**Caulobacter crescentus** Hfq structure reveals a conserved mechanism of RNA annealing regulation

Andrew Santiago-Frango* a,b,1, Kathrin S. Fröhlich c,f, Jeliazko R. Jeliazkova d, Ewelina M. Malecka b, Giada Marino*, Jeffrey J. Gray d,f, Ben F. Luis i,j, Sarah A. Woodson a,b,3 and Steven W. Hardwick b,3

aCell, Molecular and Developmental Biology and Biophysics Program, Johns Hopkins University, Baltimore, MD 21218; bDepartment of Biophysics, Johns Hopkins University, Baltimore, MD 21218; cDepartment of Biology I, Microbiology, Ludwig Maximilian University Munich, 82152 Martinsried, Germany; dProgram in Molecular Biophysics, Johns Hopkins University, Baltimore, MD 21218; eDepartment of Biology I, Plant Molecular Biology, Ludwig Maximilian University Munich, 82152 Martinsried, Germany; fDepartment of Chemical and Biomolecular Engineering, Johns Hopkins University, Baltimore, MD 21218; and gDepartment of Biochemistry, University of Cambridge, CB2 1GA Cambridge, United Kingdom

Edited by Gisel Storz, National Institute of Child Health and Human Development, Bethesda, MD, and approved April 10, 2019 (received for review August 21, 2018)

We have solved the X-ray crystal structure of the RNA chaperone protein Hfq from the alpha-proteobacterium *Caulobacter crescentus* to 2.15Å resolution, resolving the conserved core of the protein and the entire C-terminal domain (CTD). The structure reveals that the CTD of neighboring hexamers pack in crystal contacts, and that the acidic residues at the C-terminal tip of the protein interact with positive residues on the rim of Hfq, as has been recently proposed for a mechanism of modulating RNA binding. De novo computational models predict a similar docking of the acidic tip residues against the core of Hfq. We also show that *C. crescentus* Hfq has sRNA binding and RNA annealing activities and is capable of facilitating the annealing of certain *Escherichia coli* sRNA:mRNA pairs in vivo. Finally, we describe how the Hfq CTD and its acidic tip residues provide a mechanism to modulate annealing activity and substrate specificity in various bacteria.

**Significance**

In many bacteria, the RNA chaperone protein Hfq binds to hundreds of small noncoding RNAs and improves their efficacy by aiding base pairing to target mRNAs. Hfq proteins contain a variable C-terminal domain (CTD), usually structurally disordered, which was recently demonstrated to inhibit Hfq from mediating nonspecific RNA annealing. We obtained a new structure that shows how this inhibition is achieved in *Caulobacter crescentus* Hfq. The structural data and chaperone assays provide an initial view of the little-known mechanism of small RNA regulation in *Caulobacter*. In addition, this work demonstrates how the Hfq CTD has evolved to meet the needs for species-specific selectivity in RNA binding and pairing of regulatory RNAs with cognate targets.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

This open access article is distributed under Creative Commons Attribution License 4.0 (CC BY).

Data deposition: The atomic coordinates and structure factors have been deposited in the Protein Data Bank, www.wwpdb.org (PDB ID code 6GWK).

1A.S.-F. and K.S.F. contributed equally to this work.

*Author contributions: A.S.-F., K.S.F., J.R.J., E.M.M., G.M., and S.W.H. performed research; S.A.W. analyzed data; and A.S.-F., B.F.L., S.A.W., and S.W.H. wrote the paper.*

The authors declare no conflict of interest.

This article is a PNAS Direct Submission. This open access article is distributed under Creative Commons Attribution License 4.0 (CC BY).

Data deposition: The atomic coordinates and structure factors have been deposited in the Protein Data Bank, www.wwpdb.org (PDB ID code 6GWK).

1A.S.-F. and K.S.F. contributed equally to this work.

*Author contributions: A.S.-F., K.S.F., J.R.J., E.M.M., G.M., and S.W.H. performed research; S.A.W. analyzed data; and A.S.-F., B.F.L., S.A.W., and S.W.H. wrote the paper.*

This open access article is distributed under Creative Commons Attribution License 4.0 (CC BY).

Data deposition: The atomic coordinates and structure factors have been deposited in the Protein Data Bank, www.wwpdb.org (PDB ID code 6GWK).

1A.S.-F. and K.S.F. contributed equally to this work.
and the structure was solved using the CTD of Hfq in Ec. The CTD of Hfq in the CTD, the roles of these Hfq appendages are less well understood. The CTD of Hfq in Ec has recently been proposed to autoinhibit RNA annealing activity, as acidic residues at the very C-terminal tip of the protein mimic the nucleic acid phosphodiester backbone to displace RNA from the rim (26, 27). Like other disordered protein regions, the CTD of Hfq diverges at a faster rate than the structured Sm-like core, via nonconservative substitutions and indels (28-30). The Cc Hfq CTD is less than one-half the length of the Ec Hfq CTD (15 aa vs. 38 aa) yet retains an acidic motif at the CTD tip (-DADD), similar to that seen in Ec Hfq (-DSEETE) (Fig. 1).

We have solved the X-ray crystal structure of Cc Hfq to 2.15-Å resolution, including the structure of the entire length of the Hfq CTD. The structure reveals how the acidic tip residues pack against the positive core residues, in agreement with predictions (27), and shows the CTD packing against the CTD of neighboring hexamers in an antiparallel arrangement. We have used this crystal structure to validate Rosetta FloppyTail (31), a de novo modeling algorithm for disordered regions of proteins. Finally, we evaluate the RNA binding and annealing activity of Cc Hfq and show structurally and biochemically that the CTD of Hfq provides a mechanism to regulate annealing activity and substrate specificity across bacterial species.

Results
Crystal Structure of Cc Hfq. Purified full-length Cc Hfq was crystallized, and the structure was solved using the Ec Hfq core hexamer as a molecular replacement search model (22). Four Hfq hexamers could be modeled in the crystal asymmetric unit, packing in an unusual arrangement with the hexamers positioned orthogonal to one another to generate an “open square” (SI Appendix, Fig. S1A). From the electron density maps, residues in the short CTD of Cc Hfq could be visualized for several protomers within the asymmetric unit, in some cases allowing for modeling of the entire length of the CTD with a high degree of confidence (Fig. 1B). On modeling the Cc Hfq CTD, it became apparent that these extensions pack against the CTD of neighboring hexamers in an antiparallel manner (SI Appendix, Fig. S1B). The Cc Hfq CTDs that are not directly involved in contacting neighboring hexamers are partially disordered.

From the structure of Cc Hfq, it is clear that the acidic tip residues (DADD; residues 78-82) pack against the positive rim residues of the proximal face (R18, K19, and K21), with few direct contacts to the arginine patch on the distal side of the rim (R49 and R50) (Fig. 1B). Specifically, aspartates 81 and 82 are positioned to make electrostatic interactions with lysine 19 and arginine 18, respectively. Interestingly, the side chain of aspartate 79 is facing away from the positive rim residues on the neighboring protomer, but instead is within hydrogen-bonding distance of arginine 49 on a neighboring hexamer (SI Appendix, Fig. S1B). It should also be noted that lysine 21 from the positive rim patch does not directly interact with the acidic tip residues in the crystal structure, but instead makes a hydrogen-bonding interaction with arginine 49 within the same protomer (SI Appendix, Fig. S1C). Finally, glutamate 75 from within the CTD forms a hydrogen bond with arginine 49 and serine 39 on the neighboring protomer within the Hfq hexamer (SI Appendix, Fig. S1C). These interactions agree with the proposed autoinhibitory...
mechanism in which acidic tip residues pack against conserved positive rim residues on the core of Hfq to compete with nucleic acid binding to the rim and disfavor spurious RNA annealing events (27).

The well-ordered CTD–rim interactions observed in our structure suggest that the short Cc Hfq CTD may provide even stronger autoinhibition than the longer and more flexible Ec Hfq CTD.

De Novo Modeling of the Cc Hfq Tail. To better understand whether the CTD–core interactions captured in our crystals could occur in solution, we modeled the flexible NTEs and CTDs of Cc Hfq using Rosetta FloppyTail, a de novo computational modeling algorithm (31) that has been previously used to predict interaction energies in the Ec Hfq CTD with good correlations to in vitro mutagenesis assays (27). To ensure that the results were not biased, the calculations were performed without prior knowledge of the Cc Hfq crystal structure.

The overall basicity of the rim is conserved between Cc and Ec Hfqs (Fig. L4); however, the arginines are distributed more toward the rim-distal face in Cc Hfq (SI Appendix, Fig. S2). Among the lowest-energy fraction of models generated in our simulations (1% of all models, sorted by energy), acidic residues at the CTD of Cc Hfq were found to frequently form energetically favorable contacts with basic residues (arginine 18, lysine 19, and lysine 21 on the proximal-rim interface (Fig. 2A)), even when the basic NTEs were excluded from the model. In contrast, few contacts were observed to arginines 49 and 50, which are solvent-exposed but lie toward the distal side of the rim. Nearly all of the lowest-energy fraction of models had at least one CTD in contact with the rim, and many of the modeled CTD conformations closely resembled (∼2 Å rmsd) the crystallized CTD–rim interaction (Fig. 2B and SI Appendix, Figs. S3 and S4). In silico mutation of both arginine 18 and lysine 19 to alanine ablated interactions of acidic CTD residues with these positions on the core (SI Appendix, Fig. S5). These results illustrate the predictive power of Rosetta FloppyTail and suggest that the crystallized CTD–rim interaction is likely to occur in solution rather than being an artifact of crystal lattice packing.

Ec Hfq binds sRNAs more avidly than Cc Hfq. The sequence-specific RNA binding sites of the Sm-like core of Ec Hfq bear strong sequence conservation to Ec Hfq (Fig. L4) and are expected to bind U-rich RNA at the proximal face and A-rich RNA at the distal face. Docking of poly(A) RNA bound to Ec Hfq (6) onto the structure of Cc Hfq shows that the RNA can be accommodated on the distal face of the protein. Using fluorescence anisotropy, we observed that the Cc Hfq protein can indeed bind to A18-FAM RNA with high affinity (Kd ≤ 2 nM Hfq; SI Appendix, Fig. S6), comparable to that of Ec Hfq.

The distantly related E. coli and C. crescentus genomes possess very different GC content, of 51% (32) and 67% (33) respectively. Therefore, the C. crescentus genome is expected to encode sRNAs with more stable secondary structures, which may impact Hfq–sRNA interactions. We first compared the binding of Cc Hfq and Ec Hfq proteins to three sRNAs from C. crescentus. The genome of C. crescentus encodes four homologs (CCNA_R0014, CCNA_R0133, CCNA_R0143, and CCNA_R0157) of the alphaproteobacterial or15 family of noncoding RNAs, which also includes the Hfq-binding sRNAs AbcR1/2 in Agrobacterium tumefaciens and Sinorhizobium meliloti, respectively (34, 35). Likewise, CCNA_R0014 has recently been recovered in coimmunoprecipitation experiments with 3x FLAG-tagged Cc Hfq (36), and the paralog sRNAs CCNA_R0157 (R157) and CCNA_R0133 (R133) were expected to bind Hfq with high affinity. These GC-rich (60–72%) C. crescentus sRNAs are predicted to form highly stable stem-loops (Fig. 3A) and possess sequence and structural characteristics of E. coli class 1 sRNAs that predominantly engage the rim and proximal face via a U/A-rich region upstream of the terminator stem-loop and a single-stranded U-rich 3′ end (14, 37–40). A third sRNA, CCNA_R0100 (ChvR), was not coprecipitated with 3x FLAG-tagged Cc Hfq (36), and so was included as a negative control. ChvR is predicted to form a stable secondary structure but lacks U/A- and U-rich sequence motifs.

Native gel electrophoretic mobility shift assays (EMSAs) showed that Cc Hfq formed stable complexes with R157 and R133 sRNAs (Fig. 3B and SI Appendix, Table S1). However, Cc Hfq had very weak affinity for ChvR, as expected (Fig. 3B). In contrast, Ec Hfq bound all three C. crescentus sRNAs more tightly than Cc Hfq and exhibited little preference for R157 and R133 compared with the “nonspecific” sRNA ChvR (Fig. 3B and SI Appendix, Table S1). Thus, Cc Hfq binds C. crescentus sRNAs less avidly than Ec Hfq, and with greater discrimination. We noted that Ec Hfq also has a greater propensity than Cc Hfq to form higher-order complexes with sRNAs, which are thought to contain a second Hfq hexamer (Fig. 3B). This observation is addressed in Discussion.

To further explore how Ec Hfq and Cc Hfq proteins recognize their endogenous sRNAs, we examined the binding of Ec and Cc Hfq variants to the three C. crescentus sRNAs above and to four E. coli sRNAs: RydC, DsrA, RybB, and Chix (SI Appendix, Fig. S7). Ec Hfq had a greater affinity than Cc Hfq for all seven sRNAs tested (Fig. 4 and SI Appendix, Fig. S8 and Table S1). However, each sRNA interacted slightly differently with the Ec and Cc Hfq...
proteins. The *E. coli* sRNAs RydC, DsrA, and RybB are class I sRNAs that form sequence-specific contacts with the proximal face and rim of Hfq (39). Cc Hfq formed particularly weak complexes with RydC sRNA (*Kd* = 206 nM; red open symbols in Fig. 4C), an sRNA that must partially unfold to interact with the Hfq rim (41). The poor binding of Cc Hfq to RydC may be explained structurally by a clash of the RNA with the acidic CTD tip (SI Appendix, Fig. S9). In contrast, Cc Hfq and Ec Hfq bound ChiX sRNA with similar affinity. ChiX is a class II sRNA that contains an ARN motif and makes sequence-specific contacts with both the proximal and distal faces of Hfq (38–40), and thus is likely to be less susceptible to competition for the rim by the CTDs.

**Cc Hfq CTD Increases the Selectivity of sRNA Binding.** We next asked how the CTD contributes to these differences by examining sRNA binding of Ec Hfq core (Ec65; residues 1–65 only), Cc Hfq lacking the acidic tip residues (Cc78; residues 1–78 only), and a chimera of Ec Hfq core and *C. crescentus* Hfq CTD (EcCc Hfq; residues 1–67 of Ec Hfq fused to residues 71–82 of Cc Hfq) (Fig. 4A). We found that truncation of either the entire CTD (Ec65) or the acidic CTD tip (Cc78) decreased the apparent affinities and cooperativity of Hfq binding to all the sRNAs with single-stranded Hfq binding sites (R157, DsrA, RybB, and ChiX). Conversely, removal of the CTD increased the binding of Hfq to RydC and to the non–Hfq-specific ChvR sRNA (blue symbols in Fig. 4B and SI Appendix, Table S1).

When the autoinhibitory Cc Hfq CTD was appended to the Ec Hfq core (compare EcCc Hfq to Ec65 and Ec Hfqs), the average affinity for sRNAs was reduced and comparable to that of *C. crescentus* proteins (compare EcCc Hfq to Cc Hfq). The Cc Hfq CTD also restored the apparent cooperativity of Hfq binding, as shown by the transition slopes in Fig. 4B and C. Thus, the Cc Hfq CTD appears to inhibit sRNA binding more strongly than the Ec Hfq CTD. This increased autoinhibition by the shorter Cc Hfq CTD agrees with our previous observation that a shortened CTD increases autoinhibition due to a higher local concentration of the acidic CTD tip around the rim of Hfq (27).

It was initially surprising that removal of the C-terminal residues in Cc78 made binding of Hfq-specific R157 and R133 sRNAs less favorable, while making binding of non–Hfq-specific ChvR more favorable. Further inspection of the EMSA results showed that Cc78 Hfq complexes were more heterogeneous than Cc Hfq complexes (SI Appendix, Fig. S8). For example, the band for the ChvR-Cc78 Hfq complex was diffuse rather than sharp, and RydC and DsrA formed a triplet of complexes with differing affinities.

**Fig. 4.** Equilibrium interactions of Hfq and sRNAs. (A) Hfq variants used in the assays. Ec Hfqs (solid lines; filled circles) or Cc Hfqs (dotted lines; open circles) bearing no acidic CTD tip (light blue), a CTD with a distantly tethered acidic tip residues (orange), or a CTD with a closely tethered acidic tip residues (red). All Hfq variants used in for RNA binding and RNA annealing experiments lack exogenous affinity tags. (B and C) Binding curves for Ec Hfq (solid lines) or Cc Hfq (dotted lines) variants with *C. crescentus* and *E. coli* sRNAs, respectively. Symbols for single representative trials are shown. The fraction bound was fit to single binding isotherms, normalized for the maximum fraction bound at saturating Hfq. *Kd* values and Hill coefficients for each fit are listed in SI Appendix, Table S1.
mobilities, perhaps reflecting alternative RNA or protein conformations. In addition, we observed diffuse $^{32}$P below the main Hfq-sRNA band that presumably arose from dissociation of Hfq during electrophoresis. The quantity of $^{32}$P-labeled RNA from these dissociated complexes was measurably higher for binding reactions with Ec65 and Cc78 lacking the acidic CTD than for binding reactions with full-length Ec, EcCc, and Cc Hfq proteins (SI Appendix, Fig. S10). As discussed below, these observations suggest that the Hfq CTD favors specific modes of sRNA binding at the expense of nonspecific sRNA-Hfq interactions, improving the kinetic stability of the remaining Hfq-sRNA complexes and increasing the selectivity for sRNAs with Hfq recognition motifs, the kinetic stability of the remaining Hfq-sRNA complexes and that are represented in the EMSA data.

**CTD of Cc Hfq Is a Potent Autoinhibitor of RNA Annealing.** The CTD of Ec Hfq limits its annealing of minimal RNAs, which we propose is due to competition between the CTD and RNA binding to the basic patches on the rim of the protein (26). According to our model (27), the shorter CTD of Cc Hfq should inhibit annealing more strongly than the longer Ec Hfq CTD. To evaluate the contributions of the CTD to RNA annealing, we carried out RNA annealing assays with the same Cc and Ec Hfq variants used for the sRNA binding experiments (Fig. 5 and SI Appendix, Fig. S13). The ability of Hfq to accelerate base pairing between an RNA molecular beacon and three different complementary RNAs (Fig. 5A) was measured by stopped-flow fluorescence spectroscopy (42). These minimal RNAs, which lack secondary structure but retain Hfq binding sequences, have been used to compare the basal annealing activities of other bacterial Hfqs (16). The shortest (16 nt) target RNA binds the rim of Hfq weakly and nonspecifically (43). Target-U6 RNA binds tightly and specifically to the proximal face of Hfq, whereas Target-A18 RNA specifically binds to the distal face of Hfq (Fig. 5B).

**Cc Hfq was far less active than Ec Hfq on all the minimal RNAs tested (Fig. 5C). Whereas Ec Hfq accelerated annealing by approximately 10-fold, Cc Hfq barely increased the annealing rates of Target-U6 and Target-A18 when Cc Hfq hexamer and RNA beacon were present in equimolar amounts (Fig. 4B; orange and red vs. gray). The low activity of Cc Hfq was not due to poor RNA binding, as Cc Hfq can recognize single-stranded U6 and A18 tails (Fig. 3 and SI Appendix, Fig. S6).**

We next considered whether the annealing activity of Cc Hfq was inhibited by its CTD. We visualized the effect of the CTD on the chaperone activity of Hfq by comparing the annealing rates of the CTD variants with their respective wild-type parents at three different Hfq concentrations (Fig. 5D and SI Appendix, Fig. S13). Removal of the acidic CTD residues from either protein (Ec65 or Cc78) increased the annealing activity by 2- to 10-fold on these minimal RNAs (Fig. 5D; solid blue and dotted blue lines). Cc78 generally exhibited a greater relief of autoinhibition than Ec65. Although the Ec65 core was more active than Cc78 for annealing Target and Target-U6, the chimeric EcCc Hfq was less active than the wild-type Ec Hfq and produced less dsRNA product (red, Fig. 5D). An exception to this trend is that Ec65 forms inactive complexes when premixed with a twofold molar excess of an A-rich target (SI Appendix, Fig. S11). Thus, Cc Hfq is capable of annealing minimal RNAs, but this activity is inhibited by its C-terminal acidic residues. In addition, this comparison suggests that the Cc Hfq CTD inhibits nonspecific RNA annealing more strongly than Ec Hfq CTD (27).

**Cc Hfq Can Facilitate sRNA-mRNA Regulation in Vivo.** As there are currently no sRNA-mRNA regulatory pathways in C. crescentus known to rely on Hfq, we examined whether Cc Hfq could be transferred into E. coli to facilitate interactions between sRNA-mRNA pairs in vivo. As examples, we investigated down-regulation of ompF mRNA by RybB sRNA, down-regulation of hns mRNA by DsrA sRNA, and activation of cfa mRNA by the sRNA RydC. Using λRed recombination (44), we truncated

---

Fig. 5. In vitro RNA annealing activity of Cc Hfq. (A) Molecular beacon assay for annealing of 16 nt synthetic RNA (Target) that binds Hfq nonspecifically (1 μM). (B) Target RNA was extended with U6 or A18 (Target-U6 or Target-A18) that specifically bind the proximal or distal face of Hfq, respectively. (C) Stopped-flow fluorescence progress curves for annealing reactions of Target (Left), Target-A18 (Center), or Target-U6 (Right) in 50 nM Hfq hexamer. Ec-Hfq (solid lines) and Cc Hfq (dotted lines) variants are indicated in the key; the basal, no Hfq reaction is indicated by gray shading. Progress curves were normalized to the maximum fluorescence change after a maximum of 500 s. Raw data were down-sampled to enable plotting C. crescentus datasets as dotted lines. An exception is that Ec65 sequesters Target-A18 in an inactive complex in this assay (SI Appendix, Fig. S11), as explained in SI Appendix. (D) Effect of the CTD on annealing. Observed annealing rate constants ($k_{obs}$) in the presence of EcCc, Ec65, and Cc78 relative to $k_{obs}$ for the parental WT protein (Ec or Cc gray shading). Kinetics were measured in 16.7, 50, or 150 nM Hfq hexamer for three targets as in B. Rate constants are the average of at least five replicates; propagated error bars are smaller than the symbols plotted.
the coding sequence of Ec hfq (Ec65) or replaced it completely with the corresponding C. crescentus sequence (referred to as Cc Hfq) on the chromosome of E. coli MC4100 under control of the native Ec hfq promoter.

We then compared sRNA-mediated regulation of the cognate mRNA targets in the presence of Ec, Ec65, or Cc Hfq to an isogenic hfq deletion mutant (Fig. 6B). To this end, we transformed strains expressing the different Hfq variants with translational fusions of sRNA target transcripts to gfp (45). In these reporters (i.e., ompF::gfp, hns::gfp, and cfa::gfp), the 5′ UTR and the first codons of the mRNA targets are fused to the second codon of gfp, and expression is driven from the constitutive P\textsubscript{BAD} promoter. We grew all strains to midexponential phase (OD\textsubscript{600} of 1), collected total RNA samples before and at 15 min after the addition of arabinose, and monitored sRNA induction and target::gfp mRNA expression by Northern blot analysis.

Three Hfq-dependent sRNAs—RybB, MicA, and MicL—function as the repressory arms of the RpoE-controlled response to perturbations of outer membrane homeostasis (46–48). Collectively, these sRNAs repress >30 mRNAs in E. coli by direct base-pairing, including many transcripts encoding outer membrane porins (OMPs). We monitored down-regulation of the well-characterized target ompF::gfp in response to RybB induction, and observed that both Ec and Cc Hfq mediated repression of the reporter (Fig. 6C and E). In contrast, ompF::gfp mRNA abundance barely changed in cells expressing truncated Ec65 Hfq or the hfq null mutant, even though RybB was induced to comparable levels in all strains.

We next analyzed the regulatory capacity of DsrA in our different Hfq variant strains. Induced under low-temperature growth conditions, DsrA represses translation of hns mRNA in an Hfq-dependent manner (49–51). Our reporter assay confirmed the requirement of Hfq, as we observed down-regulation of hns::gfp mRNA in the presence of Ec Hfq but not in the hfq null strain (Fig. 6D and F). Both truncated Ec65 Hfq and Cc Hfq were able to mediate DsrA activity, although hns::gfp repression was slightly reduced compared with Ec Hfq (twofold and threefold repression vs. fourfold repression).

---

**Fig. 6.** In vivo RNA annealing assays. (A) Schematic of E. coli mutant strains used in RNA annealing assays. All Hfq variants are expressed from the MC4100 hfq locus under control of the native Ec hfq promoter. Expression of the Hfq variants was confirmed by Northern blot analysis and LC-MS/MS (SI Appendix, Fig. S14) (B) Reporter assay to monitor posttranscriptional activity of sRNAs. E. coli strains were cotransformed with a translational fusion of sRNA target transcripts to gfp (target::gfp; under control of the constitutive P\textsubscript{BAD} promoter) and either an empty control vector or a plasmid expressing an sRNA from the arabinose-inducible P\textsubscript{BAD} promoter. RNA collected before and at 10 min postinduction with L-arabinose (final concentration 0.01%) was analyzed on Northern blots, and target::gfp mRNA levels were quantified. (C and D) Northern blot analysis of ompF::gfp mRNA (C) and hns::gfp mRNA (D) in strains carrying a control plasmid or in response to pulse induction of RybB (C) or DsrA (D) sRNAs. Expression of sRNAs was validated using gene-specific probes; 5S rRNA served as a loading control. (E and F) Quantification of ompF::gfp (E) or hns::gfp mRNA (F) repression at the 10-min time point) from Northern blot analysis as shown in C and D. Error bars represent the deviation of three biological replicates.
We also examined the ability of Cc Hfq to complement annealing of RydC and its target cfa mRNA in vivo. RydC activates expression of the cfa transcript in an Hfq-dependent manner via formation of a base-pairing interaction of its conserved 5' end to an upstream site within the cfa 5' UTR to interfere with mRNA decay (52). Our in vitro binding assays suggested that Cc Hfq binds to RydC very poorly compared with Ec Hfq (Fig. 4C), and in accordance with these results, RydC was not able to activate cfa::gfp expression in our reporter assay in the absence of Ec Hfq, even though its expression was only mildly affected in the mutant strains (Fig. 7A). We also observed that activation of cfa::gfp expression by RydC was less effective in the presence of the truncated EcΔ65, suggesting an important contribution of the Hfq CTD in mediating this regulation. Taken together, these results indicate that Cc Hfq is able to facilitate the annealing of certain sRNA/mRNA pairs in E. coli, such as RybB/ompF mRNA and DsrA/ins mRNA, which have an accessible Hfq-binding site, but not RydC, which must partly unfold to interact with Hfq. Our crystal structure and in vitro binding results suggest that this difference arises from increased interactions between the Cc Hfq CTD and basic residues on the rim of the Hfq hexamer.

Discussion

The crystal structure of Cc Hfq has revealed detailed structural information for the divergent CTDs of bacterial Hfq proteins. Previous sequence analysis, computational modeling, and biochemical assays have suggested that many bacterial Hfq CTDs contain a flexible “linker” of variable length followed by a highly acidic motif that mimics nucleic acid (27). Cc Hfq contains a CTD with a naturally short linker but still contains a nucleic acid mimic motif at the C-terminal end (Fig. 1A). In the crystal, the majority of the linker region within the CTD is involved in a lattice packing interaction with the linker region of neighboring Hfq hexamers (SI Appendix, Fig. S1B). Although this may be a favored interaction within the context of the crystal, the self-association of the CTD from Ec Hfq has previously been reported in solution (53). Further evidence for the conformation of the CTD seen in the Cc Hfq crystal structure being more than a crystal packing artifact came from our de novo blind computational modeling, which showed good agreement in the position of the CTD in relation to the core of Hfq. In this example, the FloppyTail Rosetta modeling algorithm was tested against a crystal structure in a blind test.

In this study, we also assessed the RNA binding and annealing activity of Cc Hfq, and these experiments in themselves significantly advance our understanding of the biological role of Hfq across the bacterial kingdom. Our comparative binding results with Cc Hfq and Ec Hfq toward endogenous and nonendogenous RNA substrates show that Cc Hfq recognizes sRNAs comparably to Ec Hfq, but with a slightly lower affinity for all RNAs tested. (Binding affinities are summarized in SI Appendix, Table S1.) One might assume that Hfq proteins would have evolved to bind with higher affinity to their endogenous substrates, but this does not seem to be the case, as Ec Hfq binds with higher affinity than Cc Hfq to the specific C. crescentus sRNAs tested here. This difference in sRNA affinity is due in part to differences in the Hfq core, because EcΔ65 Hfq binds all sRNAs tested more tightly than Cc78 Hfq except for ChiX, which also interacts with the distal face (compare EcΔ65 and Cc78 in Fig. 4). Our results also suggest that the CTD is important for ensuring the specificity of C. crescentus sRNA-Hfq complexes, as Cc78 lacking the C-terminal acidic residues was much more likely to form short-lived complexes that dissociate during gel electrophoresis and complexes with different mobilities. By displacing weakly bound sRNAs, the CTDs not only make binding more selective, but also may reduce the number of allowed binding configurations so that the remaining complexes are conformationally homogeneous.

Although the seven sRNAs tested in our study represent only a small fraction of RNAs that interact with Hfq in either species, the results indicate that differences in the Hfq cores and their CTDs both influence the recognition of individual sRNAs. It is also apparent that each Hfq variant used in this study binds RNA ligands somewhat differently, and the determinants for sRNA binding are likely to be complex and remain to be fully understood. Ec Hfq is known to form complexes in which two hexamers bind one RNA, which can tether the two hexamers (41, 54, 55). Interestingly, Cc Hfq forms only a single complex on the native gels with all RNAs tested. We envisage that the longer CTD of Ec Hfq allows for a “sandwich” complex of Hfq-RNA-Hfq, whereas the limited length of the Cc Hfq CTD would prohibit formation of such a complex. It has been previously suggested that these supershifted species are less active than 1:1 Hfq:RNA complexes (56, 57); therefore, Hfq sequences that are less prone to forming these large complexes may retain a higher active fraction of proteins at high Hfq concentrations.

As with previous experiments to shorten the linker of Ec Hfq (27), we show that the naturally short linker of Cc Hfq imposes a greater degree of autoinhibition of RNA binding.

Fig. 7. Activation of cfa::gfp by RydC depends on Ec Hfq. (A) Northern blot analysis of cfa::gfp mRNA levels in strains carrying a control plasmid or in response to pulse induction of RydC. Expression of RydC was validated using a gene-specific probe; 5S rRNA served as a loading control. (B) Quantification of cfa::gfp mRNA induction (at the 10-min time point) from Northern blot analysis as shown in A. Error bars represent the deviation of three biological replicates.
It is important to also note that the trade-off between specificity and complementation between Hfq proteins from different species correlates with our in vivo annealing data. In the example of RybB sRNA regulation of ompF mRNA, as well as DsrA sRNA regulation of hns mRNA, Cc Hfq is able to complement the in vivo annealing effect of the endogenous Ec Hfq protein (Fig. 6 C and D). However, Cc Hfq is unable to support annealing of RydC and cfa mRNA (Fig. 7A). This apparent specificity in annealing activity likely reflects poor binding of Cc Hfq to RydC (Fig. 4). In line with these observations, previous work has suggested that the C-terminal tails of Salmonella Hfq are important for facilitating formation of the RydC/cfa:Hfq ternary complex (41).

Our structural, functional, and computational results show how the short CTD of Cc Hfq provides specificity for RNA binding and annealing. The role of the Hfq CTD is depicted schematically in Fig. 8. We have demonstrated that although Cc Hfq is able to bind RNA in vitro with an affinity close to that of Ec Hfq, the RNA annealing activity of the C. crescentus protein is greatly reduced. Through the construction of truncated and chimeric proteins, we have shown that the increased inhibition of RNA annealing can be attributed to the acidic residues at the tip of the shorter CTD in Cc Hfq. There is an abundance of potential nucleic acid substrates for Hfq in the cell (58). As RNA chaperones with broad substrate specificities, Hfq proteins across the bacterial kingdom seem to be under similar evolutionary pressures to balance electrostatic interactions to RNAs at the basic rims, necessary for RNA annealing and restructuring.

and annealing compared with the CTD of Ec Hfq, presumably by increasing the local concentration of the acidic motif around the arginine-rich active sites of the Hfq core. The increased specificity imparted by the Cc Hfq CTD comes at the cost of decreased RNA annealing activity, at least toward minimal unstructured RNAs (Fig. 5B and SI Appendix, Fig. S13). To what degree natural C. crescentus sRNA and mRNA targets overcome the stronger autoinhibition imposed by the Cc Hfq CTD remains to be learned.

Our successful construction of a chimeric Hfq protein composed of the core of Ec Hfq and the CTD of Cc Hfq showed that the CTD performs the same function of competing for non-specific RNA interactions in the context of Hfq from the α-proteobacterium C. crescentus or the γ-proteobacterium E. coli. This modular property of the Cc Hfq CTD is due in part to the nature of the CTD–rim interaction, which is rather non-specific and dominated by electrostatics, reducing the stringency of the CTD–rim interaction in the face of sequence divergence. We believe this observation is of particular importance as it suggests that Hfq CTDs may rapidly diverge to accommodate the acquisition of new sRNAs that could change the organization of regulatory networks.

Table 1. X-ray data collection and refinement statistics

| Parameters/statistics | Cc Hfq – 6GWK |
|------------------------|----------------|
| Wavelength             | 70.35–2.15 (2.227–2.15) |
| Space group            | P4₁            |
| Unit cell              | 97.32, 97.32, 203.64–90, 90, 90 |
| Total reflections      | 168,042 (16,595) |
| Unique reflections     | 82,963 (10,091) |
| Multiplicity           | 2.0 (1.9)      |
| Completeness, %        | 97.42 (99.06)  |
| Mean I/σI, I           | 9.44 (0.46)    |
| Wilson B-factor        | 36.75          |
| Rmerge                 | 0.1228 (0.6281) |
| Rfree                  | 0.1548 (0.823) |
| Rwork                  | 0.0932 (0.5247) |
| CC1/2                  | 0.99 (0.618)   |
| CC*                    | 0.998 (0.874)  |
| Reflections used in refinement | 99,756 (10,091) |
| Reflection used for Rfree | 4,978 (526) |
| Rwork                  | 0.2858         |
| Rfree                  | 0.3096         |
| No. of nonhydrogen atoms | 12,914         |
| macromolecules         | 12,804         |
| Solvent                | 110            |
| Protein residues       | 1,651          |
| rmsd, bonds            | 0.003          |
| rmsd, angles           | 0.54           |
| Ramachandran favored, %| 95.18          |
| Ramachandran allowed, %| 4.82           |
| Ramachandran outliers, %| 0.00           |
| Rotamer outliers, %    | 0.28           |
| Clashscore             | 7.61           |
| Average B-factor       | 54.02          |
| Macromolecules         | 54.18          |
| Solvent                | 35.90          |

Statistics for the highest-resolution shell are in parentheses. CC*, crystallographic refinement statistic.
Materials and Methods

Crystalization and X-Ray Data Collection. Cc Hfq purified from the N-terminal GST tagd construct was concentrated to 7 mg/mL, and crystals were grown via hanging-drop vapor diffusion by adding an equal volume of crystallization buffer (0.2 M sodium iodoacetate, 17% PEG 3350) to the protein sample. Crystals were harvested using 25% glycerol as a cryoprotectant and then flash-frozen in liquid nitrogen. X-ray data were collected at Diamond Light Source, beamline I03.

Crystal Structure Determination. The structure was solved by molecular replacement with PHASER (61) using Ec Hfq core [Protein Data Bank (PDB) ID code 4PNO] as a search model. Initial space group assignment and molecular replacement efforts were hampered by the presence of translational non-crystallographic symmetry. PHASER indicated the presence of two nonorinig peaks of 71% and 41% of the origin peak. Ultimately a translational correction was applied, and four copies of the Hfq hexamer were correctly placed in the asymmetric unit of the X-ray data processed in the P4 space group. Following refinement with REFMAC (62) electron density corresponding to the CTD was apparent for several protomers, and the amino acids for this region were modeled manually using Coot (63). Data collection and refinement statistics are summarized in Table 1. The coordinates and structure factors have been deposited in the PDB (ID code 6GWX).

Nucleic Acid Preparation. The sequences of RNA substrates are listed in SI Appendix, Table S2. Synthetic target RNAs, the molecular beacon (43), and FAM-labeled A18 (A18-FAM) have been described previously (42). The RNAs R157, R133, Chr, ChiK, Drsα, Rybβ, RyDC, cfa, and hns were transcribed in vitro as described previously (56).

RNA Binding. The affinities of all Hfq variants for ~1 nM 32P-labeled sRNAs at 30 °C in reaction buffer (34 mM Tris-HCl pH 7.5, 50 mM NaCl, 50 mM KCl, 50 mM NaH2, 11.4 mM EDTA, 12% glycerol, 0.005% bromophenol blue, 0.005% xylene cyanol FF) for 30 min were measured by native gel EMSAs in 1× TBE (89 mM Tris, 89 mM boric acid, 2 mM EDTA, pH 7.6) as described previously (56), except for Ec Hfq complexes with C. crescentus sRNAs, which were better resolved in 1× Tris-glycine running buffer (25 mM Tris, 192 mM glycine, pH 8.3). The fraction of 32P-labeled RNA bound to one or more Hfq hexamers, after background subtraction, was fit to a two-state binding isotherm, \[ \text{F}_{\text{obs}} = \left( \frac{[\text{Hfq}]^2}{K_{d} + [\text{Hfq}]^2} \right) (1 + \left( [\text{Hfq}]^2/K_{d} \right)) \], in which \( n \) is the Hill coefficient or steepness of the binding transition. 32P-sRNA traveling between the free sRNA and sRNA-Hfq complex was quantified (SI Appendix, Fig. S8) and treated as a separate binding transition if this fraction exceeded 10%. In this case, the fraction of detached complexes was fit using \( \text{F}_{\text{obs}} = \left( \frac{[\text{Hfq}]^2}{K_{d} + [\text{Hfq}]^2} \right) + \left( [\text{Hfq}]^2 + K_{d} \right) \), in which \( K_{d} \) is the binding constant for the dissociated complexes and \( K_{d} \) is the binding constant for the tight complexes. Binding constants for A18-FAM (5 nM) were measured in 1× TBE (89 mM Tris, 89 mM boric acid, 2 mM EDTA, pH 7.6) with the Ambion MAXiscript kit (Thermo Fisher Scientific) according to the manufacturer’s recommendations.

Acknowledgments. We thank the staff at Diamond Light Source for their help with collection of X-ray diffraction data and thank Kasia Bandry, Tom Dendooven, and Kai Papenfort for their critical reading of the manuscript and advice. S.W.H. and B.F.L. are funded by the Wellcome Trust (200873/Z/18/Z). K.S.F. receives funding from the mentoring program of the Ludwig Maximilian University Munich Faculty of Biology. This work was also supported by the National Institutes of Health (Grants R01 GM120425, to S.A.W.; F31 GM123616, to J.R.J.; and R01 GM078221, to J.J.G.).

1. Kavita K, de Mets F, Gottesman S (2018) New aspects of RNA-based regulation by Hfq and its partner sRNAs. Curr Opin Microbiol 42:53–61.
2. Updegrove TB, Zhang A, Storz G (2016) Hfq: The flexible RNA matchmaker. Curr Opin Microbiol 42:133–138.
3. Imanov L et al. (2017) Crossover between the tricarboxylic acid cycle and peptidoglycan biosynthesis in Caulobacter crescentus through the homeostatic control of the ketoligas.
4. da Silva CA, Veillette CE, Jian KG, Banerjee R, Ribeiro RR, Ramo MF (2016) Transcrip. Crystallization and characterization of the functional state response regulator sprf in Caulobacter crescentus. BMC Microbiol 16:66.
5. Santiago-Frangos A, Woodson SA (2018) Hfq chaperone brings speed dating to bacteria. RNA 24:1427–1435.
6. Link TM, Valentín-Hansen P, Brennen RG (2009) Structure of Escherichia coli Hfq bound to polynucleotide RNA. Proc Natl Acad Sci USA 106:19292–19297.
7. Schumacher MA, Pearson RF, Moller T, Valentín-Hansen P, Brennen RG (2002) Structure of the pleiotropic translational regulator hfq and an hfq RNA complex: A bacterial Sm-like protein. Nucleic Acids Res 30:133–138.
8. Kocher KE, Canty JT, Orans J, Brennen RG (2014) Recognition of U-rich RNA by Hfq from the Gram-positive pathogen Listeria monocytogenes. RNA 20:1548–1559.
9. Weichenrieder O (2014) RNA binding by Hfq and ring-forming (LSm proteins: A trade-off between optimal sequence readout and RNA backbone confirmation. RNA 20:1337–1349.
10. Nikulin A et al. (2017) Characterization of RNA-binding properties of the archaeal Hfq-like protein from Methanothermus jannaschii. J Biol Chem 342:16515–16528.
11. Horstmann N, Orans J, Valentín-Hansen P, Shellburne SA, 3rd, Brennen RG (2012) Structural mechanism of Staphylococcus aureus Hfq binding to an RNA A-tract. Nucleic Acids Res 40:11023–11035.
12. Someya T, et al. (2012) Crystal structure of Hfq from Bacillus subtilis in complex with SELEX-derived RNA aptamer: Insight into RNA-binding properties of bacterial Hfq. Nucleic Acids Res 40:1856–1867.

13. Nielsen JS, et al. (2007) An Hfq-like protein in archaea: Crystal structure and functional characterization of the 5m protein from Methanococcus jannaschii. RNA 13:341–353.

14. Sauer E, Weichenrieder O (2011) Structural basis for RNA 3′-end recognition by Hfq. Proc Natl Acad Sci USA 108:13065–13070.

15. Panja S, Schu DJ, Woodson SA (2013) Conserved arginines on the rim of Hfq catalyze base pair formation and exchange. Nucleic Acids Res 41:7536–7546.

16. Zheng A, Panja S, Woodson SA (2016) Arginine patch predicts the RNA annealing activity of Hfq from Gram-negative and Gram-positive bacteria. J Mol Biol 428:2259–2264.

17. Els MJ, Trusler RS, Hanford DB (2015) Hfq binds directly to the ribosome-binding site of 5S transposase mRNA to inhibit translation. Mol Microbiol 96:633–650.

18. Sonnleitner E, Bläsi U (2014) Regulation of Hfq by the RNA CrcZ in Escherichia coli. Nucleic Acids Res 42:4069–4081.

19. Fröhlich KS, Förstner KU, Gitai Z (2018) Post-transcriptional gene regulation by an RNA chaperone Hfq: Integrative experimental and computational insights into the role of the Hfq C-terminal region. Acta Crystallogr Sect F Struct Biol Commun 74:536–540.

20. Wen B, et al. (2017) Structural and dynamic properties of the C-terminal domain of bacterial sRNAs consists of a double or single hairpin preceded by a U-rich poly(U) tail. Proc Natl Acad Sci USA 114:2075–2080.

21. Sun X, Zhulin I, Wartell RM (2002) Predicted structure and phyletic distribution of the RNA chaperone Hfq: Integrative experimental and computational analysis of full-length Hfq from Escherichia coli. Acta Crystallogr Sect F Struct Biol Commun 58:915–920.

22. Beich-Frandsen M, Ve M, Madsen KS, Tjaden BC, Storz G, Gottesman S (2013) Mutations in interaction surfaces differentially impair Hfq association with small RNAs and their miRNA targets. J Mol Biol 425:3678–3697.

23. Ishikawa H, Otaka H, Maki K, Morita T, Aiba H (2012) The functional Hfq-binding domain of the RNA chaperone Hfq drives sRNA competition and release of target regulatory RNAs determines their performance in competition for the chaperone protein Hfq. Nucleic Acids Res 40:10982–10998.

24. Sonnleitner E, Bläsi U (2014) Regulation of Hfq by the RNA CrcZ in Escherichia coli. Nucleic Acids Res 42:4069–4081.

25. Vincent HA, et al. (2012) Characterization of Primarily due to indels in intrinsically disordered regions. Nucleic Acids Res 38:e27049.

26. Sauer E, Schmidt S, Weichenrieder O (2012) Small RNA binding to the lateral surface of Hfq: New insights. Nucleic Acids Res 40:e100440.

27. Panja S, Woodson SA (2011) Hfq proximity and orientation controls RNA annealing. Nucleic Acids Res 39:5193–5202.

28. Sauer E, Weichenrieder O (2011) Structural basis for RNA 3′-end recognition by Hfq. Proc Natl Acad Sci USA 108:13065–13070.

29. Light S, Sagit R, Sachenkova O, Ekman D, Elofsson A (2013) Protein expansion is mediated autoregulate the RNA chaperone Hfq. Nucleic Acids Res 41:7536–7546.

30. Sun X, Zhulin I, Wartell RM (2002) Predicted structure and phyletic distribution of the RNA-binding protein Hfq. Nucleic Acids Res 30:3662–3671.

31. Schulz EC, Barabas O (2014) Structure of an Escherichia coli Hfq:RNA complex at 0.97 Å resolution. Acta Crystallogr Sect F Struct Biol Commun 70:1492–1507.

32. Reich-Brandes M, Veleret B, Björblom B, Bläsi U, Djimouji-Caruge K (2011) Structural analysis of full-length Hfq from Escherichia coli. Acta Crystallogr Sect F Struct Biol Commun 67:536–540.

33. Vincent HA, et al. (2012) Characterization of Primarily due to indels in intrinsically disordered regions. Nucleic Acids Res 38:e27049.

34. Möller P, et al. (2014) Profound impact of Hfq on nutrient acquisition, metabolism and global stress regulon. Proc Natl Acad Sci USA 111:6098–6103.

35. Möller P, et al. (2014) Profound impact of Hfq on nutrient acquisition, metabolism and global stress regulon. Proc Natl Acad Sci USA 111:6098–6103.

36. Fröhlich KS, Förstner KU, Gitai Z (2011) Structural analysis of full-length Hfq from Escherichia coli. Acta Crystallogr Sect F Struct Biol Commun 67:536–540.

37. Ishikawa H, Otaka H, Maki K, Morita T, Aiba H (2012) The functional Hfq-binding domain of the RNA chaperone Hfq drives sRNA competition and release of target regulatory RNAs determines their performance in competition for the chaperone protein Hfq. Nucleic Acids Res 40:10982–10998.

38. Møller P, et al. (2014) Profound impact of Hfq on nutrient acquisition, metabolism and global stress regulon. Proc Natl Acad Sci USA 111:6098–6103.

39. Ishikawa H, Otaka H, Maki K, Morita T, Aiba H (2012) The functional Hfq-binding module of bacterial sRNAs consists of a double or single hairpin preceded by a U-rich poly(U) tail. RNA 18:1062–1074.

40. Malecka EM, Stróżeczka J, Sobasanka D, Olejnicki M (2015) Structure of bacterial regulatory RNAs determines their performance in competition for the chaperone protein Hfq. Biochemistry 54:1157–1170.