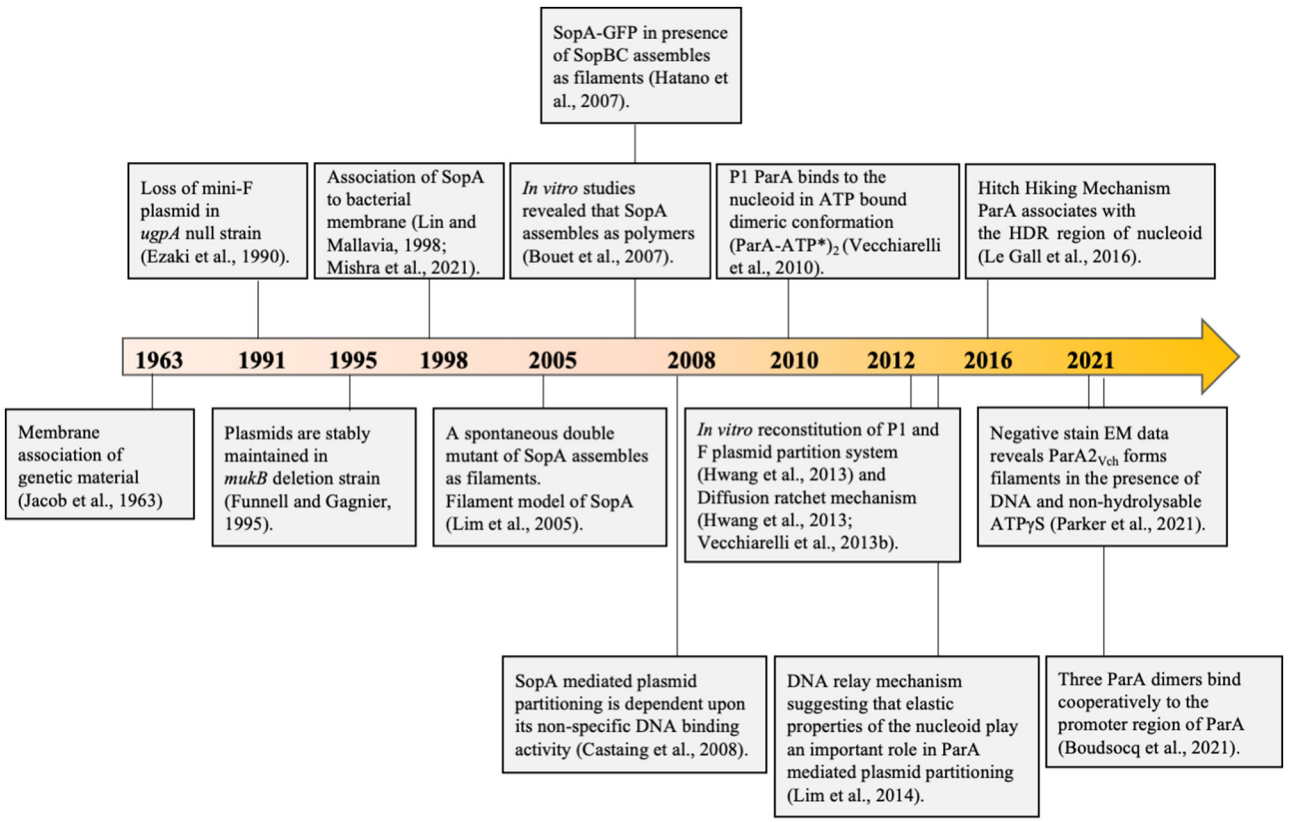


FIGURE 3 | A timeline depicting important milestones in our understanding of the models proposed for the mechanism of DNA segregation. In 1963, Francois Jacob proposed the membrane tethering model wherein the chromosome was believed to be associated with the bacterial membrane. The chromosomes would be pulled along as the cell grows and eventually separated. This was further supported by Ezaki et al. (1991) and Lin and Mallavia (1998) suggesting that the DNA was probably linked to the membrane through the partitioning protein ParA$_F$ (SopA). Toward the early part of the twenty-first century, with the emerging concept of the cytoskeleton in bacteria, the first polymerisation based filament models of segregation replaced the erstwhile membrane model, wherein it was proposed that the plasmid partitioning protein ParA polymerizes to push apart the plasmids to the two opposite ends of the cell. This model was supported by the findings of Lim et al. (2005), Bouet et al. (2007), and Hatano et al. (2007). Recent findings by Parker et al. (2021) suggest oligomerisation on DNA. Castaing et al. (2008) reported the association of the ParA with the nucleoid and, together with several other findings from many labs across the globe, led to abandoning the popular polymerisation model and models accounting for nucleoid association like the diffusion-ratchet mechanism and Hitch-Hiking mechanism were put forth. These form the current models and are the most accepted models of plasmid and chromosome segregation.

the R1 plasmid by the actin-like ParM protein (described above). Further, members of the smaller ParA family, like Soj (ParA$_{Bsu}$ and ParA$_{Tth}$), have been reported to undergo polymerisation in the presence of ATP and DNA (Leonard et al., 2004, 2005a; Barillà et al., 2005; Pratto et al., 2008; Soberón et al., 2011; Schumacher et al., 2012; Volante and Alonso, 2015).

However, several research laboratories working on various ParA proteins from diverse bacterial genera found the polymerisation and filament-pulling model inconsistent with the emerging biochemical and cell biological evidence in the following decade. Thus, the models based on polymerisation-mediated DNA segregation by ParA proteins of the Type-I system were soon superseded by models favoring chemophoretic gradients formed by ParA on the bacterial nucleoids along the cell's long axis (**Figure 4C**).

An overwhelming amount of literature on ParA from several species and exhaustive biochemistry and super-resolution imaging argue against polymerisation mediated partitioning. Instead, they support a diffusion-ratchet mechanism (Hatano and Niki, 2010; Hwang et al., 2013; Vecchiarelli et al., 2013b; Hu et al., 2017; reviewed in Brooks and Hwang, 2017) or DNA-relay mechanism (Lim et al., 2014; Surovtsev et al., 2016) for the movement of plasmids or chromosomal *oriC* proximal ParB-*parS* complexes toward the cell poles. The diffusion ratchet models are principally derived from the *in vitro* reconstitution experiments which more or less replicate the *in vivo* conditions (Vecchiarelli et al., 2010, 2013b, 2015; Hwang et al., 2013; reviewed in Brooks and Hwang, 2017). Mizuuchi and the group utilized *in vitro* reconstitution experiments mimicking the minimalistic plasmid partitioning apparatus on a glass slide coated with DNA to resemble the bacterial nucleoid. They were able to observe the dynamics of the partitioning machinery and the relevance of non-specific DNA binding in the ParA mediated movement of the partitioning complex using TIRF microscopy (Hwang et al., 2013; Vecchiarelli et al., 2013b, 2014). Moreover, this was also supported by *in vivo* imaging data reported for other members of the ParA superfamily (Fogel and Waldor, 2005; Lim et al., 2005; Hatano et al., 2007; Hester and Lutkenhaus, 2007; Hatano and

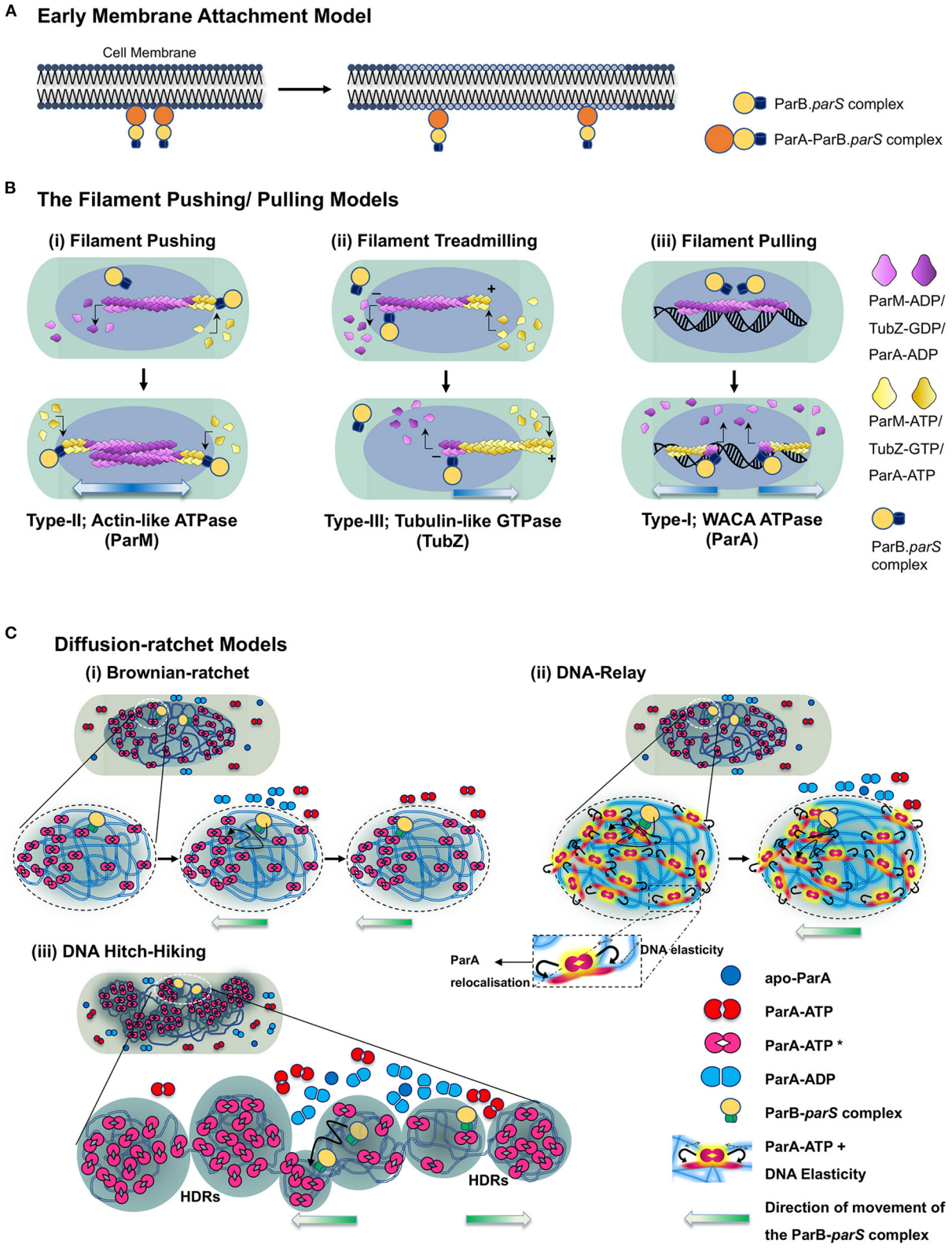


FIGURE 4 | Model depicting the mechanisms of ParA proteins in DNA partitioning. **(A)** Early membrane attachment model. An early model for DNA partitioning had suggested membrane tethering of the genetic material. As per this model, DNA is associated with the membrane, and cell elongation pulls the replicated DNA apart, which physically separates into the newly formed daughter cells upon cell division. It was originally proposed for chromosomal DNA and later proposed for F plasmids

(Continued)

FIGURE 4 | as well. $ParA_F$ (SopA) was suggested to be the membrane tether for the plasmids. **(B)** Filament pushing/pulling model. A filament model invokes the formation of long-range polymeric structures (μm scale). The filament polymerisation mechanisms are utilized by Type-II and Type-III systems. In the case of the Type-II system, the polymerisation of an actin-like ATPase (ParM) generates the pushing forces to segregate the plasmids, whereas treadmilling of the polymers (TubZ) drives the plasmids to cell poles in the Type-III system. Although a filament model in which the plasmids are pulled to the cell poles has been proposed for ParA (Type-I) as well, the existence of continuous filaments of ParA *in vivo* is debatable. **(C)** Diffusion-ratchet models. The diffusion-ratchet models posit that a chemophoretic gradient formed by ParA-ATP molecules bound to the nucleoid drives the movement of the plasmid bound partitioning complex up the gradient. The ParB-*parS* partition complex, on encountering DNA bound ParA-ATP dimers (ParA-ATP* state), stimulates the ATP hydrolysis of ParA, converting it into ParA-ADP and can no longer be associated with the nucleoid. The released ParB-*parS* complex undergoes Brownian motion but is constrained by the cell membrane and rebinds the nucleoid-associated ParA-ATP complex in its vicinity. Cycles of such release, diffusion and capture allow the partition complex to move up the ParA-ATP gradient on the nucleoid to achieve a unidirectional motion. (i) In a simplistic Brownian diffusion and capture of the partition complex (ParB-*parS*) by ParA-ATP dimers on the surface of the nucleoid, the plasmid surfs on the nucleoid and migrates toward the cell poles. (ii) In the DNA-relay models, the elasticity of the chromosomal DNA (indicated by haze around the DNA and double arrows in the inset) supports the movement of the partition complex, and simple diffusion by Brownian motion alone is not sufficient to achieve the long-distance directional motion observed in cells. Both the above models imply that the movement of the partition complex occurs on the surface of the nucleoid. (iii) The DNA Hitch-Hiking model instead suggests that the DNA segregation complex undergoes diffusive movement deep within the chromosomes in the High-density regions (HDRs). The direction of movement of the partitioning complex is indicated by gradient arrows (blue in the case of filament models and green in the case of diffusion-ratchet models).

Niki, 2010; Le Gall et al., 2016; McLeod et al., 2017). One of the main attributes of this model was that the ParA proteins bind the non-specific DNA (nucleoid) as dimers in the presence of ATP and do not assemble into polymers or filamentous structures. DNA-relay mechanism was subsequently proposed to explain the movement and dynamics of *C. crescentus* ParA observed during the relocation of the replicated origin from the old pole to the new pole (Lim et al., 2014). The model also considered the elastic properties of the bacterial nucleoid that helps in segregating the plasmids. The initial asymmetry in ParA localization is brought about by the stochasticity in the ATP hydrolysis in the region surrounding the partition complex. This asymmetry creates an imbalance in the elastic forces experienced by the partitioning complex and results in an incremental movement of the complex in the direction opposite to the site of ATP hydrolysis. When two partition complexes are in close proximity, a large zone of depletion of ParA dimers occurs and results in an apparent repulsion and change in the direction of motion of the partition complexes (Surovtsev et al., 2016). Fluctuations in the DNA positions caused by stretching forces and the elastic dynamics of the nucleoid act to position the partitioning complex at a new location and relay it across the length of the cell (Lim et al., 2014; Surovtsev et al., 2016). The diffusion-ratchet and DNA-relay models assume the DNA partitioning process ensues on the surface of the nucleoid. Further, the diffusion-ratchet mechanism invokes confinement by the inner membrane to prevent diffusion of the free plasmid in three dimensions and limit movement to almost a 2D surface (Vecchiarelli et al., 2014). However, recent super-resolution imaging, using structured illumination (SIM) and multi-focus microscopy (MFM), of $ParA_F$ and $ParF_{TP228}$ strongly suggest that the movement of the plasmid (the ParB-*parS* complex) is not the surface but rather appears to be deep within the nucleoid space (Le Gall et al., 2016; McLeod et al., 2017). These data have led to the proposition of a DNA Hitch-Hiking mechanism, a model that is fully consistent with the diffusion-ratchet and DNA-relay mechanisms (Le Gall et al., 2016) (**Figure 4C**). Our current understanding of the ParA mediated DNA segregation can thus be summarized as below:

The accurate positioning and partitioning of DNA during each round of cell division begins with its replication. The active form of ParA, i.e., the ParA-ATP* state, binds nsDNA as a dimer $(ParA\text{-}ATP^*)_2$ and remains associated within the high-density regions (HDR) within the bacterial nucleoid. The repetitive centromeric sequence, *parS*, in the plasmids or in the chromosome that lies proximal to the origin (*oriC*), is bound by ParB and results in clustering or spreading of ParB around *parS* sequences. This ParB-*parS* complex then hovers around the cell space of the bacterium, searching for the nucleoid bound ParA-ATP* dimers. The ParB-centromere complex, upon encountering $(ParA\text{-}ATP^*)_2$, stimulates the ATPase activity of ParA. ATP hydrolysis results in the conversion of ParA-ATP* to ParA-ADP, which can no longer remain bound to the bacterial nucleoid and is thus released into the cytosol. The release of ParA molecules from the nucleoid creates a zone of depletion of DNA-bound ParA-ATP in the vicinity of the ParB-centromere complex. This results in a local gradient of $(ParA\text{-}ATP^*)_2$ on the nucleoid, and the ParB-centromere complex, driven by Brownian motion, moves up the concentration gradient toward the DNA-bound ParA-ATP. The randomly chosen initial direction of motion of the ParB-*parS* complex can bias the movement in the same direction driving the long-range unidirectional movement of the plasmid. In the DNA Hitch-Hiking model, ParA-ATP molecules are bound to the nucleoid in regions of high density (HDR), and the partition complex hops from one HDR to another resulting in the progressive directional motion of the ParB-bound DNA complex.

Meanwhile, the ParA-ADP or the apo-ParA can possibly associate with ATP to form the ParA-ATP dimer $[(ParA\text{-}ATP)_2]$ almost instantly, given the relatively higher concentrations of ATP in the cell, which are usually in the millimolar range. However, the ParA-ATP dimer $[(ParA\text{-}ATP)_2]$ cannot immediately bind DNA and is slowly converted to a $(ParA\text{-}ATP^*)_2$ state, which can re-associate with the bacterial nucleoid and contribute to the gradient. This time-delay in the nucleoid association of ParA-ATP is crucial to the sustenance of the ParA-ATP gradient and the unidirectional motion of the partitioning complex in all diffusion-ratchet based models (Vecchiarelli et al., 2010, 2013a; Hu et al., 2015, 2017; Le Gall et al., 2016; Surovtsev et al., 2016; McLeod et al., 2017). Cycles of such binding and release of the ParA-ATP within the nucleoid eventually mediate displacement of the replicated plasmids away from the cell division site, ensuring equipartitioning of the genetic material.

Interestingly, recent electron microscopy studies on *V. cholerae* ParA2 and archaeal SegAB complex show that $ParA2_{Vch}$ and SegA assemble into filaments in the presence of DNA and ATP or non-hydrolysable ATP (Hui et al., 2010; Parker et al., 2021; Yen et al., 2021). Further, electron microscopy and chromatography of purified $ParA_{P1}$ in the presence of ATP and ns-DNA indicate the assembly of $ParA_{P1}$ into polymeric structures at very high concentrations of the protein. While its physiological relevance is yet to be determined, they suggest that the filament structures are also likely in the ParA-ATP* state (Dunham et al., 2009; Vecchiarelli et al., 2010). Although, these findings could revitalize the polymeric cytoskeleton models for ParA mediated DNA segregation, $ParA2_{Vch}$ does not form stable polymers and is likely to bind DNA cooperatively as higher-order oligomers, which could be compatible with Brownian-ratchet models and not filament pulling models (Chodha et al., 2021).

WALKER A CYTOSKELETAL ATPases IN BACTERIAL CHROMOSOME SEGREGATION

Soj in *Bacillus subtilis*

The partitioning locus in the case of *B. subtilis* consists of two proteins, Soj and SpoOJ and a centromere-like sequence *parS* (Ireton et al., 1994; Leonard et al., 2005a; Lee and Grossman, 2006). While Soj belongs to the ParA family, SpoOJ is a ParB-like protein, both of which were initially isolated as sporulation genes and hence the names Soj-SpoOJ. We refer to Soj and SpoOJ proteins as $ParA_{Bsu}$ and $ParB_{Bsu}$, respectively. Overexpression of $ParA_{Bsu}$ or deletion of $ParB_{Bsu}$ leads to the formation of aberrant nucleoid morphology and anucleate cells (Lee and Grossman, 2006). $ParA_{Bsu}$ is a smaller Type-Ib Walker A ATPase with only 21% sequence similarity to $ParA_F$ protein. $ParA_{Bsu}$ works cooperatively with $ParB_{Bsu}$ to enable chromosome segregation. $ParB_{Bsu}$ binds to the eight repetitive sequences on *parS* located on both sides of the origin of replication *oriC*. When visualized by fluorescence microscopy, this complex appears as discrete spots on binding $ParA_{Bsu}$ proteins. A green fluorescent protein (GFP) fusion of $ParA_{Bsu}$ shows nucleoid localization of the protein. However, when co-expressed with $ParA_{Bsu}$, it leads to the formation of discrete spots that colocalise with the ParB-*parS* complex in the cell (Leonard et al., 2005a; Scholefield et al., 2011). GFP fusions of $ParA_{Bsu}$ seem to undergo oscillatory movement (Leonard et al., 2005a), a characteristic feature of the ParA superfamily (Raskin and de Boer, 1999; Hatano et al., 2007; McLeod et al., 2017). While $ParA_{Bsu}$ controls the condensation of $ParB_{Bsu}$ foci, $ParB_{Bsu}$ regulates the dynamic behavior of $ParA_{Bsu}$, and thus both the proteins are functionally interlinked. Biochemical studies show that ATP bound $ParA_{Bsu}$ exists as a dimer, and an ATPase mutant D44A was crystallized in dimeric form (Leonard et al., 2005a). Non-specific DNA binding and nucleoid association of ATP-bound $ParA_{Bsu}$ is mainly mediated by surface-exposed arginine residues (Hester and Lutkenhaus, 2007). Further, electron microscopy data suggest that, like *Thermus thermophilus* ParA ($ParA_{Tth}$/TthSoj), $ParA_{Bsu}$ also forms a nucleoprotein filament, showing that the dimeric form of $ParA_{Bsu}$ recruits other $ParA_{Bsu}$ dimers facilitating the formation of a polymeric structure (Leonard et al., 2004, 2005a). Interaction with $ParB_{Bsu}$ stimulates the ATPase activity of $ParA_{Bsu,}$ and ADP-bound forms dissociate from DNA (Scholefield et al., 2011). In addition to its role in DNA segregation, the $ParA_{BSu}$ is also a transcriptional regulator that controls the expression of many sporulation genes, including *spo0A*, *spoIIA*, *spoIIG*, and *spoIIE* (McLeod and Spiegelman, 2005).

MipZ and ParA in *Caulobacter crescents*—Two WACA Proteins Coordinate DNA Segregation

In *C. crescentus*, two WACA family members coordinate to promote segregation (Mohl and Gober, 1997; Thanbichler and Shapiro, 2006; Toro et al., 2008; Corrales-Guerrero et al., 2020). Both proteins can interact with the centromere binding protein, ParB, which binds to the *parS* DNA sites that lie close to the origin of replication (*oriC*) in *C. crescentus*. At the beginning of the cell cycle, *parS* sites are anchored to the cell poles by the interaction of ParB with PopZ, a polarly localized protein. However, at the onset of the S-phase, the chromosome duplicates and one copy of the ParB-*parS* complex moves to the other pole of the cell driven by the nucleoid bound ParA-ATP gradient (Thanbichler and Shapiro, 2006). MipZ, another ParA member, binds to ParB at both the cell poles and forms a concentration gradient along the cell length, with the highest concentration being on the poles and the lowest at the mid-cell site (Easter and Gober, 2002; Thanbichler and Shapiro, 2006). MipZ is an antagonist of FtsZ, and a high concentration of MipZ prevents the formation of Z-ring at the poles but promotes its assembly at the mid-cell site (Thanbichler and Shapiro, 2006; Du and Lutkenhaus, 2012). Although ParB can bind both ParA and MipZ, the roles played by ParB at the two poles are different. At the old pole, where ParB is bound to *parS* and interacts with ParA, ParB acts to stimulate the ATP hydrolysis of ParA-ATP. This stimulation of the ATPase activity of ParA by the ParB-*parS* complex aids in generating the ParA gradient for the effective movement of the ParB-*parS* complex to the new pole of the cell (Corrales-Guerrero et al., 2020).

In contrast, MipZ binds ParB at both the old and new poles; ParB plays a role of a catalyst and promotes dimerization and binding of MipZ to DNA by recruiting monomers. However, ATP hydrolysis due to the intrinsic ATPase activity of MipZ results in a release of the MipZ monomers from the nucleoids at ParB distal regions. Since ParB is localized at cell poles, the association of MipZ with the nucleoid is at its lowest concentration at the mid-cell position (Kiekebusch et al., 2012). Both MipZ and ParA have conserved residues that bind nsDNA and have nucleoid binding activities that help DNA segregation (Du and Lutkenhaus, 2012). Thus, both MipZ and ParA work synergistically in *C. crescentus* DNA segregation, and both are indispensable for accurate chromosome partitioning. Further, MipZ serves another critical function of precisely positioning the Z-ring for cell division in *C. crescentus.*

Chromosome Segregation Machinery in Bacterial Pathogens

Vibrio cholerae

V. cholerae, the causative agent of cholera, is a gram-negative bacterium with two chromosomes, each with its *par* locus (Heidelberg et al., 2000). Although both the chromosomes encode the Type-I mechanism of segregation (Fogel and Waldor, 2005, 2006; Yamaichi et al., 2007), they employ different strategies for their transmission to daughter cells (Venkova-Canova et al., 2013). The 2.4-Mb chromosome I makes use of the Type-Ib system, while the smaller 1.6-Mb Chromosome II uses the Type-Ia segregation mechanism. Studies on the partitioning of chromosome II have indicated a role for $ParA2_{Vch}$. Genetic studies have shown that the deletion of *parAB2* causes loss of Chromosome II from the cells, and thus *parAB2* locus seems essential for Chromosome II segregation (Yamaichi et al., 2007). Further, it has been shown that $ParA2_{Vch}$ binds to the non-specific DNA like many other members of the superfamily, as mentioned above (Hui et al., 2010; Chodha et al., 2021). Interestingly, recent cryo-EM studies on $ParA2_{Vch}$ have reported its assembly into a polymeric nucleoprotein complex in the presence of ATPγS and DNA (Parker et al., 2021).

Helicobacter pylori

While Soj in *H. pylori* is a member of the ParA superfamily of proteins and bears 22–48% sequence similarity to $ParA_F$ protein, SpoOJ (HP1138 gene) is the ParB homolog. The *soj* (HP1139) and *spoOJ* (HP1138) genes, together with the two putative *parS* sites, are located within 20–30% of the origin-proximal region of the circular chromosome of *H. pylori*, as observed in other species (Lee et al., 2006; Chu et al., 2019). On binding to nucleotide, $HpSoj/ParA_{Hpy}$ forms a dimer, a feature that has also been observed in its crystal structure. $ParA_{Hpy}$ is a weak ATPase, and its activity is stimulated by non-specific DNA and $ParB_{Hpy}$/HpSpoOJ, a typical trait of the ParA family of proteins. However, unlike $ParA_{Tth}$ or $ParA_{Vch}$, electron microscopy failed to detect any $ParA_{Hpy}$ filaments and thus excluded the possibility of ParA polymers driving DNA segregation. Moreover, a stretch of basic residues that allows $ParA_{Hpy}$ to bind nsDNA was identified, suggesting that a diffusion-ratchet mechanism (described below) might be operative to partition DNA (Chu et al., 2019).

Pseudomonas aeruginosa

P. aeruginosa, a γ-proteobacterium, is an opportunistic pathogen and is the causative agent of morbidity in cystic fibrosis patients (Stover et al., 2000). It encodes a *parABS* system in its genome that helps segregate its large 6.3 Mb chromosome (Lagage et al., 2016). The *par* genes are located ∼8 kb from *oriC*, and the deletion of ParA or ParB leads to chromosomal segregation defects resulting in an increased production of anucleate cells (Lasocki et al., 2007). ParB of Pseudomonas has been postulated to be a NAP (Nucleoid Associated Protein) exhibiting specific interaction with *parS* sites (Kusiak et al., 2011), interaction with several hepta-nucleotide sequences in the genome and regulating transcription of various genes (Kawalek et al., 2018). Although ten 16-bp palindromic *parS* sites have been reported in the genome (Bartosik et al., 2004; Jecz et al., 2015), only four are proximal to the origin, *oriC* (Livny et al., 2007). Although, ParB binds to these *oriC* proximal four *parS* sites (Kusiak et al., 2011; Lagage et al., 2016), just one of them is sufficient for ParB binding and accurate chromosome segregation (Jecz et al., 2015). Surprisingly, displacement of *parS* from its native site does not affect chromosome segregation. The *parABS* locus remains the first locus to be segregated following replication, suggesting parS action is independent of its location in the chromosome (Lagage et al., 2016).

Mycobacterium tuberculosis

M. tuberculosis is the causative agent of tuberculosis, and its genome also encodes a parABS system that includes a parS site, parA (Rv3918) and parB (Rv3917) genes, in addition to two other parA homologs (Rv1708 *and* Rv3213). A study using Transposon site hybridization (TraSH) indicated that the parAB genes within the parABS loci are essential for growth and viability and could not be deleted in *M. tuberculosis* (Sassetti and Rubin, 2003). However, in *M. smegmatis*, ParA is not essential but is required for normal growth (Ginda et al., 2013). While all the three ParA homologs are localized at the cell poles, $ParA_{Mtb\_Rv3918}$ colocalises with ParB at either of the poles and exhibits localization at the polar or mid-cell position that overlaps with nucleoid or inter-nucleoid regions, which suggests oscillations but require further time-lapse imaging to ascertain foci movements (Maloney et al., 2009). In *M. tuberculosis*, ParB is phosphorylated by a Ser/Thr Protein Kinase (STPK) and negatively regulates ParB activity by inhibiting interaction with parS and ParA. Mutations that mimic phosphorylation affect the cellular localization of $ParB_{Mtb}$ and lead to impaired chromosome segregation (Baronian et al., 2015). Further, many studies utilizing *M. smegmatis* as a model organism confirm the dynamic localization of the ParA and ParB-parS complexes in mycobacteria (Jakimowicz et al., 2007; Ginda et al., 2013; Santi and McKinney, 2015; Trojanowski et al., 2015; Uhía et al., 2018). Interestingly ParA also associates with Wag31/DivIVA, a protein that dictates polar growth in Streptomyces and *Mycobacterium smegmatis* (Hempel et al., 2008; Ginda et al., 2013).

WALKER A TYPE CYTOSKELETAL ATPases IN ARCHAEAL DNA SEGREGATION

Recent work on Archaea, the third branch of life, has led to a better understanding of the cellular functions and intracellular dynamics of cytoskeletal proteins in this group. In the thermophilic crenarchaeon *Sulfolobus solfataricus*, *segA* encodes for a ParA-like protein with a weak ATPase activity (She et al., 2001; Schumacher et al., 2015). SegB, on the other hand, is an archaeon specific protein that shares homology with proteins found in crenarchaea and euryarchaeal but has no sequence similarity to ParB or any other bacterial or eukaryotic proteins (Kalliomaa-Sanford et al., 2012). SegB is a DNA binding protein that binds specifically to palindromic sequences that constitute the centromeric sites, S1 and S2, found upstream of *segAB* genes

in the *S. solfataricus* chromosome (Kalliomaa-Sanford et al., 2012; Barillà, 2016). While SegA assembles into polymers in the presence of ATP *in vitro*, SegB interacts with SegA in the presence of nucleotides and influences SegA polymerisation, which drives chromosome segregation (Kalliomaa-Sanford et al., 2012; Yen et al., 2021). Recent crystal structures of the SegA and SegB proteins complexed with DNA have revealed new insights into this archaeal DNA partitioning system.

Interestingly, SegA forms a non-canonical dimer that does not resemble the typical ATP-sandwich dimers observed in many ParA structures (Yen et al., 2021). Although SegB lacks any sequence similarity to bacterial or eukaryotic proteins (Schumacher et al., 2015), the structure of SegB reveals the presence of a ribbon-helix-helix (RHH) motif that is important for binding to the centromere-like DNA sites (Yen et al., 2021) and is reminiscent of many bacterial plasmid-encoded CBPs (Schumacher, 2008). Furthermore, SegB forms a superhelical chromatin-like structure and wraps around the DNA, giving rise to a left-handed helix as has been observed for the omega protein (ParB) of the *S. pyogenes* plasmid pSM19035. However, significant differences have been noted in regions that form the protein dimer interfaces (Weihofen et al., 2006; Soberón et al., 2011; Yen et al., 2021). Thus, the organization of archaeal chromosomal segregation genes and functions bear a close resemblance to bacterial *parABS* systems. Moreover, electron microscopy of SegA-SegB-DNA complexes in the presence of ATP revealed short oligomers that were rod-like or arc-shaped. The rod-like structures were only 30–40 nm in length. However, the formation of such oligomers only in the presence of ATP but not ADP suggests a role for SegA oligomers in segresome formation, chromosome organization and segregation in archaea (Yen et al., 2021).

Like many bacteria, archaea too rely on ParA proteins to segregate and maintain plasmids. pNOB8 is an archaeal plasmid from *Sulfolobus* and contains a unique partitioning system comprising three proteins and a centromeric site (She et al., 1998; Schumacher et al., 2015). While $ParA_{NOB8}$ forms the NTPase motor protein with a Walker A motif, AspA is the centromere binding protein and thus performs the analogous function of a typical ParB. The archaeal plasmid, pNOB8, also carries an atypical ParB whose N-terminal domain shares homology with bacterial ParBs, but the C-terminal domain bears structural similarity to CENP-A, a eukaryotic protein involved in chromosome segregation (Schumacher et al., 2015). Nonetheless, $ParB_{NOB8}$ acts as the adaptor protein that links the DNA to $ParA_{NOB8}$ by binding to the AspA-centromere complex. AspA binds to the centromere sequence and creates a superhelical structure for $ParB_{NOB8}$ binding. $ParA_{NOB8}$ binds to this ParB-AspA-centromeric complex and facilitates the partitioning of the plasmid into the daughter cells (Schumacher et al., 2015). Recent crystal structures of ParA-AMPPNP-DNA complexes have revealed the presence of a multifaceted nsDNA binding site in $ParA_{NOB8}$ (Zhang and Schumacher, 2017). The structural resemblance of the C-terminal domain of $ParB_{NOB8}$ to CENP-A and its binding to ParA reveal a unifying theme that underlies the DNA segregation process in the three domains of life.

SPECIALIZED POLAR TETHERING PROTEINS—ACCESSORY FACTORS FOR ParA MEDIATED CHROMOSOME SEGREGATION IN BACTERIA

Bactofilin

Bactofilins belong to a group of polymeric cytoskeletal proteins defined by a conserved DUF583 domain, also known as the bactofilin domain (Kühn et al., 2010; Punta et al., 2012). They are characterized by extended beta-sheet structures and are ubiquitous in both Gram-positive and Gram-negative bacteria (Lin and Thanbichler, 2013; Zuckerman et al., 2015). In the rod-shaped bacterium *Myxococcus xanthus*, four bactofilin paralogs have been identified. They serve diverse functions like cell-shape maintenance (Koch et al., 2011), cell motility *via* polar localization of a small GTPase, SofG (Bulyha et al., 2013) and chromosome segregation (Lin et al., 2017; Anand et al., 2020). Three paralogs of bactofilin, BacNOP assemble into a polymeric structure and form a complex with the ParB-like protein, PadC, to restrict ParABS machinery to a well-defined position in the subpolar region of the cell. The $ParA_{Mxa}$ ATPase decorates the entire length of the bactofilin filament using the ParB-like adaptor protein PadC. The pole-distal ends of the bactofilin filaments are bound by the centromeric DNA (*parS*) binding protein $ParB_{Mxa}$ ensuring that the genome is equitably transferred after each round of cell division (Lin et al., 2017; Anand et al., 2020).

DivIVA

Discovered in *B. subtilis*, DivIVA was named so due to the defects in septum placement in its absence (Edwards et al., 2000; Perry and Edwards, 2006; Oliva et al., 2010; van Baarle et al., 2013). It is a tetrameric coiled-coil protein that binds to the membrane by sensing negative curvature (Lenarcic et al., 2009; Ramamurthi and Losick, 2009; Oliva et al., 2010) and is mainly found in Gram-positive bacteria (Lin and Thanbichler, 2013). During sporulation, DivIVA localizes to the poles of the cell and binds to RacA and Soj/SpoOJ ($ParAB_{Bsu}$) complex (Wu and Errington, 2003; van Baarle et al., 2013). In addition, DivIVA associates with the peptidoglycan synthesis proteins in *Streptomyces*, *Corynebacterium* and *Mycobacterium*, organizes the growth directing tip-complex and interacts with ParB (Hempel et al., 2008, 2012; Donovan et al., 2012; Ginda et al., 2013; Holmes et al., 2013). Interestingly, a DivIVA bound protein, Scy, associates with the ParA ATPase in *S. coelicolor* (Ditkowski et al., 2013).

PopZ

While DivIVA is restricted to Gram-positive bacteria, PopZ (polar organizer protein Z), an evolutionarily unrelated small, acidic protein with an alpha-helical structure present in many Gram-negative species of α-proteobacteria, serves a similar function (Bowman et al., 2008; Ebersbach et al., 2008; Holmes et al., 2016). PopZ has a highly conserved N-terminal and a C-terminal domain, analogous to DivIVA, required for higher-order assembly (Oliva et al., 2010; Bowman et al., 2013). PopZ assembles into polymers that form a network with gel-like

properties, which is impermeable to macromolecules (Bowman et al., 2010). However, this 3D network of PopZ at the cell pole captures the ParA released from the nucleoid by the migrating ParB-*parS* complex. This capture of released ParA is mediated by direct interactions with PopZ (Ptacin et al., 2014). The increased concentration of ParA at the PopZ proximal new pole triggers the ParA DNA binding activity, which reinforces the movement of the ParB-*parS* complex in the same direction, prevents reversals and immobilizes the origins at cell poles *via* interactions with ParB. In *C. crescentus*, cells lacking PopZ have chromosome segregation defects, highlighting the important role played by PopZ in regulating ParA mediated chromosome segregation (Bowman et al., 2008, 2010; Ebersbach et al., 2008; Schofield et al., 2010; Ptacin et al., 2014; Holmes et al., 2016).

TipN

In *C. crescentus*, TipN (Tip of New Pole) is a polarly localized landmark protein and contains a large C-terminal coiled-coiled domain and two transmembrane domains at its N-terminus. TipN relocalises at the cell division site in an FtsZ and FtsI dependent manner late during the pre-divisional stage and is thus present at the new pole in both the daughter cells (Huitema et al., 2006; Lam et al., 2006). At the new pole, TipN recruits TipF, a receptor for the second messenger c-di-GMP [bis-(3′-5′)-cyclic dimeric guanosine monophosphate] and a positive regulator of flagella assembly. However, this interaction between TipN and TipF is restricted in the swarmer cells due to the low level of c-di-GMP but allows the formation of a "flagella organizing center" at the new pole of the stalked cell due to the high levels of c-di-GMP in stalked cells (Davis et al., 2013). Interestingly, while Δ*tipN* cells exhibit mild chromosome segregation defects (Ptacin et al., 2010), deletion of *popZ* leads to synthetic lethality in Δ*tipN* cells due to severe DNA segregation and cell division defects (Schofield et al., 2010). TipN directly interacts with $ParA_{Ccr}$ at the new pole and regulates the movement and overall speed of the ParB-*parS* complex (Ptacin et al., 2010; Schofield et al., 2010).

DIVERSE BIOLOGICAL FUNCTIONS OF ParA FAMILY OF ATPases

Proteins with the deviant Walker A motif, or the P-loop, serve diverse functions in different life forms (Walker et al., 1982; Koonin, 1993). These include DNA replication and partitioning, cell cycle and division, subcellular localization and spatial organization. Examples include the ParA family of proteins found in bacterial genomes and plasmids, which play a role in DNA segregation. A few that are involved in cell division, like MinD and MipZ have been extensively reviewed (Motallebi-Veshareh et al., 1990; Ebersbach and Gerdes, 2005; Leonard et al., 2005b; Hayes and Barillà, 2006; Michie and Löwe, 2006; Thanbichler and Shapiro, 2006; Schumacher, 2007; Du and Lutkenhaus, 2012; Lutkenhaus, 2012; Barillà, 2016; Jalal and Le, 2020). At the same time, some others play an important role in positioning large macromolecular complexes such as carboxysomes within cells (Savage et al., 2010). These deviant Walker A motif-containing proteins are found in all life forms, ranging from Archaea to Bacteria and constitute a versatile system to build the spatial organization in biological systems (**Figure 5**). Thus, members of the ParA superfamily are not only involved in plasmid and chromosome partitioning but are also involved in the maintenance of other cellular cargo in many bacteria. We outline a few examples here that highlight the diversity of this superfamily of proteins within the domain Bacteria.

MinD—A Spatial Regulator of Cell Division Site

MinD, a component of the *min* system, is a Walker A ATPase member majorly recruited for regulating cell division in bacteria (Lutkenhaus and Sundaramoorthy, 2003; Lutkenhaus, 2012). The structure of the MinD dimer reveals its similarity to the ParA proteins such as Soj ($ParA_{Bsu}$) (Wu et al., 2011). The *min* locus consists of two other genes, *minC* and *minE*, wherein MinC functions as an inhibitor of FtsZ ring assembly *via* its interaction with MinD (Hu et al., 1999; Hu and Lutkenhaus, 2000; Dajkovic et al., 2008). MinD, along with MinE, undergoes pole-to-pole oscillations and produces a gradient of MinCD division inhibitory complex (Hu and Lutkenhaus, 1999; Raskin and de Boer, 1999; Hale et al., 2001; Meinhardt and de Boer, 2001) with the maximum time-averaged concentration at the cell poles (Lutkenhaus, 2012). MinD, in its ATP bound dimeric form, binds to the membrane *via* a C-terminal amphipathic helix (Hu et al., 2002; Szeto et al., 2002; Hu and Lutkenhaus, 2003; Zhou and Lutkenhaus, 2003). Following this membrane association, MinE is recruited, which stimulates the ATPase activity of MinD and thus releases it from the bacterial membrane (Hu and Lutkenhaus, 2001; Hu et al., 2002; Park et al., 2011; Wu et al., 2011). Crystal structures of the MinD-MinE complexes reveal the conformational changes in MinE upon interaction with MinD. The structures also show how two released β-sheets in a four-stranded β-sheet MinE dimer convert into α-helices and suggests how MinE remains associated with the MinD-membrane complex (Park et al., 2011). Repeated binding and release cycles of MinD driving oscillation of MinCD complex that acts as a spatial regulator of FtsZ ring in bacteria. The period of the oscillations is dictated by the built-in delays in the system, such as nucleotide hydrolysis and exchange rates in MinD (Howard et al., 2001; Meinhardt and de Boer, 2001; Huang et al., 2003; Kretschmer and Schwille, 2016). Numerous studies have reconstituted the dynamics of the MinDE system *in vitro* on supported membrane bilayers and observed standing waves of MinD being chased by MinE and recapitulate oscillations within confinement (Loose et al., 2008, 2011; Ivanov and Mizuuchi, 2010; Zieske and Schwille, 2013, 2014; Vecchiarelli et al., 2016).

Further, the deletion of MinD results in the production of anucleate cells suggesting that MinD plays a critical role in chromosome segregation (Di Ventura et al., 2013). Numerical computer simulations have proposed that the *min* system may contribute to the movement of the chromosome from the mid-cell to the poles and might be mediated by the binding of MinD to DNA (Di Ventura et al., 2013). However, recent *in vitro* reconstitution experiments of MinDE on membrane bilayers failed to detect any direct recruitment of DNA to

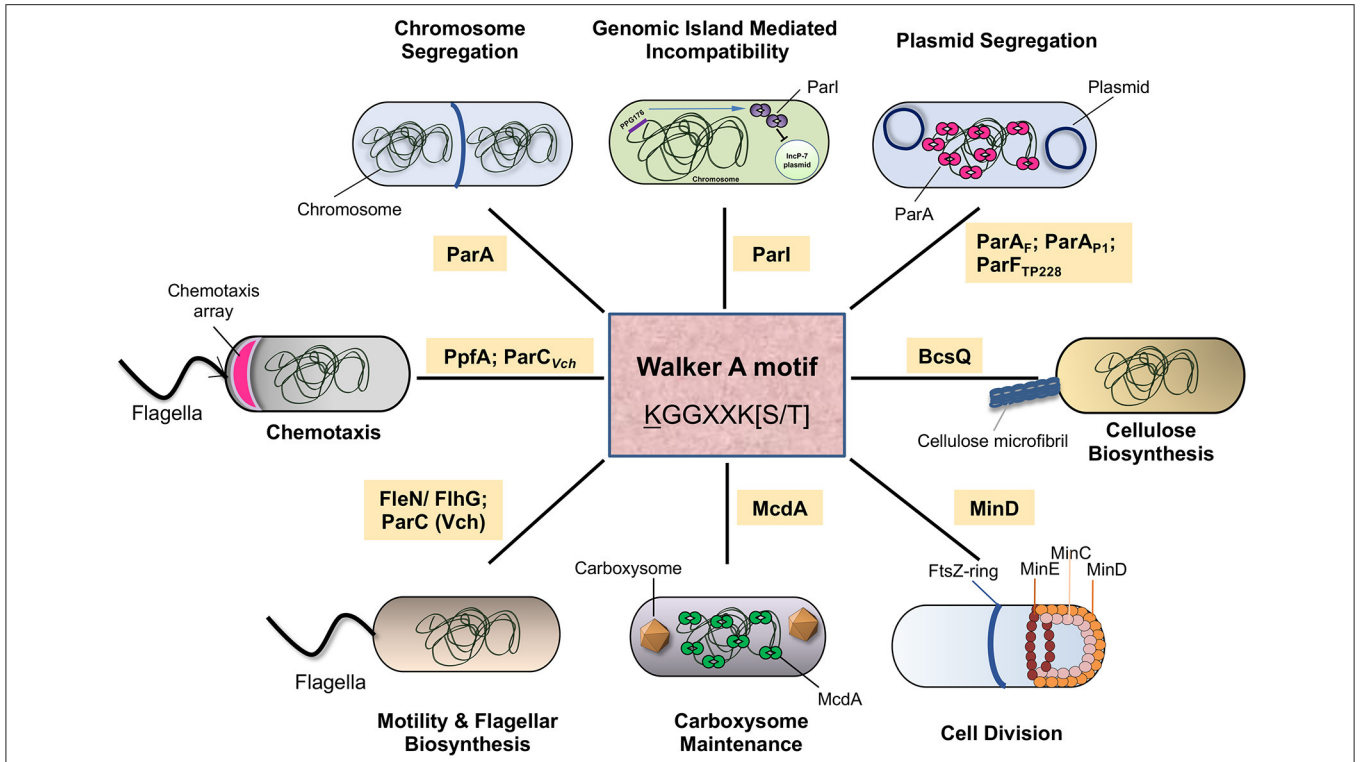


FIGURE 5 | The diverse functions of the ParA family of ATPases in bacteria. The schematic representation depicts the various cellular functions performed by the ParA superfamily of ATPases in bacteria. Chromosome segregation is carried out by ParA carried in bacterial chromosomes. Plasmid segregation is carried out by ParA$_F$ (in F plasmid) or ParA$_{P1}$ (P1 bacteriophage) ParF$_{TP228}$ (TP228, a plasmid from *S. newport*), FleN/FlgH regulates flagellar biosynthesis. McdA plays a role in carboxysome maintenance, and MinD controls cell division site positioning. ParI, an orphan ParA, regulates genomic island mediated incompatibility. BcsQ helps in cellulose biosynthesis. PpfA and ParC$_{Vch}$ (ParA-like ATPase in *V. cholerae*) control polar localization of chemotaxis clusters in *Rhodobacter* and *Vibrio,* respectively.

MinD, suggesting that the chromosome segregation defects in the absence of Min proteins might be due to indirect effects (Ramm et al., 2018).

McdA—ParA-Like Proteins in Carboxysome Maintenance in Cyanobacteria

Carboxysomes are membrane-bound organelles in Cyanobacteria that help in carbon fixation. These proteinaceous micro-compartments exist in low-copy numbers in the cells and thus depend on partitioning machinery to ensure their transmission during cell division (MacCready et al., 2018, 2020). Fluorescent labeling of the nucleoid, carboxysomes and McdA inside these cells has enabled the tracking of carboxysomes in live cells. Fluorescence imaging of carboxysomes has shown that these organelles are arranged linearly and segregate equally during cell division that was dependent upon a ParA homolog in Synechococcus elongatus (Savage et al., 2010). Recent studies have further shown that the ParA-like Walker A ATPase partitioning machinery, now named McdA (for Maintenance of Carboxysome Distribution), mediates Carboxysome maintenance (MacCready et al., 2018). Although McdA lacks the signature amino-terminal lysine residue (**Table 1**), it has a strong ATPase and non-specific DNA binding activity. Moreover, another small protein McdB, unrelated to ParB, has been shown to stimulate the ATPase activity of McdA and regulate carboxysome positioning. Although McdB shows no sequence similarity to ParB, it can form higher-order oligomers like ParB (Schumacher et al., 2019). McdB (like ParB) stimulates ATPase activity of McdA driving the directed movement of carboxysome toward a higher concentration of McdA on the nucleoid by a diffusion ratchet mechanism. Thus, Carboxysomes also employ a McdAB protein complex in a manner very similar to the ParAB complex (MacCready et al., 2018).

PpfA—An Orphan ParA Promoting Chemoreceptor Cluster Formation

Certain *par* loci in bacterial chromosomes contain only *parA* sequences and lack *parB* and centromeric sequence *parS*. These additional *par* loci are located outside the *parAB* operon (found close to the *oriC*) and are referred to as orphan ParA systems. These orphan ParAs can be found in many metabolic operons of bacterial genomes. Also, reports suggest that bacterial genomes encode multiple orphan ParAs. One such orphan ParA, called PpfA, is involved in chemotactic signaling in *Rhodobacter sphaeroides*. The chemotaxis protein cluster is formed by the partner proteins TlpT and CheW proteins. Mutants in the conserved Walker A motif are known to affect cluster formation.

Thus, PpfA helps in the dispersion and segregation of the chemoreceptor clusters (Thompson et al., 2006; Roberts et al., 2012). Remarkably, a distinct clade of ParA-like ATPase is encoded within the chemotaxis operon in *V. cholerae* (named ParC) and many γ-proteobacteria that have polar flagella (Ringgaard et al., 2011). In *V. cholerae*, the ParA-like ATPase, ParC, regulates the polar localization of the chemotaxis proteins CheW1 and CheY3 (Ringgaard et al., 2011). The subcellular localization dynamics of ParC are altered by its binding partner, ParP, which promotes the polar assembly of the chemotaxis array (Ringgaard et al., 2011, 2018; Alvarado et al., 2017).

ParI—An Incompatibility Factor Residing on a Genomic Island

ParI is an Orphan ParA member of the Walker A ATPase family present in the genomic island of *Pseudomonas putida* (Miyakoshi et al., 2012). The expression of ParI negatively regulates the maintenance of IncP-7 plasmids and results in their loss from cells. Studies on ParI have revealed that mutations in the conserved Walker A motif region (mainly the ATPase domain) of ParI fail to destabilize IncP-7 plasmids. ParI is an example of plasmid-mediated incompatibility residing within a genomic island.

BscQ—An Orphan ParA Involved in Cellulose Biosynthesis

Bacteria produce cellulose as a biofilm matrix polymer to enable the cohesion of biofilms. BcsQ (bacterial cellulose synthesis) proteins help produce cellulose in enterobacteria. BcsQ is a homolog of the ParA/MinD family of ATPase and is activated by cyclic di-GMP mediated signalling. Using fluorescently tagged BcsQ, it has been confirmed that this protein mainly localizes to the cell poles in bacteria and cell-cell adhesion mainly occurs *via* cellulose production at the cell poles. Thus, a ParA/MinD family of ATPase controls cell-cell adhesion and biofilm formation by regulating cellulose biosynthesis (Le Quéré and Ghigo, 2009).

FleN/FlhG/MinD2—A ParA Protein Regulating Flagellar Biosynthesis

Flagella serves as a locomotory organ in many organisms. The bacterial flagella is a complex structure requiring ∼40 genes for its assembly (Dasgupta et al., 2003). The flagellar genes in bacteria are involved in locomotion, biofilm formation, and pathogenesis (O'Toole and Kolter, 1998; Gellatly and Hancock, 2013; Guttenplan and Kearns, 2013; Mukherjee and Kearns, 2014). The flagella number and distribution are characteristic features of each organism. In *P. aeruginosa*, the number of flagella is regulated by FleN (Dasgupta et al., 2000, 2003; Köhler et al., 2000; Dasgupta and Ramphal, 2001; van Ditmarsch et al., 2013). FleN, also called FlhG or MinD2, is a ParA/MinD superfamily protein (**Table 1**), whose absence leads to the upregulation of genes involved in the synthesis of the flagellar motor, basal body, hook proteins etc. and conferring the multi-flagellate phenotype (Dasgupta et al., 2003; Chanchal et al., 2017). FleN acts antagonistic to FleQ and is known to inhibit FleQ dependent transcription. Such effect is essential for maintaining the correct number of flagella in the cell (Dasgupta and Ramphal, 2001; Chanchal et al., 2017). Homologs of FleN are also present in many other bacteria, including *B. subtilis*, *V. cholerae*, *Geobacillus thermodenitrificans*, and *Campylobacter jejuni* (Correa et al., 2005; Guttenplan and Kearns, 2013; Schuhmacher et al., 2015; Gulbronson et al., 2016).

CONCLUDING REMARKS

Central to all life is the duplication and equal partitioning of the genetic material into the daughter cells. Studies over the last century have revealed that complex dynamic structures such as the spindle microtubules assembled during a specific time in the cell cycle govern these processes in eukaryotic cells. On the contrary, how simple unicellular bacterial cells achieve DNA partitioning has remained elusive for a long time. It is now appreciated how several bacterial genomes and multi-drug resistance carrying plasmids, including the historic Fertility factor or the F plasmid, utilize a Walker A class of cytoskeletal ATPases (WACA) protein ParA (present in Bacteria and Archaea) to partition DNA. While ParA functions as an ATP-dependent motor protein, ParB constitutes the adaptor protein that stimulates the ATPase activity ParA and drives the dynamicity of the system.

Studies on various members of the ParA superfamily of proteins have established that the binding of ParA to non-specific DNA is vital for its function in plasmid maintenance. A chemophoretic gradient of ParA across the bacterial chromosome (nsDNA) is thought to drive the unidirectional movement of the plasmid DNA toward the cell poles. The nsDNA binding activity of ParA itself is regulated by ATP binding and a conformational switch. In the future, it would be important to establish the existence of the conformational switch to the ParA-ATP* state for multiple members of the ParA family of proteins. Further, deciphering the structural changes associated with the conformational switch and kinetics of substrate binding should lead to a better understanding of ParA binding to nsDNA and stimulation of its ATPase activity by ParB. Overall, the structures of ParA proteins are highly conserved, and a unified molecular understanding of the mechanism by which ParA protein functions has emerged. However, given the wide range of cellular functions and evolutionary divergence, it would be fascinating to probe the subtle differences in the molecular mechanisms employed by various ParA proteins.

Moreover, recent EM studies of the $ParA2_{Vch}$ and archaeal SegAB complex revealing a DNA-ParA filament complex are compelling and have reignited the idea that ParA functions require its polymerisation or oligomerisation (Parker et al., 2021; Yen et al., 2021). These studies could revive the polymeric cytoskeleton models proposed for ParA mediated DNA partitioning, especially given the earlier observations on ATP-dependent aggregation or polymerisation of some ParA members (Ebersbach and Gerdes, 2001; Suefuji et al., 2002; Leonard et al., 2004, 2005a; Barillà et al., 2005, 2007; Lim et al., 2005; Bouet et al., 2007; Dunham et al., 2009; Schumacher et al., 2012; Volante and Alonso, 2015). While the physiological

relevance of such polymeric structures formed by ParA *in vitro* is still debated, unequivocal evidence for the existence of such polymeric or oligomeric structures of ParA *in vivo* could be challenging. However, experiments using mutants that have distinct activities of nsDNA binding and polymerisation might help reveal the role of polymerisation/oligomerisation in ParA functions. Moreover, given the overarching conserved design in the structure of the ParA proteins, subtle changes to the structure and conformation might drive these diverse mechanisms. Extensive biochemical and structural studies leveraging the advances in single-molecule techniques, single-particle and cellular correlative-electron microscopy should reveal the underpinnings of mechanisms by which the ParA superfamily of proteins function across domains of life.

It is evident that ParA proteins are vital for bacterial survival and maintenance of virulence factors and antibiotic resistance on plasmids. Since the ParA family of proteins is unique to bacteria and archaea, they could potentially be considered targets for developing new antibiotics. Insights into subtle differences in the interactions of ParA proteins with nsDNA, ParB and itself could be significant for such studies. However, a major challenge could be the development of specific assays and high-throughput screening platforms. The positive selection assay for loss of plasmids developed by Swaine Chen's group could be a way forward (Chen et al., 2017). Understanding the evolution of variations and diverse mechanisms in ParA functions could also be leveraged in synthetic biology applications to design and dictate spatial organization in artificial cell-like systems.

Finally, the recent findings on the CTP binding, hydrolysis and formation of LLPS condensates by ParB have indeed added an entirely new dimension and complexity to our understanding of the plasmid and chromosomal segregation mechanisms. It will be interesting to see the influence of ParB-CTP on ParA functions, especially on the conformational changes to the proposed ParA-ATP* state, nucleoid association and dynamics and polymerisation. Reconstitution experiments and *in vivo* live-cell imaging at high-resolution of the segresome complex from diverse species, together with structural studies employing cryo-electron microscopy and cellular tomography, should help provide more insights into the functioning of these amazing and unique family of proteins.

AUTHOR CONTRIBUTIONS

DM wrote the initial draft. RS wrote sections of the manuscript and revised the manuscript. DM and RS edited the manuscript. Both authors contributed to the article and approved the submitted version.

ACKNOWLEDGMENTS

The authors thank the members of the RS lab and Tirumal K. Chowdary for comments on the manuscript. The authors are particularly grateful to Irene Aniyan Puthenthu for substantial help with editing the manuscript. Support from SERB grant (CRG/2021/000337) and DBT grant (BT/INF/22/SP33046/2019) to RS is acknowledged. The authors also thank the support from DAE. DM acknowledges financial support and fellowship from DAE.

REFERENCES

Abeles, A. L., Friedman, S. A., and Austin, S. J. (1985). Partition of unit-copy miniplasmids to daughter cells. *J. Mol. Biol.* 185, 261–272. doi: 10.1016/0022-2836(85)90402-4

Ah-Seng, Y., Lopez, F., Pasta, F., Lane, D., and Bouet, J.-Y. (2009). Dual role of DNA in regulating ATP hydrolysis by the SopA partition protein. *J. Biol. Chem.* 284, 30067–30075. doi: 10.1074/jbc.M109.044800

Ah-Seng, Y., Rech, J., Lane, D., and Bouet, J.-Y. (2013). Defining the role of ATP hydrolysis in mitotic segregation of bacterial plasmids. *PLoS Genet.* 9, e1003956. doi: 10.1371/journal.pgen.1003956

Ali Azam, T., Iwata, A., Nishimura, A., Ueda, S., and Ishihama, A. (1999). Growth phase-dependent variation in protein composition of the *Escherichia coli* nucleoid. *J. Bacteriol.* 181, 6361–6370. doi: 10.1128/JB.181.20.6361-6370.1999

Alvarado, A., Kjær, A., Yang, W., Mann, P., Briegel, A., Waldor, M. K., et al. (2017). Coupling chemosensory array formation and localization. *eLife* 6, e31058. doi: 10.7554/eLife.31058

Anand, D., Schumacher, D., and Søgaard-Andersen, L. (2020). SMC and the bactofilin/PadC scaffold have distinct yet redundant functions in chromosome segregation and organization in *Myxococcus xanthus*. *Mol. Microbiol.* 114, 839–856. doi: 10.1111/mmi.14583

Anand, S. P., Akhtar, P., Tinsley, E., Watkins, S. C., and Khan, S. A. (2008). GTP-dependent polymerization of the tubulin-like RepX replication protein encoded by the pXO1 plasmid of *Bacillus anthracis*. *Mol. Microbiol.* 67, 881–890. doi: 10.1111/j.1365-2958.2007.06100.x

Austin, S., and Abeles, A. (1983). Partition of unit-copy miniplasmids to daughter cells. *J. Mol. Biol.* 169, 373–387. doi: 10.1016/S0022-2836(83)80056-4

Autret, S., Nair, R., and Errington, J. (2001). Genetic analysis of the chromosome segregation protein Spo0J of *Bacillus subtilis*: evidence for separate domains involved in DNA binding and interactions with Soj protein. *Mol. Microbiol.* 41, 743–755. doi: 10.1046/j.1365-2958.2001.02551.x

Aylett, C. H. S., Izoré, T., Amos, L. A., and Löwe, J. (2013). Structure of the tubulin/FtsZ-like protein TubZ from Pseudomonas bacteriophage *Φ*KZ. *J. Mol. Biol.* 425, 2164–2173. doi: 10.1016/j.jmb.2013.03.019

Aylett, C. H. S., and Löwe, J. (2012). Superstructure of the centromeric complex of TubZRC plasmid partitioning systems. *Proc. Natl. Acad. Sci. U.S.A.* 109, 16522–16527. doi: 10.1073/pnas.1210899109

Aylett, C. H. S., Wang, Q., Michie, K. A., Amos, L. A., and Löwe, J. (2010). Filament structure of bacterial tubulin homologue TubZ. *Proc. Natl. Acad. Sci. U.S.A.* 107, 19766–19771. doi: 10.1073/pnas.1010176107

Babl, L., Giacomelli, G., Ramm, B., Gelmroth, A.-K., Bramkamp, M., and Schwille, P. (2022). CTP-controlled liquid-liquid phase separation of ParB. *J. Mol. Biol.* 434, 167401. doi: 10.1016/j.jmb.2021.167401

Badrinarayanan, A., Le, T. B. K., and Laub, M. T. (2015). Bacterial chromosome organization and segregation. *Annu. Rev. Cell Dev. Biol.* 31, 171–199. doi: 10.1146/annurev-cellbio-100814-125211

Balaguer, F., de, A., Aicart-Ramos, C., Fisher, G. L., de Bragança, S., Martin-Cuevas, E. M., et al. (2021). CTP promotes efficient ParB-dependent DNA condensation by facilitating one-dimensional diffusion from parS. *eLife* 10, e67554. doi: 10.7554/eLife.67554

Barillà, D. (2016). Driving apart and segregating genomes in archaea. *Trends Microbiol.* 24, 957–967. doi: 10.1016/j.tim.2016.07.001

Barillà, D., Carmelo, E., and Hayes, F. (2007). The tail of the ParG DNA segregation protein remodels ParF polymers and enhances ATP hydrolysis via an arginine finger-like motif. *Proc. Natl. Acad. Sci. U.S.A.* 104, 1811–1816. doi: 10.1073/pnas.0607216104

Barillà, D., Rosenberg, M. F., Nobbmann, U., and Hayes, F. (2005). Bacterial DNA segregation dynamics mediated by the polymerizing protein ParF. *EMBO J.* 24, 1453–1464. doi: 10.1038/sj.emboj.7600619

Baronian, G., Ginda, K., Berry, L., Cohen-Gonsaud, M., Zakrzewska-Czerwińska, J., Jakimowicz, D., et al. (2015). Phosphorylation of *Mycobacterium tuberculosis* ParB participates in regulating the ParABS chromosome segregation system. *PLoS ONE* 10, e0119907. doi: 10.1371/journal.pone.0119907

Bartosik, A. A., Lasocki, K., Mierzejewska, J., Thomas, C. M., and Jagura-Burdzy, G. (2004). ParB of *Pseudomonas aeruginosa*: interactions with its partner ParA and its target parS and specific effects on bacterial growth. *J. Bacteriol.* 186, 6983–6998. doi: 10.1128/JB.186.20.6983-6998.2004

Baxter, J. C., Waples, W. G., and Funnell, B. E. (2020). Nonspecific DNA binding by P1 ParA determines the distribution of plasmid partition and repressor activities. *J. Biol. Chem.* 295, 17298–17309. doi: 10.1074/jbc.RA120.015642

Becker, E., Herrera, N. C., Gunderson, F. Q., Derman, A. I., Dance, A. L., Sims, J., et al. (2006). DNA segregation by the bacterial actin AlfA during *Bacillus subtilis* growth and development. *EMBO J.* 25, 5919–5931. doi: 10.1038/sj.emboj.7601443

Bernhardt, T. G., and de Boer, P. A. J. (2005). SlmA, a nucleoid-associated, FtsZ binding protein required for blocking septal ring assembly over Chromosomes in *E. coli*. *Mol. Cell* 18, 555–564. doi: 10.1016/j.molcel.2005.04.012

Birge, E. A. (2006). "Plasmid molecular biology," in *Bacterial and Bacteriophage Genetics* (New York, NY: Springer-Verlag), 417–441. doi: 10.1007/0-387-31489-X_13

Bork, P., Sander, C., and Valencia, A. (1992). An ATPase domain common to prokaryotic cell cycle proteins, sugar kinases, actin, and hsp70 heat shock proteins. *Proc. Natl. Acad. Sci. U.S.A.* 89, 7290–7294. doi: 10.1073/pnas.89.16.7290

Boudsocq, F., Salhi, M., Barbe, S., and Bouet, J.-Y. (2021). Three para dimers cooperatively assemble on type ia partition promoters. *Genes* 12, 1345. doi: 10.3390/genes12091345

Bouet, J.-Y., Ah-Seng, Y., Benmeradi, N., and Lane, D. (2007). Polymerization of SopA partition ATPase: regulation by DNA binding and SopB. *Mol. Microbiol.* 63, 468–481. doi: 10.1111/j.1365-2958.2006.05537.x

Bouet, J. Y., and Funnell, B. E. (1999). P1 ParA interacts with the P1 partition complex at parS and an ATP-ADP switch controls ParA activities. *EMBO J.* 18, 1415–1424. doi: 10.1093/emboj/18.5.1415

Bowman, G. R., Comolli, L. R., Gaietta, G. M., Fero, M., Hong, S.-H., Jones, Y., et al. (2010). Caulobacter PopZ forms a polar subdomain dictating sequential changes in pole composition and function. *Mol. Microbiol.* 76, 173–189. doi: 10.1111/j.1365-2958.2010.07088.x

Bowman, G. R., Comolli, L. R., Zhu, J., Eckart, M., Koenig, M., Downing, K. H., et al. (2008). A polymeric protein anchors the chromosomal origin/ParB complex at a bacterial cell pole. *Cell* 134, 945–955. doi: 10.1016/j.cell.2008.07.015

Bowman, G. R., Perez, A. M., Ptacin, J. L., Ighodaro, E., Folta-Stogniew, E., Comolli, L. R., et al. (2013). Oligomerization and higher-order assembly contribute to sub-cellular localization of a bacterial scaffold. *Mol. Microbiol.* 90, 776–795. doi: 10.1111/mmi.12398

Brangwynne, C. P., Eckmann, C. R., Courson, D. S., Rybarska, A., Hoege, C., Gharakhani, J., et al. (2009). Germline P granules are liquid droplets that localize by controlled dissolution/condensation. *Science* 324, 1729–1732. doi: 10.1126/science.1172046

Breier, A. M., and Grossman, A. D. (2007). Whole-genome analysis of the chromosome partitioning and sporulation protein Spo0J (ParB) reveals spreading and origin-distal sites on the *Bacillus subtilis* chromosome. *Mol. Microbiol.* 64, 703–718. doi: 10.1111/j.1365-2958.2007.05690.x

Brooks, A. C., and Hwang, L. C. (2017). Reconstitutions of plasmid partition systems and their mechanisms. *Plasmid* 91, 37–41. doi: 10.1016/j.plasmid.2017.03.004

Bulyha, I., Lindow, S., Lin, L., Bolte, K., Wuichet, K., Kahnt, J., et al. (2013). Two small GTPases act in concert with the bactofilin cytoskeleton to regulate dynamic bacterial cell polarity. *Dev. Cell* 25, 119–131. doi: 10.1016/j.devcel.2013.02.017

Caccamo, M., Dobruk-Serkowska, A., Rodríguez-Castañeda, F., Pennica, C., Barillà, D., and Hayes, F. (2020). Genome segregation by the venus flytrap mechanism: probing the interaction between the ParF ATPase and the parG centromere binding protein. *Front. Mol. Biosci.* 7, 108. doi: 10.3389/fmolb.2020.00108

Campbell, C. S., and Mullins, R. D. (2007). *In vivo* visualization of type II plasmid segregation: bacterial actin filaments pushing plasmids. *J. Cell Biol.* 179, 1059–1066. doi: 10.1083/jcb.200708206

Castaing, J.-P., Bouet, J.-Y., and Lane, D. (2008). F plasmid partition depends on interaction of SopA with non-specific DNA. *Mol. Microbiol.* 70, 1000–1011. doi: 10.1111/j.1365-2958.2008.06465.x

Chanchal, B. P., and Jain, D. (2017). ATP-induced structural remodeling in the antiactivator FleN enables formation of the functional dimeric form. *Structure* 25, 243–252. doi: 10.1016/j.str.2016.11.022

Chen, B.-W., Lin, M.-H., Chu, C.-H., Hsu, C.-E., and Sun, Y.-J. (2015). Insights into ParB spreading from the complex structure of Spo0J and parS. *Proc. Natl. Acad. Sci. U.S.A.* 112, 6613–6618. doi: 10.1073/pnas.1421927112

Chen, S., Larsson, M., Robinson, R. C., and Chen, S. L. (2017). Direct and convenient measurement of plasmid stability in lab and clinical isolates of *E. coli*. *Sci. Rep.* 7, 4788. doi: 10.1038/s41598-017-05219-x

Chen, Y., and Erickson, H. P. (2008). *In vitro* assembly studies of FtsZ/tubulin-like proteins (TubZ) from Bacillus plasmids: evidence for a capping mechanism. *J. Biol. Chem.* 283, 8102–8109. doi: 10.1074/jbc.M709163200

Chodha, S. S., Brooks, A. C., Davis, P., Ramachandran, R., Chattoraj, D. K., and Hwang, L. C. (2021). Kinetic pathway of ATP-induced DNA interactions of ParA2, a protein essential for segregation of *Vibrio cholerae* chromosome 2. *BioRxiv [Preprint]*. doi: 10.1101/2021.02.27.433207

Chu, C.-H., Yen, C.-Y., Chen, B.-W., Lin, M.-G., Wang, L.-H., Tang, K.-Z., et al. (2019). Crystal structures of HpSoj-DNA complexes and the nucleoid-adaptor complex formation in chromosome segregation. *Nucleic Acids Res.* 47, 2113–2129. doi: 10.1093/nar/gky1251

Cooper, G. M. (2000). "The Eukaryotic Cell Cycle," in *The Cell: A Molecular Approach* (Sunderland: Sinauer Associates). Available online at: https://www.ncbi.nlm.nih.gov/books/NBK9876/

Corrales-Guerrero, L., He, B., Refes, Y., Panis, G., Bange, G., Viollier, P. H., et al. (2020). Molecular architecture of the DNA-binding sites of the P-loop ATPases MipZ and ParA from *Caulobacter crescentus*. *Nucleic Acids Res.* 48, 4769–4779. doi: 10.1093/nar/gkaa192

Correa, N. E., Peng, F., and Klose, K. E. (2005). Roles of the regulatory proteins FlhF and FlhG in the *Vibrio cholerae* flagellar transcription hierarchy. *J. Bacteriol.* 187, 6324–6332. doi: 10.1128/JB.187.18.6324-6332.2005

Dajkovic, A., Lan, G., Sun, S. X., Wirtz, D., and Lutkenhaus, J. (2008). MinC spatially controls bacterial cytokinesis by antagonizing the scaffolding function of FtsZ. *Curr. Biol.* 18, 235–244. doi: 10.1016/j.cub.2008.01.042

Dasgupta, N., Arora, S. K., and Ramphal, R. (2000). *fleN*, a gene that regulates flagellar number in *Pseudomonas aeruginosa*. *J. Bacteriol.* 182, 357–364. doi: 10.1128/JB.182.2.357-364.2000

Dasgupta, N., and Ramphal, R. (2001). Interaction of the antiactivator FleN with the transcriptional activator FleQ regulates flagellar number in *Pseudomonas aeruginosa*. *J. Bacteriol.* 183, 6636–6644. doi: 10.1128/JB.183.22.6636-6644.2001

Dasgupta, N., Wolfgang, M. C., Goodman, A. L., Arora, S. K., Jyot, J., Lory, S., et al. (2003). A four-tiered transcriptional regulatory circuit controls flagellar biogenesis in *Pseudomonas aeruginosa*. *Mol. Microbiol.* 50, 809–824. doi: 10.1046/j.1365-2958.2003.03740.x

Davey, M. J., and Funnell, B. E. (1997). Modulation of the P1 plasmid partition protein ParA by ATP, ADP, and P1 ParB. *J. Biol. Chem.* 272, 15286–15292. doi: 10.1074/jbc.272.24.15286

Davis, M. A., Martin, K. A., and Austin, S. J. (1992). Biochemical activities of the parA partition protein of the P1 plasmid. *Mol. Microbiol.* 6, 1141–1147. doi: 10.1111/j.1365-2958.1992.tb01552.x

Davis, M. A., Radnedge, L., Martin, K. A., Hayes, F., Youngren, B., and Austin, S. J. (1996). The P1 ParA protein and its ATPase activity play a direct role in the segregation of plasmid copies to daughter cells. *Mol. Microbiol.* 21, 1029–1036. doi: 10.1046/j.1365-2958.1996.721423.x

Davis, N. J., Cohen, Y., Sanselicio, S., Fumeaux, C., Ozaki, S., Luciano, J., et al. (2013). De- and repolarization mechanism of flagellar morphogenesis during a bacterial cell cycle. *Genes Dev.* 27, 2049–2062. doi: 10.1101/gad.222679.113

de la Hoz, A. B., Pratto, F., Misselwitz, R., Speck, C., Weihofen, W., Welfle, K., et al. (2004). Recognition of DNA by omega protein from the broad-host range *Streptococcus pyogenes* plasmid pSM19035: analysis of binding to operator

DNA with one to four heptad repeats. *Nucleic Acids Res.* 32, 3136–3147. doi: 10.1093/nar/gkh633

Derman, A. I., Becker, E. C., Truong, B. D., Fujioka, A., Tucey, T. M., Erb, M. L., et al. (2009). Phylogenetic analysis identifies many uncharacterized actin-like proteins (Alps) in bacteria: regulated polymerization, dynamic instability and treadmilling in Alp7A. *Mol. Microbiol.* 73, 534–552. doi: 10.1111/j.1365-2958.2009.06771.x

Di Ventura, B., Knecht, B., Andreas, H., Godinez, W. J., Fritsche, M., Rohr, K., et al. (2013). Chromosome segregation by the *Escherichia coli* Min system. *Mol. Syst. Biol.* 9, 686. doi: 10.1038/msb.2013.44

Ditkowski, B., Holmes, N., Rydzak, J., Donczew, M., Bezulska, M., Ginda, K., et al. (2013). Dynamic interplay of ParA with the polarity protein, Scy, coordinates the growth with chromosome segregation in Streptomyces coelicolor. *Open Biol.* 3, 130006. doi: 10.1098/rsob.130006

Dobruk-Serkowska, A., Caccamo, M., Rodríguez-Castañeda, F., Wu, M., Bryce, K., Ng, I., et al. (2012). Uncoupling of nucleotide hydrolysis and polymerization in the ParA protein superfamily disrupts DNA segregation dynamics. *J. Biol. Chem.* 287, 42545–42553. doi: 10.1074/jbc.M112.410324

Donovan, C., Sieger, B., Krämer, R., and Bramkamp, M. (2012). A synthetic *Escherichia coli* system identifies a conserved origin tethering factor in Actinobacteria. *Mol. Microbiol.* 84, 105–116. doi: 10.1111/j.1365-2958.2012.08011.x

Du, S., and Lutkenhaus, J. (2012). MipZ: one for the pole, two for the DNA. *Mol. Cell* 46, 239–240. doi: 10.1016/j.molcel.2012.04.024

Dunham, T. D., Xu, W., Funnell, B. E., and Schumacher, M. A. (2009). Structural basis for ADP-mediated transcriptional regulation by P1 and P7 ParA. *EMBO J.* 28, 1792–1802. doi: 10.1038/emboj.2009.120

Easter, J., and Gober, J. W. (2002). ParB-stimulated nucleotide exchange regulates a switch in functionally distinct ParA activities. *Mol. Cell* 10, 427–434. doi: 10.1016/s1097-2765(02)00594-4

Ebersbach, G., Briegel, A., Jensen, G. J., and Jacobs-Wagner, C. (2008). A self-associating protein critical for chromosome attachment, division, and polar organization in caulobacter. *Cell* 134, 956–968. doi: 10.1016/j.cell.2008.07.016

Ebersbach, G., and Gerdes, K. (2001). The double par locus of virulence factor pB171: DNA segregation is correlated with oscillation of ParA. *Proc. Natl. Acad. Sci. U.S.A.* 98, 15078–15083. doi: 10.1073/pnas.261569598

Ebersbach, G., and Gerdes, K. (2005). Plasmid segregation mechanisms. *Annu. Rev. Genet.* 39, 453–479. doi: 10.1146/annurev.genet.38.072902.091252

Edwards, D. H., Thomaides, H. B., and Errington, J. (2000). Promiscuous targeting of *Bacillus subtilis* cell division protein DivIVA to division sites in *Escherichia coli* and fission yeast. *EMBO J.* 19, 2719–2727. doi: 10.1093/emboj/19.11.2719

Erb, M. L., Kraemer, J. A., Coker, J. K. C., Chaikeeratisak, V., Nonejuie, P., Agard, D. A., et al. (2014). A bacteriophage tubulin harnesses dynamic instability to center DNA in infected cells. *eLife* 3, e03197. doi: 10.7554/eLife.03197

Ezaki, B., Ogura, T., Niki, H., and Hiraga, S. (1991). Partitioning of a mini-F plasmid into anucleate cells of the mukB null mutant. *J. Bacteriol.* 173, 6643–6646. doi: 10.1128/jb.173.20.6643-6646.1991

Fink, G., and Löwe, J. (2015). Reconstitution of a prokaryotic minus end-tracking system using TubRC centromeric complexes and tubulin-like protein TubZ filaments. *Proc. Natl. Acad. Sci. U.S.A.* 112, E1845–E1850. doi: 10.1073/pnas.1423746112

Fisher, G. L., Pastrana, C. L., Higman, V. A., Koh, A., Taylor, J. A., Butterer, A., et al. (2017). The structural basis for dynamic DNA binding and bridging interactions which condense the bacterial centromere. *eLife* 6, e28086. doi: 10.7554/eLife.28086

Fisher, J. K., Bourniquel, A., Witz, G., Weiner, B., Prentiss, M., and Kleckner, N. (2013). Four-dimensional imaging of *E. coli* nucleoid organization and dynamics in living cells. *Cell* 153, 882–895. doi: 10.1016/j.cell.2013.04.006

Flemming, W. (1882). *Zellsubstanz, Kern und Zelltheilung*. Leipzig: F.C.W. Vogel.

Fogel, M. A., and Waldor, M. K. (2005). Distinct segregation dynamics of the two *Vibrio cholerae* chromosomes. *Mol. Microbiol.* 55, 125–136. doi: 10.1111/j.1365-2958.2004.04379.x

Fogel, M. A., and Waldor, M. K. (2006). A dynamic, mitotic-like mechanism for bacterial chromosome segregation. *Genes Dev.* 20, 3269–3282. doi: 10.1101/gad.1496506

Fothergill, T. J. G., Barillà, D., and Hayes, F. (2005). Protein diversity confers specificity in plasmid segregation. *J. Bacteriol.* 187, 2651–2661. doi: 10.1128/JB.187.8.2651-2661.2005

Friedman, S. A., and Austin, S. J. (1988). The P1 plasmid-partition system synthesizes two essential proteins from an autoregulated operon. *Plasmid* 19, 103–112. doi: 10.1016/0147-619X(88)90049-2

Fung, E., Bouet, J. Y., and Funnell, B. E. (2001). Probing the ATP-binding site of P1 ParA: partition and repression have different requirements for ATP binding and hydrolysis. *EMBO J.* 20, 4901–4911. doi: 10.1093/emboj/20.17.4901

Funnell, B. E. (1991). The P1 plasmid partition complex at parS. The influence of *Escherichia coli* integration host factor and of substrate topology. *J. Biol. Chem.* 266, 14328–14337. doi: 10.1016/S0021-9258(18)98688-6

Funnell, B. E. (2016). Parb partition proteins: complex formation and spreading at bacterial and plasmid centromeres. *Front. Mol. Biosci.* 3, 44. doi: 10.3389/fmolb.2016.00044

Galkin, V. E., Orlova, A., Rivera, C., Mullins, R. D., and Egelman, E. H. (2009). Structural polymorphism of the ParM filament and dynamic instability. *Structure* 17, 1253–1264. doi: 10.1016/j.str.2009.07.008

Garner, E. C., Campbell, C. S., and Mullins, R. D. (2004). Dynamic instability in a DNA-segregating prokaryotic actin homolog. *Science* 306, 1021–1025. doi: 10.1126/science.1101313

Garner, E. C., Campbell, C. S., Weibel, D. B., and Mullins, R. D. (2007). Reconstitution of DNA segregation driven by assembly of a prokaryotic actin homolog. *Science* 315, 1270–1274. doi: 10.1126/science.1138527

Gayathri, P., Fujii, T., Møller-Jensen, J., van den Ent, F., Namba, K., and Löwe, J. (2012). A bipolar spindle of antiparallel ParM filaments drives bacterial plasmid segregation. *Science* 338, 1334–1337. doi: 10.1126/science.1229091

Gayathri, P., Fujii, T., Namba, K., and Löwe, J. (2013). Structure of the ParM filament at 8.5Å resolution. *J. Struct. Biol.* 184, 33–42. doi: 10.1016/j.jsb.2013.02.010

Gellatly, S. L., and Hancock, R. E. W. (2013). *Pseudomonas aeruginosa*: new insights into pathogenesis and host defenses. *Pathog. Dis.* 67, 159–173. doi: 10.1111/2049-632X.12033

Gerdes, K., Howard, M., and Szardenings, F. (2010). Pushing and pulling in prokaryotic DNA segregation. *Cell* 141, 927–942. doi: 10.1016/j.cell.2010.05.033

Gerdes, K., Møller-Jensen, J., and Bugge Jensen, R. (2000). Plasmid and chromosome partitioning: surprises from phylogeny. *Mol. Microbiol.* 37, 455–466. doi: 10.1046/j.1365-2958.2000.01975.x

Ginda, K., Bezulska, M., Ziółkiewicz, M., Dziadek, J., Zakrzewska-Czerwińska, J., and Jakimowicz, D. (2013). ParA of *Mycobacterium smegmatis* co-ordinates chromosome segregation with the cell cycle and interacts with the polar growth determinant DivIVA. *Mol. Microbiol.* 87, 998–1012. doi: 10.1111/mmi.12146

Graham, T. G. W., Wang, X., Song, D., Etson, C. M., van Oijen, A. M., Rudner, D. Z., et al. (2014). ParB spreading requires DNA bridging. *Genes Dev.* 28, 1228–1238. doi: 10.1101/gad.242206.114

Gruber, S., and Errington, J. (2009). Recruitment of condensin to replication origin regions by ParB/SpoOJ promotes chromosome segregation in *B. subtilis*. *Cell* 137, 685–696. doi: 10.1016/j.cell.2009.02.035

Guilhas, B., Walter, J.-C., Rech, J., David, G., Walliser, N. O., Palmeri, J., et al. (2020). ATP-driven separation of liquid phase condensates in bacteria. *Mol. Cell* 79, 293–303.e4. doi: 10.1016/j.molcel.2020.06.034

Gulbronson, C. J., Ribardo, D. A., Balaban, M., Knauer, C., Bange, G., and Hendrixson, D. R. (2016). FlhG employs diverse intrinsic domains and influences FlhF GTPase activity to numerically regulate polar flagellar biogenesis in *Campylobacter jejuni*. *Mol. Microbiol.* 99, 291–306. doi: 10.1111/mmi.13231

Guttenplan, S. B., and Kearns, D. B. (2013). Regulation of flagellar motility during biofilm formation. *FEMS Microbiol. Rev.* 37, 849–871. doi: 10.1111/1574-6976.12018

Hale, C. A., Meinhardt, H., and de Boer, P. A. (2001). Dynamic localization cycle of the cell division regulator MinE in *Escherichia coli*. *EMBO J.* 20, 1563–1572. doi: 10.1093/emboj/20.7.1563

Hanai, R., Liu, R., Benedetti, P., Caron, P. R., Lynch, A. S., and Wang, J. C. (1996). Molecular dissection of a protein SopB essential for *Escherichia coli* F plasmid partition. *J. Biol. Chem.* 271, 17469–17475. doi: 10.1074/jbc.271.29.17469

Hao, J.-J., and Yarmolinsky, M. (2002). Effects of the P1 plasmid centromere on expression of P1 partition genes. *J. Bacteriol.* 184, 4857–4867. doi: 10.1128/JB.184.17.4857-4867.2002

Harry, E. J. (2001). Bacterial cell division: regulating Z-ring formation. *Mol. Microbiol.* 40, 795–803. doi: 10.1046/j.1365-2958.2001.02370.x

Harry, E. J., Rodwell, J., and Wake, R. G. (1999). Co-ordinating DNA replication with cell division in bacteria: a link between the early stages of a round of replication and mid-cell Z ring assembly. *Mol. Microbiol.* 33, 33–40. doi: 10.1046/j.1365-2958.1999.01439.x

Hatano, T., and Niki, H. (2010). Partitioning of P1 plasmids by gradual distribution of the ATPase ParA. *Mol. Microbiol.* 78, 1182–1198. doi: 10.1111/j.1365-2958.2010.07398.x

Hatano, T., Yamaichi, Y., and Niki, H. (2007). Oscillating focus of SopA associated with filamentous structure guides partitioning of F plasmid. *Mol. Microbiol.* 64, 1198–1213. doi: 10.1111/j.1365-2958.2007.05728.x

Havey, J. C., Vecchiarelli, A. G., and Funnell, B. E. (2012). ATP-regulated interactions between P1 ParA, ParB and non-specific DNA that are stabilized by the plasmid partition site, parS. *Nucleic Acids Res.* 40, 801–812. doi: 10.1093/nar/gkr747

Hayakawa, Y., Murotsu, T., and Matsubara, K. (1985). Mini-F protein that binds to a unique region for partition of mini-F plasmid DNA. *J. Bacteriol.* 163, 349–354. doi: 10.1128/jb.163.1.349-354.1985

Hayes, F. (2000). The partition system of multidrug resistance plasmid TP228 includes a novel protein that epitomizes an evolutionarily distinct subgroup of the ParA superfamily. *Mol. Microbiol.* 37, 528–541. doi: 10.1046/j.1365-2958.2000.02030.x

Hayes, F., and Barillà, D. (2006). The bacterial segrosome: a dynamic nucleoprotein machine for DNA trafficking and segregation. *Nat. Rev. Microbiol.* 4, 133–143. doi: 10.1038/nrmicro1342

Hayes, F., Radnedge, L., Davis, M. A., and Austin, S. J. (1994). The homologous operons for P1 and P7 plasmid partition are autoregulated from dissimilar operator sites. *Mol. Microbiol.* 11, 249–260. doi: 10.1111/j.1365-2958.1994.tb00305.x

Heidelberg, J. F., Eisen, J. A., Nelson, W. C., Clayton, R. A., Gwinn, M. L., Dodson, R. J., et al. (2000). DNA sequence of both chromosomes of the cholera pathogen *Vibrio cholerae*. *Nature* 406, 477–483. doi: 10.1038/35020000

Hempel, A. M., Cantlay, S., Molle, V., Wang, S.-B., Naldrett, M. J., Parker, J. L., et al. (2012). The Ser/Thr protein kinase AfsK regulates polar growth and hyphal branching in the filamentous bacteria Streptomyces. *Proc. Natl. Acad. Sci. U.S.A.* 109, E2371–E2379. doi: 10.1073/pnas.1207409109

Hempel, A. M., Wang, S., Letek, M., Gil, J. A., and Flärdh, K. (2008). Assemblies of DivIVA mark sites for hyphal branching and can establish new zones of cell wall growth in Streptomyces coelicolor. *J. Bacteriol.* 190, 7579–7583. doi: 10.1128/JB.00839-08

Hester, C. M., and Lutkenhaus, J. (2007). Soj (ParA) DNA binding is mediated by conserved arginines and is essential for plasmid segregation. *Proc. Natl. Acad. Sci. U.S.A.* 104, 20326–20331. doi: 10.1073/pnas.0705196105

Hirano, M., Mori, H., Onogi, T., Yamazoe, M., Niki, H., Ogura, T., et al. (1998). Autoregulation of the partition genes of the mini-F plasmid and the intracellular localization of their products in *Escherichia coli*. *Mol. Gen. Genet.* 257, 392–403. doi: 10.1007/s004380050663

Holmes, J. A., Follett, S. E., Wang, H., Meadows, C. P., Varga, K., and Bowman, G. R. (2016). Caulobacter PopZ forms an intrinsically disordered hub in organizing bacterial cell poles. *Proc. Natl. Acad. Sci. U.S.A.* 113, 12490–12495. doi: 10.1073/pnas.1602380113

Holmes, N. A., Walshaw, J., Leggett, R. M., Thibessard, A., Dalton, K. A., Gillespie, M. D., et al. (2013). Coiled-coil protein Scy is a key component of a multiprotein assembly controlling polarized growth in Streptomyces. *Proc. Natl. Acad. Sci. U.S.A.* 110, E397–406. doi: 10.1073/pnas.1210657110

Hoshino, S., and Hayashi, I. (2012). Filament formation of the FtsZ/tubulin-like protein TubZ from the Bacillus cereus pXO1 plasmid. *J. Biol. Chem.* 287, 32103–32112. doi: 10.1074/jbc.M112.373803

Howard, M., Rutenberg, A. D., and de Vet, S. (2001). Dynamic compartmentalization of bacteria: accurate division in *E. coli*. *Phys. Rev. Lett.* 87, 278102. doi: 10.1103/PhysRevLett.87.278102

Hu, L., Vecchiarelli, A. G., Mizuuchi, K., Neuman, K. C., and Liu, J. (2015). Directed and persistent movement arises from mechanochemistry of the ParA/ParB system. *Proc. Natl. Acad. Sci. U.S.A.* 112, E7055–E7064. doi: 10.1073/pnas.1505147112

Hu, L., Vecchiarelli, A. G., Mizuuchi, K., Neuman, K. C., and Liu, J. (2017). Brownian ratchet mechanism for faithful segregation of low-copy-number plasmids. *Biophys. J.* 112, 1489–1502. doi: 10.1016/j.bpj.2017.02.039

Hu, Z., Gogol, E. P., and Lutkenhaus, J. (2002). Dynamic assembly of MinD on phospholipid vesicles regulated by ATP and MinE. *Proc. Natl. Acad. Sci. U.S.A.* 99, 6761–6766. doi: 10.1073/pnas.102059099

Hu, Z., and Lutkenhaus, J. (1999). Topological regulation of cell division in *Escherichia coli* involves rapid pole to pole oscillation of the division inhibitor MinC under the control of MinD and MinE. *Mol. Microbiol.* 34, 82–90. doi: 10.1046/j.1365-2958.1999.01575.x

Hu, Z., and Lutkenhaus, J. (2000). Analysis of MinC reveals two independent domains involved in interaction with MinD and FtsZ. *J. Bacteriol.* 182, 3965–3971. doi: 10.1128/JB.182.14.3965-3971.2000

Hu, Z., and Lutkenhaus, J. (2001). Topological regulation of cell division in *E. coli* spatiotemporal oscillation of MinD requires stimulation of its ATPase by MinE and phospholipid. *Mol. Cell* 7, 1337–1343. doi: 10.1016/s1097-2765(01)00273-8

Hu, Z., and Lutkenhaus, J. (2003). A conserved sequence at the C-terminus of MinD is required for binding to the membrane and targeting MinC to the septum. *Mol. Microbiol.* 47, 345–355. doi: 10.1046/j.1365-2958.2003.03321.x

Hu, Z., Mukherjee, A., Pichoff, S., and Lutkenhaus, J. (1999). The MinC component of the division site selection system in *Escherichia coli* interacts with FtsZ to prevent polymerization. *Proc. Natl. Acad. Sci. U.S.A.* 96, 14819–14824. doi: 10.1073/pnas.96.26.14819

Huang, K. C., Meir, Y., and Wingreen, N. S. (2003). Dynamic structures in *Escherichia coli*: spontaneous formation of MinE rings and MinD polar zones. *Proc. Natl. Acad. Sci. U.S.A.* 100, 12724–12728. doi: 10.1073/pnas.2135445100

Hui, M. P., Galkin, V. E., Yu, X., Stasiak, A. Z., Stasiak, A., Waldor, M. K., et al. (2010). ParA2, a *Vibrio cholerae* chromosome partitioning protein, forms left-handed helical filaments on DNA. *Proc. Natl. Acad. Sci. U.S.A.* 107, 4590–4595. doi: 10.1073/pnas.0913060107

Huitema, E., Pritchard, S., Matteson, D., Radhakrishnan, S. K., and Viollier, P. H. (2006). Bacterial birth scar proteins mark future flagellum assembly site. *Cell* 124, 1025–1037. doi: 10.1016/j.cell.2006.01.019

Hwang, L. C., Vecchiarelli, A. G., Han, Y.-W., Mizuuchi, M., Harada, Y., Funnell, B. E., et al. (2013). ParA-mediated plasmid partition driven by protein pattern self-organization. *EMBO J.* 32, 1238–1249. doi: 10.1038/emboj.2013.34

Hyman, A. A., Weber, C. A., and Jülicher, F. (2014). Liquid-liquid phase separation in biology. *Annu. Rev. Cell Dev. Biol.* 30, 39–58. doi: 10.1146/annurev-cellbio-100913-013325

Ireton, K., Gunther, N. W., and Grossman, A. D. (1994). spo0J is required for normal chromosome segregation as well as the initiation of sporulation in *Bacillus subtilis*. *J. Bacteriol.* 176, 5320–5329. doi: 10.1128/jb.176.17.5320-5329.1994

Ivanov, V., and Mizuuchi, K. (2010). Multiple modes of interconverting dynamic pattern formation by bacterial cell division proteins. *Proc. Natl. Acad. Sci. U.S.A.* 107, 8071–8078. doi: 10.1073/pnas.0911036107

Jacob, F., Brenner, S., and Cuzin, F. (1963). On the regulation of DNA replication in bacteria. *Cold Spring Harb. Symp. Quant. Biol.* 28, 329–348. doi: 10.1101/SQB.1963.028.01.048

Jakimowicz, D., Brzostek, A., Rumijowska-Galewicz, A., Zydek, P., Dołzbłasz, A., Smulczyk-Krawczyszyn, A., et al. (2007). Characterization of the mycobacterial chromosome segregation protein ParB and identification of its target in *Mycobacterium smegmatis*. *Microbiology* 153, 4050–4060. doi: 10.1099/mic.0.2007/011619-0

Jalal, A. S., Tran, N. T., and Le, T. B. (2020). ParB spreading on DNA requires cytidine triphosphate *in vitro*. *eLife* 9, e53515. doi: 10.7554/eLife.53515

Jalal, A. S., Tran, N. T., Stevenson, C. E., Chimthanawala, A., Badrinarayanan, A., Lawson, D. M., et al. (2021). A CTP-dependent gating mechanism enables ParB spreading on DNA. *eLife* 10, e69676. doi: 10.7554/eLife.69676

Jalal, A. S. B., and Le, T. B. K. (2020). Bacterial chromosome segregation by the ParABS system. *Open Biol.* 10, 200097. doi: 10.1098/rsob.200097

Jecz, P., Bartosik, A. A., Glabski, K., and Jagura-Burdzy, G. (2015). A single parS sequence from the cluster of four sites closest to oriC is necessary and sufficient for proper chromosome segregation in *Pseudomonas aeruginosa*. *PLoS ONE* 10, e0120867. doi: 10.1371/journal.pone.0120867

Jensen, R. B., and Gerdes, K. (1997). Partitioning of plasmid R1. The ParM protein exhibits ATPase activity and interacts with the centromere-like ParR-parC complex. *J. Mol. Biol.* 269, 505–513. doi: 10.1006/jmbi.1997.1061

Jiang, S., Narita, A., Popp, D., Ghoshdastider, U., Lee, L. J., Srinivasan, R., et al. (2016). Novel actin filaments from *Bacillus thuringiensis* form nanotubules for plasmid DNA segregation. *Proc. Natl. Acad. Sci. U.S.A.* 113, E1200–E1205. doi: 10.1073/pnas.1600129113

Kalliomaa-Sanford, A. K., Rodriguez-Castañeda, F. A., McLeod, B. N., Latorre-Roselló, V., Smith, J. H., Reimann, J., et al. (2012). Chromosome segregation in Archaea mediated by a hybrid DNA partition machine. *Proc. Natl. Acad. Sci. U.S.A.* 109, 3754–3759. doi: 10.1073/pnas.1113384109

Kar, S., Edgar, R., and Adhya, S. (2005). Nucleoid remodeling by an altered HU protein: reorganization of the transcription program. *Proc. Natl. Acad. Sci. U.S.A.* 102, 16397–16402. doi: 10.1073/pnas.0508032102

Kawalek, A., Bartosik, A. A., Glabski, K., and Jagura-Burdzy, G. (2018). *Pseudomonas aeruginosa* partitioning protein ParB acts as a nucleoid-associated protein binding to multiple copies of a parS-related motif. *Nucleic Acids Res.* 46, 4592–4606. doi: 10.1093/nar/gky257

Kawalek, A., Wawrzyniak, P., Bartosik, A. A., and Jagura-Burdzy, G. (2020). Rules and exceptions: the role of chromosomal ParB in DNA segregation and other cellular processes. *Microorganisms* 8, 105. doi: 10.3390/microorganisms8010105

Khare, D., Ziegelin, G., Lanka, E., and Heinemann, U. (2004). Sequence-specific DNA binding determined by contacts outside the helix-turn-helix motif of the ParB homolog KorB. *Nat. Struct. Mol. Biol.* 11, 656–663. doi: 10.1038/nsmb773

Kiekebusch, D., Michie, K. A., Essen, L.-O., Löwe, J., and Thanbichler, M. (2012). Localized dimerization and nucleoid binding drive gradient formation by the bacterial cell division inhibitor MipZ. *Mol. Cell* 46, 245–259. doi: 10.1016/j.molcel.2012.03.004

Kline-Smith, S. L., and Walczak, C. E. (2004). Mitotic spindle assembly and chromosome segregation: refocusing on microtubule dynamics. *Mol. Cell* 15, 317–327. doi: 10.1016/j.molcel.2004.07.012

Koch, M. K., McHugh, C. A., and Hoiczyk, E. (2011). BacM, an N-terminally processed bactofilin of *Myxococcus xanthus*, is crucial for proper cell shape. *Mol. Microbiol.* 80, 1031–1051. doi: 10.1111/j.1365-2958.2011.07629.x

Koh, F., Narita, A., Lee, L. J., Tanaka, K., Tan, Y. Z., Dandey, V. P., et al. (2019). The structure of a 15-stranded actin-like filament from *Clostridium botulinum*. *Nat. Commun.* 10, 2856. doi: 10.1038/s41467-019-10779-9

Köhler, T., Curty, L. K., Barja, F., van Delden, C., and Pechère, J. C. (2000). Swarming of *Pseudomonas aeruginosa* is dependent on cell-to-cell signaling and requires flagella and pili. *J. Bacteriol.* 182, 5990–5996. doi: 10.1128/JB.182.21.5990-5996.2000

Komai, M., Umino, M., and Hanai, R. (2011). Mode of DNA binding by SopA protein of *Escherichia coli* F plasmid. *J. Biochem.* 149, 455–461. doi: 10.1093/jb/mvq151

Koonin, E. V. (1993). A superfamily of ATPases with diverse functions containing either classical or deviant ATP-binding motif. *J. Mol. Biol.* 229, 1165–1174. doi: 10.1006/jmbi.1993.1115

Kraemer, J. A., Erb, M. L., Waddling, C. A., Montabana, E. A., Zehr, E. A., Wang, H., et al. (2012). A phage tubulin assembles dynamic filaments by an atypical mechanism to center viral DNA within the host cell. *Cell* 149, 1488–1499. doi: 10.1016/j.cell.2012.04.034

Kretschmer, S., and Schwille, P. (2016). Pattern formation on membranes and its role in bacterial cell division. *Curr. Opin. Cell Biol.* 38, 52–59. doi: 10.1016/j.ceb.2016.02.005

Kühn, J., Briegel, A., Mörschel, E., Kahnt, J., Leser, K., Wick, S., et al. (2010). Bactofilins, a ubiquitous class of cytoskeletal proteins mediating polar localization of a cell wall synthase in *Caulobacter crescentus*. *EMBO J.* 29, 327–339. doi: 10.1038/emboj.2009.358

Kusiak, M., Gapczynska, A., Plochocka, D., Thomas, C. M., and Jagura-Burdzy, G. (2011). Binding and spreading of ParB on DNA determine its biological function in *Pseudomonas aeruginosa*. *J. Bacteriol.* 193, 3342–3355. doi: 10.1128/JB.00328-11

Lagage, V., Boccard, F., and Vallet-Gely, I. (2016). Regional control of chromosome segregation in *Pseudomonas aeruginosa*. *PLoS Genet.* 12, e1006428. doi: 10.1371/journal.pgen.1006428

Lam, H., Schofield, W. B., and Jacobs-Wagner, C. (2006). A landmark protein essential for establishing and perpetuating the polarity of a bacterial cell. *Cell* 124, 1011–1023. doi: 10.1016/j.cell.2005.12.040

Lane, D., Rothenbuehler, R., Merrillat, A. M., and Aiken, C. (1987). Analysis of the F plasmid centromere. *Mol. Gen. Genet.* 207, 406–412. doi: 10.1007/bf00331608

Larsen, R. A., Cusumano, C., Fujioka, A., Lim-Fong, G., Patterson, P., and Pogliano, J. (2007). Treadmilling of a prokaryotic tubulin-like protein, TubZ, required for plasmid stability in *Bacillus thuringiensis*. *Genes Dev.* 21, 1340–1352. doi: 10.1101/gad.1546107

Lasocki, K., Bartosik, A. A., Mierzejewska, J., Thomas, C. M., and Jagura-Burdzy, G. (2007). Deletion of the parA (soj) homologue in *Pseudomonas aeruginosa* causes ParB instability and affects growth rate, chromosome segregation, and motility. *J. Bacteriol.* 189, 5762–5772. doi: 10.1128/JB.00371-07

Le Gall, A., Cattoni, D. I., Guilhas, B., Mathieu-Demazière, C., Oudjedi, L., Fiche, J.-B., et al. (2016). Bacterial partition complexes segregate within the volume of the nucleoid. *Nat. Commun.* 7, 12107. doi: 10.1038/ncomms12107

Le Quéré, B., and Ghigo, J.-M. (2009). BcsQ is an essential component of the *Escherichia coli* cellulose biosynthesis apparatus that localizes at the bacterial cell pole. *Mol. Microbiol.* 72, 724–740. doi: 10.1111/j.1365-2958.2009.06678.x

Lee, M.-J., Liu, C.-H., Wang, S.-Y., Huang, C.-T., and Huang, H. (2006). Characterization of the Soj/Spo0J chromosome segregation proteins and identification of putative parS sequences in *Helicobacter pylori*. *Biochem. Biophys. Res. Commun.* 342, 744–750. doi: 10.1016/j.bbrc.2006.01.173

Lee, P. S., and Grossman, A. D. (2006). The chromosome partitioning proteins Soj (ParA) and Spo0J (ParB) contribute to accurate chromosome partitioning, separation of replicated sister origins, and regulation of replication initiation in *Bacillus subtilis*. *Mol. Microbiol.* 60, 853–869. doi: 10.1111/j.1365-2958.2006.05140.x

Lenarcic, R., Halbedel, S., Visser, L., Shaw, M., Wu, L. J., Errington, J., et al. (2009). Localisation of DivIVA by targeting to negatively curved membranes. *EMBO J.* 28, 2272–2282. doi: 10.1038/emboj.2009.129

Leonard, T. A., Butler, P. J., and Löwe, J. (2005a). Bacterial chromosome segregation: structure and DNA binding of the Soj dimer–a conserved biological switch. *EMBO J.* 24, 270–282. doi: 10.1038/sj.emboj.7600530

Leonard, T. A., Butler, P. J. G., and Löwe, J. (2004). Structural analysis of the chromosome segregation protein Spo0J from *Thermus thermophilus*. *Mol. Microbiol.* 53, 419–432. doi: 10.1111/j.1365-2958.2004.04133.x

Leonard, T. A., Møller-Jensen, J., and Löwe, J. (2005b). Towards understanding the molecular basis of bacterial DNA segregation. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 360, 523–535. doi: 10.1098/rstb.2004.1608

Libante, V., Thion, L., and Lane, D. (2001). Role of the ATP-binding site of SopA protein in partition of the F plasmid. *J. Mol. Biol.* 314, 387–399. doi: 10.1006/jmbi.2001.5158

Lim, G. E., Derman, A. I., and Pogliano, J. (2005). Bacterial DNA segregation by dynamic SopA polymers. *Proc. Natl. Acad. Sci. U.S.A.* 102, 17658–17663. doi: 10.1073/pnas.0507222102

Lim, H. C., Surovtsev, I. V., Beltran, B. G., Huang, F., Bewersdorf, J., and Jacobs-Wagner, C. (2014). Evidence for a DNA-relay mechanism in ParABS-mediated chromosome segregation. *eLife* 3, e02758. doi: 10.7554/eLife.02758

Lin, L., Osorio Valeriano, M., Harms, A., Søgaard-Andersen, L., and Thanbichler, M. (2017). Bactofilin-mediated organization of the ParABS chromosome segregation system in *Myxococcus xanthus*. *Nat. Commun.* 8, 1817. doi: 10.1038/s41467-017-02015-z

Lin, L., and Thanbichler, M. (2013). Nucleotide-independent cytoskeletal scaffolds in bacteria. *Cytoskeleton* 70, 409–423. doi: 10.1002/cm.21126

Lin, Z., and Mallavia, L. P. (1998). Membrane association of active plasmid partitioning protein A in *Escherichia coli*. *J. Biol. Chem.* 273, 11302–11312. doi: 10.1074/jbc.273.18.11302

Livny, J., Yamaichi, Y., and Waldor, M. K. (2007). Distribution of centromere-like parS sites in bacteria: insights from comparative genomics. *J. Bacteriol.* 189, 8693–8703. doi: 10.1128/JB.01239-07

Loose, M., Fischer-Friedrich, E., Herold, C., Kruse, K., and Schwille, P. (2011). Min protein patterns emerge from rapid rebinding and membrane interaction of MinE. *Nat. Struct. Mol. Biol.* 18, 577–583. doi: 10.1038/nsmb.2037

Loose, M., Fischer-Friedrich, E., Ries, J., Kruse, K., and Schwille, P. (2008). Spatial regulators for bacterial cell division self-organize into surface waves *in vitro*. *Science* 320, 789–792. doi: 10.1126/science.1154413

Losada, A., and Hirano, T. (2005). Dynamic molecular linkers of the genome: the first decade of SMC proteins. *Genes Dev.* 19, 1269–1287. doi: 10.1101/gad.1320505

Löwe, J., and Amos, L. A. (1998). Crystal structure of the bacterial cell-division protein FtsZ. *Nature* 391, 203–206. doi: 10.1038/34472

Lutkenhaus, J. (2012). The ParA/MinD family puts things in their place. *Trends Microbiol.* 20, 411–418. doi: 10.1016/j.tim.2012.05.002

Lutkenhaus, J., and Sundaramoorthy, M. (2003). MinD and role of the deviant Walker A motif, dimerization and membrane binding in oscillation. *Mol. Microbiol.* 48, 295–303. doi: 10.1046/j.1365-2958.2003.03427.x

MacCready, J. S., Basalla, J. L., and Vecchiarelli, A. G. (2020). Origin and evolution of carboxysome positioning systems in cyanobacteria. *Mol. Biol. Evol.* 37, 1434–1451. doi: 10.1093/molbev/msz308

MacCready, J. S., Hakim, P., Young, E. J., Hu, L., Liu, J., Osteryoung, K. W., et al. (2018). Protein gradients on the nucleoid position the carbon-fixing organelles of cyanobacteria. *eLife* 7, e39723. doi: 10.7554/eLife.39723

Madariaga-Marcos, J., Pastrana, C. L., Fisher, G. L., Dillingham, M. S., and Moreno-Herrero, F. (2019). ParB dynamics and the critical role of the CTD in DNA condensation unveiled by combined force-fluorescence measurements. *eLife* 8, e43812. doi: 10.7554/eLife.43812

Maloney, E., Madiraju, M., and Rajagopalan, M. (2009). Overproduction and localization of *Mycobacterium tuberculosis* ParA and ParB proteins. *Tuberculosis* 89(Suppl. 1), S65–S69. doi: 10.1016/S1472-9792(09)70015-0

McLeod, B. N., Allison-Gamble, G. E., Barge, M. T., Tonthat, N. K., Schumacher, M. A., Hayes, F., et al. (2017). A three-dimensional ParF meshwork assembles through the nucleoid to mediate plasmid segregation. *Nucleic Acids Res.* 45, 3158–3171. doi: 10.1093/nar/gkw1302

McLeod, B. N., and Spiegelman, G. B. (2005). Soj antagonizes Spo0A activation of transcription in *Bacillus subtilis*. *J. Bacteriol.* 187, 2532–2536. doi: 10.1128/JB.187.7.2532-2536.2005

Meinhardt, H., and de Boer, P. A. (2001). Pattern formation in *Escherichia coli*: a model for the pole-to-pole oscillations of Min proteins and the localization of the division site. *Proc. Natl. Acad. Sci. U.S.A.* 98, 14202–14207. doi: 10.1073/pnas.251216598

Michie, K. A., and Löwe, J. (2006). Dynamic filaments of the bacterial cytoskeleton. *Annu. Rev. Biochem.* 75, 467–492. doi: 10.1146/annurev.biochem.75.103004.142452

Million-Weaver, S., and Camps, M. (2014). Mechanisms of plasmid segregation: have multicopy plasmids been overlooked? *Plasmid* 75, 27–36. doi: 10.1016/j.plasmid.2014.07.002

Mishra, D., Pahujani, S., Mitra, N., Srivastava, A., and Srinivasan, R. (2021). Identification of a potential membrane-targeting sequence in the C-terminus of the F plasmid segregation protein SopA. *J. Membr. Biol.* 254, 243–257. doi: 10.1007/s00232-020-00157-8

Miyakoshi, M., Shintani, M., Inoue, K., Terabayashi, T., Sai, F., Ohkuma, M., et al. (2012). ParI, an orphan ParA family protein from *Pseudomonas putida* KT2440-specific genomic island, interferes with the partition system of IncP-7 plasmids. *Environ. Microbiol.* 14, 2946–2959. doi: 10.1111/j.1462-2920.2012.02861.x

Mohl, D. A., and Gober, J. W. (1997). Cell cycle-dependent polar localization of chromosome partitioning proteins in *Caulobacter crescentus*. *Cell* 88, 675–684. doi: 10.1016/s0092-8674(00)81910-8

Møller-Jensen, J., Borch, J., Dam, M., Jensen, R. B., Roepstorff, P., and Gerdes, K. (2003). Bacterial mitosis: ParM of plasmid R1 moves plasmid DNA by an actin-like insertional polymerization mechanism. *Mol. Cell* 12, 1477–1487. doi: 10.1016/s1097-2765(03)00451-9

Montabana, E. A., and Agard, D. A. (2014). Bacterial tubulin TubZ-Bt transitions between a two-stranded intermediate and a four-stranded filament upon GTP hydrolysis. *Proc. Natl. Acad. Sci. U.S.A.* 111, 3407–3412. doi: 10.1073/pnas.1318339111

Monterroso, B., Zorrilla, S., Sobrinos-Sanguino, M., Robles-Ramos, M. A., López-Álvarez, M., Margolin, W., et al. (2019). Bacterial FtsZ protein forms phase-separated condensates with its nucleoid-associated inhibitor SlmA. *EMBO Rep.* 20:e45946. doi: 10.15252/embr.201845946

Mori, H., Kondo, A., Ohshima, A., Ogura, T., and Hiraga, S. (1986). Structure and function of the F plasmid genes essential for partitioning. *J. Mol. Biol.* 192, 1–15. doi: 10.1016/0022-2836(86)90459-6

Mori, H., Mori, Y., Ichinose, C., Niki, H., Ogura, T., Kato, A., et al. (1989). Purification and characterization of SopA and SopB proteins essential for F plasmid partitioning. *J. Biol. Chem.* 264, 15535–15541. doi: 10.1016/S0021-9258(19)84863-9

Motallebi-Veshareh, M., Rouch, D. A., and Thomas, C. M. (1990). A family of ATPases involved in active partitioning of diverse bacterial plasmids. *Mol. Microbiol.* 4, 1455–1463. doi: 10.1111/j.1365-2958.1990.tb02056.x

Mukherjee, S., and Kearns, D. B. (2014). The structure and regulation of flagella in *Bacillus subtilis*. *Annu. Rev. Genet.* 48, 319–340. doi: 10.1146/annurev-genet-120213-092406

Nasmyth, K., and Haering, C. H. (2005). The structure and function of SMC and kleisin complexes. *Annu. Rev. Biochem.* 74, 595–648. doi: 10.1146/annurev.biochem.74.082803.133219

Ni, L., Xu, W., Kumaraswami, M., and Schumacher, M. A. (2010). Plasmid protein TubR uses a distinct mode of HTH-DNA binding and recruits the prokaryotic tubulin homolog TubZ to effect DNA partition. *Proc. Natl. Acad. Sci. U.S.A.* 107, 11763–11768. doi: 10.1073/pnas.1003817107

Niki, H., Jaffé, A., Imamura, R., Ogura, T., and Hiraga, S. (1991). The new gene mukB codes for a 177 kd protein with coiled-coil domains involved in chromosome partitioning of *E. coli*. *EMBO J.* 10, 183–193.

Nogales, E., Downing, K. H., Amos, L. A., and Löwe, J. (1998). Tubulin and FtsZ form a distinct family of GTPases. *Nat. Struct. Biol.* 5, 451–458. doi: 10.1038/nsb0698-451

Ogura, T., and Hiraga, S. (1983). Partition mechanism of F plasmid: two plasmid gene-encoded products and a cis-acting region are involved in partition. *Cell* 32, 351–360. doi: 10.1016/0092-8674(83)90454-3

Oliva, M. A., Halbedel, S., Freund, S. M., Dutow, P., Leonard, T. A., Veprintsev, D. B., et al. (2010). Features critical for membrane binding revealed by DivIVA crystal structure. *EMBO J.* 29, 1988–2001. doi: 10.1038/emboj.2010.99

Oliva, M. A., Martin-Galiano, A. J., Sakaguchi, Y., and Andreu, J. M. (2012). Tubulin homolog TubZ in a phage-encoded partition system. *Proc. Natl. Acad. Sci. U.S.A.* 109, 7711–7716. doi: 10.1073/pnas.1121546109

Orlova, A., Garner, E. C., Galkin, V. E., Heuser, J., Mullins, R. D., and Egelman, E. H. (2007). The structure of bacterial ParM filaments. *Nat. Struct. Mol. Biol.* 14, 921–926. doi: 10.1038/nsmb1300

Osorio-Valeriano, M., Altegoer, F., Das, C. K., Steinchen, W., Panis, G., Connolley, L., et al. (2021). The CTPase activity of ParB determines the size and dynamics of prokaryotic DNA partition complexes. *Mol. Cell* 81, 3992–4007.e10. doi: 10.1016/j.molcel.2021.09.004

O'Toole, G. A., and Kolter, R. (1998). Flagellar and twitching motility are necessary for *Pseudomonas aeruginosa* biofilm development. *Mol. Microbiol.* 30, 295–304. doi: 10.1046/j.1365-2958.1998.01062.x

Park, K.-T., Wu, W., Battaile, K. P., Lovell, S., Holyoak, T., and Lutkenhaus, J. (2011). The Min oscillator uses MinD-dependent conformational changes in MinE to spatially regulate cytokinesis. *Cell* 146, 396–407. doi: 10.1016/j.cell.2011.06.042

Parker, A. V., Mann, D., Tzokov, S. B., Hwang, L. C., and Bergeron, J. R. C. (2021). The structure of the bacterial DNA segregation ATPase filament reveals the conformational plasticity of ParA upon DNA binding. *Nat. Commun.* 12, 5166. doi: 10.1038/s41467-021-25429-2

Perry, S. E., and Edwards, D. H. (2006). The *Bacillus subtilis* DivIVA protein has a sporulation-specific proximity to Spo0J. *J. Bacteriol.* 188, 6039–6043. doi: 10.1128/JB.01750-05

Piekarski, G. (1937). Cytologische untersuchungen an paratyphus-und colibakterien. *Archiv Für Mikrobiologie* 8, 428–439.

Popp, D., Narita, A., Lee, L. J., Ghoshdastider, U., Xue, B., Srinivasan, R., et al. (2012). Novel actin-like filament structure from *Clostridium tetani*. *J. Biol. Chem.* 287, 21121–21129. doi: 10.1074/jbc.M112.341016

Popp, D., Narita, A., Oda, T., Fujisawa, T., Matsuo, H., Nitanai, Y., et al. (2008). Molecular structure of the ParM polymer and the mechanism leading to its nucleotide-driven dynamic instability. *EMBO J.* 27, 570–579. doi: 10.1038/sj.emboj.7601978

Pratto, F., Cicek, A., Weihofen, W. A., Lurz, R., Saenger, W., and Alonso, J. C. (2008). *Streptococcus pyogenes* pSM19035 requires dynamic assembly of ATP-bound ParA and ParB on parS DNA during plasmid segregation. *Nucleic Acids Res.* 36, 3676–3689. doi: 10.1093/nar/gkn170

Protter, D. S. W., and Parker, R. (2016). Principles and properties of stress granules. *Trends Cell Biol.* 26, 668–679. doi: 10.1016/j.tcb.2016.05.004

Ptacin, J. L., Gahlmann, A., Bowman, G. R., Perez, A. M., von Diezmann, L., Eckart, M. R., et al. (2014). Bacterial scaffold directs pole-specific centromere segregation. *Proc. Natl. Acad. Sci. U.S.A.* 111, E2046–E2055. doi: 10.1073/pnas.1405188111

Ptacin, J. L., Lee, S. F., Garner, E. C., Toro, E., Eckart, M., Comolli, L. R., et al. (2010). A spindle-like apparatus guides bacterial chromosome segregation. *Nat. Cell Biol.* 12, 791–798. doi: 10.1038/ncb2083

Punta, M., Coggill, P. C., Eberhardt, R. Y., Mistry, J., Tate, J., Boursnell, C., et al. (2012). The Pfam protein families database. *Nucleic Acids Res.* 40, D290–301. doi: 10.1093/nar/gkr1065

Ramamurthi, K. S., and Losick, R. (2009). Negative membrane curvature as a cue for subcellular localization of a bacterial protein. *Proc. Natl. Acad. Sci. U.S.A.* 106, 13541–13545. doi: 10.1073/pnas.0906851106

Ramm, B., Glock, P., Mücksch, J., Blumhardt, P., García-Soriano, D. A., Heymann, M., et al. (2018). The MinDE system is a generic spatial cue for membrane protein distribution *in vitro*. *Nat. Commun.* 9, 3942. doi: 10.1038/s41467-018-06310-1

Raskin, D. M., and de Boer, P. A. (1999). Rapid pole-to-pole oscillation of a protein required for directing division to the middle of *Escherichia coli*. *Proc. Natl. Acad. Sci. U.S.A.* 96, 4971–4976. doi: 10.1073/pnas.96.9.4971

Ravin, N. V., Rech, J., and Lane, D. (2003). Mapping of functional domains in F plasmid partition proteins reveals a bipartite SopB-recognition domain in SopA. *J. Mol. Biol.* 329, 875–889. doi: 10.1016/s0022-2836(03)00525-4

Ringgaard, S., Schirner, K., Davis, B. M., and Waldor, M. K. (2011). A family of ParA-like ATPases promotes cell pole maturation by facilitating polar localization of chemotaxis proteins. *Genes Dev.* 25, 1544–1555. doi: 10.1101/gad.2061811

Ringgaard, S., van Zon, J., Howard, M., and Gerdes, K. (2009). Movement and equipositioning of plasmids by ParA filament disassembly. *Proc. Natl. Acad. Sci. U.S.A.* 106, 19369–19374. doi: 10.1073/pnas.0908347106

Ringgaard, S., Yang, W., Alvarado, A., Schirner, K., and Briegel, A. (2018). Chemotaxis arrays in Vibrio species and their intracellular positioning by the ParC/ParP system. *J. Bacteriol.* 200, e00793–e00717. doi: 10.1128/JB.00793-17

Roberts, M. A. J., Wadhams, G. H., Hadfield, K. A., Tickner, S., and Armitage, J. P. (2012). ParA-like protein uses nonspecific chromosomal DNA binding to partition protein complexes. *Proc. Natl. Acad. Sci. U.S.A.* 109, 6698–6703. doi: 10.1073/pnas.1114000109

Rodionov, O., Lobocka, M., and Yarmolinsky, M. (1999). Silencing of genes flanking the P1 plasmid centromere. *Science* 283, 546–549. doi: 10.1126/science.283.5401.546

Rothfield, L., Taghbalout, A., and Shih, Y.-L. (2005). Spatial control of bacterial division-site placement. *Nat. Rev. Microbiol.* 3, 959–968. doi: 10.1038/nrmicro1290

Salje, J., Gayathri, P., and Löwe, J. (2010). The ParMRC system: molecular mechanisms of plasmid segregation by actin-like filaments. *Nat. Rev. Microbiol.* 8, 683–692. doi: 10.1038/nrmicro2425

Salje, J., Zuber, B., and Löwe, J. (2009). Electron cryomicroscopy of *E. coli* reveals filament bundles involved in plasmid DNA segregation. *Science* 323, 509–512. doi: 10.1126/science.1164346

Santi, I., and McKinney, J. D. (2015). Chromosome organization and replisome dynamics in *Mycobacterium smegmatis*. *MBio* 6, e01999–e01914. doi: 10.1128/mBio.01999-14

Sassetti, C. M., and Rubin, E. J. (2003). Genetic requirements for mycobacterial survival during infection. *Proc. Natl. Acad. Sci. U.S.A.* 100, 12989–12994. doi: 10.1073/pnas.2134250100

Savage, D. F., Afonso, B., Chen, A. H., and Silver, P. A. (2010). Spatially ordered dynamics of the bacterial carbon fixation machinery. *Science* 327, 1258–1261. doi: 10.1126/science.1186090

Schofield, W. B., Lim, H. C., and Jacobs-Wagner, C. (2010). Cell cycle coordination and regulation of bacterial chromosome segregation dynamics by polarly localized proteins. *EMBO J.* 29, 3068–3081. doi: 10.1038/emboj.2010.207

Scholefield, G., Whiting, R., Errington, J., and Murray, H. (2011). Spo0J regulates the oligomeric state of Soj to trigger its switch from an activator to an inhibitor of DNA replication initiation. *Mol. Microbiol.* 79, 1089–1100. doi: 10.1111/j.1365-2958.2010.07507.x

Scholey, J. M., Brust-Mascher, I., and Mogilner, A. (2003). Cell division. *Nature* 422, 746–752. doi: 10.1038/nature01599

Schuhmacher, J. S., Rossmann, F., Dempwolff, F., Knauer, C., Altegoer, F., Steinchen, W., et al. (2015). MinD-like ATPase FlhG effects location and number of bacterial flagella during C-ring assembly. *Proc. Natl. Acad. Sci. U.S.A.* 112, 3092–3097. doi: 10.1073/pnas.1419388112

Schumacher, M. A. (2007). Structural biology of plasmid segregation proteins. *Curr. Opin. Struct. Biol.* 17, 103–109. doi: 10.1016/j.sbi.2006.11.005

Schumacher, M. A. (2008). Structural biology of plasmid partition: uncovering the molecular mechanisms of DNA segregation. *Biochem. J.* 412, 1–18. doi: 10.1042/BJ20080359

Schumacher, M. A., and Funnell, B. E. (2005). Structures of ParB bound to DNA reveal mechanism of partition complex formation. *Nature* 438, 516–519. doi: 10.1038/nature04149

Schumacher, M. A., Henderson, M., and Zhang, H. (2019). Structures of maintenance of carboxysome distribution Walker-box McdA and McdB adaptor homologs. *Nucleic Acids Res.* 47, 5950–5962. doi: 10.1093/nar/gkz314

Schumacher, M. A., Tonthat, N. K., Lee, J., Rodriguez-Castañeda, F. A., Chinnam, N. B., Kalliomaa-Sanford, A. K., et al. (2015). Structures of archaeal DNA segregation machinery reveal bacterial and eukaryotic linkages. *Science* 349, 1120–1124. doi: 10.1126/science.aaa9046

Schumacher, M. A., Ye, Q., Barge, M. T., Zampini, M., Barillà, D., and Hayes, F. (2012). Structural mechanism of ATP-induced polymerization of the partition factor ParF: implications for DNA segregation. *J. Biol. Chem.* 287, 26146–26154. doi: 10.1074/jbc.M112.373696

She, Q., Phan, H., Garrett, R. A., Albers, S. V., Stedman, K. M., and Zillig, W. (1998). Genetic profile of pNOB8 from Sulfolobus: the first conjugative plasmid from an archaeon. *Extremophiles* 2, 417–425. doi: 10.1007/s007920050087

She, Q., Singh, R. K., Confalonieri, F., Zivanovic, Y., Allard, G., Awayez, M. J., et al. (2001). The complete genome of the crenarchaeon Sulfolobus solfataricus P2. *Proc. Natl. Acad. Sci. U.S.A.* 98, 7835–7840. doi: 10.1073/pnas.141222098

Soberón, N. E., Lioy, V. S., Pratto, F., Volante, A., and Alonso, J. C. (2011). Molecular anatomy of the *Streptococcus pyogenes* pSM19035 partition and segrosome complexes. *Nucleic Acids Res.* 39, 2624–2637. doi: 10.1093/nar/gkq1245

Soh, Y.-M., Davidson, I. F., Zamuner, S., Basquin, J., Bock, F. P., Taschner, M., et al. (2019). Self-organization of parS centromeres by the ParB CTP hydrolase. *Science* 366, 1129–1133. doi: 10.1126/science.aay3965

Stephens, C., Arismendi, T., Wright, M., Hartman, A., Gonzalez, A., Gill, M., et al. (2020). F plasmids are the major carriers of antibiotic resistance genes in human-associated commensal *Escherichia coli*. *mSphere* 5, e00709-20. doi: 10.1128/mSphere.00709-20

Stover, C. K., Pham, X. Q., Erwin, A. L., Mizoguchi, S. D., Warrener, P., Hickey, M. J., et al. (2000). Complete genome sequence of *Pseudomonas aeruginosa* PAO1, an opportunistic pathogen. *Nature* 406, 959–964. doi: 10.1038/35023079

Suefuji, K., Valluzzi, R., and RayChaudhuri, D. (2002). Dynamic assembly of MinD into filament bundles modulated by ATP, phospholipids, and MinE. *Proc. Natl. Acad. Sci. U.S.A.* 99, 16776–16781. doi: 10.1073/pnas.262671699

Sun, Q., and Margolin, W. (2001). Influence of the nucleoid on placement of FtsZ and MinE rings in *Escherichia coli*. *J. Bacteriol.* 183, 1413–1422. doi: 10.1128/JB.183.4.1413-1422.2001

Surovtsev, I. V., Campos, M., and Jacobs-Wagner, C. (2016). DNA-relay mechanism is sufficient to explain ParA-dependent intracellular transport and patterning of single and multiple cargos. *Proc. Natl. Acad. Sci. U.S.A.* 113, E7268–E7276. doi: 10.1073/pnas.1616118113

Szeto, T. H., Rowland, S. L., Rothfield, L. I., and King, G. F. (2002). Membrane localization of MinD is mediated by a C-terminal motif that is conserved across eubacteria, archaea, and chloroplasts. *Proc. Natl. Acad. Sci. U.S.A.* 99, 15693–15698. doi: 10.1073/pnas.232590599

Tang, M., Bideshi, D. K., Park, H.-W., and Federici, B. A. (2006). Minireplicon from pBtoxis of *Bacillus thuringiensis* subsp. israelensis. *Appl. Environ. Microbiol.* 72, 6948–6954. doi: 10.1128/AEM.00976-06

Tang, M., Bideshi, D. K., Park, H.-W., and Federici, B. A. (2007). Iteron-binding ORF157 and FtsZ-like ORF156 proteins encoded by pBtoxis play a role in its replication in *Bacillus thuringiensis* subsp. *israelensis*. *J. Bacteriol.* 189, 8053–8058. doi: 10.1128/JB.00908-07

Thanbichler, M., and Shapiro, L. (2006). MipZ, a spatial regulator coordinating chromosome segregation with cell division in Caulobacter. *Cell* 126, 147–162. doi: 10.1016/j.cell.2006.05.038

Thompson, S. R., Wadhams, G. H., and Armitage, J. P. (2006). The positioning of cytoplasmic protein clusters in bacteria. *Proc. Natl. Acad. Sci. U.S.A.* 103, 8209–8214. doi: 10.1073/pnas.0600919103

Tinsley, E., and Khan, S. A. (2006). A novel FtsZ-like protein is involved in replication of the anthrax toxin-encoding pXO1 plasmid in *Bacillus anthracis*. *J. Bacteriol.* 188, 2829–2835. doi: 10.1128/JB.188.8.2829-2835.2006

Toro, E., Hong, S.-H., McAdams, H. H., and Shapiro, L. (2008). Caulobacter requires a dedicated mechanism to initiate chromosome segregation. *Proc. Natl. Acad. Sci. U.S.A.* 105, 15435–15440. doi: 10.1073/pnas.0807448105

Trojanowski, D., Ginda, K., Pióro, M., Hołówka, J., Skut, P., Jakimowicz, D., et al. (2015). Choreography of the Mycobacterium replication machinery during the cell cycle. *MBio* 6, e02125–e02114. doi: 10.1128/mBio.02125-14

Uhía, I., Priestman, M., Joyce, G., Krishnan, N., Shahrezaei, V., and Robertson, B. D. (2018). Analysis of ParAB dynamics in mycobacteria shows active movement of ParB and differential inheritance of ParA. *PLoS ONE* 13, e0199316. doi: 10.1371/journal.pone.0199316

van Baarle, S., Celik, I. N., Kaval, K. G., Bramkamp, M., Hamoen, L. W., and Halbedel, S. (2013). Protein-protein interaction domains of *Bacillus subtilis* DivIVA. *J. Bacteriol.* 195, 1012–1021. doi: 10.1128/JB.02171-12

van den Ent, F., Møller-Jensen, J., Amos, L. A., Gerdes, K., and Löwe, J. (2002). F-actin-like filaments formed by plasmid segregation protein ParM. *EMBO J.* 21, 6935–6943. doi: 10.1093/emboj/cdf672

van Ditmarsch, D., Boyle, K. E., Sakhtah, H., Oyler, J. E., Nadell, C. D., Déziel, É., et al. (2013). Convergent evolution of hyperswarming leads to impaired biofilm formation in pathogenic bacteria. *Cell Rep.* 4, 697–708. doi: 10.1016/j.celrep.2013.07.026

Vecchiarelli, A. G., Han, Y.-W., Tan, X., Mizuuchi, M., Ghirlando, R., Biertümpfel, C., et al. (2010). ATP control of dynamic P1 ParA-DNA interactions: a key role for the nucleoid in plasmid partition. *Mol. Microbiol.* 78, 78–91. doi: 10.1111/j.1365-2958.2010.07314.x

Vecchiarelli, A. G., Havey, J. C., Ing, L. L., Wong, E. O. Y., Waples, W. G., and Funnell, B. E. (2013a). Dissection of the ATPase active site of P1 ParA reveals multiple active forms essential for plasmid partition. *J. Biol. Chem.* 288, 17823–17831. doi: 10.1074/jbc.M113.469981

Vecchiarelli, A. G., Hwang, L. C., and Mizuuchi, K. (2013b). Cell-free study of F plasmid partition provides evidence for cargo transport by a diffusion-ratchet mechanism. *Proc. Natl. Acad. Sci. U.S.A.* 110, E1390–E1397. doi: 10.1073/pnas.1302745110

Vecchiarelli, A. G., Li, M., Mizuuchi, M., Hwang, L. C., Seol, Y., Neuman, K. C., et al. (2016). Membrane-bound MinDE complex acts as a toggle switch that drives Min oscillation coupled to cytoplasmic depletion of MinD. *Proc. Natl. Acad. Sci. U.S.A.* 113, E1479–E1488. doi: 10.1073/pnas.1600644113

Vecchiarelli, A. G., Mizuuchi, K., and Funnell, B. E. (2012). Surfing biological surfaces: exploiting the nucleoid for partition and transport in bacteria. *Mol. Microbiol.* 86, 513–523. doi: 10.1111/mmi.12017

Vecchiarelli, A. G., Neuman, K. C., and Mizuuchi, K. (2014). A propagating ATPase gradient drives transport of surface-confined cellular cargo. *Proc. Natl. Acad. Sci. U.S.A.* 111, 4880–4885. doi: 10.1073/pnas.1401025111

Vecchiarelli, A. G., Taylor, J. A., and Mizuuchi, K. (2015). Reconstituting ParA/ParB-mediated transport of DNA cargo. *Methods Cell Biol.* 128, 243–269. doi: 10.1016/bs.mcb.2015.01.021

Venkova-Canova, T., Baek, J. H., Fitzgerald, P. C., Blokesch, M., and Chattoraj, D. K. (2013). Evidence for two different regulatory mechanisms linking replication and segregation of *Vibrio cholerae* chromosome II. *PLoS Genet.* 9, e1003579. doi: 10.1371/journal.pgen.1003579

Verma, S. C., Qian, Z., and Adhya, S. L. (2019). Architecture of the *Escherichia coli* nucleoid. *PLoS Genet.* 15, e1008456. doi: 10.1371/journal.pgen.1008456

Volante, A., and Alonso, J. C. (2015). Molecular anatomy of ParA-ParA and ParA-ParB interactions during plasmid partitioning. *J. Biol. Chem.* 290, 18782–18795. doi: 10.1074/jbc.M115.649632

Walczak, C. E., Cai, S., and Khodjakov, A. (2010). Mechanisms of chromosome behaviour during mitosis. *Nat. Rev. Mol. Cell Biol.* 11, 91–102. doi: 10.1038/nrm2832

Walker, J. E., Saraste, M., Runswick, M. J., and Gay, N. J. (1982). Distantly related sequences in the alpha- and beta-subunits of ATP synthase, myosin, kinases and other ATP-requiring enzymes and a common nucleotide binding fold. *EMBO J.* 1, 945–951. doi: 10.1002/j.1460-2075.1982.tb01276.x

Wang, H., Yan, X., Aigner, H., Bracher, A., Nguyen, N. D., Hee, W. Y., et al. (2019). Rubisco condensate formation by CcmM in β-carboxysome biogenesis. *Nature* 566, 131–135. doi: 10.1038/s41586-019-0880-5

Watanabe, E., Wachi, M., Yamasaki, M., and Nagai, K. (1992). ATPase activity of SopA, a protein essential for active partitioning of F plasmid. *Mol. Gen. Genet.* 234, 346–352. doi: 10.1007/BF00538693

Weihofen, W. A., Cicek, A., Pratto, F., Alonso, J. C., and Saenger, W. (2006). Structures of omega repressors bound to direct and inverted DNA repeats explain modulation of transcription. *Nucleic Acids Res.* 34, 1450–1458. doi: 10.1093/nar/gkl015

Wendler, P., Ciniawsky, S., Kock, M., and Kube, S. (2012). Structure and function of the AAA+ nucleotide binding pocket. *Biochim. Biophys. Acta* 1823, 2–14. doi: 10.1016/j.bbamcr.2011.06.014

Wu, L. J., and Errington, J. (2003). RacA and the Soj-Spo0J system combine to effect polar chromosome segregation in sporulating *Bacillus subtilis*. *Mol. Microbiol.* 49, 1463–1475. doi: 10.1046/j.1365-2958.2003.03643.x

Wu, W., Park, K.-T., Holyoak, T., and Lutkenhaus, J. (2011). Determination of the structure of the MinD-ATP complex reveals the orientation of MinD on the membrane and the relative location of the binding sites for MinE and MinC. *Mol. Microbiol.* 79, 1515–1528. doi: 10.1111/j.1365-2958.2010.07536.x

Yamaichi, Y., Fogel, M. A., and Waldor, M. K. (2007). par genes and the pathology of chromosome loss in *Vibrio cholerae*. *Proc. Natl. Acad. Sci. U.S.A.* 104, 630–635. doi: 10.1073/pnas.0608341104

Yamaichi, Y., and Niki, H. (2000). Active segregation by the *Bacillus subtilis* partitioning system in *Escherichia coli*. *Proc. Natl. Acad. Sci. U.S.A.* 97, 14656–14661. doi: 10.1073/pnas.97.26.14656

Yamaichi, Y., and Niki, H. (2004). migS, a cis-acting site that affects bipolar positioning of oriC on the *Escherichia coli* chromosome. *EMBO J.* 23, 221–233. doi: 10.1038/sj.emboj.7600028

Yen, C.-Y., Lin, M.-G., Chen, B.-W., Ng, I. W., Read, N., Kabli, A. F., et al. (2021). Chromosome segregation in Archaea: SegA- and SegB-DNA complex structures provide insights into segrosome assembly. *Nucleic Acids Res.* 49, 13150–13164. doi: 10.1093/nar/gkab1155

Yu, X. C., and Margolin, W. (1999). FtsZ ring clusters in min and partition mutants: role of both the Min system and the nucleoid in regulating FtsZ ring localization. *Mol. Microbiol.* 32, 315–326. doi: 10.1046/j.1365-2958.1999.01351.x

Zehr, E. A., Kraemer, J. A., Erb, M. L., Coker, J. K. C., Montabana, E. A., Pogliano, J., et al. (2014). The structure and assembly mechanism of a novel three-stranded tubulin filament that centers phage DNA. *Structure* 22, 539–548. doi: 10.1016/j.str.2014.02.006

Zhang, H., and Schumacher, M. A. (2017). Structures of partition protein ParA with nonspecific DNA and ParB effector reveal molecular insights into principles governing Walker-box DNA segregation. *Genes Dev.* 31, 481–492. doi: 10.1101/gad.296319.117

Zhou, H., and Lutkenhaus, J. (2003). Membrane binding by MinD involves insertion of hydrophobic residues within the C-terminal amphipathic helix into the bilayer. *J. Bacteriol.* 185, 4326–4335. doi: 10.1128/JB.185.15.4326-4335.2003

Zieske, K., and Schwille, P. (2013). Reconstitution of pole-to-pole oscillations of min proteins in microengineered polydimethylsiloxane compartments. *Angew. Chem. Int. Ed. Engl.* 52, 459–462. doi: 10.1002/anie.201207078

Zieske, K., and Schwille, P. (2014). Reconstitution of self-organizing protein gradients as spatial cues in cell-free systems. *eLife* 3, e03949. doi: 10.7554/eLife.03949

Zimmerman, S. B. (2003). Underlying regularity in the shapes of nucleoids of *Escherichia coli*: implications for nucleoid organization and partition. *J. Struct. Biol.* 142, 256–265. doi: 10.1016/s1047-8477(02)00637-8

Zuckerman, D. M., Boucher, L. E., Xie, K., Engelhardt, H., Bosch, J., and Hoiczyk, E. (2015). The bactofilin cytoskeleton protein BacM of *Myxococcus xanthus* forms an extended β-sheet structure likely mediated by hydrophobic interactions. *PLoS ONE* 10, e0121074. doi: 10.1371/journal.pone.0121074

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

frontiers | Frontiers in Microbiology

ORIGINAL RESEARCH
published: 15 June 2022
doi: 10.3389/fmicb.2022.898938

Cat8 Response to Nutritional Changes and Interaction With Ehrlich Pathway Related Factors

Zhengda Du[1,2†], Hong Deng[1,2†], Yanfei Cheng[1], Zhiguang Zhai[3], Xuena Guo[1], Zhaoyue Wang[1] and Xiuping He[1,2]**

[1] CAS Key Laboratory of Microbial Physiological and Metabolic Engineering, State Key Laboratory of Mycology, Institute of Microbiology, Chinese Academy of Sciences, Beijing, China, [2] College of Life Sciences, University of the Chinese Academy of Sciences, Beijing, China, [3] Institute of Basic Theory for Chinese Medicine, China Academy of Chinese Medical Sciences, Beijing, China

OPEN ACCESS

Edited by:
Ulrike Kappler,
The University of Queensland,
Australia

Reviewed by:
Mingtao Huang,
South China University of Technology,
China
Jingwen Zhou,
Jiangnan University, China
Liang Sun,
University of Wisconsin–Madison,
United States

***Correspondence:**
Zhaoyue Wang
wangzy@im.ac.cn
Xiuping He
hexp@im.ac.cn

[†] These authors have contributed equally to this work and share first authorship

Specialty section:
This article was submitted to Microbial Physiology and Metabolism, a section of the journal Frontiers in Microbiology

Received: 18 March 2022
Accepted: 13 May 2022
Published: 15 June 2022

Citation:
Du Z, Deng H, Cheng Y, Zhai Z, Guo X, Wang Z and He X (2022) Cat8 Response to Nutritional Changes and Interaction With Ehrlich Pathway Related Factors. Front. Microbiol. 13:898938. doi: 10.3389/fmicb.2022.898938

Cat8 is an important transcription factor regulating the utilization of non-fermentative carbon sources in *Saccharomyces cerevisiae*. However, our previous studies found that Cat8 may play a critical role in nitrogen metabolism, but the regulatory mechanism has not been elucidated. In this study, the nuclear localization and analysis of regulatory activity showed that the Cat8 function relies on Snf1 kinase. In the fermentation with glucose or glycerol as carbon sources under phenylalanine (Phe) induction, by comparing the changes of cellular gene expression and Cat8 target gene binding profiles after Cat8 overexpression, enhanced transcription was shown among key genes involved in the Ehrlich pathway (e.g., *ARO9*, *ARO10*, and *ADH2*) and its upstream and downstream related factors (e.g., *GAP1*, *AGP1*, *GAT1*, *PDR12*, and *ESPB6*), indicating that Cat8 participated in the regulation of nitrogen metabolism. Moreover, highly active Cat8 interacts with transcriptional activator Aro80 and GATA activator Gat1 coordinately to regulate the transcription of *ARO10*. Altogether, our results showed that Cat8 may act as a global transcription factor in response to nutritional changes, regulating both carbon and nitrogen utilization. This provides a new insight for us to explore the regulation of cell nutrient metabolism networks in yeast.

Keywords: Cat8, Snf1, Ehrlich pathway, RNA-seq, ChIP-seq, Aro80, Gat1

INTRODUCTION

Cat8, a zinc cluster transcriptional activator, is necessary for regulating gene expression during yeast diauxic shift from fermentative metabolism to respiratory metabolism (Haurie et al., 2001). Cat8 activates many metabolic pathways such as gluconeogenesis, tricarboxylic acid (TCA) cycle, the glyoxylate cycle, β-oxidation of fatty acids, and respiratory chain to enable cells to utilize non-fermentative carbon sources (Hedges et al., 1995; Bonander et al., 2008; Laera et al., 2016). The expression and transcriptional activity of Cat8 are regulated complicatedly: In the presence of glucose, the transcription factor Mig1 (C_2H_2 zinc finger protein) becomes dephosphorylated and localizes to the nucleus, repressing the expression of *CAT8*. When glucose is depleted, the Snf1 kinase complex (AMP-activated S/T protein kinase) is activated and phosphorylates Mig1 to remove the repression of Mig1, allowing *CAT8* to initiate transcription [reviewed in

Zaman et al. (2008)]. In addition, Cat8 is controlled by the carbon source at a post-transcriptional level, which undergoes multi-step phosphorylation before activating the transcription of target genes (Randez-Gil et al., 1997), among them, the Snf1 kinase complex plays a dominant role in the phosphorylation of Cat8. However, the regulatory mode of Cat8 phosphorylation and related kinases are still unclear (Turcotte et al., 2010). It has been reported that the Cat8 middle homology region (MHR) may harm the transcriptional activation domain (TAD) (**Supplementary Figure 1**; Delaveau et al., 1994; Schjerling and Holmberg, 1996), and phosphorylation of Cat8 by kinases appears to relieve this negative effect (Charbon et al., 2004). Furthermore, overexpression of Cat8 under repressive conditions (high glucose, 2%) increases the expression of CSRE (carbon source-responsive element)-dependent genes, but the transcriptional activity of Cat8 is still affected by the activity of the carbon source-controlled TAD (Hedges et al., 1995; Rahner et al., 1996, 1999). The post-translational modification of Cat8 regulated by carbon source produces different degrees of activation of its transcriptional activity (Charbon et al., 2004), which may be the reason why yeast can perform refined transcriptional reprogramming in response to different types of carbon sources. Collectively, Cat8 transcription and transcriptional activity co-regulate cellular responses to changes in carbon sources in the environment.

Cat8 is an important transcription factor for the response to the carbon source, some researchers ever predicted Cat8 and Mig1 as key transcriptional regulators controlling the differential expression of the genes affected by aromatic alcohol communication (Wuster and Babu, 2010). Aromatic alcohols 2-phenylethanol (2-PE) and tryptophol are quorum-sensing signaling molecules for fungal filamentous growth (Chen and Fink, 2006), which are synthesized from Phe and tryptophan (Trp) in a three-step reaction of transamination, decarboxylation, and reduction via Ehrlich pathway (Etschmann et al., 2002). Up to now, the regulation mechanism of the Ehrlich pathway is not yet very clear and only a few regulatory factors have been reported. It has been reported that the transcription factors Aro80 and GATA activators (Gln3 and Gat1) are involved in the regulation of the Ehrlich pathway in *Saccharomyces cerevisiae* (MacPherson et al., 2006; Eden et al., 2007; Lee and Hahn, 2013). In our previous study on the regulation of 2-PE synthesis, we found that Cat8 could promote yeast cells to synthesize 2-PE, moreover, Cat8 overexpression strain in Phe medium significantly up-regulated the expression of key enzymes in the Ehrlich pathway (Wang et al., 2017), suggesting that Cat8 is not only significant in carbon metabolism regulation but also may play an essential role in regulating the Ehrlich pathway and nitrogen source metabolism.

For a deeper understanding of Cat8 role in nitrogen metabolism response, we carried out the investigation on the differences between the expression, functional activity, and regulation of Cat8 under two conditions of fermentation and respiration in yeast in response to nitrogen source metabolism. Our results suggest that highly active Cat8 bound to the promoter region and enhanced the transcription of key genes involved in the Ehrlich pathway and its upstream and downstream related genes. Moreover, Cat8 interacts with transcriptional activator Aro80 and GATA activator Gat1 coordinately to regulate the transcription of *ARO10*. Altogether, the data support the view that Cat8 is involved in the regulation of nitrogen source metabolism.

MATERIALS AND METHODS

Media and Yeast Cultivation Conditions

Yeast cells were grown in a standard YPD or SC medium at 30°C. The solid medium contains 20 g/L agar. YPD medium consisted of 10 g/L yeast extract, 20 g/L peptones, and 20 g/L glucose. The defined synthetic complete (SC) medium consisted of 6.7 g/L yeast nitrogen base (YNB), 20 g/L carbon sources (glucose or glycerol), and 20 mg/L uracil (Ura), and 40 mg/L each of Trp, leucine (Leu), and histidine (His). The fermentation medium is obtained by adding 2 g/L Phe to the SC medium containing 20 g/L glucose (Glc) or glycerol (Gly).

Strain Construction

Yeast strains used in this work are derived from *S. cerevisiae* YS58 and listed in **Supplementary Table 1**. The *S. cerevisiae* YS58 strain was kindly given by Prof. Teunissen and was derived from diploid yeast YS60 (Teunissen et al., 1993). All DNA fragments integrated into the genome were done by the method of homologous recombination. The 3 × HA tag was connected to the carboxyl terminus of Cat8 by introducing Zeocin resistance to obtain the YS58-HA strain, and the promoter of *CAT8* was replaced with the *TEF2* promoter by introducing G418 resistance to obtain YS58-CAT8-HA strain. Based on YS58-HA and YS58-CAT8-HA strains, YS58-HA Δsnf1 and YS58-CAT8-HA Δsnf1 strains were obtained by introducing hygromycin resistance and deleting *SNF1*. The following strains with gene deletion were obtained using the scarless knockout method created by our laboratory (Zhou et al., 2021), including the strains YS58-Δcat8, YS58-Δaro80, and YS58-Δgln3gat1. *CAT8* self-promoter was further replaced with *TEF2* promoter by the seamless replacement method to obtain YS58-CAT8 Δaro80 and YS58-CAT8 Δgln3gat1 strains. The YCpA-CAT8-GFP plasmid was introduced into YS58-Δcat8 to obtain the CAT8-EGFP strain. CAT8-EGFP Δsnf1 strain was then obtained by using hygromycin to knock out *SNF1* in the CAT8-EGFP strain. Plasmids YEpA-ARO80-Myc, YEpA-GLN3-Myc, and YEpA-GAT1-Myc were introduced into the YS58-CAT8-HA strain, respectively, then YS58-CAT8-HA Aro80-Myc, YS58-CAT8-HA Gln3-Myc, and YS58-CAT8-HA Gat1-Myc strains were obtained.

Gene Cloning and Plasmids Construction

Plasmids used in this study were listed in **Supplementary Table 2**. YCpA-CAT8-EGFP backbone was derived from pYC-AGA constructed in our laboratory. The enzyme digestion at the *Bam*HI site linearized pYC-AGA. 4.3 k *CAT8* fragment was amplified from the YS58 genome. The linearized plasmid and *CAT8* fragment were introduced into YS58-Δcat8 strain and homologous recombination resulted in the YCpA-CAT8-EGFP plasmid. *ADH1* promoter and terminator were inserted into

the multi-clone sites of YEp352 to obtain YEpA. The YEpA plasmid was linearized, PCR amplified 2.8, 2.2, and 1.5 k *ARO80*, *GLN3*, and *GAT1* fragments from the YS58 genome, and also amplified Myc tag from the commercial plasmid. The linearized plasmid, Myc tag, and gene fragment were mixed and cloned with ClonExpress Ultra One Step Cloning Kit (Vazyme Biotech Co., Ltd., China) to obtain YEpA-ARO80-Myc, YEpA-GLN3-Myc, and YEpA-GAT1-Myc plasmids.

2-Phenylethanol Detection

Cell density was tested at OD_{600}. 2-PE was quantified by HPLC (Agilent 1260 HPLC, United States) equipped with a diode array detector (DAD) using a 250 mm C-18 column (guige; Agilent Technologies, United States). The mobile phase was 50% methanol, the column temperature was 30°C, the flow rate was 1.0 mL/min, the detection wavelength was 206 nm, and the peak time of 2-PE was 7.3 min.

Microscopic Analysis of Protein Localization

Yeast cells were cultured in YPD until the early log phase (0.8–1 OD), then transferred to 250 ml induction medium containing 2% Glc or Gly for 6 h. In all subsequent experiments, unless specific, the strain culture method was the same as that mentioned here. Fluorescence signals were detected by confocal microscope (Leica SP8), and localization of nuclei was detected by staining DNA with 2.5 μg/ml of 4, 6 diamidino-2-phenylindol (DAPI).

Western Blotting Analysis

Cells were washed 3 times with cold PBS (pH 7.4) [137 mM NaCl, 2.7 mM KCl, 8 mM Na_2HPO_4, and 1.8 mM KH_2PO_4], collected by rapid centrifugation at 4°C, and immediately frozen in liquid nitrogen. The cell pellets were ground with liquid nitrogen and added 150 μl PBS [Before using, add 1% PMSF (Thermo Fisher Scientific Inc., United States) and 1% PIC (Beijing Solarbio Science & Technology Co., Ltd., China)] for ice bathing for about 10 min, and then centrifuged at 10,000 rpm for 10 min. The appropriate amount of supernatant was mixed with 5 × SDS loading buffer, boiled for 10 min, and centrifuged at 10,000 rpm for 1 min, supernatants were stored at −80°C. Western blot was performed on 6 or 8% polyacrylamide gels as previously described (Laemmli, 1970; Towbin et al., 1979). Cat8 was detected by incubation with monoclonal anti-HA tag antibody (dilution of 1:5,000; Abcam, United States) and peroxidase-conjugated goat anti-rabbit antibody (dilution of 1:4,000; EASYBIO, China). Aro80, Gln3, and Gat1 were detected by incubation with monoclonal anti-Myc tag antibody (dilution of 1:4,000; Mei5 Biotechnology Co., Ltd., China) and peroxidase-conjugated goat anti-mouse antibody (dilution of 1:3,000; EASYBIO, China). β-Actin was detected by incubation with a β-Actin rabbit polyclonal antibody (dilution of 1:4,000; Mei5 Biotechnology Co., Ltd., China). Bound antibody was detected by using the Pierce ECL Western Blotting Substrate (Thermo Fisher Scientific Inc., United States) as instructed by the manufacturer.

RNA Preparation and Quantitative RT-PCR (qRT-PCR)

The yeast cells were collected by centrifuging at 4°C and washed with pre-cooled PBS and then crushed by liquid nitrogen grinding. The total RNA was obtained by using the GeneJET RNA Purification Kit (Thermo Fisher Scientific Inc., United States) and the relative amount of specific mRNA was determined by qRT-PCR. Briefly, 2 μg of total RNA was subjected to reverse transcription in a 20 μl 1 × FastKing-RT SuperMix reaction mixture (TIANGEN Biotech Co., Ltd., Beijing, China), and then reacted in 1 × SuperReal PreMix Plus (SYBR Green) reaction mixture (TIANGEN Biotech Co., Ltd., Beijing, China) using Roche LightCycler 96 real-time PCR system (Roche Diagnostics, Germany). Expression data were processed by the second-derivative maximum method of LightCycler 96 software SW1.1. The reference gene was *ACT1* and fold changes were calculated using the $2^{-\Delta\Delta CT}$ method (Livak and Schmittgen, 2001).

RNA-Seq, Gene Expression Profiling, and Gene Ontology (GO) Enrichment Analysis

RNA samples were analyzed by RNA-seq (Novogene Co., Ltd., China) on an Illumina NovaSeq 6000 platform (Illumina Inc., United States), and at least 2 Gb clean reads were obtained for each sample. The detailed data was deposited in National Microbiology Data Center (NMDC) with accession numbers NMDC10018022. The FASTX toolkit[1] was used to evaluate the raw data and discard low-quality reads. The clean reads were aligned to the reference genome of strain S288C using HISAT2 (Pertea et al., 2016), and the numbers of reads mapped to each gene were counted by HTSeq (Anders et al., 2015). The Q30 was > 94% of the reads were mapped to the reference genome. The differentially expressed genes were screened and annotated with DESeq2 (Anders and Huber, 2010; Anders et al., 2015). A corrected $p \leq 0.05$ and a log2 fold change ≤ -1 or ≥ 1 were set as the threshold for significantly differential expression. GO functional annotation and the KEGG pathway enrichment analysis were performed by the Database for Annotation, Visualization and Integrated Discovery (DAVID) v6.8 (Huang et al., 2009). To understand the relationships between rich clusters, we use Metascape to build the network of selected enriched terms (Zhou et al., 2019).

ChIP-Seq and Bioinformatic Analysis

Chromatin Immunoprecipitation (ChIP) experiment was conducted as previously described with some modifications (Cam and Whitehall, 2016). The YS58-CAT8-HA strain was cultured in YPD until the early log phase (0.8–1 OD), then transferred to a Phe induction medium containing 2% Glc or Gly for 6 h. Cells were cross-linked with 1% formaldehyde for 20 min and quenched with glycine (final concentration 125 mM) for 15 min at 30°C. Cells were washed 3 times with cold PBS (pH 7.4), quickly frozen in liquid nitrogen, and stored at −80°C for later use. The cell pellet was ground with liquid nitrogen, and 0.6 g of powder was added to 12 ml of pre-chilled ChIP Buffer I

[1]https://github.com/agordon/fastx_toolkit/

[50 mM HEPES-KOH (pH 7.5), 140 mM NaCl, 1% Triton X-100, 1 mM EDTA, and 0.1% Sodium Deoxycholate] [Before using, add 1% PMSF (Thermo Fisher Scientific Inc., United States) and 1% PIC (Beijing Solarbio Science & Technology Co., Ltd., China)] and mixed, and the lysate was divided into six copies into 15 ml of ultrasound tubes, 2 ml each. The chromatin was interrupted by ultrasound with Bioruptor Plus (Diagenode, Belgium), high-power ultrasound for 30 s, and 30 s pause for 1 time, a total of 30 times. The best DNA size after the ultrasound was 250–500 bp. About 20 μl Dynabeads® Protein A (Thermo Fisher Scientific Inc., United States) was washed with ChIP Buffer I, added to the chromatin solution prepared above, and incubated at 4°C for 2–3 h for pre-washing, separated and discarded the magnetic beads. About 500 μl of supernatant was transferred to a new centrifuge tube labeled "input" as the input group and stored at 4°C. 6 μg of monoclonal anti-HA (Abcam, United States) was added to the pre-washed chromatin and incubated overnight at 4°C with rotation, then centrifuged at 12,000 rpm at 4°C for 10 min, and the supernatant was transferred to a new centrifuge tube. The washed magnetic beads (40 μl) were added to the supernatant and incubated at 4°C with rotation for 2–3 h, then the magnetic beads were separated and the supernatant was discarded. Pre-chilled 1 ml ChIP Buffer I, ChIP Buffer II [50 mM HEPES-KOH (pH 7.5), 500 mM NaCl, 1% Triton X-100, 1 mM EDTA, and 0.1% Sodium Deoxycholate], and ChIP Buffer III [10 mM Tris–HCl (pH 8.0), 0.25 M Lithium chloride monohydrate, 1% NP-40, 1 mM EDTA, and 0.1% Sodium Deoxycholate] were used to wash the magnetic beads and each buffer was used to wash the beads at 4°C for 15, 5, and 5 min, respectively. The magnetic beads were washed with 1 ml of TE buffer [100 mM Tris–HCl (pH 8.0) and 10 mM EDTA (pH 8.0)] and the supernatant was discarded. The beads were centrifuged at 3,000 rpm for 1 min at room temperature to remove residual TE buffer. The magnetic beads were incubated with 250 μl Elution buffer [50 mM Tris–HCl (pH 8.0), 10 mM EDTA (pH 8.0), and 1% SDS] at 65°C for 10 min to elute the protein complexes, eluted twice and transferred the supernatant to a new centrifuge tube, labeled "ChIP." About 20 μl of 5 M NaCl (final concentration 0.2 M) was added to "ChIP" and "input" tubes and incubated overnight at 65°C to decrosslink. High-quality DNA was purified and concentrated by ChIP DNA Clean & Concentrator (Zymo Research Corporation, United States).

The samples were sequenced using the Illumina NovaSeq 6000 platform (Illumina Inc., United States) for high-throughput sequencing (Annoroad Gene Technology Beijing Co., Ltd., China), and the sequencing read length was PE150. A total of 30.8 G data volume was obtained for sequencing, with an average Q30 of 93.25%. The detailed data was deposited in National Microbiology Data Center (NMDC) with accession number: NMDC10018022. The raw data was quality controlled by FastQC, and then Bowtie2 was used for sequence alignment (Langmead and Salzberg, 2012). ChIP peaks were visualized with Integrative Genomics Viewer (IGV) version 2.7.2, a genomic data set viewer for visualization of genomic features (Robinson et al., 2011). We used peak-calling algorithm MACS (Model-based Analysis for ChIP-seq) version 2.2.7.1 (Zhang et al., 2008; Park, 2009), a commonly used peak-caller with more accurate results than competing peak-callers (Thomas et al., 2017). ChIP peaks were annotated by ChIPseeker (Yu et al., 2015), which integrates ChIP annotation, comparison, and visualization and serves as a toolbox for the analysis of ChIP-seq data.

Co-immunoprecipitation Assay

Co-Immunoprecipitation (Co-IP) was performed as previously described (Björck and Kronvall, 1984; Harlow and Lane, 1988; Wikström et al., 1995), YS58-CAT8-HA Aro80-Myc strain, YS58-CAT8-HA Gln3-Myc strain, and YS58-CAT8-HA Gat1-Myc strain were cultured in YPD until early log phase (0.8–1 OD), then transferred to Phe induction medium containing 2% Gly for 6 h. A total of 50 mg cells were collected, ground with liquid nitrogen and added IP Lysis/Wash Buffer (pH7.4) [25 mM Tris, 150 mM NaCl, 1 mM EDTA, 1% NP-40, and 5% glycerol] 1.5 ml [Before use, add 1% PMSF (Thermo Fisher Scientific Inc., United States) and 1% PIC (Beijing Solarbio Science & Technology Co., Ltd., China)]. After centrifugation at 10,000 rpm for 10 min, lysate supernatant containing 1–2 mg total protein was incubated with 20 μl magnetic beads (washed twice with IP Lysis/Wash Buffer) for pre-cleaning at 4°C for 2 h. The magnetic beads were separated and some of the supernatants were preserved as Input at −20°C. The remaining supernatants were added with an appropriate amount of antibodies and rotated at 4°C for overnight incubation. About 20 μl magnetic beads were added to the supernatant containing antigen-antibody complex and incubated at 4°C for 1 h. Separate the magnetic beads from the supernatant (the supernatant can be preserved until the IP experiment is successful), then wash the magnetic beads with 1 ml IP Lysis/Wash Buffer three times rotating for 10 min at 4°C each time. The magnetic beads were transferred to a new centrifugal tube and 500 μl Saline Solution [150 mM NaCl] was added to clean the magnetic beads for 10 min at 4°C. The separated magnetic beads were mixed with 50 μl 2 × SDS PAGE Sample Buffer [100 mM Tris (pH 6.8), 40 mM DTT, 2% SDS, 20% glycerol, and 0.2% bromophenol blue] and boiled for 5–8 min at 100°C. The supernatant was taken and the isolated binding protein was analyzed by Western blot.

Mass Spectrometry and Post-translational Modification Analysis

The proteins were digested with trypsin and subjected to an EASY-nLC 1000 interfaced via a Nanospray Flex ion source to an Orbitrap Fusion Tribrid mass spectrometer (Thermo Fisher Scientific, United States) (nano-LC-MS/MS) analysis at the Technological Platform of Mass Spectrum Centre of Institute of Microbiology, Chinese Academy of Sciences. The peptides were loaded onto a trap column (C18, 3 μm particles, 100 μm ID, 3 cm length, Dr. Maisch GmbH) and separated using an analytical column (C18, 1.9 μm particles, 150 μm ID, 15 cm length, Dr. Maisch GmbH) at a flow rate of 500 nl/min with a 60 min LC gradient composed of Solvent A (0.1% formic acid) and Solvent B (acetonitrile, 0.1% formic acid). The gradient was 3–8% B for 5 min, 8–22% B for 40 min, 22–35% B for 10 min, 35–90% B for 3 min, and finally 90% B for 2 min. The mass spectrometer was operated in a data-dependent acquisition mode, in which

the precursor MS1 scan (m/z 350–1,550) was acquired in the Orbitrap at a resolution setting of 120,000, followed by Orbitrap HCD-MS/MS and OTHCD-MS/MS of the 20 most abundant multiply charged precursors in the MS1 spectrum. MS2 spectra were acquired at a resolution of 30,000.

MS/MS data was processed using Mascot search engine (v.2.8.0, 2021[2]; Matrix Science Ltd., London, United Kingdom). Tandem mass spectra were searched against protein sequence, and Trypsin/P was specified as a cleavage enzyme allowing up to two missing cleavages. For precursor ions, the mass error was set to 10 ppm, and for fragment ions, the mass error is set to 0.02 Da. Carbamidomethylation on Cys was specified as a fixed modification and oxidation on Met and Phosphorylation on Ser and Thr were specified as variable modifications. False discovery rate thresholds for protein, peptide, and modification sites were specified at 1%. All the other parameters in Mascot were set to default values.

[2]http://www.matrixscience.com

RESULTS

Overexpressed Cat8 Enhances 2-PE Production by Improving Cell Growth in the Presence of Glucose

We characterized the related fermentation phenotypes of the wild-type strain YS58, *CAT8* knock-out strain (YS58-Δcat8), and *CAT8* overexpression strain (YS58-CAT8) in the fermentation

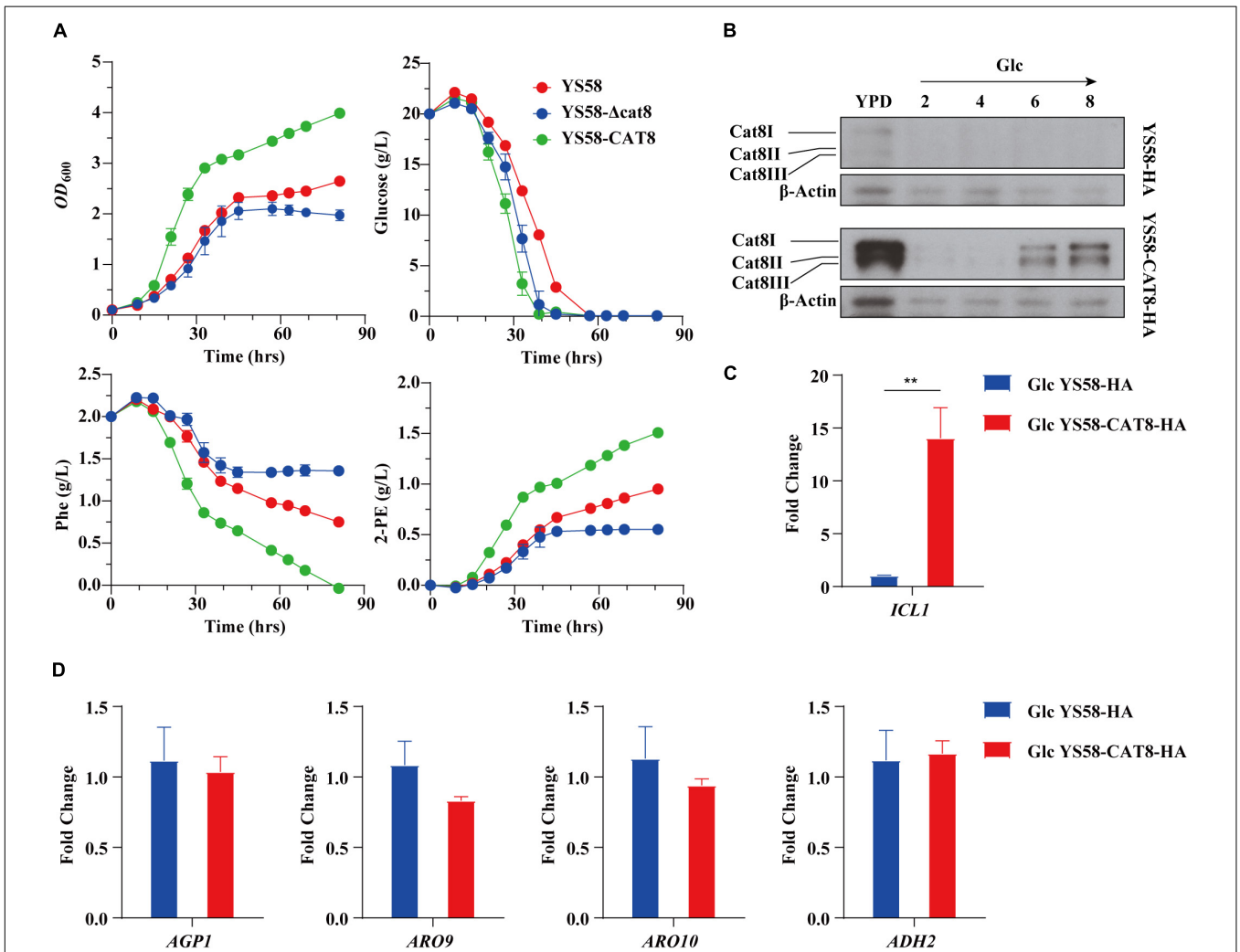


FIGURE 1 | Effect of Cat8 on 2-PE biosynthesis and the transcriptional activity of Cat8 in the presence of glucose. **(A)** Cell growth, residual sugar, Phe, and 2-PE contents. Yeast strains were cultured in fermentation medium at 30°C and 200 rpm. **(B)** Cat8 level assay. Cells were cultured in YPD overnight and transferred to fermentation medium with glucose as the only carbon sources for 6 h at 30°C and 200 rpm. Samples were taken at indicated times. Cell free extracts (2 μg protein) were subjected to immunoblot analysis using polyclonal anti-HA and β-Actin antibodies. **(C)** Transcriptional levels of Cat8 target gene *ICL1*. **(D)** Transcriptional levels of genes involved in Ehrlich pathway. For **(C,D)**, yeast cells were cultured in fermentation medium with glucose as the only carbon source at 30°C and 200 rpm for 6 h. Data shown are mean values ± SDs of triplicates. Statistical difference was determined by two-tailed Student's *t*-test, $^{**}p < 0.01$.

medium (**Figure 1A**). In the overexpression strain, strong constitutive promoter pTEF2 was used in order to obtain the stable and high-level transcription of *CAT8*. The cell growth, 2-PE yield, and glucose and Phe consumption of YS58-CAT8 were higher than those of YS58 and YS58-Δcat8 during the whole growth process. To fully understand the role of Cat8 in the biosynthesis of 2-PE, the protein levels of Cat8 and its transcriptional activation activity were determined. A significant increase in Cat8 levels was observed in strain YS58-CAT8-HA compared to the control strain YS58-HA, confirming the enhanced expression of Cat8 (**Figure 1B**). The transcriptional levels of *ICL1*, a key target gene of Cat8 in the glyoxylic acid pathway (Rahner et al., 1999) in strain YS58-CAT8-HA were 13-fold higher than that in strain YS58-HA (**Figure 1C**), suggesting that constitutively overexpressed Cat8 possess transcriptional activation activity under glucose condition. mRNA levels of genes such as amino acid permease gene *AGP1*, transaminase gene *ARO9*, decarboxylase gene *ARO10*, and reductase gene *ADH2* had no significant differences between strain YS58-CAT8-HA and YS58-HA at the early stage of exponential growth under glucose condition (**Figure 1D**), indicating that enhanced 2-PE in glucose medium might result mainly from increased cell growth.

Snf1 Is Essential for Cat8 Nucleic Localization and Enhances the Regulatory Activity of Cat8

To explore the role of Cat8 in the regulation of 2-PE synthesis, different carbon sources were further adopted, and the subcellular localization of Cat8 was detected. GFP-tagged Cat8 was mainly concentrated in the nucleus whether fermented carbon source glucose or non-fermented carbon source glycerol was the sole carbon source (**Figure 2A**). However, the deletion of protein kinase Snf1 caused the accumulation of Cat8 in the cytoplasm even under glycerol conditions (**Figure 2B**). In our constitutively overexpressed Cat8 strain, Snf1 might be active but with weak activity under glucose conditions, and Cat8 was phosphorylated to a low degree. The transcriptional levels of Cat8 target genes including *ICL1*, *MLS1*, and *MDH2* in glycerol conditions were 5, 36, and 10 times higher than those in glucose conditions, respectively, while their mRNA levels decreased significantly in *SNF1* knock out strain (**Figure 2C**), indicating the partial activity of Cat8 under glucose condition and the significance of Snf1 for the Cat8 activity. Strain YS58-HA Δsnf1 could not use glycerol to grow, while strain YS58-CAT8-HA Δsnf1 showed a weak phenotype of glycerol utilization (**Figure 2D**), suggesting that when Cat8 overexpressing in Snf1 deletion strain, other kinases might participate in the phosphorylation of Cat8, which were weak and not the central kinases for Cat8 compared to Snf1.

Cat8 Plays an Essential Role in Response to External Nutritional Changes

RNA-seq was performed to explore the regulatory network of Cat8 in response to Phe induction under glucose or glycerol condition. Gene expression heat map showed that the two carbon sources greatly influenced the transcript spectrums of yeast cells after Cat8 overexpression. The transcription levels of 2.3% of genes on a genome-wide scale changed more than two times between the wild-type strain and *CAT8* overexpressing strain under glucose conditions, while the differential genes accounted for 28.8% of the total genes under glycerol conditions (**Supplementary Figure 2**), which was consistent with the changes of more than 1,700 genes during the diauxic shift of *S. cerevisiae* reported by DeRisi et al. (1997). Gene Ontology (GO) analysis showed that Cat8 overexpression enhanced the TCA cycle, respiratory metabolism, and glyoxylic acid cycle rather than metabolic pathways related to 2-PE synthesis under glucose conditions. In the TCA cycle, the transcription of isocitrate lyase gene *ICL1*, cytoplasmic malate dehydrogenase gene *MDH2*, isocitrate dehydrogenase gene *IDP2*, and malate synthase gene *MLS1* were up-regulated 32.8-fold, 2.7-fold, 2.4-fold, and 2.1-fold, respectively, (**Figure 3A** and **Supplementary Dataset 1**). Under glycerol conditions, activities of various metabolic pathways, such as glycerol metabolism, respiratory metabolism, fatty acid metabolism in the peroxisome, amino acids (arginine, proline, and phenylalanine) catabolism, were enhanced by overexpressing *CAT8*, while glycolysis, gluconeogenesis, biosynthesis of amino acids (lysine, phenylalanine, tyrosine, and tryptophan) were inhibited (**Figure 3A**). Moreover, a variety of transmembrane transport pathways including amino acid transport, ammonium transport, ion transport, and drug transport were activated in *CAT8* overexpressing strain (**Figure 3A** and **Supplementary Dataset 2**), which is partly related to substrate uptake and efflux of by-products in the process of 2-PE synthesis.

About 17 actively expressed genes under glucose or glycerol condition were randomly selected and determined their transcription levels by qRT-PCR. A strong correlation between RNA-seq and qRT-PCR for the transcription levels of 17 genes was observed (**Figure 3B** and **Supplementary Table 3**), indicating the reliability of RNA-seq results. Among the 17 genes, the transcription of amino acid permease genes *GAP1* and *AGP1*, transcriptional activator gene *GAT1*, Ehrlich pathway genes *ARO9*, *BAT2*, *ARO10*, and *ADH2* in *CAT8* overexpressing strain were up-regulated 36-fold, 7-fold, 18-fold, 7-fold, 9-fold, 8-fold, and 7-fold, respectively, under glycerol condition compared to that of the wild-type strain (**Figure 3C** and **Supplementary Table 3**). The transcription levels of *PDR12* and *ESBP6* for fusel acids export increased by 2.3-fold and 3.4-fold after Cat8 overexpression (**Supplementary Table 3**) in glycerol conditions. These results suggested that highly active Cat8 in *S. cerevisiae* not only activated carbon metabolism-related pathways but also enhanced the transcriptional levels of key enzymes in the Ehrlich pathway, upstream amino acid transporters, downstream metabolites fusel acid transporters, and transcriptional activator Gat1, showing that Cat8 regulation on 2-PE synthesis is comprehensive and extensive.

The up-regulated genes in Cat8 overexpressing strain under glycerol conditions were enriched in many terms related to "sequence-specific DNA binding," "transcriptional activator activity," and "sequence-specific DNA binding of RNA polymerase II regulatory region" in the molecular function analysis of GO. Based on the quantitative measurement of

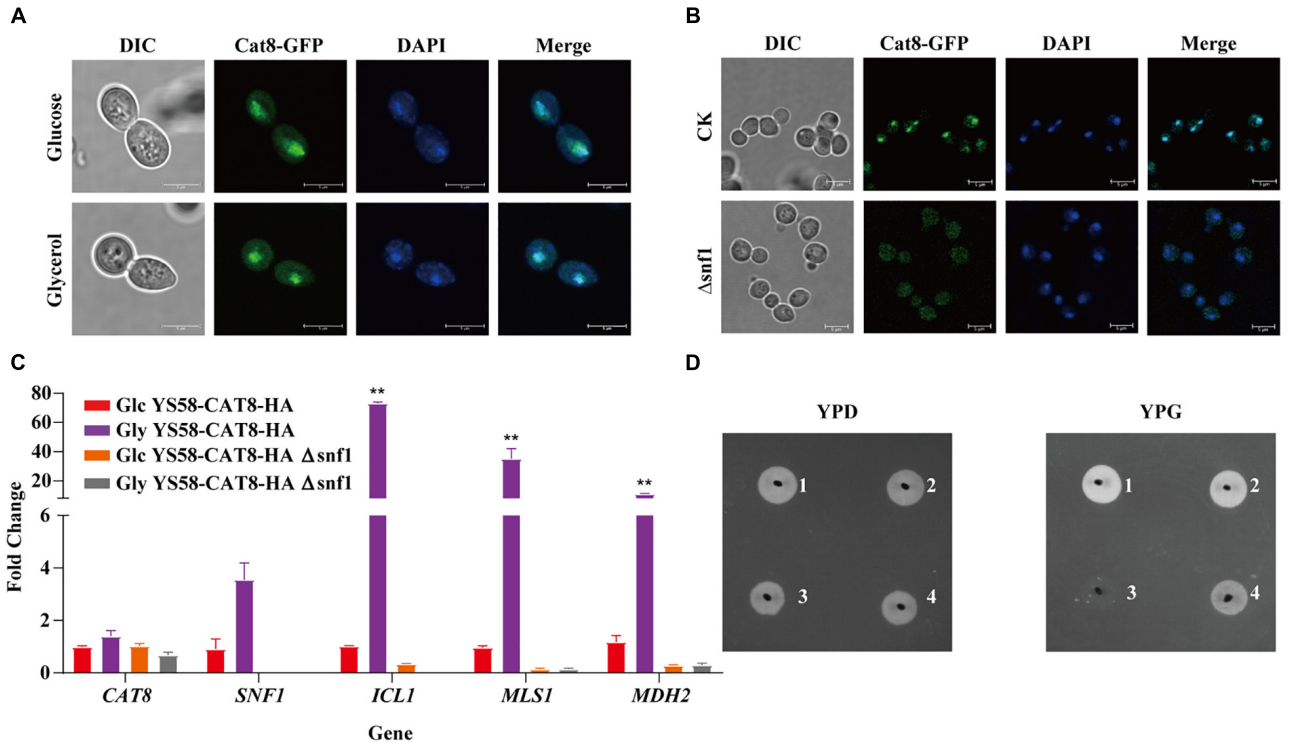


FIGURE 2 | Cat8 intracellular localization and transcriptional regulatory activity in response to different carbon sources. **(A)** Cat8 intracellular localization under glucose or glycerol condition. **(B)** Effect of Snf1 on Cat8 subcellular localization under glycerol condition. For **(A,B)**, the white bar represents 5 μm. **(C)** Transcriptional levels of genes *CAT8*, *SNF1*, *ICL1*, *MLS1*, and *MDH2* normalized to *ACT1*. **(D)** Growth assay to indicate the effect of Snf1 deletion and Cat8 overexpression, YS58-HA (1), YS58-CAT8-HA (2), YS58-HA Δ snf1 (3), and YS58-CAT8-HA Δ snf1 (4). Data shown are mean values ± SDs of triplicates. Statistical difference was determined by two-tailed Student's *t*-test, $^{**}p < 0.01$.

functional similarity of terms (Huang et al., 2009) and using Metascape website to analyze the pathways and processes of the genes in the above terms, 13 significantly enriched ($p < 0.01$) GO terms were obtained (**Figure 3D** and **Supplementary Dataset 3**) forming different enrichment networks. The results showed that highly active Cat8 was involved in regulating cell response to carbon and nitrogen sources under glycerol conditions. In terms of carbon sources, Cat8 mainly participated in regulating non-fermenting carbon source utilization and fatty acid metabolism; in terms of nitrogen sources, Cat8 was involved in derepressing nitrogen catabolite repression (NCR).

Cat8 Regulates the Transcription of Genes Related to Carbon and Nitrogen Metabolism by Direct Binding With Promoters

To explore the interaction of Cat8 and its target genes, chromatin immunoprecipitation sequencing (ChIP-seq) was carried out for strain YS58-CAT8-HA under glucose or glycerol conditions with Phe induction for 6 h. Cat8 was mainly bound to the upstream 1,000 bp region from the transcription start site (TSS) of target genes, and the binding ability was significantly increased under glycerol conditions compared to glucose conditions (**Figure 4A**). Using Integrative Genomics Viewer (IGV) visualization tool, the binding peaks of Cat8 and ChIP-seq sequencing readings in the gene promoter region were analyzed. The results showed that the binding peak map of Cat8 with *ARO9* and *ARO10* related to nitrogen metabolism was similar to that with *FBP1*, *ICL1*, *MLS1*, and *PCK1* involved in carbon metabolism (**Figure 4B**). The motif of Cat8 target genes for non-fermentative carbon source utilization was CCGGN (Badis et al., 2008). We combined the central sequence of binding peaks displayed by IGV to determine the binding elements of Cat8. As a result, the transcriptional binding sites of Cat8 on the key enzyme genes *ARO9*, *BAT2*, *ARO10*, and *ADH2* in the Ehrlich pathway, amino acid transporter genes *GAP1*, *AGP1*, and *BAP2*, and transcriptional activator *GAT1* were identified, whose sequences and respective positions were fairly coincident with the above-reported motif (**Figure 4C**).

Analysis of Cat8 interacting genes showed that there were 380 genes under glucose conditions with 209 unique binding genes, while there were 458 genes directly interacting with Cat8 under glycerol conditions with 287 unique binding genes (**Supplementary Figure 3**). Interestingly, we noticed that a large number of Cat8 binding peaks could be gotten under glucose conditions, which might be due to the low level of phosphorylation of Cat8 by weak Snf1 activity or by other kinases in Cat8 overexpression strain. In the presence of glycerol, Cat8 activity was enhanced by the phosphorylation catalyzed

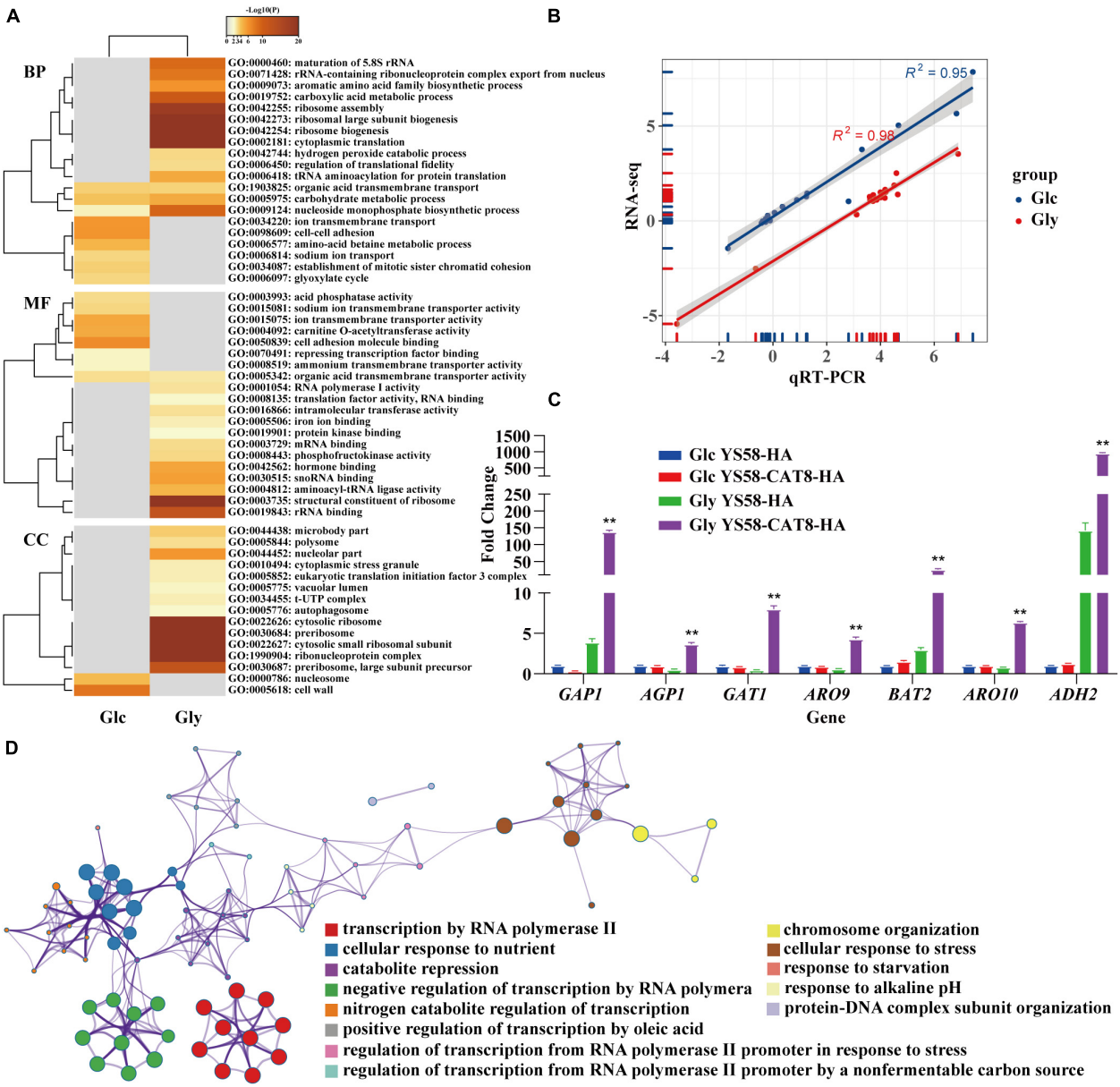


FIGURE 3 | RNA-seq analysis of yeast strain in response to external nutritional changes. **(A)** The list of differential genes in glucose (Glc) and glycerol (Gly) after overexpression of Cat8 was enriched by GO differential heat map of biological process (BP), molecular function (MF), and cellular component (CC). Heatmap showing the top enrichment clusters, one row per cluster, using a discrete color scale to represent statistical significance. Gray color indicates a lack of significance. **(B)** The correlation between RNA-seq and qRT-PCR results. **(C)** The transcriptional levels of genes related to fusel alcohols synthesis via Ehrlich pathway in RNA-seq data. **(D)** Metascape visualization of the interactome network formed by all 78 gene candidates from differential expression of transcription factors after Cat8 overexpression under Gly condition, where each node represented an enriched term and was colored by its cluster-ID. The strains used were YS58-HA and YS58-CAT8-HA. Data shown are mean values ± SDs of triplicates. Statistical difference was determined by two-tailed Student's *t*-test, $^{**}p < 0.01$.

by completely active Snf1 to make Cat8 bind with much more genes, resulting in sharper and higher binding peaks than that under glucose conditions (**Figure 4B**). In addition, we noticed that there were prominent Cat8 binding peaks in the promoter region of some genes. However, the transcriptional levels of these genes were down-regulated under glucose and glycerol conditions (**Supplementary Dataset 4**), which might be due to the competition between Cat8 and other transcription activators for the same *cis*-regulatory element. Most of the previous studies explored Cat8 regulation on carbon metabolism under glucose limitation. However, according to our transcriptome data and ChIP-seq data, Cat8 had transcriptional activation function when Cat8 was constitutively strongly expressed under glucose conditions, inferring that the Snf1 signal pathway might be partially activated; it was also speculated that other kinases in yeast cells might phosphorylate Cat8.

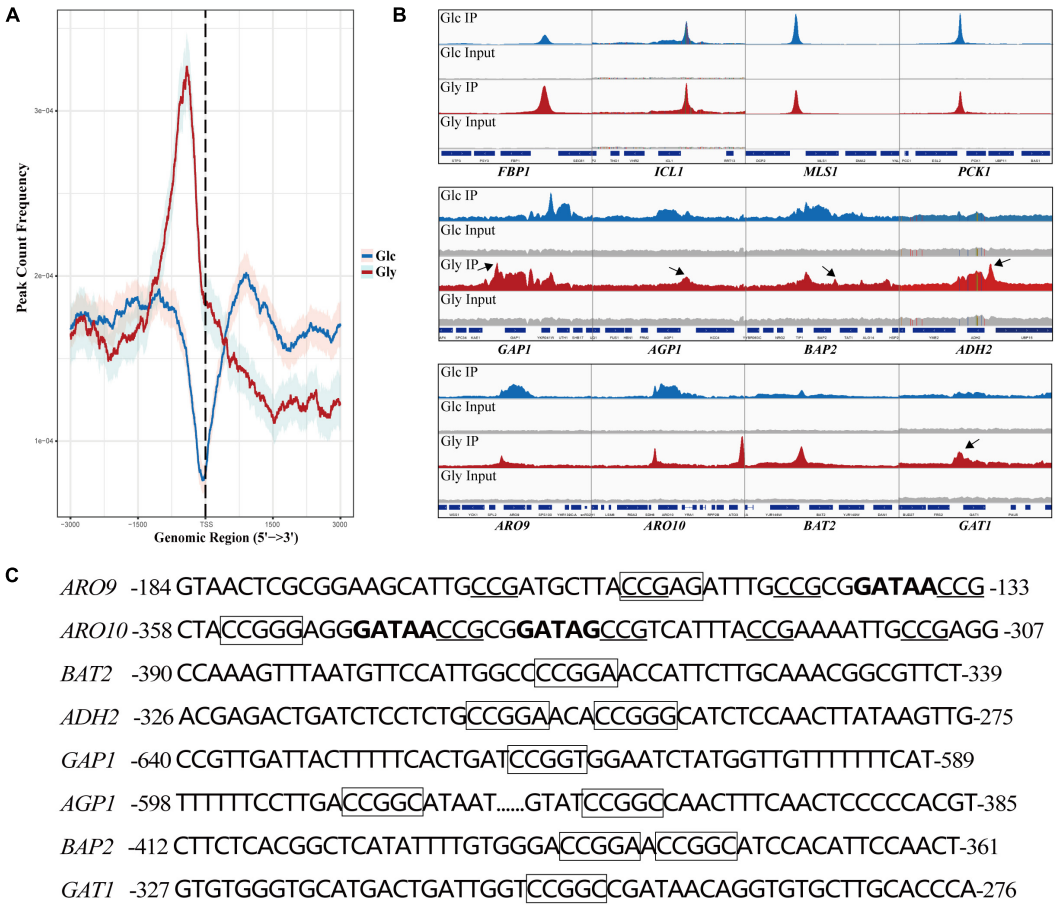


FIGURE 4 | Analysis of binding peaks of transcription factor Cat8 under two carbon sources by ChIP-seq. **(A)** The distribution of reads in the peak region (3,000 bp upstream and downstream of TSS) on all transcripts under the glucose (Glc) or glycerol (Gly) condition was calculated to reflect the binding intensity of Cat8 to DNA at 0.95 confidence interval. **(B)** IGV examined the peaks enriched by ChIP-seq. The blue and red peaks represented the binding peaks of Cat8 under the conditions of Glc and Gly, respectively, and the gray was input. **(C)** Promoter sequences of Cat8 target genes include *ARO9*, *ARO10*, *BAT2*, *ADH2*, *GAP1*, *AGP1*, *BAP2*, and *GAT1*. Cat8 binding sites (CCGGN) were marked with a box. CCG triplets, the binding site of Aro80, were underlined, and GATA factor binding sites (GATAA/G) were indicated in bold.

Furthermore, Cat8 binding genes under two conditions were enriched by GO (BP\CC\MF) and KEGG (**Supplementary Dataset 5**), and the enrichment network was also analyzed (**Supplementary Figure 4**). The result is consistent with that in our transcriptome data analysis, which confirmed again that transcriptional activator Cat8 could not only regulate carbon metabolism but also play an important role in responding to nitrogen starvation and regulating cellular nitrogen metabolism. Biosynthesis regulation was dominant under glucose conditions, while catabolism regulation was prominent under glycerol conditions.

Cat8 Interacts Directly With Transcription Factor Aro80 and GATA Activating Factor Gat1

Based on the analysis of ChIP-seq, it was found that the binding sites of Cat8 on the key enzyme genes *ARO9*/*ARO10* of the Ehrlich pathway were close to those of transcriptional activator Aro80 and GATA activators Gln3 and Gat1 (**Figure 4C**). We speculated that Cat8 might interact with them and coordinately regulate gene transcription. To detect the intracellular interaction of Cat8 with Aro80, Gln3, and Gat1, the above three proteins with Myc tags were overexpressed in the HA tagged Cat8 strain (YS58-CAT8-HA), then Co-Immunoprecipitation (Co-IP) assay for the protein complexes with anti-HA, anti-Myc, and IgG antibodies, respectively, were carried out. The results showed protein-level interaction between Cat8 and Aro80 and Gat1, while there was no such interaction between Cat8 and Gln3 (**Figure 5A**).

The transcription level of *ARO10* was analyzed by qRT-PCR after deleting Aro80 in the wild-type strain and Cat8 overexpression strain. The results showed that the *ARO10* expression levels in yeast cells with Aro80 deletion decreased to 2.5% of that without Aro80 deletion, indicating that Aro80 was the main transcription factor for activating *ARO10*, which was consistent with the previous report on Aro80 (Lee and Hahn, 2013). However, in Cat8 overexpression strain, the transcriptional level of *ARO10* increased six-fold regardless of

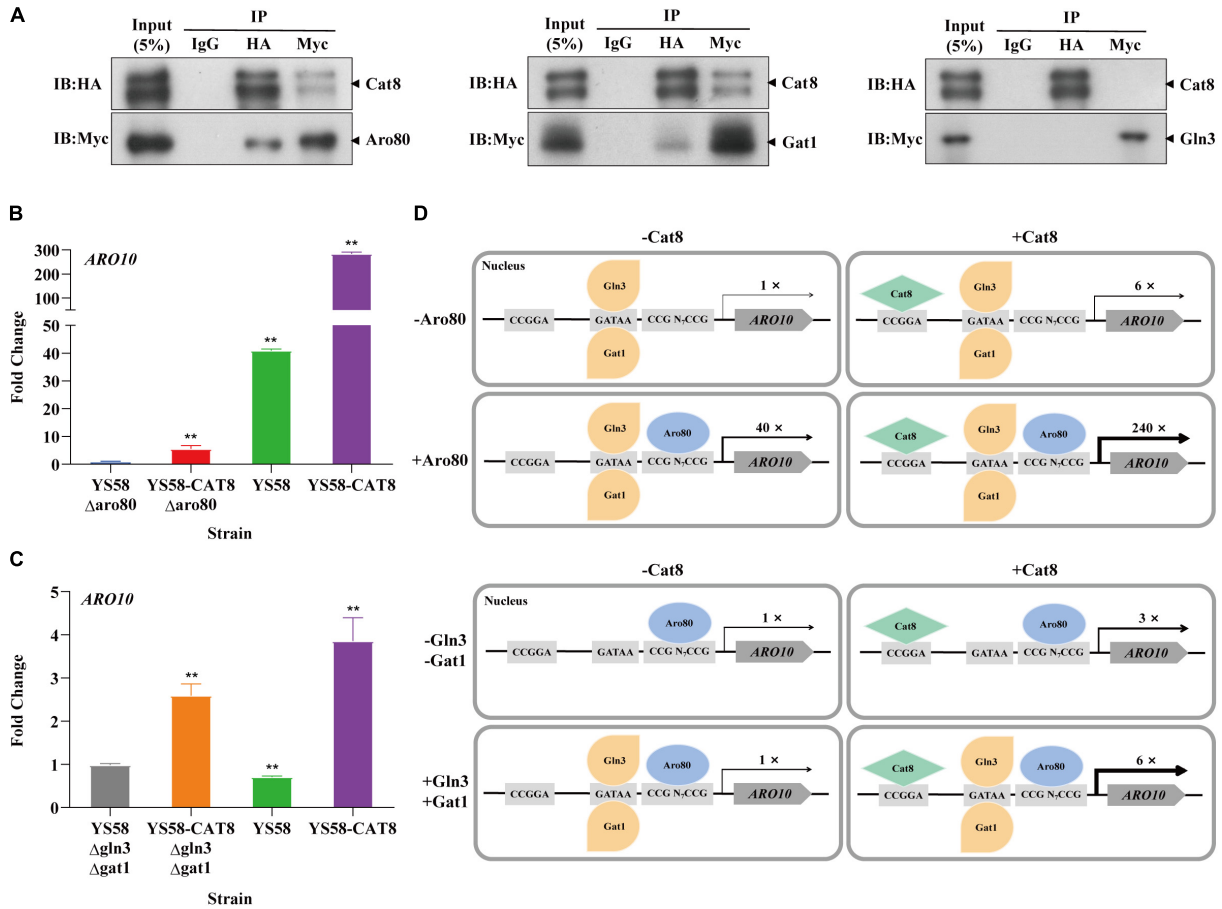


FIGURE 5 | The interaction of Cat8 with Aro80 and GATA activating factors. **(A)** The Co-immunoprecipitation of total protein extracted from strains YS58-CAT8-HA Aro80-Myc, YS58-CAT8-HA Gat1-Myc, and YS58-CAT8-HA Gln3-Myc from left to right. Input (5%) was used as a positive control and IgG as a negative control. Immunoprecipitated proteins (IPs) were analyzed by immunoblotting and probing with HA and Myc antibodies. **(B)** Effect of Aro80 on the *ARO10* transcription level of wild-type strain and *CAT8* overexpression strain. **(C)** Effect of GATA factors on the *ARO10* transcription level of wild-type strain and *CAT8* overexpression strain. **(D)** The interaction pattern of Cat8 with Aro80 and GATA activator factors. Yeast cells were cultivated in fermentation medium with glycerol as the only carbon source for 6 h. *ACT1* was used as the reference gene, and wild-type strains with *ARO80* or *GLN3/GAT1* deletion were used as controls. Data shown are mean values ± SDs of triplicates. Statistical difference between strains was determined by two-tailed Student's *t*-test, $^{**}p < 0.01$.

whether Aro80 deletion or not (**Figure 5B**). This indicated that Cat8 itself could activate transcription of *ARO10* in the absence of Aro80. When Aro80 exists, the interaction between Cat8 and Aro80 may make them bind to the promoter region of *ARO10* more tightly and bring a significant increase in the transcription of *ARO10* (**Figure 5D**).

There is no significant difference in *ARO10* level between the wild-type strain and the strain with GATA activator Gln3 and Gat1 deletion (**Figure 5C**). It had been reported that when yeast cells grew in YPD medium to the early logarithmic stage and were treated with rapamycin, the *ARO10* transcription level in GATA activator deletion strains decreased by 50% (Lee and Hahn, 2013). The inconsistency between this study and the literature may be caused by the different cultural conditions or strain differences. When Cat8 was overexpressed in Gln3 and Gat1 deletion strain, the *ARO10* level increased by three times. When Gln3 and Gat1 existed, Cat8 overexpression enhanced *ARO10* transcription by six times. The difference between the above two group strains was two times. Considering the Co-IP results, it was suggested that the interaction and coordination between Cat8 and Gat1 are also involved in the transcriptional regulation of *ARO10*.

DISCUSSION

Cat8 is a crucial factor for non-fermentable carbon source utilization. During the analysis of growth and 2-PE synthesis, we found that diauxic growth disappeared in YS58-Δcat8, in the meantime, the 2-PE titer in this strain did not change either at the later stage, hinting that the synthesis of 2-PE is coupled with cell growth. Combined with the transcriptome data, we speculated that overexpression of Cat8 increases the metabolic activity of the cellular TCA cycle and the glyoxylate cycle under glucose conditions, this enabled continuous active production of α-ketoglutarate (the direct precursor to glutamate), and glutamate is the end product of the biosynthetic nitrogen source

(Butow and Avadhani, 2004); furthermore, α-ketoglutarate is an intermediate receptor for the α-amino generated by transamination in the Ehrlich pathway (Etschmann et al., 2002). This allows cells to obtain higher biomass and 2-PE yield.

Cat8 relies on phosphorylation to play its function, and Snf1 is the main kinase of Cat8. It was reported that Snf1 phosphorylating Cat8 was inhibited in the presence of glucose and could not activate Cat8 in the wild-type yeast (Busti et al., 2010; Turcotte et al., 2010). However, overexpressed Cat8 in our study played the role in the whole process of yeast cell growth, and the results of RNA-seq and ChIP-seq also showed that overexpressed Cat8 could bind to the target gene and activate the related pathway under the glucose condition. Recently, some researchers found that an unusual ratio of nutrients could trigger some signal pathways in yeast resulting in unexpected activation of Snf1 under high reducing sugars conditions (Vallejo et al., 2020). Hence, we surmised that when Cat8 was constitutively overexpressed, the Snf1 signal pathway might be triggered and Snf1 was partially activated to phosphorylate Cat8 at a low level. The constitutive overexpression of *CAT8* also can restore the phenotype of the snf1-deletion strain that originally cannot use non-fermented carbon sources. This implies that there may be other kinases involved in the phosphorylation of Cat8 in yeast, which might be weaker and not the main kinases for Cat8 compared with Snf1.

The residue site Ser562 was reported to be the key phosphorylation site of Cat8 directly acted by Snf1 (Charbon et al., 2004), while the SGD database contains 13 of the Cat8 phosphorylation sites reported in the previous literature (Albuquerque et al., 2008; Holt et al., 2009; Swaney et al., 2013; Lanz et al., 2021) but does not contain Ser562, suggesting that differences of phosphorylation sites in Cat8 might exist in different strains. In this study, three bands of Cat8I∼III were presented after Western Blot (**Supplementary Figure 5A**). Mass spectrometry detection showed the differences in post-translational modification (PTM) such as phosphorylation, methylation, and acetylation among these three bands. For example, phosphorylation modification at Thr53 or Ser55, methylation at Arg84, and acetylation at Lys507 and Lys936 were different (**Supplementary Table 4**). The strain in our research did not show Ser562 phosphorylation either, further reflecting different key phosphorylation sites in Cat8 in different yeast strains. During Western Blot analysis, we also found the differences in Cat8 bands pattern (**Supplementary Figures 5B,C**) compared to that in the previous report (Randez-Gil et al., 1997), for example, the Cat8 band pattern did not change after the Lambda-phosphatase treatment. Amino acid sequence alignment showed quite a few differences in the sequences between our strain and the previously reported strain, including multiple amino acid site mutations and a fragment deletion/insertion (InDel) (**Supplementary Figure 6A**). Selecting the reported Cat8 protein sequences of *S. cerevisiae* in various habitats for phylogenetic analyses, it was found that the strains used in the previous literature (CEN.PK2-1C) and in this study (YS58) were located in the wild/transitional group and liquid-state fermentation group (LSF), respectively (**Supplementary Figure 6B**), which further confirmed the above differences.

The domesticated lineages of yeast *S. cerevisiae* documented worldwide so far belong to two major monophyletic groups associated with solid-and liquid-state fermentation (SSF and LSF), respectively (Duan et al., 2018). Interestingly, we note that the molecular phylogenetic tree constructed from Cat8 protein sequences clearly distinguishes between the LSF and the SSF in the domestication lineages. Previous studies on Cat8 mainly focused on the regulation of carbon source metabolism (Turcotte et al., 2010), and this study proposed that Cat8 also plays a regulatory role in nitrogen source metabolism such as amino acids catabolism. Perhaps the obvious difference in the Cat8 protein sequence among different strains is due to adaptive evolution. These results imply that Cat8 may play an essential role in *S. cerevisiae*'s adaptation to different fermentation types as a global transcriptional regulator, and may further help the yeast better adapt to different nutritional environments during the evolutionary process through phosphorylation, acetylation, methylation, and other modifications.

The combined analysis of RNA-seq and ChIP-seq of overexpression Cat8 in fermentative and non-fermentative carbon sources conditions provided large-scale information on cell reprogramming responses to nutritional changes. Cat8 regulates carbon metabolism, such as activating the glyoxalic acid cycle and gluconeogenesis pathway and simultaneously participates in responding to nitrogen hunger and regulates nitrogen metabolism. Overexpression of Cat8 activated multiple transcription factors, most of which participated in the cellular response to nutrients. More Cat8 binding peaks were pulled under glycerol conditions, and the peak type was sharper and taller than that under glucose conditions. Combined with the results of **Figure 4A**, we speculated that the difference in binding peaks might be related to the phosphorylation of Cat8 on the one hand; On the other hand, it may be related to the level of cell autophosphorylation. When cells are exposed to non-fermentative carbon sources, the Snf1 phosphokinase complex is activated completely, and transcriptional activation is promoted through the activation of transcription factors and chromatin structure reconstruction of the target gene promoter region (Agricola et al., 2004; Tachibana et al., 2005; Biddick et al., 2008; Broach, 2012). Therefore, Cat8 has a strong ability to bind DNA under glycerol conditions. These results suggest that Cat8 has a more comprehensive regulation in nutritional response, and plays a similar role as a global regulator of energy homeostasis or an inhibitor of the anabolic process. Cat8 can regulate the transcription of target genes in different metabolic pathways through sensing carbon source signals to control the direction of cell metabolism.

In this study, it was demonstrated that Cat8 directly binds and regulates key enzyme genes of the Ehrlich pathway (*ARO9*, *BAT2*, *ARO10*, and *ADH2*), amino acid transporter gene (*GAP1* and *AGP1*), and transcriptional activator (*GAT1*). At the same time, the results proved that Cat8 interacts with GATA activator Gat1 and Ehrlich pathway-specific transcriptional activator Aro80 at the protein level, and the level of transcription activation of target genes between them can be superimposed on each other. The nature of transcription factors' protein-protein interactions and their affinity for DNA binding motifs define gene target selectivity and chromatin binding dynamics

(Francois et al., 2020). Therefore, the binding of the transcription factors to the promoter region is more stable, which improves the transcription level of the target gene. This study indicates that Cat8 directly regulates the critical steps of yeast synthesis of 2-PE, which helps to improve the efficiency of 2-PE synthesis. It has been suggested that Gat1 plays a core role in NCR control, and small changes in Gat1 expression have a high impact on the expression of NCR genes (Georis et al., 2009). Cat8 not only regulates the transcription of Gat1 but also interacts with Gat1. Therefore, Cat8 overexpression may participate in the regulation of nitrogen metabolism by regulating the GATA factor and making cells respond to starvation. In addition, we found that overexpression of Cat8 resulted in obvious cell flocculation (results not shown) and enrichment in the pathways related to flocculation and cell adhesion. Flocculation was related to stress resistance (Soares, 2011), which is also a coping method for yeast to adapt to the changes in different environments.

In summary, Cat8 participates in both non-fermentative carbon and nitrogen utilization. In the regulation of aromatic alcohol synthesis, Cat8 can promote Ehrlich pathway flux by directly binding and activating the related genes, it also can interact with other transcription factors such as Aro80 and Gat1 to coordinate enhance the Ehrlich pathway. As a global transcription factor, Cat8 plays a significant role in response to nutritional changes and regulates both carbon and nitrogen metabolism.

DATA AVAILABILITY STATEMENT

The data presented in the study are deposited in the National Microbiology Data Center (NMDC) repository, accession number NMDC10018022, available at the link: https://nmdc.cn/resource/genomics/sra?keyword=NMDC10018022.

AUTHOR CONTRIBUTIONS

ZD: designing experiments, implementation, data curation, and writing – original draft, review and editing HD: experiments implementation and data curation. YC, ZZ, and XG: methodology, data analysis, and experimental assistance. ZW: conceptualization, supervision, funding acquisition, and writing – original draft, review and editing. XH: conceptualization, supervision, and writing – review and editing All authors agreed to be accountable for the content of the work.

FUNDING

This study was financially supported by the National Natural Science Foundation of China (Grant Number 31771959).

ACKNOWLEDGMENTS

We would like to thank Jingfang Liu and Weilin Li (Institutional Center for Shared Technologies and Facilities of Institute of Microbiology, Chinese Academy of Sciences), for their work in identification of phosphorylation sites with Mass Spectrometry in this manuscript. We would also like to thank Ripeng Zhang for his help with bioinformatics. We are very grateful to all the lab members for their helpful discussion and support.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmicb.2022.898938/full#supplementary-material

Supplementary Dataset 1 | Results of differential gene analysis under Glc and Gly conditions (transcriptome data).

Supplementary Dataset 2 | Results of GO enrichment analysis under Glc and Gly conditions (transcriptome data).

Supplementary Dataset 3 | Information and GO enrichment results of 78 selected transcription factors.

Supplementary Dataset 4 | Enrichment rate and fold change of Cat8 binding genes under Glc and Gly conditions.

Supplementary Dataset 5 | Enrichment analysis results of GO and KEGG pathways of Cat8 binding genes under Glc and Gly conditions (ChIP-seq data).

REFERENCES

Agricola, E., Verdone, L., Xella, B., Di Mauro, E., and Caserta, M. (2004). Common chromatin architecture, common chromatin remodeling, and common transcription kinetics of Adr1-dependent genes in *Saccharomyces cerevisiae*. *Biochemistry* 43, 8878–8884. doi: 10.1021/bi049577

Albuquerque, C. P., Smolka, M. B., Payne, S. H., Bafna, V., Eng, J., and Zhou, H. (2008). A multidimensional chromatography technology for in-depth phosphoproteome analysis. *Mol. Cell Prot.* 7, 1389–1396. doi: 10.1074/mcp.M700468-MCP200

Anders, S., and Huber, W. (2010). Differential expression analysis for sequence count data. *Genome Biol.* 11:R106. doi: 10.1186/gb-2010-11-10-r106

Anders, S., Pyl, P. T., and Huber, W. (2015). HTSeq–a Python framework to work with high-throughput sequencing data. *Bioinformatics* 31, 166–169. doi: 10.1093/bioinformatics/btu638

Badis, G., Chan, E. T., van Bakel, H., Pena-Castillo, L., Tillo, D., Tsui, K., et al. (2008). A library of yeast transcription factor motifs reveals a widespread function for Rsc3 in targeting nucleosome exclusion at promoters. *Mol. Cell.* 32, 878–887. doi: 10.1016/j.molcel.2008.11.020

Biddick, R. K., Law, G. L., and Young, E. T. (2008). Adr1 and Cat8 mediate coactivator recruitment and chromatin remodeling at glucose-regulated genes. *PLoS One* 3:e1436. doi: 10.1371/journal.pone.0001436

Björck, L., and Kronvall, G. (1984). Purification and some properties of streptococcal protein G, a novel IgG-binding reagent. *J. Immunol.* 133, 969–974.

Bonander, N., Ferndahl, C., Mostad, P., Wilks, M. D., Chang, C., Showe, L., et al. (2008). Transcriptome analysis of a respiratory *Saccharomyces cerevisiae* strain suggests the expression of its phenotype is glucose insensitive and predominantly controlled by Hap4, Cat8 and Mig1. *BMC Gen.* 9:365. doi: 10.1186/1471-2164-9-365

Broach, J. R. (2012). Nutritional control of growth and development in yeast. *Genetics* 192, 73–105. doi: 10.1534/genetics.111.135731

Busti, S., Coccetti, P., Alberghina, L., and Vanoni, M. (2010). Glucose signaling-mediated coordination of cell growth and cell cycle in *Saccharomyces cerevisiae*. *Sensors* 10, 6195–6240. doi: 10.3390/s100606195

Butow, R. A., and Avadhani, N. G. (2004). Mitochondrial signaling: the retrograde response. *Mol. Cell* 14, 1–15. doi: 10.1016/s1097-2765(04)00179-0

Cam, H. P., and Whitehall, S. (2016). Chromatin Immunoprecipitation (ChIP) in *Schizosaccharomyces pombe*. *Cold Spring Harb. Protoc.* 2016:11. doi: 10.1101/pdb.prot091546

Charbon, G., Breunig, K. D., Wattiez, R., Vandenhaute, J., and Noël-Georis, I. (2004). Key role of Ser562/661 in Snf1-dependent regulation of Cat8p in *Saccharomyces cerevisiae* and *Kluyveromyces lactis*. *Mol. Cell Biol.* 24, 4083–4091. doi: 10.1128/MCB.24.10.4083-4091.2004

Chen, H., and Fink, G. R. (2006). Feedback control of morphogenesis in fungi by aromatic alcohols. *Genes Dev.* 20, 1150–1161. doi: 10.1101/gad.1411806

Delaveau, T., Delahodde, A., Carvajal, E., Subik, J., and Jacq, C. (1994). *PDR3*, a new yeast regulatory gene, is homologous to *PDR1* and controls the multidrug resistance phenomenon. *Mol. Gen. Genet.* 244, 501–511. doi: 10.1007/BF00583901

DeRisi, J. L., Iyer, V. R., and Brown, P. O. (1997). Exploring the metabolic and genetic control of gene expression on a genomic scale. *Science* 278, 680–686. doi: 10.1126/science.278.5338.680

Duan, S. F., Han, P. J., Wang, Q. M., Liu, W. Q., Shi, J. Y., Li, K., et al. (2018). The origin and adaptive evolution of domesticated populations of yeast from Far East Asia. *Nat. Commun.* 9:2690. doi: 10.1038/s41467-018-05106-7

Eden, E., Lipson, D., Yogev, S., and Yakhini, Z. (2007). Discovering motifs in ranked lists of DNA sequences. *PLoS Comput. Biol.* 3:e39. doi: 10.1371/journal.pcbi.0030039

Etschmann, M. M., Bluemke, W., Sell, D., and Schrader, J. (2002). Biotechnological production of 2-phenylethanol. *Appl. Microbiol. Biotechnol.* 59, 1–8. doi: 10.1007/s00253-002-0992-x

Francois, M., Donovan, P., and Fontaine, F. (2020). Modulating transcription factor activity: interfering with protein-protein interaction networks. *Semin. Cell Dev. Biol.* 99, 12–19. doi: 10.1016/j.semcdb.2018.07.019

Georis, I., Feller, A., Vierendeels, F., and Dubois, E. (2009). The yeast GATA factor Gat1 occupies a central position in nitrogen catabolite repression-sensitive gene activation. *Mol. Cell Biol.* 29, 3803–3815. doi: 10.1128/MCB.00399-09

Harlow, E., and Lane, D. (1988). *Antibodies: A laboratory manual*. New York, NY: Cold Spring Harbor Laboratory Press.

Haurie, V., Perrot, M., Mini, T., Jenö, P., Sagliocco, F., and Boucherie, H. (2001). The transcriptional activator Cat8p provides a major contribution to the reprogramming of carbon metabolism during the diauxic shift in Saccharomyces cerevisiae. *J. Biol. Chem.* 276, 76–85. doi: 10.1074/jbc.M008752200

Hedges, D., Proft, M., and Entian, K. D. (1995). CAT8, a new zinc cluster-encoding gene necessary for derepression of gluconeogenic enzymes in the yeast *Saccharomyces cerevisiae*. *Mol. Cell Biol.* 15, 1915–1922. doi: 10.1128/MCB.15.4.1915

Holt, L. J., Tuch, B. B., Villén, J., Johnson, A. D., Gygi, S. P., and Morgan, D. O. (2009). Global analysis of Cdk1 substrate phosphorylation sites provides insights into evolution. *Science* 325, 1682–1686. doi: 10.1126/science.1172867

Huang, D., Sherman, B. T., and Lempicki, R. A. (2009). Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat. Protoc.* 4, 44–57. doi: 10.1038/nprot.2008.211

Laemmli, U. K. (1970). Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature* 227, 680–685. doi: 10.1038/227680a0

Laera, L., Guaragnella, N., Ždraleviæ, M., Marzulli, D., Liu, Z., and Giannattasio, S. (2016). The transcription factors *ADR1* or *CAT8* are required for RTG pathway activation and evasion from yeast acetic acid-induced programmed cell death in raffinose. *Microbial. Cell* 3, 621–631. doi: 10.15698/mic2016.12.549

Langmead, B., and Salzberg, S. L. (2012). Fast gapped-read alignment with Bowtie 2. *Nat Methods* 9, 357–359. doi: 10.1038/nmeth.1923

Lanz, M. C., Yugandhar, K., Gupta, S., Sanford, E. J., Faça, V. M., Vega, S., et al. (2021). In-depth and 3-dimensional exploration of the budding yeast phosphoproteome. *EMBO Rep.* 22:e51121. doi: 10.15252/embr.202051121

Lee, K., and Hahn, J. S. (2013). Interplay of Aro80 and GATA activators in regulation of genes for catabolism of aromatic amino acids in *Saccharomyces cerevisiae*. *Mol. Microbiol.* 88, 1120–1134. doi: 10.1111/mmi.12246

Livak, K. J., and Schmittgen, T. D. (2001). Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method. *Methods* 25, 402–408. doi: 10.1006/meth.2001.1262

MacPherson, S., Larochelle, M., and Turcotte, B. (2006). A fungal family of transcriptional regulators: the zinc cluster proteins. *Microbiol. Mol. Biol. Rev.* 70, 583–604. doi: 10.1128/MMBR.00015-06

Park, P. J. (2009). ChIP-seq: advantages and challenges of a maturing technology. *Nat. Rev. Genet.* 10, 669–680. doi: 10.1038/nrg2641

Pertea, M., Kim, D., Pertea, G. M., Leek, J. T., and Salzberg, S. L. (2016). Transcript-level expression analysis of RNA-seq experiments with HISAT, StringTie and Ballgown. *Nat. Protoc.* 11, 1650–1667. doi: 10.1038/nprot.2016.095

Rahner, A., Hiesinger, M., and Schüller, H. J. (1999). Deregulation of gluconeogenic structural genes by variants of the transcriptional activator Cat8p of the yeast *Saccharomyces cerevisiae*. *Mol. Microbiol.* 34, 146–156. doi: 10.1046/j.1365-2958.1999.01588.x

Rahner, A., Schöler, A., Martens, E., Gollwitzer, B., and Schüller, H. J. (1996). Dual influence of the yeast Cat1p (Snf1p) protein kinase on carbon source-dependent transcriptional activation of gluconeogenic genes by the regulatory gene *CAT8*. *Nucleic Acids Res.* 24, 2331–2337. doi: 10.1093/nar/24.12.2331

Randez-Gil, F., Bojunga, N., Proft, M., and Entian, K. D. (1997). Glucose derepression of gluconeogenic enzymes in *Saccharomyces cerevisiae* correlates with phosphorylation of the gene activator Cat8p. *Mol. Cell Biol.* 17, 2502–2510. doi: 10.1128/MCB.17.5.2502

Robinson, J. T., Thorvaldsdóttir, H., Winckler, W., Guttman, M., Lander, E. S., Getz, G., et al. (2011). Integrative genomics viewer. *Nat. Biotechnol.* 29, 24–26. doi: 10.1038/nbt.1754

Schjerling, P., and Holmberg, S. (1996). Comparative amino acid sequence analysis of the C6 zinc cluster family of transcriptional regulators. *Nucleic Acids Res.* 24, 4599–4607. doi: 10.1093/nar/24.23.4599

Soares, E. V. (2011). Flocculation in *Saccharomyces cerevisiae*: a review. *J. Appl. Microbiol.* 110, 1–18. doi: 10.1111/j.1365-2672.2010.04897.x

Swaney, D. L., Beltrao, P., Starita, L., Guo, A., Rush, J., Fields, S., et al. (2013). Global analysis of phosphorylation and ubiquitylation cross-talk in protein degradation. *Nat. Methods* 10, 676–682. doi: 10.1038/nmeth.2519

Tachibana, C., Yoo, J. Y., Tagne, J. B., Kacherovsky, N., Lee, T. I., and Young, E. T. (2005). Combined global localization analysis and transcriptome data identify genes that are directly coregulated by Adr1 and Cat8. *Mol. Cell Biol.* 25, 2138–2146. doi: 10.1128/MCB.25.6.2138-2146.2005

Teunissen, A. W., van den Berg, J. A., and Steensma, H. Y. (1993). Physical localization of the flocculation gene *FLO1* on chromosome I of *Saccharomyces cerevisiae*. *Yeast* 9, 1–10. doi: 10.1002/yea.320090102

Thomas, R., Thomas, S., Holloway, A. K., and Pollard, K. S. (2017). Features that define the best ChIP-seq peak calling algorithms. *Brief Bioinform.* 18, 441–450. doi: 10.1093/bib/bbw035

Towbin, H., Staehelin, T., and Gordon, J. (1979). Electrophoretic transfer of proteins from polyacrylamide gels to nitrocellulose sheets: procedure and some applications. *Proc. Natl. Acad. Sci. USA.* 76, 4350–4354. doi: 10.1073/pnas.76.9.4350

Turcotte, B., Liang, X. B., Robert, F., and Soontorngun, N. (2010). Transcriptional regulation of nonfermentable carbon utilization in budding yeast. *FEMS Yeast Res.* 10, 2–13. doi: 10.1111/j.1567-1364.2009.00555.x

Vallejo, B., Matallana, E., and Aranda, A. (2020). *Saccharomyces cerevisiae* nutrient signaling pathways show an unexpected early activation pattern during winemaking. *Microb. Cell Fact.* 19:124. doi: 10.1186/s12934-020-01381-6

Wang, Z., Bai, X., Guo, X., and He, X. (2017). Regulation of crucial enzymes and transcription factors on 2-phenylethanol biosynthesis via Ehrlich pathway in *Saccharomyces cerevisiae*. *J. Ind. Microbiol. Biotechnol.* 44, 129–139. doi: 10.1007/s10295-016-1852-5

Wikström, M., Sjöbring, U., Drakenberg, T., Forsén, S., and Björck, L. (1995). Mapping of the immunoglobulin light chain-binding site of protein L. *J. Mol. Biol.* 250, 128–133. doi: 10.1006/jmbi.1995.0364

Wuster, A., and Babu, M. M. (2010). Transcriptional control of the quorum sensing response in yeast. *Mol. Biosyst.* 6, 134–141. doi: 10.1039/b913579k

Yu, G., Wang, L. G., and He, Q. Y. (2015). ChIPseeker: an R/Bioconductor package for ChIP peak annotation, comparison and visualization. *Bioinformatics* 31, 2382–2383. doi: 10.1093/bioinformatics/btv145

Zaman, S., Lippman, S. I., Zhao, X., and Broach, J. R. (2008). How *Saccharomyces* responds to nutrients. *Annu. Rev. Genet.* 42, 27–81. doi: 10.1146/annurev.genet.41.110306.130206

Zhang, Y., Liu, T., Meyer, C. A., Eeckhoute, J., Johnson, D. S., Bernstein, B. E., et al. (2008). Model-based analysis of ChIP-Seq (MACS). *Genome Biol.* 9:R137. doi: 10.1186/gb-2008-9-9-r137

Zhou, C., Li, M., Lu, S., Cheng, Y., Guo, X., He, X., et al. (2021). Engineering of cis-element in *Saccharomyces cerevisiae* for efficient accumulation of value-added compound squalene via downregulation of the downstream metabolic flux. *J. Agric. Food Chem.* 69, 12474–12484. doi: 10.1021/acs.jafc.1c04978

Zhou, Y., Zhou, B., Pache, L., Chang, M., Khodabakhshi, A. H., Tanaseichuk, O., et al. (2019). Metascape provides a biologist-oriented resource for the analysis of systems-level datasets. *Nat. Commun.* 10:1523.

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

frontiers | Frontiers in Microbiology

ORIGINAL RESEARCH
published: 20 June 2022
doi: 10.3389/fmicb.2022.890856

Arginine Catabolism and Polyamine Biosynthesis Pathway Disparities Within *Francisella tularensis* Subpopulations

*Yinshi Yue[1], Bhanwar Lal Puniya[2], Tomáš Helikar[2], Benjamin Girardo[1], Steven H. Hinrichs[1] and Marilynn A. Larson[1]**

[1]Department of Pathology and Microbiology, University of Nebraska Medical Center, Omaha, NE, United States,
[2]Department of Biochemistry, University of Nebraska–Lincoln, Lincoln, NE, United States

Francisella tularensis is a highly infectious zoonotic pathogen with as few as 10 organisms causing tularemia, a disease that is fatal if untreated. Although *F. tularensis* subspecies *tularensis* (type A) and subspecies *holarctica* (type B) share over 99.5% average nucleotide identity, notable differences exist in genomic organization and pathogenicity. The type A clade has been further divided into subtypes A.I and A.II, with A.I strains being recognized as some of the most virulent bacterial pathogens known. In this study, we report on major disparities that exist between the *F. tularensis* subpopulations in arginine catabolism and subsequent polyamine biosynthesis. The genes involved in these pathways include the *speHEA* and *aguAB* operons, along with *metK*. In the hypervirulent *F. tularensis* A.I clade, such as the A.I prototype strain SCHU S4, these genes were found to be intact and highly transcribed. In contrast, both subtype A.II and type B strains have a truncated *speA* gene, while the type B clade also has a disrupted *aguA* and truncated *aguB*. Ablation of the chromosomal *speE* gene that encodes a spermidine synthase reduced subtype A.I SCHU S4 growth rate, whereas the growth rate of type B LVS was enhanced. These results demonstrate that spermine synthase SpeE promotes faster replication in the *F. tularensis* A.I clade, whereas type B strains do not rely on this enzyme for *in vitro* fitness. Our ongoing studies on amino acid and polyamine flux within hypervirulent A.I strains should provide a better understanding of the factors that contribute to *F. tularensis* pathogenicity.

Keywords: *Francisella tularensis*, tularemia, metabolism, amino acid metabolism, polyamine biosynthesis

OPEN ACCESS

Edited by:
Xinqing Zhao,
Shanghai Jiao Tong University, China

Reviewed by:
Sezer Okay,
Hacettepe University, Turkey
Karsten R. O. Hazlett,
Albany Medical College,
United States

****Correspondence:***
Marilynn A. Larson
malarson@unmc.edu

Specialty section:
This article was submitted to Microbial Physiology and Metabolism, a section of the journal Frontiers in Microbiology

Received: *06 March 2022*
Accepted: *27 May 2022*
Published: *20 June 2022*

Citation:
Yue Y, Puniya BL, Helikar T, Girardo B, Hinrichs SH and Larson MA (2022) Arginine Catabolism and Polyamine Biosynthesis Pathway Disparities Within Francisella tularensis Subpopulations. Front. Microbiol. 13:890856. doi: 10.3389/fmicb.2022.890856

INTRODUCTION

Francisella tularensis is a highly infectious pathogen with as few as 10 organisms causing the fatal zoonotic disease tularemia if left untreated, and infects over 250 species, including humans, rodents, lagomorphs, and arthropods (Sjostedt, 2007). This highly virulent and easily aerosolized facultative intracellular pathogen is classified as a Tier 1 category A select agent by the Centers for Disease Control and Prevention (CDC; Bhattacharjee, 2011). Although the subspecies and subtypes within *F. tularensis* share over 99.5% average nucleotide identity, major disparities in virulence exist between these subpopulations. These pathogenicity differences are due in part

to genomic rearrangements, which are primarily due to the transposition of insertion sequence (IS) elements and direct repeats, that affect gene integrity and expression (Larson et al., 2011, 2015). IS elements typically encode transposases that are flanked by small inverted repeats (Mahillon and Chandler, 1998). As has been observed for other highly virulent intracellular pathogens, *F. tularensis* genomes are small (<2 Mbp) and contain numerous IS elements with IS*Ftu1* being the most abundant mobile element (**Table 1**; Rohmer et al., 2007; Larson et al., 2015). Consequently, IS*Ftu1* is often responsible for the observed gene disruptions that result in a pseudogene, and nucleotide polymorphisms can create a premature stop codon that produce a truncated and non-functional gene product (Rohmer et al., 2007; Carlson et al., 2009).

The clinically relevant *F. tularensis* subpopulations that cause tularemia throughout the northern hemisphere include subspecies *tularensis* (or type A) and subspecies *holarctica* (or type B) (Larsson et al., 2009). Due to differences in genomic organization and virulence, the type A clade has been further divided into subtype A.I and A.II, with the A.I strains being recognized as some of the most pathogenic bacteria known (Molins et al., 2010; Birdsell et al., 2014). Like all facultative intracellular pathogens, *F. tularensis* must be able to rapidly gauge the nutritional environment and appropriately adapt. Amino acids have been proven to be critical compounds that allow this pathogen to survive and proliferate (Ziveri et al., 2017). These metabolic compounds contribute to a wide variety of roles beyond protein assimilation, including the production of energy, sulfate assimilation, purine production, carbon acquisition, and cell wall synthesis.

The initial studies in the development of medium with defined components that would support moderate growth of the fastidious pathogen *F. tularensis* established that 13 amino acids, along with spermine or spermidine were needed (Traub et al., 1955; Nagle et al., 1960). Chamberlain further optimized this chemically defined medium (CDM) for enhanced *F. tularensis in vitro* growth (Chamberlain, 1965), which is frequently used for controlled and specialized studies. However, most of the information about amino acid metabolism in the hypervirulent A.I prototype strain SCHU S4 has derived from *in silico* genome analyses, which predicted the absence or incomplete synthesis pathways for at least six amino acids, specifically arginine, cysteine, histidine, lysine, methionine, and tyrosine (Larsson et al., 2005; Rohmer et al., 2007; Meibom and Charbit, 2010). Although attenuated type B LVS is predicted to share similar amino acid auxotrophies with the hypervirulent A.I clade, with the exception of tyrosine in which LVS is capable of synthesizing (Meibom and Charbit, 2010), confirmation of amino acid metabolism in the hypervirulent A.I strains such as the prototype SCHU S4 is needed.

The current study describes major disparities between the *F. tularensis* subpopulations in the metabolic enzymes required for arginine catabolism and subsequent polyamine biosynthesis. Polyamines have been determined to contribute to the ability of bacterial pathogens to adapt and survive, and the processes of polyamine synthesis, uptake, and degradation are coordinated to stringently regulate intracellular levels (Shah and Swiatlo, 2008). We report for the first time that the hypervirulent *F. tularensis* A.I strains can synthesize the polyamines agmatine, putrescine, and spermidine *de*

TABLE 1 | Genomic features in representative strains from each *Francisella tularensis* subpopulation.

***Francisella tularensis* features**[a]	**A.I**	**A.I**	**A.II**	**A.II**	**B**	**B**
	SCHU S4 (1941, Ohio)[b]	**MA00-2987 (2000, Massachusetts)**[b]	**WY96-3418 (1996, Wyoming)**[b]	**WY-00W4114 (2000, Wyoming)**[b]	**FSC200 (1998, Sweden)**[b]	**LVS (1930, Russia)**[b]
Length (bp)	1,892,775	1,892,645	1,898,476	1,899,252	1,894,157	1,895,994
GC content (%)	32.3	32.3	32.3	32.3	32.2	32.2
Total protein ORFs	1,660	1,773	1,761	1,757	1,604	1,689
Disrupted ORFs/ pseudogenes	188	124	125	133	276	238
Large duplicated regions (>5 kbp)	3	3	3	3	3	3
Francisella pathogenicity islands	2	2	2	2	2	2
IS*Ftu1* transposase/IS elements[c]	47	47	48	48	58	59
All IS elements (full-length and remnants)[d]	74	74	101	101	107	109
Structural tRNA	38	38	38	38	38	38
Structural rRNA	10	10	10	10	10	10
Noncoding RNA	4	4	4	4	4	4

[a]*Features were based on NCBI annotations for SCHU S4 NC_006570.2, MA00-2987 NZ_CP012372.1, WY96-3418 NC_009257.1, WY-00W4114 NZ_CP009753.1, FSC200 NC_019551.1, and LVS NC_07880.1, as well as other references for more detailed IS element content (Rohmer et al., 2007; Svensson et al., 2012; Larson et al., 2015).*

[b]*Information within the parentheses denotes the year that the respective F. tularensis strain was isolated and the location.*

[c]*Includes only full-length ISFtu1 ORFs, which is the most abundant transposase/insertion sequence (IS) element in the F. tularensis genomes.*

[d]*Includes both full-length and remnants of all IS elements in the respective F. tularensis genome, including ISFtu1 (IS630 family), ISFtu2 (IS5 family), ISFtu3 (ISNCY family, ISHpal-IS1016), ISFtu4 (IS982 family), ISFtu5 (IS4 family), ISFtu6 (IS1595 family), and ISSod13 (IS3 family).*

novo, unlike the A.II and B clades, and that this inherent trait promotes a faster replication rate.

MATERIALS AND METHODS

Bacterial Strains and Growth Conditions

The *F. tularensis* strains used in this study included hypervirulent A.I strains SCHU S4 and MA00-2987, virulent A.II WY96-3418, and attenuated B strain LVS, and were all obtained from the Biodefense and Emerging Infections (BEI) Resources, which was established by the National Institute of Allergy and Infectious Diseases (NIAID). Select agent *F. tularensis* strains were transferred to the University of Nebraska Medical Center in Omaha following the requirements of the Federal Select Agent Program as outlined in the Animal and Plant Health Inspection Service/Centers for Disease Control and Prevention (CDC) Form 2, Guidance Document for Request to Transfer Select Agents and Toxin. Manipulation of viable culture material was performed by authorized individuals within a biosafety level 3 (BSL-3) laboratory certified for select agent work by the United States Department of Health and Human Services using laboratory biosafety criteria, according to requirements of the Federal Select Agent Program. For each experiment, *F. tularensis* strains were cultured from a master stock onto Remel Chocolate agar plates (Lenexa, KS) and incubated at 37°C with 5% CO_2 for 2 days before further subculturing or processing. Growth curves of *F. tularensis* were obtained by culturing in brain heart infusion broth (BHI) (Becton, Dickinson and Company, Sparks, MD) or Chamberlain's CDM (Chamberlain, 1965) at 37°C with shaking in unbaffled flasks or in a Tecan Spark system in which OD_{600} readings were recorded every hour.

RNA Preparation and RNA-Seq

For the RNA-Seq libraries, *F. tularensis* SCHU S4, MA00-2987, WY96-3418, and LVS were grown in BHI to mid-exponential growth phase in triplicate and in four independent experiments. *Francisella tularensis* cells were then treated with RNAprotect Bacteria Reagent (Invitrogen) to retain high RNA integrity. Next, FastRNA Pro Blue Kit (MP Biomedicals) or TRIzol (Invitrogen) was used to isolate total RNA, as recommended by the manufacturer. RNA was treated with Baseline-ZERO™ DNase (Epicentre Biotechnologies, San Diego, CA, United States) and the absence of residual genomic DNA was confirmed using control reactions without the addition of reverse transcriptase and subsequent PCR amplification with *Francisella*-specific primer, as previously described (Larson et al., 2011). To determine the RNA concentration, a Qubit fluorometer with the Qubit RNA HS Assay Kit (Invitrogen) were used, as described by the manufacturer. The integrity of the isolated RNA was checked by fractionation in an agarose gel containing ethidium bromide and visualization using a UV transilluminator.

The ScriptSeq Complete Kit for bacteria (Epicentre) was used to generate a ribodepleted, stranded transcriptome of coding and noncoding RNA for each of the four *F. tularensis* strains, specifically SCHU S4, MA00-2987, WY96-3418, and LVS. To monitor the ribosomal RNA removal process, RNA samples were assessed on a Fragment Analyzer (Agilent Technologies, Santa Clara, CA, United States). Twenty-four RNA-Seq libraries were prepared for each of the *F. tularensis* strains, totaling 96 RNA-Seq libraries for statistical power. Deep RNA sequencing of the *F. tularensis* libraries was performed by the University of Nebraska Sequencing Core facility using the Illumina NextSeq platform. Multiplexing of 48 samples per group and two groups in total were each ran on a single Mid-output Illumina flow cell, generating approximately 2.5 million 150 bp paired end reads per sample in 300 cycles.

Bioinformatic Analyses of RNA-Seq Data

In silico analysis of the RNA-Seq data was performed using genomic sequences for the aforementioned *F. tularensis* strains deposited in the National Institute of Health (NIH) GenBank database. The *F. tularensis* genome sequences were obtained from NCBI microbial genome database and had the following accession numbers: SCHU S4 (NC_006570.2), MA00-2987 (NZ_CP012372.1), WY96-3418 (NC_009257.1), WY-00W4114 (NZ_CP009753.1), FSC200 (NC_019551.1), and LVS (NC_07880.1).

Both FastQC and Trimmomatic were used for testing and adjusting the quality of the samples reads. FastQC was initially used for each read to provide quality data, similar to the average Phred score. The Phred score or quality (Q) score of a base is an integer value representing the estimated probability of an error (P) due to an incorrect base. Q and P were defined as follows:

$$P = 10^{-Q/10}$$

$$Q = -10 \log_{10} P$$

Trimmomatic was utilized to remove the low-quality reads obtained by FastQC. Overrepresented sequences and an average Phred score of less than 20 per base (or 1% probability of an incorrect base) indicates low quality sequences. Therefore, the sequences with the Phred score of less than 20 or overrepresented sequences were eliminated from the reads. After the Trimmomatic modifications, reads with lengths of less than 35 bases were removed from the analysis, and only sequences 35 bases in length or longer were retained. To further ensure quality sequences, FastQC was again applied on the trimmed output results. To align the quality reads for each sample, the splice junction mapper TopHat2 was used (Kim et al., 2013). Bowtie 2 was utilized for indexing the genome assembly (Langmead and Salzberg, 2012). Cufflink identified expressed genes and calculated transcript levels in the aligned reads within the TopHat2 generated Sequence Alignment/Map (SAM) formatted files. The measurement of gene expression was determined using fragments per kilobase of exon per million (FPKM) reads, which quantifies the amount of data that matches a given coding sequence per kilobase length for a given gene or coding sequence per million reads analyzed (Croucher and Thomson, 2010; Giannoukos et al., 2012).

The expression level of the same gene from the different *F. tularensis* strains were then compared to the normalized transcriptome data.

F. tularensis nucleotide and protein sequences of interest were obtained from the GenBank database. The Basic Local Alignment Search Tool (BLAST), which is provided by the NCBI within the National Library of Medicine (NLM), was used to identify regions of similarity between *F. tularensis* nucleotide or protein sequences. Nucleotide and protein alignments were performed using Clustal Omega (Sievers and Higgins, 2014). Metabolic pathways for gene products of interest were assessed using the Kyoto Encyclopedia of Genes and Genomes (KEGG) analysis tool. Note that as the NCBI annotations continue to evolve with intermittent iterations, *speH* may also be referred to as *speD*.

Reverse Transcription Quantitative Real-Time PCR

For reverse transcription quantitative real-time PCR (RT-qPCR), *F. tularensis* strains were grown in BHI or CDM as specified in triplicate and in three independent experiments. RNA was isolated, DNase treated, and checked for integrity and concentrations, as described above. To confirm the transcriptional expression of the genes of interest, first-strand cDNA was prepared using the SuperScript IV First-Strand Synthesis System and gene-specific primers, as recommended by the manufacturer (Invitrogen). Next, cDNA was amplified with appropriate gene-specific primer pairs. If conventional RT-qPCR was performed, products were evaluated by agarose gel fractionation, ethidium bromide staining, and densitometry, using ImageJ software (NIH). PowerUp SYBR Green Master Mix (Applied Biosystems) was also used with a QuantStudio 3 Real-time PCR System (Applied Biosystem), according to the manufacturer's recommendations. Relative gene expression was determined by normalization to the internal control *lpnA* transcript and triplicate samples were evaluated in three independent experiments. Primers utilized for PCR and RT-qPCR are shown in **Supplementary Table S1**.

DNA Manipulations and Reagents

All oligonucleotide primers used in this study were synthesized by Thermo Fisher Scientific and are listed in **Supplementary Table S1**. Genomic DNA was isolated using the Gentra Puregene Cell Kit (Qiagen) or cetyl trimethylammonium bromide (CTAB), according to recommended procedures. Plasmid DNA was prepared with Zymoclean Gel DNA Recovery Kit (Zymo Research Corporation, Irvine, CA, United States), as recommended by the manufacturer. DNA restriction enzyme digests, cloning, and electrophoresis were performed according to standard protocols. PCR was performed using Platinum Taq DNA Polymerase High Fidelity enzyme and associated components (Invitrogen). Ligations were performed using a T4 DNA ligase (Roche) or NEBuilder HiFi DNA Assembly Master Mix (New England Biolabs, Ipswich, MA, United States). DNA fragments were purified using either a QIAquick PCR Purification or QIAquick Gel Extraction Kit (Qiagen). DNA concentrations were obtained using a Nanodrop spectrophotometer or Qubit fluorometer (Invitrogen). DNA sequencing was performed by the University of Nebraska Medical Center Genomics Core Facility using Sanger sequencing. Electroporation of plasmid DNA into *F. tularensis* or *Escherichia coli* was conducted as previously described (Horzempa et al., 2010), and an ECM 630 BTX Electroporation System (BTX Harvard Apparatus) was utilized for these procedures.

Construction of *Francisella tularensis speE* Deletion Mutants and Expression Plasmids

In-frame and markerless *speE* deletion mutants were generated in *F. tularensis* LVS and SCHU S4 using previously described procedures (Horzempa et al., 2010). To preserve operon integrity and any potential regulatory sites within the target gene sequence, nine consecutive codons at the beginning and end of *speE* were retained in the construct, along with three consecutive stop codons that were inserted after the last of the nine codons at the 5′ region of this gene. Fragments that include these regions and ~500 bp upstream and downstream of the *speE* coding sequence were amplified by PCR, digested, and ligated into PstI/SmaI digested pJH1. The resulting pJH1/*speH_speA* construct was DNA sequenced to confirm content and then transferred into *F. tularensis* LVS and SCHU S4 by tri-parental mating, as described previously (Horzempa et al., 2010). Merodiploid strains were recovered and transformed with pGUTS by electroporation for the expression of I-SceI (Horzempa et al., 2010). Colonies resistant to kanamycin were screened by PCR for deletion of the *speE* gene. To cure the *F. tularensis* Δ*speE* mutant of pGUTS, the strains were passaged several times in tryptic soy broth containing cysteine, diluted, and plated on chocolate agar II plates, to obtain isolated CFUs. Plates were incubated for at least 3 days at 37°C with 5% CO_2 and colonies that formed were replica plated onto chocolate II agar plates with and without kanamycin. Colonies sensitive to kanamycin were isolated and retested for sensitivity to this antibiotic. Deletion of *speE* in the resulting *F. tularensis* LVS Δ*speE* and SCHU S4 Δ*speE* mutants were confirmed by PCR amplification of the relevant chromosomal locus and subsequent fractionation in an agarose gel and staining for comparison to the associated wild-type strain amplicon derived with the same primer pair. Content of this chromosomal locus in the *F. tularensis* LVS Δ*speE* and SCHU S4 Δ*speE* mutants was verified by DNA sequencing. RT-qPCR was used to confirm that *speE* was deleted in the Δ*speE* mutant and that the adjacent genes were transcribed at levels similar to wild-type transcripts.

To complement the Δ*speE* mutant, PCR amplification of the full-length *speE* gene was performed and the resulting amplicon was digested with NheI and BamHI. This PCR product was then ligated into pFNLTP, which was digested with the same restriction endonucleases. The *groESL* or native *spe* promoter were cloned upstream of *speE* into the KpnI and XhoI site of the pFNLTP plasmid containing *speE*. Both the promoter region and the *speE* gene in the pFNLTP expression plasmid were sequenced to verify content. The resulting pFN/Pr_*groESL*/*speE* and pFN/Pr_*spe*/*speE* expression constructs

were then electroporated into the *F. tularensis* Δ*speE* mutant for trans-complementation. Expression of *speE* in the complemented Δ*speE* mutant was confirmed by RT-qPCR for triplicate samples in three independent experiments.

TMT-Labeled Untargeted Mass Spectrometry

Bicinchoninic acid (BCA) assays were performed to determine the concentration of protein in the lysates obtained from *F. tularensis* SCHU S4, MA00-2987, WY96-3418, and LVS. Next, equivalent amounts of protein from the four different *F. tularensis* strains were reduced with DTT, alkylated with iodoacetamide, and digested overnight with sequencing-grade trypsin (Promega). Tryptic peptides were labeled using the Tandem Mass Tag (TMT) 6-plex Reagents (Thermo Fisher Scientific), pooled, and concentrated to 20 µl by vacuum centrifugation. The TMT-labeled peptides were then analyzed using a high-resolution nano-liquid chromatography with tandem mass spectrometry (LC–MS/MS) Tribrid system that included an Orbitrap Fusion with a Lumos coupled to an UltiMate 3,000 HPLC system (Thermo Fisher Scientific). Peptides (500 ng) were run on a pre-column (Acclaim PepMap 100, 75 µm × 2 cm, nanoViper, Thermo Fisher Scientific) and then on an analytical column (Acclaim PepMap RSCL, 75 µm × 50 cm, nanoViper, Thermo Fisher Scientific). The samples were eluted using a 90-min linear gradient of acetonitrile (4%–99%) in 0.1% formic acid.

All LC–MS/MS samples were analyzed using the Protein Discoverer software, version 2.1 (Thermo Fisher Scientific). Sequest HT was set up to search the Swiss-Prot database for both reviewed and unreviewed *F. tularensis* entries. The parameters for Sequest HT were set as follows: enzyme was trypsin, maximum missed cleavage was 2, precursor mass tolerance was 10 ppm, peptide tolerance was ±0.6 Da, fixed modifications were carbamidomethyl (C) and TMT sixplex (any N-terminus), and dynamic modifications was oxidation (M) and TMT sixplex (K). The parameters for Reporter Ion Quantifier were set as follows: integration tolerance was 20 ppm, integration method was most confident centroid, mass analyzer was FTMS, MS order was MS3, activation type was HCD, minimum collision energy was 0, and maximum collision energy was 1,000. The parameters for Percolator were set as follows: target FDR (strict) was 0.01, target FDR (relaxed) was 0.05, and validation was based on *q*-value.

The Proteome Discoverer software version 2.1 (Thermo Fisher Scientific) was again used to normalize the total peptide count for each sample. The algorithm summarizes the peptide group abundance for each sample and determines the maximum sum for all files to obtain a normalization factor. After normalization, the Reporter Ion Quantifier node performs scaling so that the average of all channels is obtained, and the node scales the abundance values of each sample so that the average of all control samples is 100. All other samples are then scaled up or down relative to 100 and when using multiplexed files, the node processes the samples from each file separately. Protein samples were prepared for each of the four *F. tularensis* strains in two independent experiments.

Statistical Analysis

To determine if differential gene expression was significant, three statistical tests were applied amongst the four *F. tularensis* strains, specifically the two-sample T-test, the Mann–Whitney U-test, and the two-sample Kolmogorov–Smirnov test. The 24 sets of RNA-seq data for each gene in each strain was compared pairwise to each of the other RNA-Seq sets for the same gene in other strains, using the above three statistical tests. A differentially expressed gene was defined as significant if the value of p value of all statistical tests for all the pairwise comparisons was less than 0.01 (value of $p < 0.01$).

For RT-qPCR analyses, two-way ANOVA followed by Tukey's *post hoc* tests were used for statistical analysis when multiple groups were analyzed. Data are presented as means ± SEM and are representative of triplicates in not less than three independent experiments. GraphPad Prism was used for the statistical analyses.

RESULTS

Differential Expression of *speHEA* and *aguAB* in *Francisella tularensis* Subpopulations

To characterize potential differences in the expression of metabolic enzymes within the *F. tularensis* subpopulations, deep RNA sequencing was performed for all three clades. RNA-Seq was performed on representative select agent *F. tularensis* strains that included hypervirulent subtype A.I strains SCHU S4 (A.I prototype) and MA00-2987 and virulent subtype A.II strain WY96-3418 (A.II prototype). Attenuated type B LVS was also evaluated since this non-select agent strain is often studied as a surrogate for this species and may serve as a reference. These strains were cultivated to mid-log in BHI, since growth in this medium recapitulates gene expression by this intracellular pathogen during a macrophage infection (Hazlett et al., 2008). Analysis of the 24 RNA-Seq libraries for each of the four *F. tularensis* strains revealed that most of the genes within the *speHEA* and *aguAB* operons were transcribed at higher levels in the hypervirulent A.I strains compared to subtype A.II WY96-3418 and attenuated type B LVS (**Figure 1A**). Examination of the genomic organization and content of *speHEA* and *aguAB* in the *F. tularensis* A.I, A.II, and B clades revealed that *speHEA* and *aguAB* were encoded in two adjacent operons (**Figure 1A**; **Supplementary Figure S1**). Further, only the A.I genomes contained intact *speHEA* and *aguAB* operons in which all the genes were full-length (**Figure 1A**; **Supplementary Figure S1**). More specifically, in the A.II and B genomes, *speA* was truncated. In type B strains, *aguA* and *aguB* were not full-length genes; *aguA* was disrupted by the insertion sequence element IS*Ftu1* and *aguB* had a premature stop codon (**Figure 1A**; **Supplementary Figure S1**). Together these analyses revealed that the genomic content of *speHEA*/*aguAB* differed between the *F. tularensis* A.I, A.II, and B clades, but was conserved within each subpopulation, and that only the A.I strains contained all intact genes within these operons.

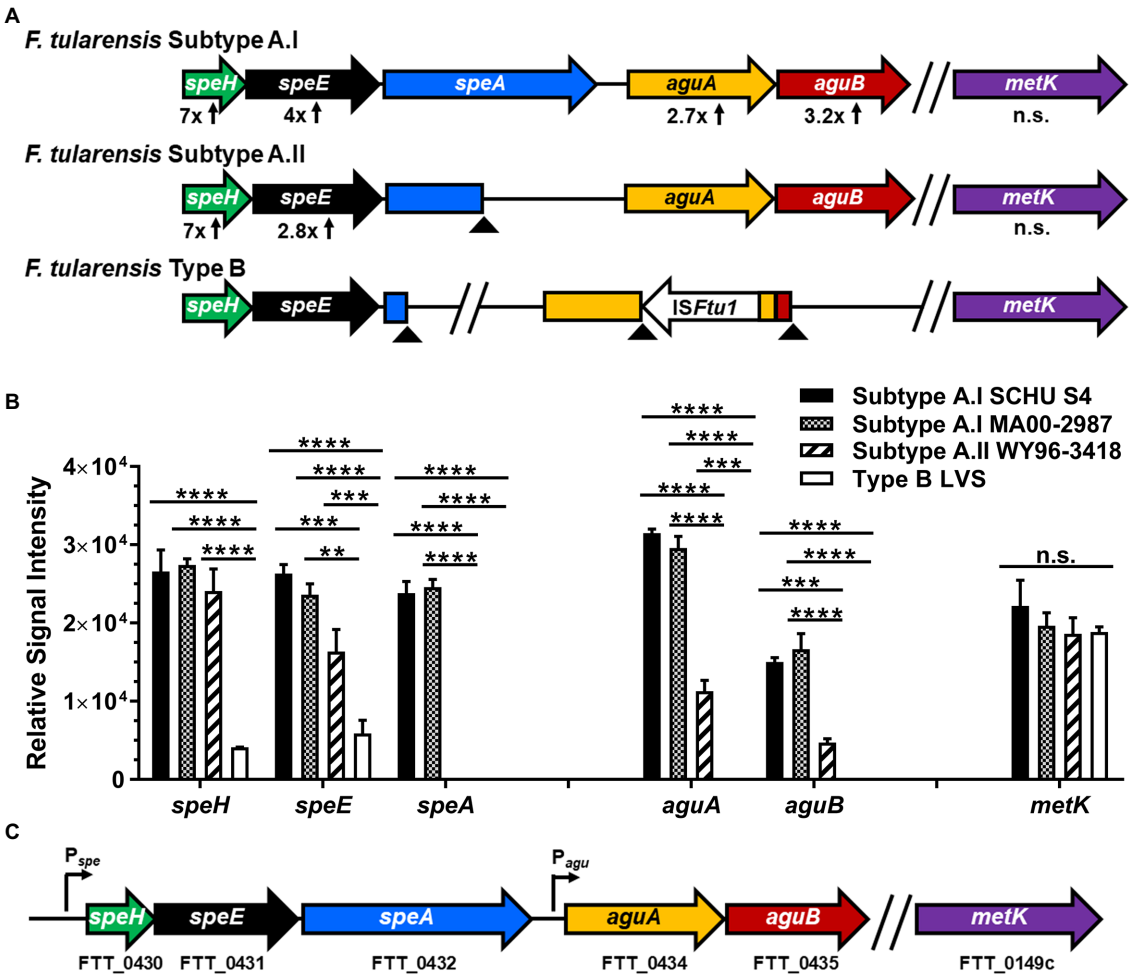


FIGURE 1 | Genetic organization of genes involved in arginine and methionine catabolism and subsequent polyamine biosynthesis within the different *F. tularensis* subpopulations, along with associated transcript levels. **(A)** Diagram depicting the *speHEA* and *aguAB* operons and *metK* gene in the genomes of *F. tularensis* subtype A.I, subtype A.II, and type B clades. **(B)** RT-qPCR results confirming differential expression of the genes within the *speHEA* and *aguAB* operons and *metK* transcript levels in A.I strains SCHU S4 and MA00-2987, A.II strain WY96-3418, and type B strain LVS, during mid-log growth in brain heart infusion broth (BHI). **(C)** Diagram showing the locus tags for prototype A.I strain SCHU S4 associated with *speH*, *speE*, *speA*, *aguA*, *aguB*, and *metK*, as well as the *speHEA* operon promoter (P_{spe}) and *aguAB* operon promoter (P_{agu}). In panel **A**, the fold-change increase in mRNA abundance obtained in RT-qPCR relative to the transcript from intact genes in LVS or alternatively WY96-3418 as appropriate, are shown below the associated gene. Arrowheads denote a premature stop codon that results in a truncated gene. In panel **B**, transcript levels obtained in RT-qPCR were normalization to *lpnA* mRNA and the data shown are cumulative of three independent experiments, each conducted with three technical replicates, and are expressed as the mean ± SEM. Data were analyzed using two-way ANOVA with multiple comparisons and Tukey's *post hoc* tests. Value of $p > 0.05$ were considered not significant (n.s.), and in panel **B**, comparisons that were n.s. are only shown for *metK*. $^{**}p < 0.01$; $^{***}p < 0.001$; and $^{****}p < 0.0001$.

To verify that the *speHEA* and *aguAB* genes were differentially transcribed in the A.I strains SCHU S4 and MA00-2987, A.II WY96-3418, and type B LVS, as was initially shown in the deep RNA sequencing data, RT-qPCR was performed. These results showed that the *F. tularensis speH* transcript, which is the first gene in the *spe* operon was expressed at a 7-fold higher level in the *F. tularensis* A.I and A.II strains relative to this gene in type B LVS (**Figures 1A,B**). The adjacent *speE* gene was expressed at a 4- and 2.8-fold higher level in the *F. tularensis* A.I and A.II strains, respectively, compared to type B LVS. The transcripts for *aguA* and *aguB* were 2.7- and 3.2-fold higher in abundance in the A.I strains, respectively, compared to the A.II strain, whereas both *aguA* and *aguB* are disrupted in all type B strains including LVS. Collectively, these data confirmed the differential expression of *F. tularensis speHEA* and *aguAB* and a variable trait between the A.I, A.II, and B subpopulations.

In the *F. tularensis* A.I strains, the stop codon and start codons for *speH* and *speE* overlap by a single bp, *speE* and *speA* are separated by 25 bp, and *aguA* and *aguB* are separated by a single bp, whereas 218 bp separate *speA* and *aguA*. Canonical −35 and −10 promoter elements were only identified upstream of *speH* and *aguA*. More importantly, our RT-qPCR results confirmed that *speHEA* and *aguAB* are two separate operons, each with a separate promoter (**Figure 1C**).

Arginine and Methionine Catabolic Pathways

An examination of the predicted function of *F. tularensis speHEA* and *aguAB* revealed that these genes, along with *metK* that is located at a different chromosomal location, encode enzymes involved in methionine and arginine catabolism and subsequent polyamine biosynthesis (**Figure 2**). However, unlike the *speHEA* and *aguAB* genes, *metK* transcript levels were similar and only slightly higher in the *F. tularensis* A.I SCHU S4 and MA00-2987strains relative to A.II WY96-3418 and type B LVS (**Figure 1B**). The genes within the *speHEA* and *aguAB* operons, along with *metK* in the A.I strains encode enzymes that contribute to the catabolism of arginine and methionine and the subsequent biosynthesis of the polyamines agmatine, putrescine, and spermidine (**Figure 2**). The *metK* gene product S-adenosylmethionine synthetase and ATP are required for the first step in converting methionine to S-adenosylmethionine (SAM; **Figure 2**). SAM is the principal methyl donor in transmethylation and an aminopropyl donor in polyamine synthesis (Chiang et al., 1996). The next metabolic step for the conversion of SAM to S-adenosylmethioninamine in *F. tularensis* requires the enzyme S-adenosylmethionine decarboxylase, which is encoded by *speH* (**Figure 2**). This process releases carbon dioxide while producing S-adenosylmethioninamine, one of two substrates required by spermidine synthase SpeE for spermidine biosynthesis in *F. tularensis* (**Figure 2**). The methionine catabolic enzymes MetK, SpeH, and SpeE are all full-length in the *F. tularensis* A.I, A.II, and B clades, in contrast to the arginine catabolic enzymes SpeA in both the A.II and B subpopulations and AguA and AguB in the type B clade (**Figures 1, 2**).

In *F. tularensis*, arginine decarboxylase SpeA is responsible for the catabolism of arginine to agmatine and is only full-length in the A.I strains due to premature stop codons within this gene in the A.II and B clades (**Figures 1, 2**). This catabolic step releases carbon dioxide while producing the substrate agmatine. Next, agmatine deiminase AguA catabolizes agmatine to produce N-carbamoylputrescine in both the A.I and A.II strains, but not the type B strains since this gene is disrupted by the IS element IS*Ftu1* (**Figures 1, 2**). N-carbamoylputrescine amidase AguB in the *F. tularensis* type A strains has dual enzymatic activity that can catabolizes both N-carbamoylputrescine and citrulline, producing putrescine and ornithine, respectively (**Figure 2**). The catabolism of these two substrates by AguB results in the production of carbon dioxide and ammonium in both metabolic reactions.

A comparison of the primary sequence for these arginine and methionine metabolic enzymes in the *F. tularensis* A.II and B clades relative to the associated enzyme in the A.I subpopulation was next assessed. SpeH had one conserved residue difference in which the A.I and B strains had a methionine at residue 19 and the A.II strains had a valine at this position (M19V). SpeE had two conserved residues (I186L in both A.II and B strains and K266R in A.II strains), one non-conserved amino acid in A.II strains (G189D), and one non-conserved residue in the type B clade (S107F), relative to this enzyme in A.I subpopulation. Since SpeA was only produced by the A.I clade and not the A.II and B subpopulations, no comparisons could be made for this enzyme. A comparison of AguA and AguB in the A.I and A.II strains revealed that AguA in A.II strains had three non-conservative residue differences (K62N, R199H, and Q253K), whereas AguB was identical in the two type A subpopulations. Relative to S-adenosylmethionine synthetase MetK in the A.I strains, A.II strains had two non-conserved amino acids (E67K and G98D) and type B strains had one non-conserved residue (T105A). Together these results indicated that S-adenosylmethionine decarboxylase SpeH was the most conserved enzyme in all three *F. tularensis* subpopulations in these comparisons, suggesting an important role in SAM catabolism. SpeH also provides one of the two substrates required for spermidine biosynthesis in the A.I strains (**Figure 2**), indicating an additional vital role within this species.

Chromosomal Location of *Francisella tularensis spe/agu* Operons and *metK*

The chromosomal location and directionality of the *speHEA*/*aguAB* operons and *metK* were compared in the *F. tularensis* A.I, A.II, and B strains. These results revealed that the spatial location and directionality of the *spe*/*agu* operons

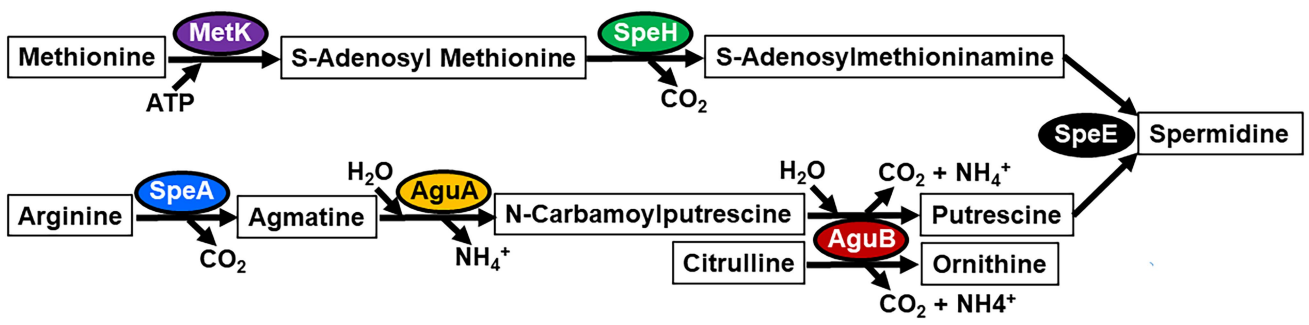


FIGURE 2 | Metabolic pathways for methionine and arginine catabolism and subsequent polyamine biosynthesis in hypervirulent *F. tularensis* A.I strains. Full-length metabolic enzymes in the *F. tularensis* A.I clade include S-adenosylmethionine synthetase (MetK), S-adenosylmethionine decarboxylase (SpeH), arginine decarboxylase (SpeA), agmatine deiminase (AguA), N-carbamoylputrescine amidase (AguB), and spermidine synthase (SpeE).

differed the most between the A.I and A.II strains and were the most similar in the A.I and B clades, even though only *speH* and *speE* are intact in the type B strains (**Figure 3**). In contrast, *metK* in the A.I and A.II clades was positioned in a similar chromosomal location and direction, and the type B *metK* differed in both aspects relative to the type A strains. However, in all three *F. tularensis* subpopulations, *metK* was positioned near the gene encoding the bacterial replication initiator protein DnaA, with the A.I *metK* being the closest to *dnaA* (**Figure 3**). Although all *F. tularensis* subpopulations share numerous features and have numerous IS elements, the type B strains have considerably more IS elements, which has ostensibly contributed to the higher abundance of pseudogenes in this clade (**Table 1**). Moreover, the most abundant IS element, specifically IS*Ftu1* from the IS630 family of transposases was responsible for the disruption of *aguA* in the type B clade.

The genes adjacent to the *speHEA*/*aguAB* operons in the *F. tularensis* A.I, A.II, and B clades were conserved in organization and directionality, despite the genomic, transcriptomic, and proteomic differences within the *spe*/*agu* operons. In all three subpopulations, threonine synthase *thrC1* was adjacent and 5′ to *speH*, while UDP-2,3-diacylglucosamine hydrolase *lpxH* was adjacent and 3′ to the *agu* operon (**Table 2**; **Supplementary Figure S2A**). The proteins encoding by *thrC1* and *lpxH* both provide critical cellular functions; ThrC1 catalyzes the production of threonine and LpxH is involved in lipid A biosynthesis. The *thrC1* gene was transcribed in the same direction as the *speHEA*/*aguAB* operons, whereas *lpxH* was expressed in the opposite orientation. A remnant of a beta-galactosidase gene encoding only the first 76 amino acids of the 656-residue full-length enzyme also separated *thrC1* and *speH* and was transcribed in the opposite direction of these two genes. This beta-galactosidase gene remnant shared the highest homology to associated enzyme in *Burkholderia* and *Neisseria*.

The genes flanking *metK* were also highly conserved within the *F. tularensis* A.I, A.II, and B subpopulations. A gene encoding a fatty acid desaturase was located 5′ to *metK*, but in the opposite orientation, while *rpsP* was positioned 3′ to *metK* and also oriented in the opposing direction, indicating that *metK* is monocistronic (**Table 2**; **Supplementary Figure S2B**). The *rpsP* gene encodes the 30S ribosomal protein S16 that is highly expressed and required for protein translation. The conservation of the genes and associated regulatory regions adjacent to the *speHEA*/*aguAB* operons and *metK* gene in the different *F. tularensis* subpopulations may provide a fitness advantage to this facultative intracellular pathogen.

Transcriptional Expression of *Francisella tularensis speHEA*, *aguAB*, and *metK*

The *F. tularensis* promoter region for the *speHEA* operon was 586 bp in length in the A.I and A.II clades and 564 bp in the B strains due to a 22 bp deletion. The same 22 bp sequence was located directly after the threonine synthase *thrC* coding sequence in the type A strains, whereas in the type B clade this sequence was incorporated into the 3′ end of the slightly shorter *thrC* open reading frame. The type B strains also had a single adenine insertion and a SNP that was a thymine instead of a guanine compared to the type A strains in the *spe* promoter region. A putative OmpR-like response regulator binding sequence (5′-TTGTAGCA-3′) was located between the −35 and −10 region in the *spe* promoter and was conserved within the A.I, A.II, and B clades. Although type A strains encode two OmpR-like orphan response regulators that are 228 and 229 residues in length with 54% amino acid similarity, type B strains only encode one of these regulatory factors. These and other differences in regulatory capabilities between these subpopulations may have contributed to the reduced expression levels of *speH* and *speE* in type B LVS relative to the A.I and A.II strains (**Figure 1B**).

A comparison of the nucleotide content in the 218 bp region separating the *spe* and *agu* operons in the A.I, A.II, and B clades revealed four SNPs. These SNPs included three guanines and one thymine in subtype A.I; two guanines and two thymines in subtype A.II; and one guanine, two adenines, and one thymine in type B strains. One SNP comprised the last 3′ nucleotide in the Pribnow box/−10 region, one SNP followed

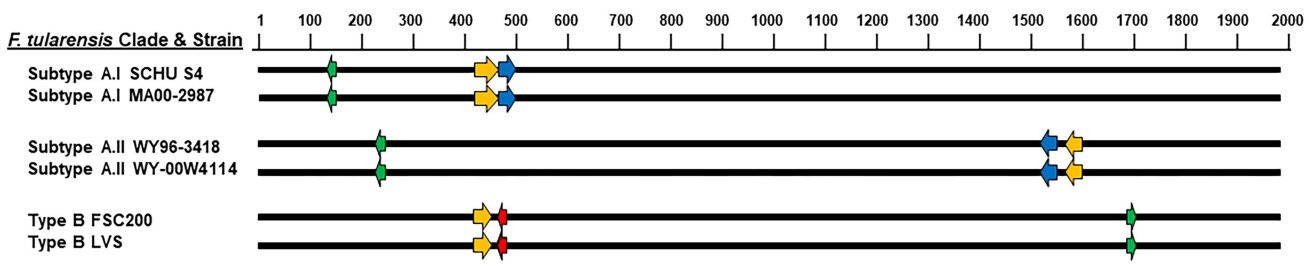


FIGURE 3 | Diagram depicting chromosomal location and directionality of the *speHEA*/*aguAB* operons and the *metK* gene in representative *F. tularensis* subtype A.I, subtype A.II, and type B strains. The relative nucleotide position in kilobase pairs of the double-stranded DNA genome in *F. tularensis* is shown at the top, and the circular chromosome in *F. tularensis* was linearized for this figure using *dnaA* as the initial coding sequence after the first base pair. Representatives from the *F. tularensis* subpopulations include subtype A.I strains SCHU S4 and MA00-2987, subtype A.II strains WY96-3418 and WY-00W4114, and type B strains FSC200 and LVS. The *speHEA* and *aguAB* operons are represented by the adjacent orange and blue arrows, respectively, and *metK* is represented with a green arrow. The IS*Ftu1* insertion sequence element in the type B *aguA* gene is denoted with a red arrow. The arrows show the direction of the operon and gene coding sequences and were enlarged for visualization. Nucleotide sequences in the NCBI database were used to position the operons and genes of interest in the *F. tularensis* genomes, and the associated accession numbers are described in the Materials and Methods section.

TABLE 2 | Genes adjacent to *speHEA/aguAB* operons and *metK* in *F. tularensis* subpopulations.[a]

Clade	Gene(s)[a]	5′ Adjacent gene	3′ Adjacent gene
Subtype A.I	*speHEA/aguAB*	Threonine synthase *thrC*	UDP-2,3-diacylglucosamine hydrolase *lpxH*
Subtype A.I	*metK*	Fatty acid desaturase	30S Ribosomal protein S16 *rpsP*
Subtype A.II	*speHEA/aguAB*	Threonine synthase *thrC*	UDP-2,3-diacylglucosamine hydrolase *lpxH*
Subtype A.II	*metK*	Fatty acid desaturase	30S Ribosomal protein S16 *rpsP*
Type B	*speHEA/aguAB*	Threonine synthase *thrC*	UDP-2,3-diacylglucosamine hydrolase *lpxH*
Type B	*metK*	Fatty acid desaturase	30S Ribosomal protein S16 *rpsP*

[a]**Supplementary Figure S2** *shows a diagram with the directionality of the 5′ and 3′ adjacent genes to the speHEA/aguAB operons and metK in F. tularensis.*

the Shine-Dalgarno/ribosomal binding site, and the remaining SNPs were in the intergenic region between the *spe* and *agu* operons. Overall, the A.II and B clades have more A/T enrichment in the intergenic nucleotides between the *spe* and *agu* operons than the A.I strains.

The *F. tularensis metK* promoter region was 293 bp and highly conserved in the A.I, A.II, and B subpopulations with only two SNPs, neither of which were located in the predicted Pribnow box/−10 region, −35 region, nor Shine-Dalgarno/ribosomal binding site. One SNP was a thymine in the A.I and B clades and a cytosine in the A.II strains. The other SNP was an adenine in the A.I clade, whereas the A.II and B strains had a guanine in this position. Interestingly, the promoter for *metK* and the adjacent *rpsP* gene, which encodes the 16S ribosomal protein, share the same predicted Pribnow box/−10 region that was located approximately in the middle of this promoter region. In addition to *rpsP*, other essential genes that are highly expressed during bacterial replication were located nearby to *metK*, including genes that encode 50S ribosomal proteins, tRNAs, and DNA-directed RNA polymerase subunits. These analyses support the premise that *metK* and *rpsP* are co-regulated and transcribed during favorable growth conditions for rapid replication.

A global RNA-Seq analysis of the immortalized mouse monocyte cell line P388D1 infected with *F. tularensis* LVS reported a 6.2- and 4.8-fold increase in *speH* and *speE* at 4 h post infection (hpi), respectively, and then a decrease in their abundance by 8 hpi (Bent et al., 2013). These results suggest that *speH* and *speE* contribute to the phagosomal escape of this intracellular pathogen, even though *speA*, *aguA*, and *aguB* are not intact in type B strains. Another notable global transcriptome study showed that the expression of the polyamine synthesis genes *speHEA*, *aguAB*, and *metK* were upregulated in primary mouse macrophages infected with *F. tularensis* SCHU S4 (Wehrly et al., 2009). Collectively, these data demonstrate that these metabolic gene products provide an important role in the persistence of *F. tularensis*, particularly for hypervirulent A.I SCHU S4.

Protein Expression of *Francisella tularensis speHEA*, *agu*, and *metK*

Untargeted proteomic data from subtype A.I strains SCHU S4 and MA00-2987, subtype A.II strain WY96-3418, and type B LVS grown in BHI to mid-log growth phase were obtained using TMT mass spectrometry. These results revealed that AguB was 3.7- and 4-fold more abundant in the A.I strains SCHU S4 and MA00-2987, respectively, relative to this enzyme in A.II WY96-3418. MetK levels were similar in the A.I strains SCHU S4 and MA00-2987 and were only 1.7- and 1.8-fold higher when compared to the abundance of this enzyme in the A.II and B strains, respectively. In SCHU S4 and MA00-2987, MetK was 2-fold more abundant than AguB. SpeH, SpeE, SpeA, and AguA, however, were not detected in these analyses.

A comparison of proteins produced by *F. tularensis* LVS grown in several different culture broths showed similar MetK levels, while SpeE abundance was low or undetectable (Holland et al., 2017). In another global proteomic investigation, MetK was detected in both LVS and SCHU S4, whereas AguA, AguB, SpeE were only identified in SCHU S4 with just one of the two biological replicates detecting SpeE (Lenco et al., 2009). Collectively, these results indicated that although all the *F. tularensis* clades can metabolize methionine to produce SAM and S-adenosylmethioninamine, only the hypervirulent A.I strains can produce the polyamines agmatine, putrescine, and spermidine *de novo via* the enzymes encoded by *speHEA*, *aguAB*, and *metK*.

Construction and Characterization of *Francisella tularensis* Δ*speE* Mutants

To determine the contribution of spermidine synthase SpeE in *F. tularensis*, chromosomal *speE* deletion mutants were generated in hypervirulent A.I SCHU S4 and type B LVS, since this enzyme was intact in both clades. The methods used to generate markerless and in-frame *speE* deletion mutants (Δ*speE*) in *F. tularensis* were carried out as previously described (Horzempa et al., 2010). PCR amplification of the *speHEA* chromosomal locus in the *F. tularensis* SCHU S4 Δ*speE* and LVS Δ*speE* mutants confirmed that the 0.8 kbp *speE* gene was deleted from the genome of the respective wild-type strain (**Figure 4**; **Supplementary Figure S3**). DNA sequencing of this region confirmed the expected content and verified that the adjacent genes *speH* and *speA* were not disrupted. RT-qPCR validated that the absence of *speE* mRNA in the *F. tularensis* Δ*speE* mutants and that no polar effects occurred in the transcription of the adjacent upstream and downstream genes in the SCHU S4 Δ*speE* mutant relative to wild-type SCHU S4 (**Supplementary Figure S4**).

To assess if the deletion of *speE* altered growth, a comparison of wild-type SCHU S4 and wild-type LVS to the associated and isogenic Δ*speE* mutant was evaluated. These results showed that SCHU S4 Δ*speE* grew slower than wild-type SCHU S4, whereas the LVS Δ*speE* mutant grew faster than wild-type LVS in

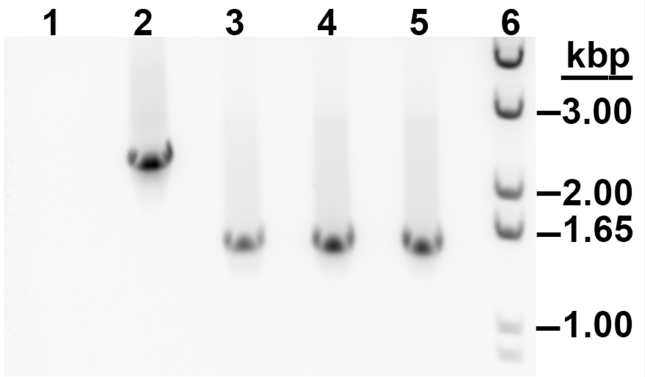


FIGURE 4 | PCR amplification of the *speHEA* chromosomal locus in *F. tularensis* A.I SCHU S4 wildtype and the isogenic Δ*speE* mutants. Shown are SCHU S4 wildtype (lane 2) and the isogenic Δ*speE* mutant used for this study (lane 3), along with two additional Δ*speE* mutants (lanes 4 and 5). The no template control (lane 1) and dsDNA marker (lane 6) are also shown.

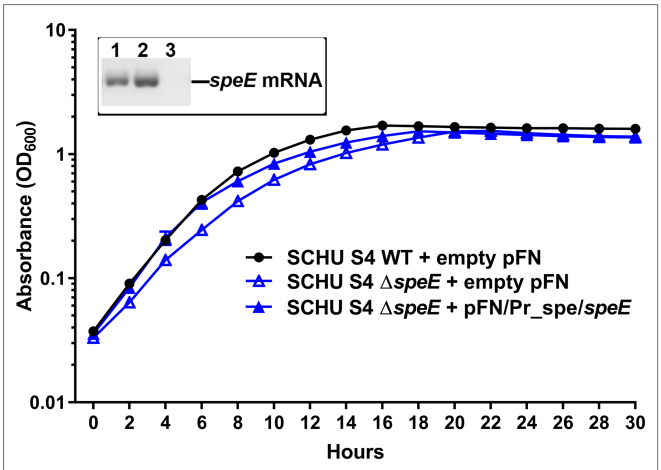


FIGURE 5 | Trans-complementation of the *F. tularensis* SCHU S4 Δ*speE* mutant. Growth of wild-type SCHU S4 and the isogenic SCHU S4 Δ*speE* mutant that contained an empty pFN plasmid was compared to the in *trans* complemented SCHU S4 Δ*speE* mutant containing the *speE* expression plasmid pFN/Pr_*spe*/*speE* during growth in Chamberlain's chemically defined medium (CDM). Insert shows representative RT-qPCR products obtained with *speE*-specific primers for the presence or absence of *speE* transcripts as appropriate for wild-type SCHU S4 containing empty plasmid (lane 1), complemented SCHU S4 Δ*speE* mutant containing expression plasmid pFN/Pr_*spe*/*speE* (lane 2), and SCHU S4 Δ*speE* mutant containing empty plasmid (lane 3). The mean with ± SEM is shown for the growth curves obtained from triplicate samples in three independent experiments.

Chamberlain's CDM, suggesting that spermidine synthase SpeE provides a fitness advantage to SCHU S4 and not LVS. To ensure that no polar and off-target effects occurred during the process of producing the SCHU S4 Δ*speE* mutant that contributed to this phenotype, trans-complementation was performed. For these assessments, the SCHU S4 Δ*speE* mutant was complemented in *trans* with the *Francisella* expression plasmid pFNLTP containing the constitutive *groESL* promoter upstream of *speE* (pFN/Pr_*gro*/*speE*). However, these complementation experiments were unsuccessful, so the native *spe* promoter was then utilized to regulate *speE* expression. The usage of the native *spe* promoter in pFNLTP (pFN/Pr_*spe*/*speE*) successfully complemented the SCHU S4 Δ*speE* mutant in *trans*, resulting in growth comparable to wild-type SCHU S4 (**Figure 5**). As shown in the **Figure 5** insert, *speE* transcripts were present in wild-type SCHU S4 with empty plasmid and the complemented SCHU S4Δ*speE* mutant containing the pFN/Pr_*spe*/*speE* expression plasmid, but not the SCHU S4Δ*speE* mutant with empty plasmid. These data validated that *speE* is not transcribed by *F. tularensis* SCHU S4 Δ*speE* and that the reduced growth rate of this mutant can be complemented.

Transcriptional Regulation of *speHEA*, *aguAB*, and *metK* in Wildtype and Δ*speE* Mutant

To determine if arginine and methionine concentrations contributed to the regulation of the transcriptional expression of the *speHEA*/*aguAB* operons and *metK* by *F. tularensis* SCHU S4 and the isogenic Δ*speE* mutant, RT-qPCR was performed. For these experiments, the *F. tularensis* strains were cultured in modified CDM containing 0, 0.3, 3, or 6 mM of arginine or methionine. When the bacteria were in mid-log growth phase, RNA was isolated and prepared for RT-qPCR, as described in the Materials and Methods section. No significant change in transcript abundance was observed for the *spe*, *agu*, and *metK* genes (data not shown). Further, an examination of the transcript levels for the proposed arginine and methionine importers ArgP and MetNIQ, respectively, did not change during growth with different concentrations of these essential amino acids (data not shown). These results indicated that the transcriptional expression of *F. tularensis* A.I *speHEA*/*aguAB* operons and *metK*, as well as the amino acid transporters ArgP and MetNIQ is not regulated by just the levels of these essential amino acids.

Next, we sought to determine if stress conditions altered the transcriptional expression of the genes in the polyamine biosynthesis pathway in *F. tularensis* SCHU S4 and the isogenic Δ*speE* mutant when exposed to acidic conditions, hydrogen peroxide, and nitric oxide. After exposure to a low pH of 5.2 in CDM adjusted with hydrochloric acid for 1 h, 1 mM hydrogen peroxide for 1 h, and 100 μM of the nitric oxide-generating compound S-nitroso-N-acetylpenicillamine (SNAP) for 0.5 h, RNA was extracted and evaluated by RT-qPCR. Again, these conditions did not significantly increase nor decrease the transcript levels for these genes (data not shown). Therefore, a combination of stress factors and/or different conditions may be needed to modify the transcriptional expression of the *speHEA*, *aguAB*, and *metK* genes in *F. tularensis* SCHU S4 and the isogenic Δ*speE* mutant. Alternatively, these gene products may be regulated at the translational or post-translational level to promptly respond to internal and external nutritional status and stress.

Charity and associates showed that LVS *speH* transcript levels were 3.3- and 3.9-fold lower in Δ*mglA* and Δ*sspA* deletion mutants, respectively (Charity et al., 2007). Since MglA and SspA facilitate sigma factor RpoD (σ^{70}) binding to DNA to regulate virulence and virulence-enhancing genes (Travis et al.,

2021), these findings suggest that these transcription factors regulate *spe* operon expression and that these genes may contribute to the virulence of *F. tularensis*. Therefore, our ongoing studies will investigate the various factors that regulate the polyamine biosynthesis genes *speHEA*, *aguAB*, and *metK*, along with the role of the associated metabolic enzymes in *F. tularensis* pathogenicity.

Growth Comparison of *Francisella tularensis* Wildtype and Δ*speE* Mutant Strains

To evaluate the growth rate and yield of *F. tularensis* SCHU S4 and LVS relative to the isogenic Δ*speE* mutants in different media, these strains were grown in Chamberlain's CDM and in BHI. These assessments demonstrated that all four strains grew to a higher cell density in Chamberlain's CDM, which contains the polyamine spermidine or spermine (Chamberlain, 1965), in comparison to BHI (**Figure 6**). These results validated that the components within Chamberlain's CDM provide the appropriate provisions for more efficient *in vitro* growth by *F. tularensis*, as well as corroborates with the data obtained by others (Mc Gann et al., 2010).

In BHI, *F. tularensis* SCHU S4 and the SCHU S4 Δ*speE* mutant differed in growth rate and maximum cell density (**Figure 6A**), whereas the growth of wild-type LVS and the LVS Δ*speE* mutant was similar (**Figure 6B**). Unlike wild-type SCHU S4, the SCHU S4 Δ*speE* mutant exhibited diauxic growth in BHI and grew at a faster rate than wild-type SCHU S4 during these two exponential phases. The SCHU S4 Δ*speE* mutant also reached a slightly higher cell density (maximum OD_{600} of 1.2) than wild-type SCHU S4 (maximum OD_{600} of 1.0), which occurred after 25 h of growth in BHI (**Figure 6A**). However, after SCHU S4 Δ*speE* reached maximum cell density, cell death occurred with a concomitant and continuous decrease in cell density. Unlike the SCHU S4 Δ*speE* mutant, wild-type SCHU S4 did not grow in a diauxic manner in BHI, but rather exhibited the canonical bacterial lag, exponential growth, and stationary phases. After 26 h of growth in BHI, SCHU S4 cell density reached a maximum OD_{600} of 1.0 and remained mainly unchanged in stationary phase for at least 48 h when the experiments were terminated. Total cell density based on the area under the growth curve for SCHU S4 and SCHU S4 Δ*speE* grown in BHI was 32.13 ± 0.26 and 33.89 ± 0.20, respectively.

For both wild-type LVS and LVS Δ*speE* cultured in BHI, the growth rate slightly decreased after 13 h until maximum cell density was obtained (OD_{600} of 0.8), which occurred after a total of 22 h when the cells were entering stationary phase (**Figure 6B**). The overall cell density based on the area under the growth curve for LVS and LVS Δ*speE* grown in BHI was 26.02 ± 0.22 and 25.28 ± 0.20, respectively. Since growth rate and cell density were similar between wild-type LVS and the isogenic Δ*speE* mutant, these results indicated that SpeE spermidine synthase does not substantially contribute to the fitness of these type B strains when cultured in BHI. In contrast, these findings indicated that spermidine synthase SpeE promotes steady state growth in SCHU S4 and implicates that spermidine production by this enzyme may also provide a protective function against factors that promote cell death during growth in BHI.

In Chamberlain's CDM, both *F. tularensis* SCHU S4 and SCHU S4 Δ*speE* grew substantially faster than LVS and LVS Δ*speE*, with wild-type SCHU S4 growing 3.8-fold faster than wild-type LVS. The generation time of 1.22 h was obtained for SCHU S4, whereas LVS had a doubling time of 4.66 h (**Figure 6B**). However, wild-type LVS and the LVS Δ*speE* mutant eventually reached a similar maximum cell density after 33 h (OD_{600} of 1.4) and 27 h (OD_{600} of 1.53), respectively (**Figure 6B**). In comparison, wild-type SCHU S4 and the SCHU S4 Δ*speE* mutant grown in CDM reached the highest cell density after 12 h (OD_{600} of 1.6) and 17 h (OD_{600} of 1.54), respectively (**Figure 6A**). Further, in CDM both LVS and the LVS Δ*speE* mutant showed diauxic growth (**Figure 6B**), which was not the case for SCHU S4 and the SCHU S4 Δ*speE* mutant (**Figure 6A**). In CDM, the total cell density based on the area under the growth curves for SCHU S4 and SCHU S4 Δ*speE* was 50.07 ± 0.20 and 44.98 ± 0.12, respectively. For LVS and LVS Δ*speE* grown in CDM, the overall cell density based on the area under the growth curves was 31.29 ± 0.33 and 36.61 ± 0.17, respectively. In general, these results demonstrated that spermidine synthase SpeE in wild-type SCHU S4 contributes to an enhanced growth rate and cell yield in CDM, whereas the deletion of *speE* in LVS augmented the growth rate and slightly increased cell yield in CDM (**Figure 6**).

DISCUSSION

In the current study, we showed that considerable differences exist between *F. tularensis* subpopulations in the metabolic pathways associated with arginine and polyamine metabolism, but not methionine catabolism. Hypervirulent subtype A.I strains such as SCHU S4 and MA00-2987 have intact *speHEA* and *aguAB* operons, whereas subtype A.II strains have a prematurely terminated *speA* gene and type B strains have truncated *speA* and *aguAB* genes (**Figure 1A**; **Supplementary Figure S1**). Moreover, the *F. tularensis* A.I strains SCHU S4 and MA00-2987 expressed higher levels of all the genes involved in arginine and polyamine metabolism within these two operons. Inactivation of a gene(s) in an operon or pathway will most likely promote an increase in the frequency of A/T mutations and ongoing erosion (Ochman, 2003; Bobay and Ochman, 2017), which appears to be the case within these altered regions in the *F. tularensis* A.II and B clades. These modifications are common in bacteria whose ecology includes an intracellular niche (Balbi et al., 2009). Implication of lineage derivations and genetic polymorphisms has been previously discussed (Svensson et al., 2005; Larson et al., 2015). Nevertheless, the disparities in the *spe*/*agu* arginine and polyamine metabolic pathways within the *F. tularensis* subpopulations suggests that the A.II and B clades must utilize alternative and available mechanisms to acquire essential polyamines.

The initial enzyme involved in the *spe*/*agu* arginine catabolism and polyamine biosynthesis pathway is arginine decarboxylase SpeA, which was only intact in *F. tularensis* A.I strains. Since *F. tularensis* does not have the biosynthetic pathways to synthesize arginine, this essential amino acid must be imported for numerous

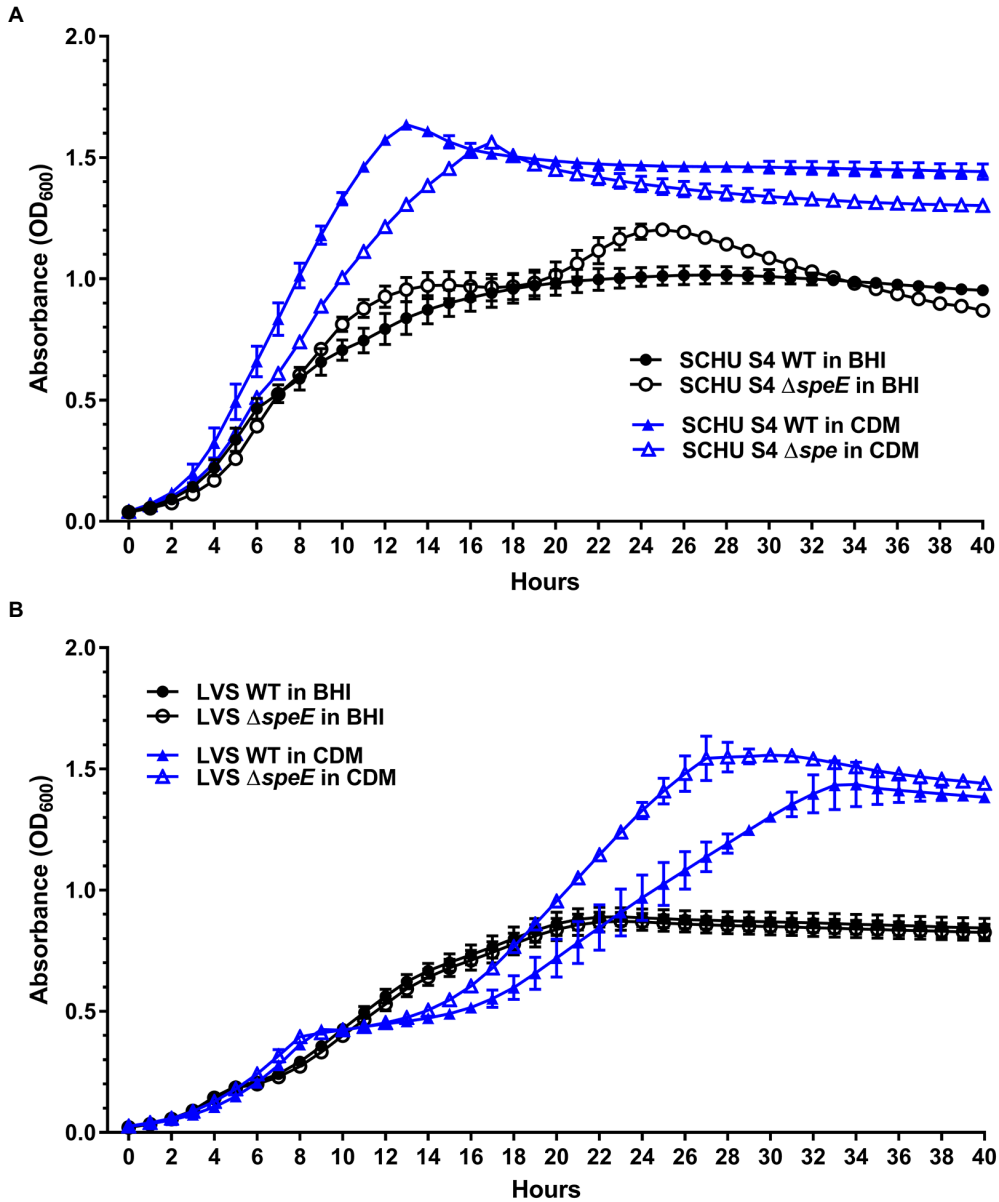


FIGURE 6 | Growth comparison of *F. tularensis* subtype A.I SCHU S4 and type B LVS versus the associated and isogenic Δ*speE* mutant in brain heart infusion broth (BHI) and Chamberlain's chemically defined medium (CDM). **(A)** *F. tularensis* SCHU S4 wildtype (WT) growth versus the isogenic Δ*speE* mutant and **(B)** *F. tularensis* LVS wildtype (WT) versus the isogenic Δ*speE* mutant in BHI (black circles) and Chamberlain's CDM (blue arrowheads). WT strains are denoted with filled symbols and the Δ*speE* mutants are indicated with unfilled symbols. The mean with ±SEM for triplicate samples in three independent experiments is shown.

cellular functions such as protein production, as well as *de novo* polyamine biosynthesis in the A.I clade. The genome of this intracellular pathogen does not encode a protein with homology to an ArgR-like repressor for the regulation of arginine metabolism, but there may be other unidentified regulators. Further, since arginine abundance and the stress factors examined did not substantially alter the transcriptional expression of the *spe*/*agu* genes nor *argP*, which encodes an arginine importer, regulation may be occurring post-transcriptionally and/or post-translationally. Alternatively, additional factors are required or *F. tularensis* SCHU S4 may endogenously have the capability to import this critical amino acid or peptides containing this residue in a steady state manner.

In addition to arginine catabolism by arginine decarboxylase SpeA, there are three additional enzymes encoded in the *F. tularensis* A.I genome that can hydrolyze arginine. These enzymes are annotated in the National Center for Biotechnology Information (NCBI) database as arginine-like deiminases and catalyze the conversion of arginine to the non-proteinogenic amino acid citrulline, as well as ammonium. The importance of arginine

deiminases in a bacterial species with a highly reduced genome was shown to facilitate adaptation from the environment to a mammalian niche, due to the ability of these enzymes to confer protection against acid stress through ammonia production (Tian et al., 2022). In *F. tularensis* A.I genomes, these genes encode a 304-residue and a 307-residue arginine-like deiminase, which have 48% identity in 146 residues and 64% similarity in 195 residues to each other, whereas the third gene encodes a 183-residue protein with no homology to the two larger enzymes. The smallest 183-residue arginine-like deiminase shares the highest homology to a 261-residue amidinotransferase in *Candidatus Roizmanbacteria* (40% identity in 73 out of 183 residues and 62% similarity in 113 out of 183 residues) and another water-associated bacterium, specifically *Methylocystis* (39% identity in 72 out of 187 residues and 58% similarity in 110 out of 187 residues). *Francisella tularensis* A.II genomes contain the 183-residue and 304-residue arginine-like deiminases while the type B genomes encode only the 307-residue arginine-like deiminase, exemplifying more metabolic differences within these subpopulations.

AguB N-carbamoylputrescine amidase, which is only intact in the *F. tularensis* A.I and A.II strains and not the type B clade, hydrolyzes citrulline and N-carbamoylputrescine, producing ornithine and the polyamine putrescine, respectively. Citrullinase activity has been historically used to differentiate *F. tularensis* type A strains from type B strains due to the differentiating capability of AguB in the type A clade to catabolize citrulline to ornithine. An investigation by others evaluated a SCHU S4 Δ*aguB* mutant and reported that no growth defect in culture nor infected macrophages was observed; however, mice infected with this mutant lived longer and had a significantly reduced bacterial burden in the lungs, liver, and spleen relative to wildtype infected mice, indicating that *aguB* contributes to intracellular persistence and pathogenesis (Mahawar et al., 2009). Catabolism by arginine deiminase and N-carbamoylputrescine amidase AguB, as well as agmatine deiminase AguA all generate ammonium, which may contribute to basification of the surrounding environment and/or to a metabolic signaling network for an appropriate cellular response. However, the functional contribution of these enzymes and resulting products to the fitness of *F. tularensis* will require further study.

Methionine serves as an initiation amino acid in protein translation in all domains of life (Kozak, 1999). This essential amino acid can be converted to SAM (Cantoni, 1953), and in *F. tularensis*, SAM is derived from the catabolism of methionine by SAM synthetase MetK (**Figure 2**). SAM provides an activated methyl group donor for many fundamental cellular processes, including the methylation of proteins and nucleic acids and serves as a substrate for spermidine biosynthesis. This metabolite is the second most used biological compound after ATP (Cantoni, 1975). Analysis of a *F. tularensis* SCHU S4 transposon library revealed that *metK* contributed to fitness and was essential (Ireland et al., 2019). Our results corroborate with these findings and show that the transcriptional and translational expression of *metK* is similar in the *F. tularensis* A.I, A.II, and B clades with only slightly higher levels in the A.I strains. An evaluation of the *metK* promoter in representative strains from each of these subpopulations revealed a highly conserved 293 bp region with only two nucleotide differences. This promoter region also controls the adjacent and highly transcribed gene *rpsP*, which encodes the ribosomal protein S16 in the 30S small subunit of the ribosome that is required to stabilize this complex. The conservation of this promoter region and shared Pribnow/−10 region that regulates both the expression of the SAM synthetase *metK* gene and the divergently transcribed *rpsP* gene implicates the presence of highly regulated *cis*-acting elements that sense metabolite abundance to co-regulate these vital gene products. However, neither low nor high methionine abundance nor the stress conditions examined substantially altered the transcriptional expression of *metK* or the genes that encode the methionine importer MetNIQ in wild-type SCHU S4 or the isogenic Δ*speE* mutant. These data suggest that *F. tularensis metK* expression may be needed to produce SAM even when conditions are not optimum and that the abundance of this transcript is tightly regulated, supporting the critical role of this multifunctional metabolite that includes serving as a substrate for spermidine biosynthesis in the A.I strains.

In the *F. tularensis* A.I clade, the catabolism of arginine produces the polyamines agmatine and diamine putrescine, whereas the metabolism of both arginine and methionine is required to provide the substrates for *de novo* biosynthesis of the triamine spermidine (**Figure 2**). Polyamines are small organic polycations with amine groups that are found in nearly all living organisms and are involved in a variety of essential biological processes (Igarashi and Kashiwagi, 2000, 2010). Studies have shown that polyamines modulate chromatin organization, DNA replication, transcription, translation, ion transport, and membrane dynamics by the electrostatic interactions of these small polycationic molecules with negatively charged macromolecules, including DNA, RNA, proteins, and phospholipids (Wortham et al., 2007; Shah and Swiatlo, 2008; Michael, 2018). The ability of polyamines to enhance the replication rate of *Francisella* was discovered during the efforts of others to produce a chemically defined medium (Traub et al., 1955; Nagle et al., 1960; Chamberlain, 1965). Moreover, other studies reported on spermine-responsive regions within *F. tularensis* IS elements that regulate the expression of several adjacent genes (Carlson et al., 2009), and identified a spermine-responsive gene that was shown to be involved in virulence (Russo et al., 2011).

The results from the current study comparing *F. tularensis* wild-type A.I and B strains with the isogenic Δ*speE* mutants demonstrated that SpeE spermidine synthase contributes to an increase in the growth rate of subtype A.I SCHU S4 and conversely hinders the growth rate of type B LVS in Chamberlain's CDM. In BHI, SCHU S4 Δ*speE* grew similar to wildtype, but unlike wildtype, this mutant exhibited diauxic growth and began to lyse during stationary phase. Since BHI has a high content of essential nutrients, including polyamines and peptides as an amino acid source, this medium is often used to cultivate fastidious pathogens. Therefore, although SCHU S4 Δ*speE* can no longer synthesize spermidine *de novo*, BHI provides polyamines, as well as amino acids *via* peptides for growth. The apparent biphasic growth, however, indicates that this mutant needs to make a critical phenotypic adaption for continued replication, which appears to be energy costly and ultimately cause cell death. Nonetheless, these data suggest that the polyamine spermidine serves an important role for

the overall fitness and persistence of SCHU S4. In contrast, the deletion of *speE* from the LVS chromosome may reduce energy expenditure for the expression of this gene, since most of the genes involved in *spe*/*agu* metabolic pathways are truncated in this subpopulation. These findings indicate that type B strains have adapted to the absence of an intact *spe*/*agu* arginine catabolism pathway, whereas *de novo* biosynthesis of spermidine by *F. tularensis* A.I SCHU S4 provides a fitness advantage and promotes faster replication.

In summary, our findings show for the first time that arginine catabolism and polyamine biosynthesis differ considerably within the *F. tularensis* clades. The numerous IS elements within this species have contributed to genomic rearrangements and decay, as well as metabolic differences within the subpopulations. The broad host range and environmental conditions encountered by this facultative intracellular pathogen requires nutritional plasticity to persist. Although most intracellular pathogens have limited metabolic capacities, we propose that the retention of the polyamine biosynthesis pathway in the A.I subpopulation provides a fitness advantage that contributes to the persistence of this hypervirulent clade. Moreover, the loss of genes needed to produce agmatine, putrescine, and spermidine *de novo* in the *F. tularensis* A.II and B clades suggests that compensatory mechanisms exist to provide vital polyamines. Our current studies are focused on evaluating the contribution and regulation of arginine, methionine, and polyamine metabolism in *F. tularensis*, including during an intracellular infection, to better understand *F. tularensis* pathobiology.

DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found at: NCBI BioProject—PRJNA817369, NCBI GEO—GSE202948.

AUTHOR CONTRIBUTIONS

YY and ML designed the experiments. YY, BG, and ML performed the experiments. BP and ML analyzed the transcriptome and proteome data. ML drafted the manuscript. All authors contributed to the article and approved the submitted version.

FUNDING

This study was supported in part by a grant from the Nebraska Research Initiative (NRI) to ML and TH and NIH grant R35GM119770 to TH. The UNMC Genomics Core Facility receives partial support from the National Institute for General Medical Science (NIGMS) INBRE P20GM103427-19, as well as the National Cancer Institute and The Fred and Pamela Buffett Cancer Center Support grant P30CA036727. The contents of this publication are the sole responsibility of the authors and do not necessarily represent the views of any institution or agency.

ACKNOWLEDGMENTS

We thank Joseph Horzempa, Gerald Nau, Anders Sjöstedt, and Vinai Thomas for the valuable discussions and sharing their technical expertise, as well as Robert Moore, Khalid Sayood, and Amirsalar Mansouri for their assistance with the bioinformatics.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmicb.2022.890856/full#supplementary-material

REFERENCES

Balbi, K. J., Rocha, E. P., and Feil, E. J. (2009). The temporal dynamics of slightly deleterious mutations in *Escherichia coli* and *Shigella* spp. *Mol. Biol. Evol.* 26, 345–355. doi: 10.1093/molbev/msn252

Bent, Z. W., Brazel, D. M., Tran-Gyamfi, M. B., Hamblin, R. Y., VanderNoot, V. A., and Branda, S. S. (2013). Use of a capture-based pathogen transcript enrichment strategy for RNA-Seq analysis of the *Francisella tularensis* LVS transcriptome during infection of murine macrophages. *PLoS One* 8:e77834. doi: 10.1371/journal.pone.0077834

Bhattacharjee, Y. (2011). Biosecurity. Panel selects most dangerous select agents. *Science* 332, 1491–1492. doi: 10.1126/science.332.6037.1491

Birdsell, D. N., Johansson, A., Ohrman, C., Kaufman, E., Molins, C., Pearson, T., et al. (2014). *Francisella tularensis* subsp. *tularensis* group A.I, United States. *Emerg. Infect. Dis.* 20, 861–865. doi: 10.3201/eid2005.131559

Bobay, L. M., and Ochman, H. (2017). The evolution of bacterial genome architecture. *Front. Genet.* 8:72. doi: 10.3389/fgene.2017.00072

Cantoni, G. L. (1953). S-Adenosylmethionine; a new intermediate formed enzymatically from L-methionine and adenosinetriphosphate. *J. Biol. Chem.* 204, 403–416. doi: 10.1016/S0021-9258(18)66148-4

Cantoni, G. L. (1975). Biological methylation: selected aspects. *Annu. Rev. Biochem.* 44, 435–451. doi: 10.1146/annurev.bi.44.070175.002251

Carlson, P. E., Horzempa, J., O'Dee, D. M., Robinson, C. M., Neophytou, P., Labrinidis, A., et al. (2009). Global transcriptional response to spermine, a component of the intramacrophage environment, reveals regulation of *Francisella* gene expression through insertion sequence elements. *J. Bacteriol.* 191, 6855–6864. doi: 10.1128/JB.00995-09

Chamberlain, R. E. (1965). Evaluation of live tularemia vaccine prepared in a chemically defined medium. *Appl. Microbiol.* 13, 232–235. doi: 10.1128/am.13.2.232-235.1965

Charity, J. C., Costante-Hamm, M. M., Balon, E. L., Boyd, D. H., Rubin, E. J., and Dove, S. L. (2007). Twin RNA polymerase-associated proteins control virulence gene expression in *Francisella tularensis*. *PLoS Pathog.* 3:e84. doi: 10.1371/journal.ppat.0030084

Chiang, P. K., Gordon, R. K., Tal, J., Zeng, G. C., Doctor, B. P., Pardhasaradhi, K., et al. (1996). S-adenosylmethionine and methylation. *FASEB J.* 10, 471–480. doi: 10.1096/fasebj.10.4.8647346

Croucher, N. J., and Thomson, N. R. (2010). Studying bacterial transcriptomes using RNA-seq. *Curr. Opin. Microbiol.* 13, 619–624. doi: 10.1016/j.mib.2010.09.009

Giannoukos, G., Ciulla, D. M., Huang, K., Haas, B. J., Izard, J., Levin, J. Z., et al. (2012). Efficient and robust RNA-seq process for cultured bacteria and complex community transcriptomes. *Genome Biol.* 13:R23. doi: 10.1186/gb-2012-13-3-r23

Hazlett, K. R., Caldon, S. D., McArthur, D. G., Cirillo, K. A., Kirimanjeswara, G. S., Magguilli, M. L., et al. (2008). Adaptation of *Francisella tularensis* to the

mammalian environment is governed by cues which can be mimicked in vitro. *Infect. Immun.* 76, 4479–4488. doi: 10.1128/IAI.00610-08

Holland, K. M., Rosa, S. J., Kristjansdottir, K., Wolfgeher, D., Franz, B. J., Zarrella, T. M., et al. (2017). Differential growth of *Francisella tularensis*, which alters expression of virulence factors, dominant antigens, and surface-carbohydrate synthases, governs the apparent virulence of *Ft* SchuS4 to immunized animals. *Front. Microbiol.* 8:1158. doi: 10.3389/fmicb.2017.01158

Horzempa, J., Shanks, R. M., Brown, M. J., Russo, B. C., O'Dee, D. M., and Nau, G. J. (2010). Utilization of an unstable plasmid and the I-SceI endonuclease to generate routine markerless deletion mutants in *Francisella tularensis*. *J. Microbiol. Methods* 80, 106–108. doi: 10.1016/j.mimet.2009.10.013

Igarashi, K., and Kashiwagi, K. (2000). Polyamines: mysterious modulators of cellular functions. *Biochem. Biophys. Res. Commun.* 271, 559–564. doi: 10.1006/bbrc.2000.2601

Igarashi, K., and Kashiwagi, K. (2010). Modulation of cellular function by polyamines. *Int. J. Biochem. Cell Biol.* 42, 39–51. doi: 10.1016/j.biocel.2009.07.009

Ireland, P. M., Bullifent, H. L., Senior, N. J., Southern, S. J., Yang, Z. R., Ireland, R. E., et al. (2019). Global analysis of genes essential for *Francisella tularensis* Schu S4 growth in vitro and for fitness during competitive infection of Fischer 344 rats. *J. Bacteriol.* 201, e00630–e00718. doi: 10.1128/JB.00630-18

Kim, D., Pertea, G., Trapnell, C., Pimentel, H., Kelley, R., and Salzberg, S. L. (2013). TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. *Genome Biol.* 14:R36. doi: 10.1186/gb-2013-14-4-r36

Kozak, M. (1999). Initiation of translation in prokaryotes and eukaryotes. *Gene* 234, 187–208. doi: 10.1016/s0378-1119(99)00210-3

Langmead, B., and Salzberg, S. L. (2012). Fast gapped-read alignment with bowtie 2. *Nat. Methods* 9, 357–359. doi: 10.1038/nmeth.1923

Larson, M. A., Fey, P. D., Bartling, A. M., Iwen, P. C., Dempsey, M. P., Francesconi, S. C., et al. (2011). *Francisella tularensis* molecular typing using differential insertion sequence amplification. *J. Clin. Microbiol.* 49, 2786–2797. doi: 10.1128/JCM.00033-11

Larson, M. A., Nalbantoglu, U., Sayood, K., Zentz, E. B., Bartling, A. M., Francesconi, S. C., et al. (2015). *Francisella tularensis* subtype A.II genomic plasticity in comparison with subtype A.I. *PLoS One* 10:e0124906. doi: 10.1371/journal.pone.0124906

Larsson, P., Elfsmark, D., Svensson, K., Wikstrom, P., Forsman, M., Brettin, T., et al. (2009). Molecular evolutionary consequences of niche restriction in *Francisella tularensis*, a facultative intracellular pathogen. *PLoS Pathog.* 5:e1000472. doi: 10.1371/journal.ppat.1000472

Larsson, P., Oyston, P. C., Chain, P., Chu, M. C., Duffield, M., Fuxelius, H. H., et al. (2005). The complete genome sequence of *Francisella tularensis*, the causative agent of tularemia. *Nat. Genet.* 37, 153–159. doi: 10.1038/ng1499

Lenco, J., Link, M., Tambor, V., Zakova, J., Cerveny, L., and Stulik, A. J. (2009). iTRAQ quantitative analysis of *Francisella tularensis* ssp. *holarctica* live vaccine strain and *Francisella tularensis* ssp. *tularensis* SCHU S4 response to different temperatures and stationary phases of growth. *Proteomics* 9, 2875–2882. doi: 10.1002/pmic.200700820

Mahawar, M., Kirimanjeswara, G. S., Metzger, D. W., and Bakshi, C. S. (2009). Contribution of citrulline ureidase to *Francisella tularensis* strain Schu S4 pathogenesis. *J. Bacteriol.* 191, 4798–4806. doi: 10.1128/JB.00212-09

Mahillon, J., and Chandler, M. (1998). Insertion sequences. *Microbiol. Mol. Biol. Rev.* 62, 725–774. doi: 10.1128/MMBR.62.3.725-774.1998

Mc Gann, P., Rozak, D. A., Nikolich, M. P., Bowden, R. A., Lindler, L. E., Wolcott, M. J., et al. (2010). A novel brain heart infusion broth supports the study of common *Francisella tularensis* serotypes. *J. Microbiol. Methods* 80, 164–171. doi: 10.1016/j.mimet.2009.12.005

Meibom, K. L., and Charbit, A. (2010). *Francisella tularensis* metabolism and its relation to virulence. *Front. Microbiol.* 1:140. doi: 10.3389/fmicb.2010.00140

Michael, A. J. (2018). Polyamine function in archaea and bacteria. *J. Biol. Chem.* 293, 18693–18701. doi: 10.1074/jbc.TM118.005670

Molins, C. R., Delorey, M. J., Yockey, B. M., Young, J. W., Sheldon, S. W., Reese, S. M., et al. (2010). Virulence differences among *Francisella tularensis* subsp. *tularensis* clades in mice. *PLoS One* 5:e10205. doi: 10.1371/journal.pone.0010205

Nagle, S. C., Anderson, R. E., and Gary, N. D. (1960). Chemically defined medium for the growth of *Pasteurella tularensis*. *J. Bacteriol.* 79, 566–571. doi: 10.1128/jb.79.4.566-571.1960

Ochman, H. (2003). Neutral mutations and neutral substitutions in bacterial genomes. *Mol. Biol. Evol.* 20, 2091–2096. doi: 10.1093/molbev/msg229msg229

Rohmer, L., Fong, C., Abmayr, S., Wasnick, M., Larson Freeman, T. J., Radey, M., et al. (2007). Comparison of *Francisella tularensis* genomes reveals evolutionary events associated with the emergence of human pathogenic strains. *Genome Biol.* 8:R102. doi: 10.1186/gb-2007-8-6-r102

Russo, B. C., Horzempa, J., O'Dee, D. M., Schmitt, D. M., Brown, M. J., Carlson, P. E. Jr., et al. (2011). A *Francisella tularensis* locus required for spermine responsiveness is necessary for virulence. *Infect. Immun.* 79, 3665–3676. doi: 10.1128/IAI.00135-11

Shah, P., and Swiatlo, E. (2008). A multifaceted role for polyamines in bacterial pathogens. *Mol. Microbiol.* 68, 4–16. doi: 10.1111/j.1365-2958.2008.06126.x

Sievers, F., and Higgins, D. G. (2014). Clustal omega. *Curr. Protoc. Bioinformatics* 48, 11–16. doi: 10.1002/0471250953.bi0313s48

Sjostedt, A. (2007). Tularemia: history, epidemiology, pathogen physiology, and clinical manifestations. *Ann. N. Y. Acad. Sci.* 1105, 1–29. doi: 10.1196/annals.1409.009

Svensson, K., Larsson, P., Johansson, D., Bystrom, M., Forsman, M., and Johansson, A. (2005). Evolution of subspecies of *Francisella tularensis*. *J. Bacteriol.* 187, 3903–3908. doi: 10.1128/JB.187.11.3903-3908.2005

Svensson, K., Sjodin, A., Bystrom, M., Granberg, M., Brittnacher, M. J., Rohmer, L., et al. (2012). Genome sequence of *Francisella tularensis* subspecies *holarctica* strain FSC200, isolated from a child with tularemia. *J. Bacteriol.* 194, 6965–6966. doi: 10.1128/JB.01040-12

Tian, J., Utter, D. R., Cen, L., Dong, P. T., Shi, W., Bor, B., et al. (2022). Acquisition of the arginine deiminase system benefits epiparasitic *Saccharibacteria* and their host bacteria in a mammalian niche environment. *Proc. Natl. Acad. Sci. U. S. A.* 119:e2114909119. doi: 10.1073/pnas.2114909119

Traub, A., Mager, J., and Grossowicz, N. (1955). Studies on the nutrition of *Pasteurella tularensis*. *J. Bacteriol.* 70, 60–69. doi: 10.1128/jb.70.1.60-69.1955

Travis, B. A., Ramsey, K. M., Prezioso, S. M., Tallo, T., Wandzilak, J. M., Hsu, A., et al. (2021). Structural basis for virulence activation of *Francisella tularensis*. *Mol. Cell* 81, 139.e10–152.e10. doi: 10.1016/j.molcel.2020.10.035

Wehrly, T. D., Chong, A., Virtaneva, K., Sturdevant, D. E., Child, R., Edwards, J. A., et al. (2009). Intracellular biology and virulence determinants of *Francisella tularensis* revealed by transcriptional profiling inside macrophages. *Cell. Microbiol.* 11, 1128–1150. doi: 10.1111/j.1462-5822.2009.01316.x

Wortham, B. W., Patel, C. N., and Oliveira, M. A. (2007). Polyamines in bacteria: pleiotropic effects yet specific mechanisms. *Adv. Exp. Med. Biol.* 603, 106–115. doi: 10.1007/978-0-387-72124-8_9

Ziveri, J., Barel, M., and Charbit, A. (2017). Importance of metabolic adaptations in *Francisella* pathogenesis. *Front. Cell. Infect. Microbiol.* 7:96. doi: 10.3389/fcimb.2017.00096

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Advantages of publishing in Frontiers

OPEN ACCESS
Articles are free to read for greatest visibility and readership

FAST PUBLICATION
Around 90 days from submission to decision

HIGH QUALITY PEER-REVIEW
Rigorous, collaborative, and constructive peer-review

TRANSPARENT PEER-REVIEW
Editors and reviewers acknowledged by name on published articles

Frontiers
Avenue du Tribunal-Fédéral 34
1005 Lausanne | Switzerland

Visit us: www.frontiersin.org
Contact us: frontiersin.org/about/contact

REPRODUCIBILITY OF RESEARCH
Support open data and methods to enhance research reproducibility

DIGITAL PUBLISHING
Articles designed for optimal readership across devices

FOLLOW US
@frontiersin

IMPACT METRICS
Advanced article metrics track visibility across digital media

EXTENSIVE PROMOTION
Marketing and promotion of impactful research

LOOP RESEARCH NETWORK
Our network increases your article's readership