Received: 11 October 2021 | Revised: 3 November 2021 | Accepted: 3 November 2021

DOI: 10.1111/mmi.14842

MICROREVIEW

WILEY

KH domain proteins: Another family of bacterial RNA matchmakers?

Mikolaj Olejniczak [1] | Xiaofang Jiang [2] | Maciej M. Basczok [1] | Gisela Storz [3]

[1]Institute of Molecular Biology and Biotechnology, Faculty of Biology, Adam Mickiewicz University, Poznań, Poland

[2]Intramural Research Program, National Library of Medicine, National Institutes of Health, Bethesda, Maryland, USA

[3]Division of Molecular and Cellular Biology, *Eunice Kennedy Shriver* National Institutes of Child Health and Human Development, Bethesda, Maryland, USA

Correspondence
Mikolaj Olejniczak, Institute of Molecular Biology and Biotechnology, Faculty of Biology, Adam Mickiewicz University, Uniwersytetu Poznańskiego 6, 61-614 Poznań, Poland.
Email: mol@amu.edu.pl

Gisela Storz, Division of Molecular and Cellular Biology, *Eunice Kennedy Shriver* National Institutes of Child Health and Human Development, Bethesda, MD 20892-4417, USA.
Email: storzg@mail.nih.gov

Funding information
U.S. National Library of Medicine; *Eunice Kennedy Shriver* National Institute of Child Health and Human Development, Grant/Award Number: ZIA HD001608-30; Narodowe Centrum Nauki, Grant/Award Number: 2018/31/B/NZ1/02612 and 2020/39/O/NZ1/02448

Abstract

In many bacteria, the stabilities and functions of small regulatory RNAs (sRNAs) that act by base pairing with target RNAs most often are dependent on Hfq or ProQ/FinO-domain proteins, two classes of RNA chaperone proteins. However, while all bacteria appear to have sRNAs, many have neither Hfq nor ProQ/FinO-domain proteins raising the question of whether another factor might act as an sRNA chaperone in these organisms. Several recent studies have reported that KH domain proteins, such as KhpA and KhpB, bind sRNAs. Here we describe what is known about the distribution, structures, RNA-binding properties, and physiologic roles of KhpA and KhpB and discuss evidence for and against these proteins serving as sRNAs chaperones.

KEYWORDS
Hfq, KhpA, KhpB, small RNA

1 | INTRODUCTION

RNA-binding proteins with KH domains have been found across all kingdoms, with roles in many different processes. The *K-h*omology or KH domain was initially identified in human heterogeneous nuclear ribonucleoprotein K (hnRNP K) and a *Xenopus laevis* hnRNP K that binds cytidine-rich sequences in pre-mRNAs (Siomi et al., 1993). In KH RNA-binding domains, a conserved GXXG amino acid sequence motif located between two α-helices with accompanying β-strands serves to recognize a specific RNA sequence (Nagai, 1996; Nicastro et al., 2015). There are two types of KH domains: type I in eukaryotes, where the core KH domain is accompanied by additional α-helix and β-sheet motifs on the C-terminal side and type II in bacteria, where the KH domain is accompanied by additional α-helix and β-sheet motifs on the N-terminal side (Nicastro et al., 2015; Valverde et al., 2008). Although eukaryotic proteins often contain multiple KH domains, bacterial proteins typically contain only one or two KH domains (Nicastro et al., 2015).

KH domains have been found in bacterial proteins with a wide range of functions. These include the enzymes polynucleotide

phosphorylase (PNPase) (Dendooven et al., 2021; Hardwick et al., 2012) and RNase Y (Nagata et al., 2008; Shahbabian et al., 2009); transcription elongation factor NusA (Gopal et al., 2001; Worbs et al., 2001); GTPases Era, which binds to the 30S ribosomal subunits (Tu et al., 2009; Verstraeten et al., 2011), and Der, which is involved in the assembly of the 50S subunits (Robinson et al., 2002); cold-shock ribosome factor A (RbfA), which assists in the assembly of the 30S subunits (Huang et al., 2003; Verstraeten et al., 2011); and ribosomal protein S3 (Watson et al., 2020; Wimberly et al., 2000).

Two other KH domain proteins found in multiple bacterial species are KhpA and KhpB (the later also denoted Jag and EloR). KhpA (*KH* domain *p*rotein *A*) initially was identified as a protein involved in cell elongation in *Streptococcus pneumoniae* (Zheng et al., 2017), whereas KhpB was first identified as being encoded adjacent to a sporulation gene in *Bacillus subtilis*, where it was originally named Jag (*spoIIIJ*-*a*ssociated *g*ene) (Errington et al., 1992), but later also was shown to be involved in *S. pneumoniae* cell division (Stamsås et al., 2017; Ulrych et al., 2016; Zheng et al., 2017). Both KhpA and KhpB only have a single KH domain (Figure 1). The KhpA protein is small and comprised solely of the KH domain. In contrast, the KhpB protein has a second RNA-binding domain, an R3H domain (named for the characteristic spacing of an arginine and a histidine residue) (reviewed in Grishin, 1998) on the C-terminus, and, in many cases, a Jag-N domain of unknown function on the N-terminus. KhpA proteins have been found to homodimerize as well as form heterodimers with KhpB proteins (Winther et al., 2019).

Several recent studies have shown that KhpA and KhpB are associated with small regulatory RNAs (sRNAs) (Hör et al., 2020; Lamm-Schmidt et al., 2021; Riediger et al., 2021; Zheng et al., 2017). This is of interest because KhpA and KhpB protein family members are found in several bacteria such as *S. pneumoniae*, which do not have Hfq and ProQ/FinO-domain RNA chaperone proteins. Hfq and ProQ/FinO-domain proteins stabilize and promote the functions of sRNAs that act by base pairing (reviewed in Olejniczak & Storz, 2017; Woodson et al., 2018). These families of proteins are capable of binding both the sRNAs and their base pairing targets consistent with having multiple RNA-binding sites and thus help to promote sRNA interactions with target RNAs. For bacteria-lacking Hfq and ProQ/FinO-domain proteins altogether or where the deletion of these genes has not resulted in an sRNA phenotype, there has been the unanswered question of whether other proteins facilitate sRNA base pairing.

One source of evidence for KhpA and KhpB binding to sRNAs comes from Grad-seq experiments in which cell extracts are fractionated on gradients and the RNA and protein compositions of the individual gradient fractions are determined by RNA sequencing and mass spectrometry, respectively (Hör et al., 2020; Lamm-Schmidt et al., 2021; Riediger et al., 2021). Cofractionation of sRNAs and KhpA and KhpB proteins suggest a possible interaction, which in some species has been further verified by tagging the KhpA and KhpB proteins and identifying the RNAs that specifically copurify with these proteins (Lamm-Schmidt et al., 2021; Zheng et al., 2017).

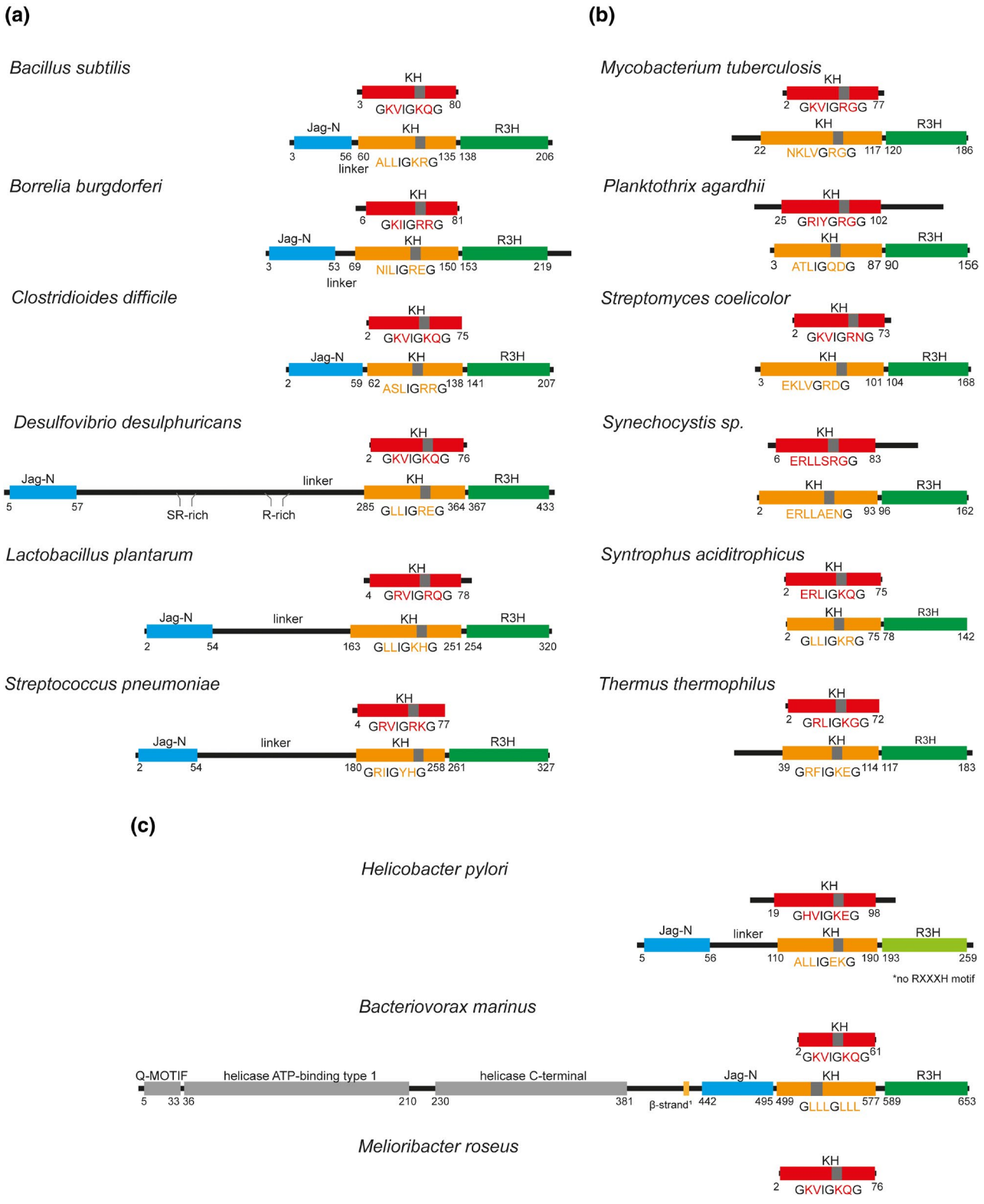
Here we describe what is known about KhpA and KhpB proteins and discuss their possible roles in facilitating the functions of sRNAs.

2 | DISTRIBUTION OF KhpA AND KhpB PROTEINS

We examined the distribution of the KhpA and KhpB proteins based on the presence of the key functional domains of KhpA and KhpB in proteins from the Genome Taxonomy Database (GTDB) predicted by in silico analysis and based on gene synteny (Table S1 and Figures S1–S3, 48% of the 45,555 species we analyzed had *khpA* or *khpB* or both). These analyses revealed that while *khpA* and *khpB* genes are quite prevalent in some phyla, such as Actinobacteria and Firmicutes, they are absent in others such as α-, β-, and γ-Proteobacteria and Bacteroidetes (Prezza et al., 2021). Several other interesting features of these gene families can be noted. First, the genes are only present in a single copy in almost all genomes. Second, although the *khpA* and *khpB* genes are not physically linked on the chromosome, there is an extremely high co-occurrence of the two genes. More than 80% of all species and more than 90% of the Firmicutes and Actinomycetes species, in which we identified *khpA* and/or *khpB*, have both genes (Table S1). There are a limited number of clades where only the *khpA* gene is present. Third, there are two different categories of KhpB proteins, some of which contain both the Jag-N and R3H domains along with the KH domain (Figure 1a) and others which only contain KH and R3H domains (Figure 1b). This observation suggests that the KH and R3H domains form a functional unit that can participate in RNA metabolism, either alone or in connection with the Jag-N domain. Interestingly, very little is known about the function of the Jag-N domain.

Despite similar overall domain composition, KhpA and KhpB from different bacterial species exhibit remarkable sequence diversity. The KH domains of KhpA and KhpB homologs differ in the composition of the variable residues in and around the GxxG motif (Figure 1). In some species, the conserved G residues even are substituted by other amino acids. Additionally, although the folding of the individual KH, R3H, and Jag-N domains of KhpB proteins is similar as judged by ColabFold-based structure predictions, the overall sequence conservation of these domains is low. For example, the sequence identity between the *S. pneumoniae* KhpB protein and those from *B. subtilis* or *Helicobacter pylori* is <30% (Figure 1).

The sequences outside of the conserved domains also vary greatly. Overall, the sequence similarity among KhpA proteins is higher than the similarity among KhpB proteins, but KhpA proteins still vary in the length of the unstructured regions adjacent to the KH domain, resulting in a range in overall size. For instance, *Bacteriovorax marinus* KhpA is 62 aa and *Planktothrix agardhii* KhpA is 149 aa. The KhpB proteins that are composed only of the KH and R3H domains are more similar in size ranging from 142 aa for *Syntrophus aciditrophicus* to 189 aa for *T. thermophilus* (Figure 1b). In contrast, for KhpB proteins that have Jag-N domains, the length of the linker between the Jag-N and the KH domain is remarkably variable resulting in a range of KhpB protein lengths from 208 aa for *B. subtilis* and 328 aa for *S. pneumoniae* to 437 aa for *Desulfovibrio desulphuricans* (Figure 1a). It remains to be seen how all the differences affect the functions of KhpA and KhpB in different bacteria.



There also are a few interesting KhpB variants (Figure 1c). For instance, while the R3H domain of KhpB proteins is defined by an RXXXH motif, this sequence is not present in the corresponding domain of the *H. pylori* protein. As another example, the KhpB homolog from *B. marinus* has a helicase domain connected N-terminally of the Jag-N domain. As will be discussed in conjunction with possible KhpA and KhpB functions, there is significant synteny in the genes surrounding *khpA* and *khpB* (Figure S3). The absence of the genes typically found adjacent to *khpB* suggests that in *B. marinus*, the *khpB* gene was fused with another gene at a new genome location.

FIGURE 1 The diversity of structures of KhpA and KhpB proteins for representative organisms. For Figures 1, S2 and S3, we selected the species that have been the subject of studies on KhpA and KhpB proteins (Errington et al., 1992; Grishin, 1998; Hare et al., 2007; Lamm-Schmidt et al., 2021; Myrbraten et al., 2019; Riediger et al., 2021; Zheng et al., 2017), and additionally included representative species that could illustrate the sequence and structural diversity of KH domain proteins. Schematic structures of KhpA proteins are presented above the schematic structures of KhpB proteins for each species. The KH domains are colored red in KhpA and orange in KhpB proteins, whereas the KhpB Jag and R3H domains are colored blue and green, respectively. The atypical R3H domain in *Helicobacter pylori*, which is devoid of an RxxxH sequence, is colored light green, and the additional helicase domains in the N-terminal part of *Bacteriovorax marinus* KhpB are in gray. The sequence of the GxxIGxxG motif at the junction of the second and third α-helix of the KH domain is provided below each KH domain. The conserved residues in this motif are in black font, while the nonconserved ones are in the color of the KH domain. The position of this sequence in a KH domain is marked with a gray bar. Numbers denote positions of the first and last residue of each domain. To identify domains in protein sequences and assign the domain borders, proteins were folded using ColabFold software (Mirdita et al., 2021) based on AlphaFold 2.0 (Jumper et al., 2021) and MMseqs2 (Steinegger & Soding, 2017). The following bacterial species are represented (the percentages after the species name provide the sequence identity of the KhpA and KhpB proteins, respectively, calculated for homologous regions, relative to the corresponding proteins in *Streptococcus pneumoniae*): (a) KhpA/KhpB sets with KhpB proteins containing the Jag-N domain from *Bacillus subtilis* (38%, 27%), *Borrelia burgdorferi* (33%, 19%), *Clostridioides difficile* (35%, 27%), *Desulfovibrio desulphuricans* (30%, 18%), *Lactobacillus plantarum* (46%, 29%), and *S. pneumoniae*; (b) KhpA/KhpB sets with KhpB proteins without the Jag-N domain from *Mycobacterium tuberculosis* (27%, 26%), *Planktothrix agardhii* (25%, 30%), *Streptomyces coelicolor* (26%, 25%), *Synechocystis* sp. PCC 6803 (23%, 20%), *Syntrophus aciditrophicus* (30%, 24%), and *Thermus thermophilus* (25%, 22%); (c) unusual KhpA/KhpB sets from *H. pylori* (24%, 19%) and *B. marinus* (35%, 24%), and the KhpA protein from *Melioribacter roseus*, which does not have a KhpB homolog (30%). The *khpA* and *khpB* gene synteny for the above species is shown in Figures S2 and S3, whereas the phylogeny of KhpA/KhpB proteins is shown in Figure S1, and the extensive list of KhpA and KhpB homologs is given in Table S1. For *B. marinus*, β-strand[1] denotes a β-strand that interacts with the β-sheet of the KH domain in the ColabFold predicted structure of the KhpB

3 | STRUCTURES OF KhpA AND KhpB PROTEINS

Given that the structures of many KH domain proteins have been determined, a fair amount is known about the basic structure of the minimal motif comprised of the two α-helices, flanked on either side by one β-strand (Siomi et al., 1993; Valverde et al., 2008). As mentioned above, this minimal KH motif is accompanied by additional α and β structures either on the C-terminus (type I domain architecture) or on the N-terminus (type II domain architecture) (Grishin, 2001). The conserved GxxG sequence motif, located between the two α-helices of the minimal KH motif, together with amino acid residues of the neighboring β-strand, are part of a cleft on the surface of the domain that provides an RNA-binding surface (Nicastro et al., 2015; Valverde et al., 2008). Although the two middle residues in the GxxG motif are varied, at least one residue is usually positively charged and arginine, lysine, or glycine residues are frequent. Double aspartate residues between the glycine residues, on the other hand, have been shown to be detrimental for RNA binding (Hollingworth et al., 2012). In type I and II domains, the additional β-strands are differently engaged with the β-sheet of the minimal KH motif, which also leads to different locations of the variable sequence loop (Valverde et al., 2008).

Previously, the structure of *S. pneumoniae* KhpA was predicted using i-Tasser (Winther et al., 2019), and the partial structures of *Clostridium symbiosum* KhpB (pdb: 3GKU) and *H. pylori* KhpB (pdb: 2PT7) (Hare et al., 2007) were solved using X-ray crystallography. We predicted the structures of KhpA and KhpB proteins from both *S. pneumoniae* and *Clostridioides difficile* (Figure 2) using ColabFold software (Mirdita et al., 2021), which is based on AlphaFold 2.0 (Jumper et al., 2021) and MMseqs2 (Steinegger & Soding, 2017). In the ColabFold predictions, the KH domains of KhpA and KhpB can be modeled with high confidence. The comparison of these four predicted KH domain structures shows a fold typical for type II KH proteins with the additional β-strand (β1) on the N-terminus antiparallel to the first β-strand (β2) of the minimal KH domain (Figure 2). Despite the similarity of the overall fold, the visualization of the electrostatic surface potential shows differences in the locations of charged amino acid side chains, which could affect interactions with RNA molecules (Figure 2).

The structures of some KH domain proteins other than KhpA and KhpB have been solved in complex with RNA by X-ray crystallography (Figure 3). In a complex of the *Aquifex aeolicus* Era GTPase with a 12-nt 3′-terminal fragment of 16S rRNA, the bases and riboses of RNA bind the KH domain using hydrogen bonding along with hydrophobic contacts with the peptide backbone and side chains at and around the GKKG sequence, in the subsequent α-helix, next β-strand, and the variable sequence loop (Tu et al., 2009). In a complex of the *Mycobacterium tuberculosis* NusA transcription factor, which has two KH domains connected by a six-amino acid linker, with a 11-nt RNA fragment of the Box C antitermination sequence, the binding clefts of both KH domains form a continuous binding site for the 11-nt RNA (Beuth et al., 2005). As in other KH domains, each cleft is formed by both α-helices surrounding the GxxG motif (GPMG in the KH1 domain and GKEG in the KH2 domain) and the subsequent β-strand, with additional contacts made with the loops between protein secondary structure motifs. In both the Era and the NusA complexes, a stretch of 4–6 nucleotides of RNA fits into the RNA-binding cleft of a single KH domain. In the structure of the trimeric *Caulobacter crescentus* PNPase copurified with an RNA from *E. coli* cells, a 12-nt sequence was well resolved in the crystal structure. Here the RNA formed hydrogen bonding contacts with KH domains of each monomer of the trimeric complex via the loops containing GSGG motifs (Hardwick et al., 2012). Hence, while RNA binding occurred at a single KH domain in the Era GTPase, the RNA-binding site consisted of more than one KH motif in both NusA and PNPase (Beuth et al., 2005; Hardwick et al., 2012; Tu et al., 2009).

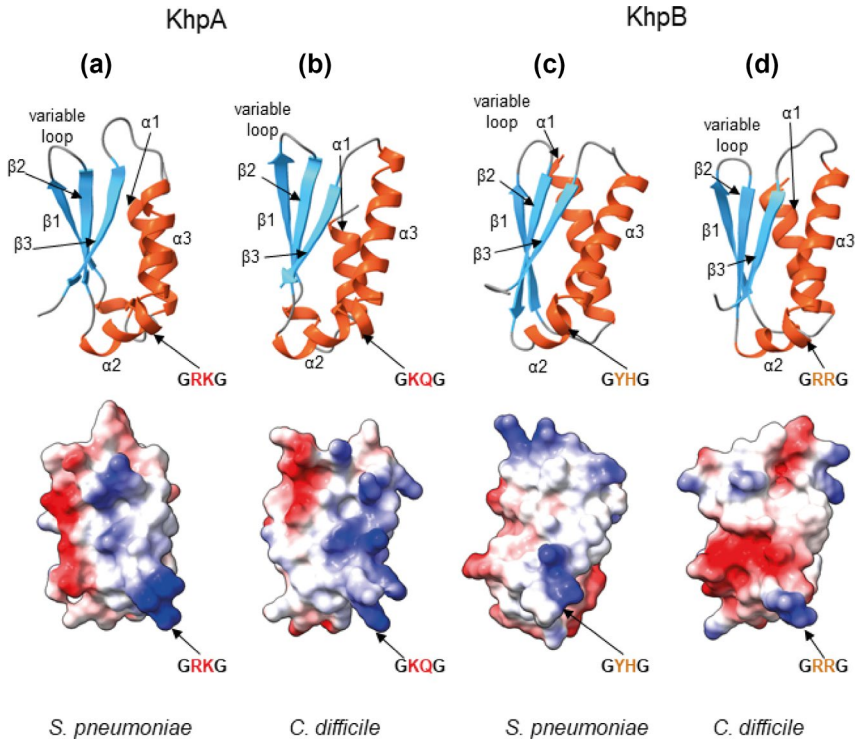


FIGURE 2 Predicted structures of the KH domains of KhpA and KhpB proteins from *Streptococcus pneumoniae* and *Clostridioides difficile* reveal the same overall fold but different electrostatic surface potential. (a) Predicted structure of KhpA from *S. pneumoniae*. (b) Predicted structure of KhpA from *C. difficile*. (c) Predicted structure of KH domain of KhpB protein from *S. pneumoniae*. (d) Predicted structure of KH domain of KhpB protein from *C. difficile*. All structures were predicted using ColabFold software (Mirdita et al., 2021) based on AlphaFold 2.0 (Jumper et al., 2021) and MMseqs2 (Steinegger & Soding, 2017) and visualized using ChimeraX (Pettersen et al., 2021). In each pair, a ribbon representation is shown above with α-helices shown in red, and β-strands in blue, and an electrostatic surface potential, calculated using ChimeraX, is shown below. The view of the KH domains is the face involved in RNA binding

Despite this structural information, the determinants of RNA-binding specificity of KH domains are not well understood (Auweter et al., 2006; Corley et al., 2020; Nicastro et al., 2015). It has been proposed that contacts within the RNA-binding groove and the shape of the groove determine the RNA-binding specificity of individual KH domains (Nicastro et al., 2015). Additionally, the variable sequence loops of KH domains have been proposed to play a role in RNA recognition for KH domains of IMP proteins (Biswas et al., 2019). Furthermore, the glycines in the GXXG motif could be important for RNA binding or just have structural roles as the GXXG loop is located at the bend between two α-helices. Indeed, for Era, one of the glycines contacts RNA, while for NusA, neither of the glycines is directly involved in RNA binding. Nevertheless, while the structures of KhpA and KhpB in complexes with RNA are not yet available, the overlay of amino acid residues contacting RNA in the structures of type II KH domains of *M. tuberculosis* Era and *A. aeolicus* NusA onto the homologous sequences of KH domains of KhpA and KhpB suggests regions that could be involved in RNA binding in KhpA and KhpB (Figure 3).

It should be noted that KhpA and KhpB might bind RNA as homo- or heterodimers, which would result in tandem KH domains (Winther et al., 2019). KhpB also has a second RNA-binding domain, R3H (Ciesla et al., 2020; Grishin, 1998). Thus, these proteins might bind RNA regions that are larger than what would fit into a single KH groove. When the binding of isolated KH domains of the FMRP protein to short RNA ligands was measured, the data showed that the RNA-binding affinities were very weak (Athar & Joseph, 2020). Hence, it is possible that the tight and specific RNA binding by KH domain-containing proteins requires the cooperation of different KH domains and possibly also other RNA-binding domains such as the R3H domain (Dagil et al., 2019; Korn et al., 2021; Schneider et al., 2019).

4 | RNAs BOUND BY KhpA AND KhpB PROTEINS

Although single-stranded CA-rich sequences and G-rich sequences have been proposed most often as RNA recognition motifs of eukaryotic KH proteins (Nicastro et al., 2015), a U-rich sequence was recently reported as a motif recognized by a KH domain of a DEAD-box helicase (Yadav et al., 2021). Thus, KH domains can bind a wide variety of sequence motifs. It is likely that the amino acid sequence and the exact structure of the recognition motif dictate the binding specificity (Dominguez et al., 2018).

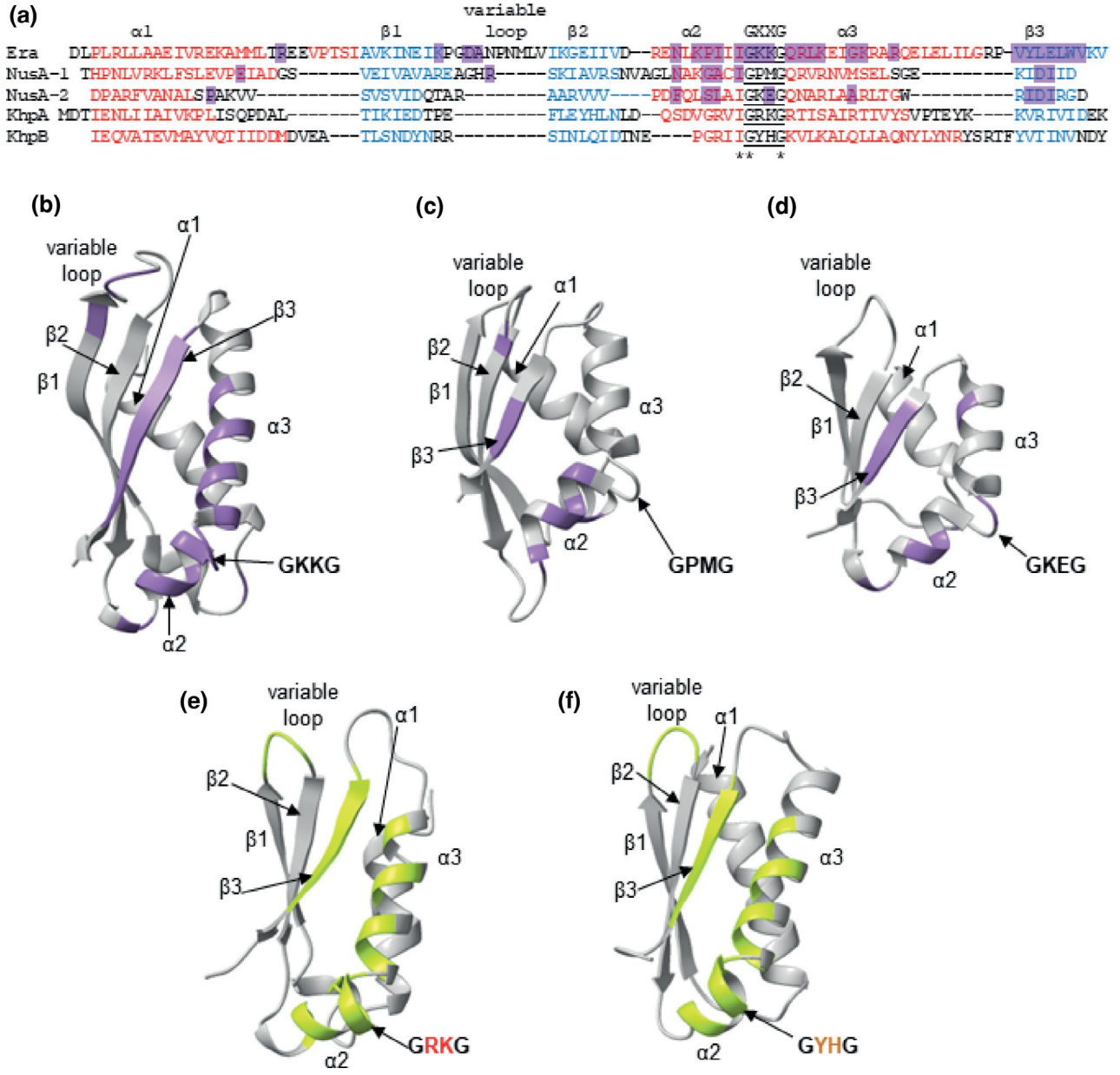


FIGURE 3 Structurally determined and predicted RNA-binding contacts in KH domains. (a) Alignment of the structurally homologous regions of the KH domain from *Aquifex aeolicus* Era protein (Tu et al., 2009), the KH domain 1 from *Mycobacterium tuberculosis* NusA protein, the KH domain 2 from *M. tuberculosis* NusA protein (Beuth et al., 2005), and KH domains from Alphafold-predicted *Streptococcus pneumoniae* KhpA and KhpB. For the alignment, homologous sequences were first aligned using Clustal Omega, and then structurally homologous regions were manually aligned based on the Era and NusA structures (Beuth et al., 2005; Tu et al., 2009) and the predicted structures of KhpA and KhpB. (b) The structure of the KH domain of *A. aeolicus* Era with RNA contacts according to Tu et al. (2009) marked purple. (c) The structure of the KH 1 domain of *M. tuberculosis* NusA with RNA contacts according to Beuth et al. (2005) marked purple. (d) The structure of the KH 2 domain of *M. tuberculosis* NusA with RNA contacts according to Beuth et al. (2005) marked purple. (e) The predicted structure of full-length *S. pneumoniae* KhpA, in which regions that could hypothetically be involved in RNA binding are marked lime green. (f) The predicted structure of the KH domain of *S. pneumoniae* KhpB, in which regions that could hypothetically be involved in RNA binding are marked lime green. The structures of *S. pneumoniae* KhpA and KhpB proteins were predicted using ColabFold (Mirdita et al., 2021), and all structures were visualized using ChimeraX (Pettersen et al., 2021)

Although there are several data sets for RNAs that copurify with KhpA and KhpB proteins, no RNA motif that is recognized by either of the two proteins has been reported. RNA immunoprecipitation (RIP) experiments for *S. pneumoniae* showed that KhpA and KhpB, each tagged with a carboxy terminal 3X FLAG tag, bind the same pool of approximately 170 RNA species, which showed at least a 4-fold enrichment upon immunoprecipitation with either KhpA or KhpB (Zheng et al., 2017). This data set includes mRNAs, two tRNAs, and some sRNAs. Another data set corresponds to RNAs that copurify with KhpB-3xFLAG in *C. difficile*

(Lamm-Schmidt et al., 2021). This study reported enrichment for about 1,400 RNAs. Among these, mRNAs were overrepresented. Although there was no enrichment of rRNAs or tRNAs, 12 sRNAs copurified with KhpB-3xFLAG. It is interesting to note that in *C. difficile*, which encodes an Hfq protein, some RNAs, including eight of the sRNAs, copurify with both Hfq and KhpB while other RNAs are only bound by one or the other chaperone suggesting overlapping as well as distinct cellular roles. Although Hfq predominantly binds to the 5′ and 3′ ends of mRNAs, the full mRNA and even full operons are enriched for KhpB for *C. difficile*. Hopefully, further analyses of the *S. pneumoniae* and *C. difficile* data sets as well as RNAs that copurify with KhpA and KhpB from other bacteria will reveal whether the proteins recognize a specific motif(s) or structure(s), that may or may not differ between bacteria, whether these recognition motifs are typically found at a specific location such as the 5′ end, middle, or 3′ end of a transcript, and whether the proteins have binding sites for more than one RNA.

A related question is what occurs to the RNAs upon binding to a KhpA or KhpB monomer or more likely KhpA homodimer or KhpA-KhpB heterodimer. A comparison of the transcriptome of wild type and a Δ*khpB* deletion strain in *C. difficile* suggests that KhpB might have both positive and negative effects on RNA levels. How this occurs and whether KhpA and KhpB proteins affect the folding of a bound RNA, recruit ribonucleases, or promote base pairing with another RNA remain to be investigated. Spectroscopic studies of NusA binding to a 43-nt RNA indicated that this RNA is unfolded upon binding by NusA, suggesting that KH domain-containing proteins could induce changes in RNA structure (Beuth et al., 2005).

5 | PHYSIOLOGIC ROLES OF KhpA AND KhpB PROTEINS

Although the physiologic roles of only a limited number of KhpA and KhpB proteins have been examined, a few themes are starting to emerge. The role characterized most extensively is one in cell elongation. KhpA and KhpB may have multiple roles in controlling cell elongation, but in *S. pneumoniae*, the absence of these RNA-binding proteins clearly leads to increased levels of transcripts in the WalRK regulon, which responds to peptidoglycan stress (Zheng et al., 2017). KhpB similarly has been shown to negatively affect the levels of mRNAs encoding virulence factors in *C. difficile* (Lamm-Schmidt et al., 2021). The precise mechanisms by which the KH proteins impact the levels of these transcripts are unknown. It is noteworthy that the *khpA* and *khpB* genes show conserved synteny with genes encoding proteins involved in RNA processing, protein synthesis, and protein translocation across membranes.

5.1 | Role in controlling cell elongation

Mutations that inactivate *khpA* and *khpB* were identified in two independent screens for suppressors of the growth defect associated with the lack of the penicillin-binding protein Pbp2b in *S. pneumoniae* (Stamsås et al., 2017; Tsui et al., 2016; Zheng et al., 2017). Pbp2b is an essential enzyme required for peptidoglycan elongation outward from the midcells of dividing *S. pneumoniae* cells (reviewed in Briggs et al., 2021). The association with elongasome proteins led to the alternate name of EloR (*elo*ngasome regulating protein) for KhpB (Stamsås et al., 2017). Consistent with a role in cell elongation, *S. pneumoniae* strains lacking either *khpA* or *khpB* have a reduced length (Stamsås et al., 2017; Ulrych et al., 2016; Zheng et al., 2017) and width (Ulrych et al., 2016; Zheng et al., 2017) compared with wild-type cells and also show slower growth (Stamsås et al., 2017; Ulrych et al., 2016; Zheng et al., 2017). Furthermore, the observation that a Δ*khpA* Δ*khpB* double mutant has the same phenotype as the single mutants indicates that the two proteins act in the same pathway (Zheng et al., 2017).

Several observations, in addition to the co-occurrence of the two genes, indicate that KhpA and KhpB act together (Zheng et al., 2017). The proteins colocalize, diffusing in early divisional cells, and enriched at the midcell in dividing cells (Stamsås et al., 2017; Winther et al., 2019; Zheng et al., 2017). Further evidence for an association between the two proteins has come from copurification and bacterial two-hybrid experiments (Winther et al., 2019; Zheng et al., 2017). The studies of different KhpB truncation mutants revealed that KhpA and KhpB heterodimerize via their KH domains and that this interaction is required for the suppression of the Δ*pbp2b* growth defect (Winther et al., 2019). KhpA also can homodimerize, and the α3 helix in the KH domain is required for both homodimerization and heterodimerization with KhpB (Winther et al., 2019).

Although the interaction studies revealed that the KH domain is required for oligomerization of KhpA and KhpB, the role of RNA binding in the elongasome is less clear. It is intriguing and consistent with a role in cell division that several mRNAs enriched by co-immunoprecipitation of KhpA or KhpB encode cell division proteins, including the cell division protein FtsA (Zheng et al., 2017). While the relative amount of *ftsA-ftsZ* mRNA transcript remains nearly the same, the cellular amount of FtsA protein increases in the *khp* single and double mutant strains. Assays of the effects of different sections of the *ftsA* gene revealed that the 5′-UTR of *ftsAZ* mRNA is required for the KhpA- and KhpB-dependent downregulation of FtsA levels (Zheng et al., 2017), but it is not known how this posttranscriptional regulation is brought about by the KH domain proteins. Suggested possibilities include direct downregulation by the KhpA and KhpB proteins or indirect regulation through an sRNA chaperoned by the KH proteins or another protein modulated by KhpA and KhpB.

It is possible that KhpA and KhpB proteins have multiple roles in cell division, only one of which might depend on RNA binding. Several studies showed that *S. pneumoniae* KhpB is phosphorylated on the linker residue threonine 89 by the StkP kinase that is also part of the elongasome network (Stamsås et al., 2017; Sun et al., 2010; Ulrych et al., 2016). However, phenotypic effects of phosphoablative or phosphomimetic mutations were not observed in some studies (Stamsås et al., 2017; Zheng et al., 2017),

and the threonine 89 residue is not conserved in all KhpB proteins. *S. pneumoniae* KhpB also has been shown to interact with the peptidoglycan muramidase MpgA (Winther et al., 2021), which was previously called MltG (Taguchi et al., 2021; Tsui et al., 2016). Interestingly, mutations in *mpgA*, like *khpA* and *khpB* mutations, suppress the Δ*pbp2b* phenotype (Tsui et al., 2016). The interaction between KhpB and MpgA involves the Jag-N domain of KhpB and is required for localization of the KhpA–KhpB complex to the midcell (Winther et al., 2021).

Although the KhpA and KhpB roles in cell division have been studied most extensively in *S. pneumoniae*, the phenotypes associated with the lack of these proteins in other bacteria also is consistent with roles in cell division. In *Lactobacillus plantarum*, a CRISPR knockdown of either *khpA* or *khpB* (*eloR*) results in cell shortening, with the effect being stronger for KhpA (Myrbraten et al., 2019). As the mechanism of cell elongation differs between cocci (peptidoglycan insertion occurs at midcell) and bacilli (peptidoglycan insertion occurs over the full cell length), these observations suggest a very general role for these proteins in cell wall synthesis. However, the KhpA and KhpB may not have a role in cell division in all organisms, given that in *C. difficile*, the size of Δ*khpA* cells is like that of wild type, though Δ*khpB* mutants had slightly increased cell length and width (Lamm-Schmidt et al., 2021).

5.2 | Role in regulating virulence gene expression

In both *S. pneumoniae* and *C. difficile*, KhpA and KhpB also bind RNAs that do not encode cell division proteins, indicating broader physiologic roles of these proteins, including roles in regulating virulence. Among the transcripts bound by KhpB in *C. difficile* is the *tcdA* mRNA encoding the clostridial toxin A (Lamm-Schmidt et al., 2021). Consistent with a KhpB role in modulating toxin A production, the levels of both the toxin A mRNA and the protein increase in a Δ*khpB* mutant. However, given only a minimal difference in *tdcA* mRNA half-life after rifampicin treatment when comparing a wild type and Δ*khpB* strain, it is not yet clear how KhpB acts as a negative regulator. The detection of *S. pneumoniae khpB* in a Tn-seq screen for reduced fitness in a mouse model of pneumonia further supports a KhpB role in regulating virulence gene expression (van Opijnen & Camilli, 2012).

5.3 | Role in protein synthesis or translocation

Another possible clue to the physiologic roles of the KhpA and KhpB proteins comes from the functions of the genes that are syntenic with *khpA* (Figure S2) and *khpB* (Figure S3). *khpA* is strongly coconserved with *rpsP* encoding the ribosomal protein S16 as well as *rimH* encoding a ribosome maturation factor and *trmD* encoding a tRNA methyltransferase, while *khpB* appears to be in an operon with *yidC* (*spoIIIJ* in *B. subtilis*) encoding a protein translocase and near *rnpA* encoding the RNA component of RNase P, among others. All the proteins encoded by syntenic genes directly or indirectly affect translation or protein translocation. Along the same lines, the *H. pylori* KhpB protein with the noncanonical R3H domain was found to bind and inhibit the HP0525 inner membrane ATPase which has a role in the transport through type IV secretion systems (Hare et al., 2007). Here the KH and R3H domains contact the ATPase, but again it is not clear if RNA binding is involved.

6 | OUTLOOK

The KhpA and KhpB protein families are clearly broadly distributed in several different phyla, suggesting important roles for these RNA-binding proteins. However, even though other KH domains have been studied extensively, relatively little is known about how KhpA and KhpB bind to RNA and the consequences of this binding. Global cosedimentation approaches and copurification approaches have revealed binding to sRNAs, but much remains to be learned about the role of this binding. Do KhpA and KhpB promote sRNA pairing with target mRNAs? How is the role in RNA binding connected to the phenotypes related to peptidoglycan synthesis and cell division? Do KhpA and KhpB proteins have different roles in different species? Do the proteins generally act together or also have separate functions? It will be exciting to see what answers future studies will provide to these and other open questions.

ACKNOWLEDGMENTS

We thank P. Adams, F. Faber, V. Lamm-Schmidt, T. Tsui, and M. Winkler for their comments. We would also like to thank W. Yan from the Jiang Lab for help with figure generation. Research in the M.O. lab is supported by National Science Centre in Poland (grants nos. 2018/31/B/NZ1/02612 and 2020/39/O/NZ1/02448). Research in the X.J. group is supported by the Intramural Research Program of the NIH, National Library of Medicine. Research in the G.S. lab is supported by the Intramural Research Program of the *Eunice Kennedy Shriver* National Institute of Child Health and Human Development.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

DATA AVAILABILITY STATEMENT

Data available in article Supporting Information.

ORCID

Mikolaj Olejniczak https://orcid.org/0000-0001-5949-3944
Gisela Storz https://orcid.org/0000-0001-6698-1241

REFERENCES

Athar, Y.M. & Joseph, S. (2020) RNA-binding specificity of the human fragile X mental retardation protein. *Journal of Molecular Biology*, *432*, 3851–3868.

Auweter, S.D., Oberstrass, F.C. & Allain, F.H. (2006) Sequence-specific binding of single-stranded RNA: is there a code for recognition? *Nucleic Acids Research*, *34*, 4943–4959.

Beuth, B., Pennell, S., Arnvig, K.B., Martin, S.R. & Taylor, I.A. (2005) Structure of a *Mycobacterium tuberculosis* NusA-RNA complex. *EMBO Journal*, *24*, 3576–3587.

Biswas, J., Patel, V.L., Bhaskar, V., Chao, J.A., Singer, R.H. & Eliscovich, C. (2019) The structural basis for RNA selectivity by the IMP family of RNA-binding proteins. *Nature Communications*, *10*, 4440.

Briggs, N.S., Bruce, K.E., Naskar, S., Winkler, M.E. & Roper, D.I. (2021) The pneumococcal divisome: dynamic control of *Streptococcus pneumoniae* cell division. *Frontiers in Microbiology*, *12*, 737396.

Ciesla, M., Turowski, T.W., Nowotny, M., Tollervey, D. & Boguta, M. (2020) The expression of Rpb10, a small subunit common to RNA polymerases, is modulated by the R3H domain-containing Rbs1 protein and the Upf1 helicase. *Nucleic Acids Research*, *48*, 12252–12268.

Corley, M., Burns, M.C. & Yeo, G.W. (2020) How RNA-binding proteins interact with RNA: molecules and mechanisms. *Molecular Cell*, *78*, 9–29.

Dagil, R., Ball, N.J., Ogrodowicz, R.W., Hobor, F., Purkiss, A.G., Kelly, G. et al. (2019) IMP1 KH1 and KH2 domains create a structural platform with unique RNA recognition and re-modelling properties. *Nucleic Acids Research*, *47*, 4334–4348.

Dendooven, T., Sinha, D., Roeselova, A., Cameron, T.A., De Lay, N.R., Luisi, B.F. et al. (2021) A cooperative PNPase-Hfq-RNA carrier complex facilitates bacterial riboregulation. *Molecular Cell*, *81*, 2901–2913.e2905.

Dominguez, D., Freese, P., Alexis, M.S., Su, A., Hochman, M., Palden, T. et al. (2018) Sequence, structure, and context preferences of human RNA binding proteins. *Molecular Cell*, *70*, 854–867.e859.

Errington, J., Appleby, L., Daniel, R.A., Goodfellow, H., Partridge, S.R. & Yudkin, M.D. (1992) Structure and function of the *spoIIIJ* gene of *Bacillus subtilis*: a vegetatively expressed gene that is essential for s^G activity at an intermediate stage of sporulation. *Journal of General Microbiology*, *138*, 2609–2618.

Gopal, B., Haire, L.F., Gamblin, S.J., Dodson, E.J., Lane, A.N., Papavinasasundaram, K.G. et al. (2001) Crystal structure of the transcription elongation/anti-termination factor NusA from *Mycobacterium tuberculosis* at 1.7 Å resolution. *Journal of Molecular Biology*, *314*, 1087–1095.

Grishin, N.V. (1998) The R3H motif: a domain that binds single-stranded nucleic acids. *Trends in Biochemical Sciences*, *23*, 329–330.

Grishin, N.V. (2001) KH domain: one motif, two folds. *Nucleic Acids Research*, *29*, 638–643.

Hardwick, S.W., Gubbey, T., Hug, I., Jenal, U. & Luisi, B.F. (2012) Crystal structure of *Caulobacter crescentus* polynucleotide phosphorylase reveals a mechanism of RNA substrate channelling and RNA degradosome assembly. *Open Biology*, *2*, 120028.

Hare, S., Fischer, W., Williams, R., Terradot, L., Bayliss, R., Haas, R. et al. (2007) Identification, structure and mode of action of a new regulator of the *Helicobacter pylori* HP0525 ATPase. *EMBO Journal*, *26*, 4926–4934.

Hollingworth, D., Candel, A.M., Nicastro, G., Martin, S.R., Briata, P., Gherzi, R. et al. (2012) KH domains with impaired nucleic acid binding as a tool for functional analysis. *Nucleic Acids Research*, *40*, 6873–6886.

Hör, J., Garriss, G., Di Giorgio, S., Hack, L.M., Vanselow, J.T., Förstner, K.U. et al. (2020) Grad-seq in a Gram-positive bacterium reveals exonucleolytic sRNA activation in competence control. *EMBO Journal*, *39*, e103852.

Huang, Y.J., Swapna, G.V., Rajan, P.K., Ke, H., Xia, B., Shukla, K. et al. (2003) Solution NMR structure of ribosome-binding factor A (RbfA), a cold-shock adaptation protein from *Escherichia coli*. *Journal of Molecular Biology*, *327*, 521–536.

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.

Korn, S.M., Ulshofer, C.J., Schneider, T. & Schlundt, A. (2021) Structures and target RNA preferences of the RNA-binding protein family of IGF2BPs: an overview. *Structure*, *29*, 787–803.

Lamm-Schmidt, V., Fuchs, M., Sulzer, J., Gerovac, M., Hör, J., Dersch, P. et al. (2021) Grad-seq identifies KhpB as a global RNA-binding protein in *Clostridioides difficile* that regulates toxin production. *microLife*, *2*, uqab004.

Mirdita, M., Ovchinnikov, S. & Steinegger, M. (2021) ColabFold - Making protein folding accessible to all. *bioRxiv*. Available from: https://doi.org/10.1101/2021.1108.1115.456425

Myrbraten, I.S., Wiull, K., Salehian, Z., Havarstein, L.S., Straume, D., Mathiesen, G. et al. (2019) CRISPR interference for rapid knockdown of essential cell cycle genes in *Lactobacillus plantarum*. *mSphere 4*, e00007-19.

Nagai, K. (1996) RNA-protein complexes. *Current Opinion in Structural Biology*, *6*, 53–61.

Nagata, M., Kaito, C. & Sekimizu, K. (2008) Phosphodiesterase activity of CvfA is required for virulence in *Staphylococcus aureus*. *Journal of Biological Chemistry*, *283*, 2176–2184.

Nicastro, G., Taylor, I.A. & Ramos, A. (2015) KH-RNA interactions: back in the groove. *Current Opinion in Structural Biology*, *30*, 63–70.

Olejniczak, M. & Storz, G. (2017) ProQ/FinO-domain proteins: another ubiquitous family of RNA matchmakers? *Molecular Microbiology*, *104*, 905–915.

Pettersen, E.F., Goddard, T.D., Huang, C.C., Meng, E.C., Couch, G.S., Croll, T.I. et al. (2021) UCSF ChimeraX: structure visualization for researchers, educators, and developers. *Protein Science*, *30*, 70–82.

Prezza, G., Ryan, D., Mädler, G., Reichardt, S., Barquist, L. & Westermann, A.J. (2022) Comparative genomics provides structural and functional insights into Bacteroides RNA biology. *Molecular Microbiology*, *117*, 67–85. https://doi.org/10.1111/mmi.14793

Riediger, M., Spät, P., Bilger, R., Voigt, K., Maček, B. & Hess, W.R. (2021) Analysis of a photosynthetic cyanobacterium rich in internal membrane systems via gradient profiling by sequencing (Grad-seq). *The Plant Cell*, *33*, 248–269.

Robinson, V.L., Hwang, J., Fox, E., Inouye, M. & Stock, A.M. (2002) Domain arrangement of Der, a switch protein containing two GTPase domains. *Structure*, *10*, 1649–1658.

Schneider, T., Hung, L.H., Aziz, M., Wilmen, A., Thaum, S., Wagner, J. et al. (2019) Combinatorial recognition of clustered RNA elements by the multidomain RNA-binding protein IMP3. *Nature Communications*, *10*, 2266.

Shahbabian, K., Jamalli, A., Zig, L. & Putzer, H. (2009) RNase Y, a novel endoribonuclease, initiates riboswitch turnover in *Bacillus subtilis*. *EMBO Journal*, *28*, 3523–3533.

Siomi, H., Matunis, M.J., Michael, W.M. & Dreyfuss, G. (1993) The pre-mRNA binding K protein contains a novel evolutionarily conserved motif. *Nucleic Acids Research*, *21*, 1193–1198.

Stamsås, G.A., Straume, D., Ruud Winther, A., Kjos, M., Frantzen, C.A. & Havarstein, L.S. (2017) Identification of EloR (Spr1851) as a regulator of cell elongation in *Streptococcus pneumoniae*. *Molecular Microbiology*, *105*, 954–967.

Steinegger, M. & Soding, J. (2017) MMseqs2 enables sensitive protein sequence searching for the analysis of massive data sets. *Nature Biotechnology*, *35*, 1026–1028.

Sun, X., Ge, F., Xiao, C.L., Yin, X.F., Ge, R., Zhang, L.H. et al. (2010) Phosphoproteomic analysis reveals the multiple roles of phosphorylation in pathogenic bacterium *Streptococcus pneumoniae*. *Journal of Proteome Research*, *9*, 275–282.

Taguchi, A., Page, J., Tsui, H.T., Winkler, M.E. & Walker, S. (2021) Biochemical reconstitution defines new functions for membrane-bound glycosidases in assembly of the bacterial cell wall. *Proceedings*

of the National Academy of Sciences of the United States of America, *118*, e2103740118.

Tsui, H.C., Zheng, J.J., Magallon, A.N., Ryan, J.D., Yunck, R., Rued, B.E. et al. (2016) Suppression of a deletion mutation in the gene encoding essential PBP2b reveals a new lytic transglycosylase involved in peripheral peptidoglycan synthesis in *Streptococcus pneumoniae* D39. *Molecular Microbiology*, *100*, 1039–1065.

Tu, C., Zhou, X., Tropea, J.E., Austin, B.P., Waugh, D.S., Court, D.L. et al. (2009) Structure of ERA in complex with the 3′ end of 16S rRNA: implications for ribosome biogenesis. *Proceedings of the National Academy of Sciences*, *106*, 14843–14848.

Ulrych, A., Holeckova, N., Goldova, J., Doubravova, L., Benada, O., Kofronova, O. et al. (2016) Characterization of pneumococcal Ser/Thr protein phosphatase *phpP* mutant and identification of a novel PhpP substrate, putative RNA binding protein Jag. *BMC Microbiology*, *16*, 247.

Valverde, R., Edwards, L. & Regan, L. (2008) Structure and function of KH domains. *FEBS Journal*, *275*, 2712–2726.

van Opijnen, T. & Camilli, A. (2012) A fine scale phenotype-genotype virulence map of a bacterial pathogen. *Genome Research*, *22*, 2541–2551.

Verstraeten, N., Fauvart, M., Versees, W. & Michiels, J. (2011) The universally conserved prokaryotic GTPases. *Microbiology and Molecular Biology Reviews*, *75*, 507–542.

Watson, Z.L., Ward, F.R., Meheust, R., Ad, O., Schepartz, A., Banfield, J.F. et al. (2020) Structure of the bacterial ribosome at 2 Å resolution. *eLife*, *9*, e60482.

Wimberly, B.T., Brodersen, D.E., Clemons, W.M. Jr, Morgan-Warren, R.J., Carter, A.P., Vonrhein, C. et al. (2000) Structure of the 30S ribosomal subunit. *Nature*, *407*, 327–339.

Winther, A.R., Kjos, M., Herigstad, M.L., Havarstein, L.S. & Straume, D. (2021) EloR interacts with the lytic transglycosylase MltG at midcell in *Streptococcus pneumoniae* R6. *Journal of Bacteriology*, *203*, e00691-20.

Winther, A.R., Kjos, M., Stamsas, G.A., Havarstein, L.S. & Straume, D. (2019) Prevention of EloR/KhpA heterodimerization by introduction of site-specific amino acid substitutions renders the essential elongasome protein PBP2b redundant in *Streptococcus pneumoniae*. *Scientific Reports*, *9*, 3681.

Woodson, S.A., Panja, S. & Santiago-Frangos, A. (2018) Proteins that chaperone RNA regulation. *Microbiology Spectrum*, *6*, RWR-0026-2018.

Worbs, M., Bourenkov, G.P., Bartunik, H.D., Huber, R. & Wahl, M.C. (2001) An extended RNA binding surface through arrayed S1 and KH domains in transcription factor NusA. *Molecular Cell*, *7*, 1177–1189. https://doi.org/10.1016/S1097-2765(01)00262-3

Yadav, M., Singh, R.S., Hogan, D., Vidhyasagar, V., Yang, S., Chung, I.Y.W. et al. (2021) The KH domain facilitates the substrate specificity and unwinding processivity of DDX43 helicase. *Journal of Biological Chemistry*, *296*, 100085.

Zheng, J.J., Perez, A.J., Tsui, H.T., Massidda, O. & Winkler, M.E. (2017) Absence of the KhpA and KhpB (JAG/EloR) RNA-binding proteins suppresses the requirement for PBP2b by overproduction of FtsA in *Streptococcus pneumoniae* D39. *Molecular Microbiology*, *106*, 793–814.

SUPPORTING INFORMATION

Additional supporting information may be found in the online version of the article at the publisher's website.

How to cite this article: Olejniczak, M., Jiang, X., Basczok, M.M. & Storz, G. (2022) KH domain proteins: Another family of bacterial RNA matchmakers? *Molecular Microbiology*, 117, 10–19. https://doi.org/10.1111/mmi.14842