bioRxiv preprint doi: https://doi.org/10.1101/150672; this version posted June 16, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Producing Hfq/Sm Proteins and sRNAs for Structural and Biophysical Studies of Ribonucleoprotein Assembly

Kimberly A. Stanek* and Cameron Mura*

Author affiliations & correspondence:

Department of Chemistry; University of Virginia; Charlottesville, VA 22904 USA

*Correspondence can be addressed to KS (kas4wk@virginia.edu) or CM (cmura@virginia.edu).

Document information:

Last modified:	15 June 2017
Running title:	*Producing Hfq•RNA Complexes for Structural Studies*
Abbreviations:	3D, three-dimensional; AU, asymmetric unit; CV, column volume; DEPC, diethyl pyrocarbonate; HDV, hepatitis δ virus; IMAC, immobilized metal affinity chromatography; MW, molecular weight; MWCO, molecular weight cut-off; nt, nucleotide; PDB, Protein Data Bank; RNP, ribonucleoprotein; RT, room temperature
Additional notes:	The main text is accompanied by 2 figures and 1 table.
Journal format:	*Methods in Molecular Biology* (*Springer Protocols* series); this volume is entitled "*Bacterial Regulatory RNA: Methods and Protocols*"; an author guide is linked at http://www.springer.com/series/7651

bioRxiv preprint doi: https://doi.org/10.1101/150672; this version posted June 16, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Abstract

Hfq is a bacterial RNA-binding protein that plays key roles in the post–transcriptional regulation of gene expression. Like other Sm proteins, Hfq assembles into toroidal discs that bind RNAs with varying affinities and degrees of sequence specificity. By simultaneously binding to a regulatory small RNA (sRNA) and an mRNA target, Hfq hexamers facilitate productive RNA···RNA interactions; the generic nature of this chaperone-like functionality makes Hfq a hub in many sRNA-based regulatory networks. That Hfq is crucial in diverse cellular pathways—including stress response, quorum sensing and biofilm formation—has motivated genetic and 'RNAomic' studies of its function and physiology (*in vivo*), as well as biochemical and structural analyses of Hfq···RNA interactions (*in vitro*). Indeed, crystallographic and biophysical studies first established Hfq as a member of the phylogenetically-conserved Sm superfamily. Crystallography and other biophysical methodologies enable the RNA-binding properties of Hfq to be elucidated in atomic detail, but such approaches have stringent sample requirements, *viz*.: reconstituting and characterizing an Hfq•RNA complex requires ample quantities of well-behaved (sufficient purity, homogeneity) specimens of Hfq and RNA (sRNA, mRNA fragments, short oligoribonucleotides, or even single nucleotides). The production of such materials is covered in this Chapter, with a particular focus on recombinant Hfq proteins for crystallization experiments.

Keywords: Hfq; Sm; sRNA; RNA chaperone; RNA-binding protein; crystallization; *in vitro* transcription

bioRxiv preprint doi: https://doi.org/10.1101/150672; this version posted June 16, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

1 **Introduction**

The bacterial protein Hfq, initially identified as a host factor required for the replication of bacteriophage Qβ RNA [1], plays a central role in RNA biology: both in RNA-based regulation of gene expression and in modulating RNA stability and lifetime *in vivo* [2]. Hfq functions broadly as a chaperone, facilitating contacts between small non-coding RNAs (sRNAs) and their cognate mRNAs [3]. The RNA interactions may either stimulate or inhibit expression, depending on the identity of the mRNA–sRNA pair and the molecular nature of the interaction (high or low affinity, stable or transient, etc.) [4,5]. In many cases, Hfq is required for these pairings to be effective [6], and knockdown of the *hfq* gene results in pleiotropic phenotypes such as increased UV sensitivity, greater susceptibility to oxidative or osmotic stress, decreased growth rates, etc. [7]. A flood of 'RNAomics'-type studies, over the past decade, has shaped what we know about Hfq-associated RNAs [2,8,9]. Hfq has been linked to many cellular pathways that rely on rapid responses at the level of post-transcriptional/mRNA regulation, including stress responses [10-12], quorum sensing [13], biofilm formation [14], and virulence factor expression [15,12].

Hfq homologs are typically ≈ 80-100 amino acids in length, with the residues folding as an α-helix followed by five β-strands arranged into a highly-bent, antiparallel β-sheet [16,17]. Hfq monomers self-assemble into a toroidal hexamer, the surface of which features at least three distinct regions that can bind RNA. The *proximal* face of the hexamer (proximal with respect to the *N*′-terminal α-helix) is known to bind U-rich sequences [16,18], while the *distal* face of the $(Hfq)_6$ ring binds preferentially to A-rich RNA elements [19,20]. Recently a third, lower-affinity, *lateral* surface on the outer rim of the Hfq ring has been shown to bind RNA [21] and aid in sRNA···mRNA annealing [22]. This lateral site likely has a preference for U-rich segments [23], but also may interact fairly non-specifically with RNA because of an arginine-rich region that is found in some homologs. While the exact mechanism by which Hfq facilitates productive RNA···RNA interactions remains unclear, it is thought that the distal face binds the 3′-poly(A) tails of mRNAs while the proximal face binds to 3′ U-rich regions of sRNAs [3]. The lateral surface may

bioRxiv preprint doi: https://doi.org/10.1101/150672; this version posted June 16, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

act either to cycle different RNAs onto Hfq [22] or as an additional surface for binding to internal, U-rich regions of an sRNA. A recent study suggested that this mechanistic model holds for only a subset of sRNAs, termed 'Class I' sRNAs [24]. A second subset of sRNAs ('Class II') appears to bind both the proximal and distal sites of Hfq; the mRNA targets of these Class II sRNAs are predicted to bind preferentially to the lateral region of the Hfq ring.

A detailed understanding of how different sRNAs interact with Hfq, and with target RNAs in a ternary RNA•Hfq•RNA complex, requires atomic-resolution structural data. While multiple structures have been determined for short (≲10-nucleotide) RNAs bound to either the proximal [16,18], distal [19,20], or lateral [23] sites of Hfq (Table 1), as of this writing only one structure of an Hfq bound to a full-length sRNA has been reported [25]. In that Hfq•RNA complex, comprised of *E. coli* Hfq bound to the *Salmonella* RydC sRNA (Fig 1A), the 3′-end of the sRNA encircles the pore, towards the proximal face of the hexamer, while an internal U–U dinucleotide binds in one of the six lateral pockets on the periphery of the Hfq ring (Fig 1B). Though other regions of the sRNA were found to further contact a neighboring Hfq ring in the lattice (Fig 1C), the stoichiometry of the Hfq•RydC complex *in vivo*, at limiting RNA concentrations, is thought to be 1:1. (Interestingly, two distinct interaction/binding modes were seen between RydC and the lateral rim of an adjacent hexamer [Fig 1C].) The Hfq•RydC complex offers a valuable window into our understanding of Hfq···sRNA interactions, limited mainly by the relatively low resolution (3.48 Å) of the refined structure. For this and other Hfq•RNA complexes, many questions can be addressed by leveraging different types of structural and biophysical approaches. Ideally, the methods used would provide a variety of complementary types of information (i.e., the underlying strategy in taking a 'hybrid methods' approach [26,27]).

Atomic-resolution information may be obtained in several ways. Historically, the premier methodologies have been X-ray crystallography and solution NMR spectroscopy; these well-established approaches are described in many texts, such as [28] and [29]. Though beyond the scope of this chapter, note that much

bioRxiv preprint doi: https://doi.org/10.1101/150672; this version posted June 16, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

progress in recent years has positioned electron cryo-microscopy (cryo-EM) as a powerful methodology for high-resolution (nearly atomic) structural studies of macromolecular assemblies [30,31], including ribonucleoprotein (RNP) complexes such as the ribosome [32-34], telomerase [35] and, most recently, the spliceosome [36,37]. Thus far, all Hfq and Hfq•RNA structures deposited in the Protein Data Bank (PDB), listed in Table 1, have been determined via X-ray crystallography. The molecular weight (MW) of a typical Hfq hexamer is ≈ 60 kDa while sRNAs, which range in length from ≈ 50–500 nucleotides (nt), have MWs of ≈ 16–1600 kDa. An RNP complex of this size is ideally suited to macromolecular crystallography.

In this Chapter, we describe how to prepare and crystallize Hfq•sRNA complexes for structure determination and analysis via the classic, single-crystal X-ray diffraction approach. However, if crystallographic efforts with a particular Hfq or Hfq•sRNA complex prove difficult because of a lack of well-diffracting crystals—or, even when such crystals can be reproducibly obtained—then one can also consider investigating the Hfq-based complex via complementary approaches. Two main families of alternative methodological approaches are available: (i) NMR and other spectroscopic methods (e.g., EPR [38]), and (ii) cryo-EM and other scattering-based approaches (e.g., SAXS [39]). NMR and cryo-EM are routinely used for smaller or larger-sized biomolecular complexes, respectively, though methodological developments are continuously redefining these limitations. The current upper size limit for *de novo* NMR structure determination is ≈ 40 kDa, this limit being reached via the application of techniques such as TROSY, as well as relatively recently developed approaches for deriving distance restraints (e.g., paramagnetic relaxation enhancement). NMR applications to RNP complexes have been recently reviewed [40]. In the reverse direction, from large to small, cryo-EM was recently used to determine the structure of a protein as small as ≈ 170 kDa [41]; the highest-resolution cryo-EM structure reported thus far has reached a near-atomic 2.2 Å [42].

bioRxiv preprint doi: https://doi.org/10.1101/150672; this version posted June 16, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

As alluded to above, RNA and RNP complexes pose particular challenges in crystallographic structure determination [43,44]. Most proteins typically adopt a discrete, well-defined three-dimensional (3D) structure, but populations of RNAs tend to sample broad ranges of conformational states, yielding greater structural heterogeneity; notably, this holds even if the sample is technically monodisperse (i.e., homogeneous in terms of MW). Therefore, RNA and RNP crystals often exhibit significant disorder and correspondingly poorer diffraction, as gauged by resolution, mosaicity, and other quality indicators [45]. Synthesis and purification of an RNA construct through *in vitro* transcription (see §3.2) generates large quantities of chemically homogeneous RNA, and also conveniently lends itself to the engineering of constructs that may be more crystallizable, or that exhibit improved diffraction. Alongside crystallization efforts, chemical probing [REF-Book-Chap?] and structure prediction/modelling methods [46] can be used to examine the secondary structure of the RNA of interest, as well as identify potential protein-binding sites. Then, in designing a more crystallizable construct, extraneous regions of RNA can be either removed or replaced with more stable secondary structures (e.g. stem-loops incorporating tetraloop/tetraloop-receptor pairs); these rigid structural elements can aid crystal contacts and enhance lattice order [44,47]. As an example of judiciously choosing (and/or designing) an RNA system for crystallographic work, the aforementioned RydC sRNA (Fig 1A) is a favorable candidate for crystallization efforts because (i) it is relatively small and compact (forming a pseudoknot), and (ii) it features multiple U-rich regions that can potentially bind to both its cognate Hfq (within a single RNP complex) as well as other Hfq proteins across the lattice. In the crystal structure (Fig 1B, C), the RNA was found to span two Hfq hexamers, forming intermolecular contacts that helped stitch together a stable crystal lattice.

Once crystals that diffract to even low resolution (e.g., $\lesssim$ 4 Å) are obtained, a native (underivatized) X-ray diffraction dataset can be collected. From this dataset alone, much can be learned [48], including the likely stoichiometry in the specimen that crystallized (e.g., 1:1 or 2:1 Hfq:RNA?) and whether or not the complex found in the crystalline asymmetric unit (AU) features any additional (non-crystallographic) symmetry. Calculation of an initial electron density map from the diffraction data requires approximate

bioRxiv preprint doi: https://doi.org/10.1101/150672; this version posted June 16, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

phases for each X-ray reflection. Such phases can be estimated, *de novo*, via a family of computational approaches based on the two fundamental ideas of multiple isomorphous replacement (MIR) or multi-wavelength anomalous dispersion (MAD); these general approaches require derivatization of native crystals, either via soaking with heavy atoms (MIR/SIR/etc.) or covalent introduction of an anomalously scattering atom (MAD/SAD/etc.) such as selenium. For more information on approaches to *de novo* estimation of initial sets of phases, see [28].

If a known 3D structure is similar to the (unknown) structure that one seeks to determine, then the phase problem can be greatly simplified. In such cases, a phasing approach known as molecular replacement (MR) can be used to estimate initial phases for the unknown structure (the 'target') using a known structure (the 'probe'), or a suitable modification thereof (e.g., a homolog model). Essentially, the MR approach can be thought of as a 'fit', via rigid-body transformations that sample the three rotational and three translational degrees of freedom, of the probe structure to the unknown phases of the diffraction data (which, in turn, directly result from the detailed 3D coordinates of the target structure). Because 3D structures are available for Hfq homologs from many species (Table 1), the phase problem is much simplified by using MR. Similarly, the phases computed in refining a given *apo* Hfq structure can then be used as an initial phase estimate for X-ray data collected for a corresponding Hfq•sRNA complex. The protocols below assume that one can successfully estimate initial phases via MR; if such is not the case, e.g., if there is an unexpected/complicated stoichiometry in the AU (say four Hfq rings and three RNAs, in an odd geometric arrangement), then one must resort to *de novo* phasing methods.

2 **Materials**

2.1 **For Hfq purification**

1. DNA sample that contains the *hfq* gene of interest (e.g., genomic DNA)

bioRxiv preprint doi: https://doi.org/10.1101/150672; this version posted June 16, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

2. Inducible expression vector capable of encoding a His6× affinity tag (e.g., pET-22b(+) or pET-28b(+))
3. Chemically competent BL21(DE3) *E. coli* cells
4. Luria-Bertani (LB) agar plate, supplemented with appropriate antibiotic (e.g., 100 µg/mL ampicillin or 50 µg/mL kanamycin)
5. LB liquid media, supplemented with a suitable antibiotic (e.g., 100 µg/mL ampicillin or 50 µg/mL kanamycin)
6. 1 mM isopropyl *β*-D-1-thiogalactopyranoside (a 1000× stock [1 M] can be prepared, partitioned into 1-mL aliquots, and stored at –20 °C)
7. Lysis buffer: 50 mM Tris pH 7.5, 750 mM NaCl
8. DNase I and RNase A enzymes
9. 1 mg/mL chicken egg-white lysozyme
10. Urea or guanidinium hydrochloride (GndCl)
11. 0.2-µm syringe filters
12. His-Trap HP pre-packed sepharose column (GE Healthcare)
13. 200 mM Ni_2SO_4
14. Wash buffer: 50 mM Tris pH 7.5, 200 mM NaCl, 10 mM imidazole
15. Elution buffer: 50 mM Tris pH 7.5, 200 mM NaCl, 600 mM imidazole
16. Thrombin
17. Benzamidine sepharose resin
18. 3-kDa molecular weight cut-off (MWCO) filter-concentrators

2.2 **For sRNA purification**

1. Template DNA (≈ 1 µg/µL for a 3-kb linearized plasmid)

bioRxiv preprint doi: https://doi.org/10.1101/150672; this version posted June 16, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

2. 10x transcription buffer: 500 mM Tris-HCl pH 8.0, 100 mM NaCl, 60 mM $MgCl_2$, 20 mM spermidine
3. 10× rNTP mix: rATP, rCTP, rUTP, rGTP, each at 20 mM
4. 100 mM dithiothreitol (DTT)
5. Diethyl pyrocarbonate (DEPC)–treated H_2O (RNase-free)
6. T7 RNA polymerase
7. RNase-free DNase I
8. Denaturing (8 M urea) 5% polyacrylamide gel
9. Elution buffer: 20 mM Tris-HCl pH 7.5, 1 mM EDTA, 1% (w/v) SDS
10. Phenol:chloroform (1:1)
11. 96% ethanol

2.3 **For co-crystallization trials**

1. Purified Hfq protein (see §3.1)
2. Purified sRNA construct (see §3.2)
3. The 'Natrix' and 'Crystal Screen' sparse-matrix crystallization screens (Hamton Research)
4. Intelliplate 96-3 Microplates (Hampton Research)
5. Sealing tape (Hampton Research)
6. 24-well VDX plates with sealant (Hampton Research)
7. Siliconized glass cover slips (Hampton Research)
8. Vacuum grease
9. Light stereomicroscope, with cross-polarizing lenses (e.g., Zeiss Discovery V20)

3 **Methods**

bioRxiv preprint doi: https://doi.org/10.1101/150672; this version posted June 16, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

3.1 **Hfq Purification**

Crystallization efforts typically require large quantities of highly-purified and concentrated material, e.g., on the scale of >100 µL at >10 mg/mL of the biomolecule. To achieve this, Hfq is often expressed in a standard *E. coli* K12 laboratory strain, using a plasmid-based construct created via standard recombinant DNA techniques. The expression vector, e.g. an inducible T7*lac*-based system (pET series), can be used to add various affinity tags to the N′ or C′ termini of the wild-type sequence. Then, over-expressed Hfq can be readily purified via affinity chromatographic means. Further steps, detailed below and in Fig 2, may be required to remove co-purifying proteins and nucleic acids. Because at least some Hfq homologs bind nucleic acids fairly indiscriminately, the ratio of absorbances at 260 nm and 280 nm (A_{260}/A_{280}) should be monitored over the various stages of purification in order to detect the presence of contaminating nucleic acids. Pure nucleic acid is characterized by an A_{260}/A_{280} ratio of ≈ 1.5-2.0, versus ≈ 0.7 for pure protein; rapid colorimetric assays can be used alongside absorbance readings to discern whether a contaminant is mostly RNA or DNA [49].

In terms of solution behavior, experience has shown that Hfq homologs generally remain soluble in aqueous buffers at temperatures >70 °C, and that they resist chemical denaturation (e.g., treatment with 6 M GndCl); in many cases, depending on the species of origin, Hfq samples are insoluble at low temperatures. We have found that the common practice of purifying/storing proteins at 4 °C can be unwise with Hfq homologs: if visible precipitation occurs at ≈ 4 °C, we recommend that purification be conducted at ambient room temperature (≈18–22 °C), and that elevated temperatures be considered for long-term storage of purified protein (e.g., ≈37–42 °C works well for an Hfq homolog from the hyperthermophile *Aquifex aeolicus*). In terms of protein expression behavior, purification strategies, solubility properties (temperature- and ionic strength–dependence), etc., we have found that the *in vitro* behavior of many Hfq constructs resembles the overall properties of Hfq orthologs (Sm proteins) from the archaeal domain of life [17].

bioRxiv preprint doi: https://doi.org/10.1101/150672; this version posted June 16, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Previously, His-tagged [50,18] and self-cleaving intein-tags [51,16,21] have been used for affinity purification of Hfq, although it has also been purified without the use of a tag. Un-tagged Hfq has been purified using immobilized-metal affinity chromatography (IMAC), as the native protein has been shown to associate with the resin [4]. Poly(A)-sepharose [10] and butyl-sepharose [52,53] columns also have been utilized to purify untagged Hfq, leveraging the RNA-binding properties and partially hydrophobic nature of the surface of the protein (respectively). Below, we outline the purification of recombinant Hfq using a His6x-tagged construct. This tag can be removed at a later step through protease treatment; this is a crucial feature, as it is possible that even a modestly-sized His6x tag can interfere with the oligomerization behavior and binding properties of Hfq [54]. The intein-mediated purification with an affinity chitin-binding tag (IMPACT) scheme, used to both clone and purify intein-tagged constructs, has been successfully applied to Hfq by multiple labs (e.g., [51,21]). This system is available as a kit from NEB, so that method will not be described herein. Note that many of the considerations and notes described below, for His6×-tagged Hfq, also apply when purifying and working with any Hfq construct, intein–based or otherwise.

1. Clone the *hfq* gene, using a genomic sample as PCR template, into an appropriate expression vector; ideally, such a vector will add a His-tag. A compatible vector from the pET series of plasmids often works well (e.g., pET-28b(+), which fuses an *N*-terminal His6× tag).
2. Transform competent BL21(DE3) *E. coli* with the recombinant Hfq plasmid and plate onto LB agar supplemented with antibiotic (e.g., 50 µg/mL kanamycin if using pET-28b(+)).
3. Grow-out the transformed cells in LB media at 37 °C with shaking (225 rpm) to an optical density at 600 nm (OD_{600}) of ≈ 0.6-0.8. Then, induce over-expression of Hfq by adding IPTG to a final concentration of 1 mM. Optionally, immediately before adding IPTG take a 1-mL aliquot of the cell culture as a $t = 0$ (pre-induction) sample; this sample can be stored at –20 °C

bioRxiv preprint doi: https://doi.org/10.1101/150672; this version posted June 16, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

and later analyzed alongside a post-induction sample (by SDS-PAGE) in order to assess over-expression levels.

4. Incubate the cell cultures for an additional 3-4 h at 37 °C, with continued shaking, and then centrifuge at 15000*g* for 5 min to pellet. Optionally, take a 1-mL aliquot of the cell culture at *t* ≈ 2–3 h post-induction; this sample can be stored at –20 °C and analyzed by SDS-PAGE.

5. Resuspend the cell pellet from the previous step in Lysis Buffer (§2.1). Optionally, DNase I and RNase A can be added at this stage in order to hydrolyze any nucleic acids, Hfq-associated or otherwise (*see* **Note 1**).

6. Incubate the lysate with 0.01 mg/mL chicken egg-white lysozyme for 30 min (if a more thorough chemical lysis is required), at either RT or 37 °C; gently shake/invert the sample a few times during this incubation.

7. Mechanically lyse the cells using a sonicator or other similar means (e.g., a microfluidizer or French press). Remove cellular debris from the lysate by centrifugation at 16000*g* for 5-10 min at RT.

8. Initial purification of Hfq can be achieved by using a heat-cut to precipitate endogenous, mesophilic *E. coli* proteins. Proceed by incubating the supernatant from the last step (i.e., clarified lysate) at ≈ 70–80 °C for ≈ 10–15 min (*see* **Note 2**); a substantial amount of white precipitate should develop within minutes. Next, use a high-speed centrifugation step (e.g., 33000*g* for 30 min) to clarify the soluble, Hfq-containing supernatant; for pilot studies, the supernatant and pellet fractions from this step can be saved in case SDS-PAGE analysis becomes necessary.

9. If nucleic acid is still present in the heat-treated sample, as assessed by A_{260}/A_{280} ratios, colorimetric assays [49], or dye-binding assays (e.g., cyanine-based stains such as PicoGreen or SYBR-Gold), then a chaotropic agent such as urea or GndCl can be added to the sample, to a concentration of up to 8 M or 6 M, respectively (*see* **Note 3**).

bioRxiv preprint doi: https://doi.org/10.1101/150672; this version posted June 16, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

10. Pass the latest Hfq-containing sample through a 0.2-µm filter (syringe or vacuum line) to remove any particulate matter, prior to applying the material to a high-performance liquid chromatography (HPLC) or fast protein liquid chromatography (FPLC) system in the next step.
11. Isolate the His6×-tagged Hfq via IMAC, using an iminodiacetic acid sepharose resin in a prepacked column connected to an HPLC or FPLC instrument. All buffers should be vacuum-filtered (0.45-µm filters) and sonicated before use. In brief, this IMAC step entails the following sub-steps:
    1. Prepare the resin by washing with 3-4 column volumes (CVs) of dH_2O, and then 3-4 CVs of Wash Buffer.
    2. Charge the resin with Ni^{2+} by loading at least 1-2 CVs of 200 mM $NiSO_4$ (*see* **Note 4**).
    3. Load the crude (unpurified) Hfq-containing protein sample (collect the flow-through), and then wash the column with several CVs of Wash Buffer (until the A_{280} trace drops near baseline).
    4. Elute the Hfq protein by applying a linear gradient of Elution Buffer, from of 0→100% over 10 CVs.
    5. Combine the fractions thought to contain Hfq (as assessed by A_{280} and the elution profile), and dialyze into 50 mM Tris pH 7.5, 200 mM NaCl, 12.5 mM EDTA in order to remove any residual Ni^{2+}.
    6. To regenerate a column for subsequent use, strip the resin with 4-5 CVs of 100 mM EDTA; remove the EDTA by washing with 5-6 CVs of dH_2O (and, for long-term storage, wash with 20% EtOH).
12. To proteolytically remove the His6×-tag (*see* **Note 5)**, incubate the sample overnight with thrombin at a 1:600 mass ratio of thrombin:Hfq.
13. To remove thrombin from the latest sample, either apply the material to a benzamidine column or mix it with a resin (in batch mode).
14. Additional chromatographic steps, such as size exclusion chromatography (Fig 2), may be necessary in order to isolate the various populations of hexameric, 'free' Hfq versus any subpopulations with RNA bound.

bioRxiv preprint doi: https://doi.org/10.1101/150672; this version posted June 16, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

3.2 **Large-scale Synthesis and Purification of the sRNA**

In addition to purified protein, crystallizing an RNP complex also requires milligram quantities of RNA of sufficient quality. Here, 'quality' means that the ideal RNA sample will be (i) *chemically uniform*, in terms of sequence, length, and phosphate end-chemistry (i.e., uniform covalent structure), and also (ii) *structurally homogeneous* (i.e., narrow distribution of conformational states in solution). The first issue—monodispersity—is a fairly straightforward matter of chemistry, and is within one's control (e.g., use an RNA synthesis scheme that minimizes heterogeneity of the 3′-termini of the product RNA molecules). The second issue, concerning structural heterogeneity, is a matter of physics: one can anneal RNAs by heating/cooling, adjusting pH, ionic strength, etc., to try and modulate the solution-state behavior of an RNA, but the intrinsic structural/dynamical properties of RNA are generally not easily regulated; ultimately, one must empirically monitor the RNA and its properties of interest (e.g., 'crystallizability').

For sRNAs, which range from ≈ 50 to 200 nt, a sufficient quantity of material can be readily synthesized via run-off transcription, *in vitro*, using phage T7 RNA polymerase and a linearized plasmid as the DNA template (*see* **Note 6**). Traditional *in vitro* transcription is limited by the fact that T7 polymerase strongly prefers guanosine at the 5′-end of the transcript [55], thus adversely affecting yields for target RNAs lacking a 5′ G. In addition, the polymerase typically incorporates a few nucleotides at the 3′-end of the transcript in a random, template-independent manner, giving an RNA population that is heterogeneous in length and 3′ sequence. Both of these limitations can be avoided by including, 5′ and 3′ to the RNA sequence of interest, a pair of *cis*-acting, self-cleaving ribozymes [56,57]. The flanking ribozymes ensure that the population of RNA products is accurate and chemically uniform, given the single-nt precision with which ribozymes self-cleave at the scissile bond. In principle, any self-cleaving ribozyme can be used (hammerhead, hairpin, hepatitis δ virus (HDV), etc.). In practice, an engineered 5′ hammerhead and 3′ HDV ribozyme have been found to work well [47], and impose virtually no sequence constraints on the target RNA product; some obligatory base-pairing interactions between the 5′ hammerhead ribozyme and

bioRxiv preprint doi: https://doi.org/10.1101/150672; this version posted June 16, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

the target RNA sequence does mean that this region will need to be re-designed for each new target RNA construct that one seeks to produce.

A DNA template suitable for the *in vitro* transcription reaction can be generated by cloning the construct into a high copy number plasmid containing the T7 promoter upstream of a multiple cloning site (MCS). The plasmid will need to be linearized using a restriction enzyme with selectivity to a site that is 3′ of the sequence of interest. The individual components for the *in vitro* transcription reaction may be prepared by the user or purchased from a manufacturer; whole kits are also commercially available (e.g., MEGAscript T7 transcription kit, Invitrogen). Large quantities of T7 RNA polymerase can be produced in-house in a cost-effective manner by using a His-tagged construct and affinity purification (similar to that described above for Hfq). In general, the concentrations of rNTPs, $MgCl_2$, and T7 polymerase in the transcription reaction will require optimization for each new RNA construct/system. General guidelines and examples can be found in [56-58]. Using the method outlined below, it is ideally possible to generate milligram quantities of RNA.

1. Clone the DNA construct into a high copy number plasmid containing a T7 promoter upstream of a MCS (e.g., the pBluescript or pGEM series; *also see* **Note 7**)
2. Linearize the plasmid using a restriction enzyme for a site 3′ to the sequence of interest.
3. Mix the following components (final concentrations are noted) in the listed order, and incubate at 37 °C for 1-2 h:
    1. 1x transcription buffer (50 mM Tris-HCl pH 8.0, 10 mM NaCl, 6 mM $MgCl_2$, 2 mM spermidine)
    2. 2 mM rNTP mix
    3. 10 mM DTT
    4. Template DNA (≈ 0.05 µg/µl for a 3-kb linearized plasmid)
    5. DEPC-treated H_2O to bring to volume

bioRxiv preprint doi: https://doi.org/10.1101/150672; this version posted June 16, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

6. 0.5 U/µl T7 RNA polymerase

4. If desired, add RNase-free DNase I and incubate for 30 min to digest the original template.
5. Purify the RNA by first separating on a denaturing (≈8 M urea) 10% w/v polyacrylamide gel and then excising the band corresponding to the transcript (*see* **Note 8**).
6. Add the gel slice to a tube containing 400 µl Elution Buffer and incubate for several hours at 4 °C. Centrifuge at 10000*g* for 10 min at 4°C and transfer the supernatant to a new tube.
7. Extract the RNA by phase separation with 1-2 V phenol:chloroform. Centrifuge at 10000*g* for 20 min at 4 °C and transfer the aqueous phase to a new tube.
8. Precipitate with 2-3 V ethanol.
9. Resuspend in dH2O or an appropriate buffer (e.g., Tris-HCl pH 7.5, 0.1 mM EDTA).

3.3 **Crystallization of the Hfq•sRNA complex**

There is no reliable way to predict the conditions that will yield well-diffracting crystals of a macromolecule or macromolecular complex, such as an Hfq•sRNA assembly: the process is almost entirely empirical. The word 'almost' appears in the last sentence because the process of crystallizing biomolecules is one of guided luck. Many of the biochemical properties and behavior of a system that are most salient to crystallization—idiosyncratic variations in solubility with pH, metal ions, presence of ligands, etc.—become manifest as the knowledge that one develops after many hours of working with a biomolecular sample at the bench. This implicit knowledge is highly system-specific (sometimes varying for even a single-residue mutant in a given system), it accumulates in a tortuously incremental manner, and it directly factors into the decision-making steps that ultimately dictate the success of a crystallization effort. Thus, the best advice for crystallizing an Hfq•sRNA complex is to work as extensively as possible to characterize the Hfq and sRNA components, as well as the assembled RNP, prior to extensive crystallization trials.

bioRxiv preprint doi: https://doi.org/10.1101/150672; this version posted June 16, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Ideally, one's samples will be structurally homogeneous, thus increasing the likelihood of successful crystallization. Even given that, still it is often necessary to empirically screen through myriad potential crystallization conditions. High-throughput kits are available for the rapid screening of conditions that have successfully yielded crystals in the past (a technique commonly referred to as sparse matrix screening [59]). We recommend the Natrix Screen (Hampton Research), as it is specifically tailored to nucleic acid and protein-nucleic acid complexes; other commonly used screens, such as the Crystal Screen and PEG-Ion Screen are also advised. These kits are available in 15-mL or 1-mL high-throughput (HT) formats. Once a potential crystallization condition is identified, further optimization is usually required in order to improve the quality of the crystalline specimen. Often, this is pursued via 'grid screens'. In grid screens, one or two free parameters are varied in a systematic manner; these parameters often include the buffer and pH, protein concentration, salts (types, concentrations), types and concentrations of other precipitants (e.g., PEGs), inclusion of small-molecule additives, temperature, etc. Further information on crystallization can be found in many excellent texts (e.g., [60]) and other resources, such as the Crystal Growth 101 literature available online (https://hamptonresearch.com/growth_101_lit.aspx).

In general, the purified Hfq and sRNA must be prepared and then assessed for homogeneity and stability before crystallization trials begin (often this is done via biophysical approaches or, ideally, via functional assays). Also, we advise adhering as closely as possible to RNase-free procedures (e.g., use DEPC-treated water, RNase Zap) both in biochemical characterization steps and in handling Hfq, sRNA and Hfq•sRNA specimens for crystallization trials. The following is a general protocol to get started:

1. Dialyze the purified Hfq into an appropriate buffer (e.g., 20 mM Tris-HCl pH 7.5, 200 mM NaCl). This should be the simplest, most minimalistic buffer in which the biomolecule is stable and soluble (at a concentration of at least 1-2 mg/mL). Because RNA is involved, inclusion of salts of divalent cations, such as $MgCl_2$, may be found to aid in crystallization and overall diffraction quality.

bioRxiv preprint doi: https://doi.org/10.1101/150672; this version posted June 16, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

2. Bring the [Hfq] to ≈ 15 mg/mL via concentration or dilution, as necessary (*see* **Note 9**); concentration is often achieved via centrifugal filtration devices with a suitable MWCO (such as those by Amicon).
3. To potentially enhance the conformational homogeneity of the sRNA via annealing, incubate at 80 °C and slowly cool to RT; another approach worth trying is to heat the RNA sample and then snap-cool to ≈4 °C on ice.
4. As a cautionary (and troubleshooting) step, one can test for background RNase activity in the crystallization sample by incubating the Hfq and RNA together for ≈ 2 weeks and assay degradation via PAGE or other methods (a molar ratio of between 1:1 and 2:1 protein:RNA is recommended as a starting point [25,47,61]).

Finally, to begin crystallization trials one should follow the manufacturer's protocol for the particular sparse-matrix screen. Crystal trays should be stored in a temperature and humidity-controlled environment. Crystals can take between hours and months to develop. We recommend checking trays for crystals relatively frequently at the start of the process—e.g., after 1, 2, 4, 8, 16, 32 days. Experience suggests that, ideally, precipitation should occur in roughly one-third to one-half of the conditions within minutes of setting-up the crystallization drop; if this is not the case, the Hfq, sRNA, or Hfq•sRNA concentrations may need to be adjusted accordingly. Once a potential crystallization condition has been identified, it should be re-made in-house (using one's own reagents) in order to ensure reproducibility. Then, large-scale grid screening and further optimization can be pursued.

Intriguingly, a survey of the PDB identifies several crystallization agents that seem to recur in the crystallization of Hfq and Hfq•RNA complexes, as detailed in Table 1. The most commonly occurring reagents are (i) sodium cacodylate and citrate buffers, (ii) PEG 3350 and 2-methyl-2,4-pentanediol (MPD) precipitants, and (iii) $MgCl_2$, $CoCl_2$, and KCl salts as additives. Other divalent cations and polyamines, such as metal hexammines (e.g., hexammine cobalt(III) chloride, $[Co(NH_3)_6]Cl_3$), spermine, and spermidine, have been found to aid in the crystallization of many protein•nucleic acid complexes [47,61].

bioRxiv preprint doi: https://doi.org/10.1101/150672; this version posted June 16, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Several methods can be applied to verify that new crystals are indeed of an Hfq•sRNA complex. An Intelli-Plate (Hampton Research, HR3-118) may be used during sparse-matrix screening in order to test, in parallel, multiple components for each crystallization condition (e.g., the Hfq•sRNA complex, the Hfq alone, and the buffer alone). If crystals appear only in the drop containing Hfq•sRNA complex, there is a high likelihood that the crystals are composed of the complex. Also, macromolecular crystals can be washed, dissolved, and run on an SDS-PAGE or native polyacrylamide gel, or one can subject them to the flame of a Bunsen burner (biomolecular crystals melt, whereas salt crystals survive this trial by fire [62]). Small-molecule dyes, such as crystal violet or methylene blue, are taken up by macromolecular crystals but not by salt crystals, and thus can be used to distinguish between the two [63]. Finally, obtaining a diffraction dataset is the ultimate way to determine if a given crystal is macromolecular and, if so, the likelihood of a successful structure determination from that specimen.

4 **Notes**

1. Hfq is known to protect RNAs [64], and nucleic acid may still remain even after nuclease treatment. This potential pitfall should be monitored by A_{260}/A_{280}. If protein degradation is detected by SDS-PAGE or other means, then the Lysis Buffer used to resuspend the frozen cell pellet should be supplemented with a protease inhibitor cocktail, either commercial or home-made (including such compounds as PMSF, AEBSF, EDTA, aprotinin, leupeptin, etc.).
2. Experience with many recombinant Sm-like protein constructs suggests that the efficacy of the heat-cut step (i.e., degree of purification achieved) can vary greatly with temperature: we have found that many (>5) more *E. coli* proteins retain solubility in the clarified lysate after a 70 °C heat-cut, versus at 75 °C, at least for the BL21(DE3) strain. In purifying a new Hfq homolog, one can test 500 μL-aliquots of the clarified lysate at a series of temperatures near this range, say 65, 70, 75 and 80 °C.

bioRxiv preprint doi: https://doi.org/10.1101/150672; this version posted June 16, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

3. In our experience, Hfq withstands treatment with conventional chaotropic agents, such as high concentrations (≈ 6-8 M) of urea or GndCl. While such treatment may not fully denature the protein, we have found that it can disrupt potential Hfq···nucleic acid interactions. Adding such denaturants to the wash and elution buffers used in the IMAC stage can help mitigate nucleic acid contamination.

4. The divalent cation Co^{2+} has a lower affinity (than Ni^{2+}) for the imidazole side-chain of histidine, but it also features less non-specific binding to arbitrary proteins; if necessary because of persistent contaminants in the Hfq eluate, one can try Co^{2+} in place of Ni^{2+} in the critical IMAC purification step.

5. Often, proteins are crystallized with an intact His6x-tag. Protein tags can potentially interfere with structure or function, although this is less likely with the small His6x-tag. His-tags can also deleteriously affect crystallizability, by increasing the length of a disordered tail or by forming spurious (and weak) crystal contacts that lead to lattice disorder. We recommend cleaving the tag if possible, as this better replicates the wild type sequence. If crystals cannot be obtained with the un-tagged protein, the tagged construct should be considered for crystallization too. As two practical anecdotes from our work with the Sm-like archaeal protein (SmAP) homologs of Hfq, we note the following: (i) with *Pyrobaculum aerophilum* SmAP1, a C-terminal His6x tag was found to interfere with oligomerization *in vitro*, and ultimately the tag was cleaved-off in order for crystallization to succeed [65], and (ii) for that same recombinant construct, attempts to remove the His-tag via treatment with thrombin failed (even though the linker between the tag and the native protein sequence was designed to include a thrombin recognition site), but the tag could be successfully removed by proteolytic treatment with trypsin. In such work, we generally use mass spectrometry (typically MALDI-TOF, sometimes electrospray) to assess the accuracy of the cut-site and completeness of proteolysis.

6. Previous work has examined the RNA-binding properties of Hfq using either (i) free ribonucleotides of various forms (e.g., rNTPs, rNMPs, etc.), (ii) short oligoribonucleotides of ≲30-nt, e.g. the rU_6 oligo co-crystallized by Stanek et al. [23], or (iii) longer, full-length sRNAs, such as the 65-nt *Sal-*

bioRxiv preprint doi: https://doi.org/10.1101/150672; this version posted June 16, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

monella RydC sRNA co-crystallized with *E. coli* Hfq [25]. The nucleotides in category (i) are readily purchased, and the oligonucleotides in category (ii) are readily obtained via stepwise, solid-phase chemical synthesis (such RNAs are available from various suppliers, e.g. Dharmacon). In contrast, such approaches are inefficient for the longer (≳30-nt) oligonucleotides of (iii), and these can be efficiently generated by enzymatic synthesis, using RNA polymerase *in vitro*.

7. The plasmids pUC18 and pUC19 are commonly used for *in vitro* transcription. The T7 promoter will need to be cloned into these plasmids as well.

8. If the RNA construct contains ribozymes, be careful to not overload gels in order to enable the correctly processed, self-cleaved RNA transcript to be isolated from other products that differ by only a few nucleotides.

9. Typically, proteins are crystallized at concentrations between ≈ 5-20 mg/mL. Nevertheless, concentrations well outside this range have been required for Hfq and other Sm proteins; for instance, *A. aeolicus* Hfq crystallized at 4 mg/mL [23], while *P. aerophilum* SmAP3 was at 85 mg/mL [66]. Hfq concentrations may well be limited by protein solubility (not just supply), and likely will need to be varied in any successful set of crystallization trials.

Acknowledgements

This work was funded by NSF CAREER award MCB–1350957.

bioRxiv preprint doi: https://doi.org/10.1101/150672; this version posted June 16, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Figure Legends

Figure 1 Crystal structure of Hfq in complex with RydC sRNA (PDB 4V2S) (A) The sequence of *S. enterica* RydC sRNA is shown. The grey residues were not discernible in the crystal structure and were manually modelled in (B) and (C). Residues that bind Hfq at the lateral and proximal sites are highlighted. (B) In this cartoon ribbon representation of the *E.coli* Hfq hexamer, alternating monomeric subunits are colored blue and cyan. N′- and C′-termini are labelled for the monomer at the 6-o'clock position. The RydC RNA backbone is shown as a tan-colored tube, with the termini labelled. The 3′ end of the RydC RNA wraps around the *proximal* pore of the Hfq ring, and an internal region of the RNA binds to the *lateral* rim (yellow arrow). Uracil bases involved in binding Hfq at the *proximal* and *lateral* sites are thickened and colored orange and yellow (respectively). (C) The RydC sRNA mediates crystal contacts via binding to the *lateral* pocket of an adjacent Hfq hexamer, as indicated by the red arrow. The same coloring scheme is used as in (B), with the uridines that facilitate crystal contacts thickened and colored red. This figure was created with PyMOL.

Figure 2 Size-exclusion chromatography of Hfq samples reveals the impact of a chaotrope such as guanidinium chloride (GndCl) on elution profiles and co-purifying nucleic acid content. In particular, high concentrations of GndCl can disrupt Hfq···RNA interactions, as shown here via preparative-scale SEC chromatograms for recombinant His-tagged *A. aeolicus* Hfq constructs that were previously purified by IMAC either in the absence (0 M) or presence (at 3 M, 6 M) of GndCl. The peak that elutes at ≈ 60 mL corresponds to Hfq associated with nucleic acids, as indicated by the higher molecular weight (versus Hfq alone) and the high A_{260}/A_{280} absorbance ratio for this eluate; the peak at ≈ 100 mL corresponds to pure Hfq protein. Note the smooth shift from nucleic acid–bound Hfq to free protein as [GndCl] increases.

bioRxiv preprint doi: https://doi.org/10.1101/150672; this version posted June 16, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

References

1. Franze de Fernandez MT, Eoyang L, August JT (1968) Factor fraction required for the synthesis of bacteriophage Qbeta-RNA. Nature 219 (5154):588-590
2. Vogel J, Luisi BF (2011) Hfq and its constellation of RNA. Nature reviews Microbiology 9 (8):578-589. doi:10.1038/nrmicro2615
3. Sauer E (2013) Structure and RNA-binding properties of the bacterial LSm protein Hfq. RNA Biol 10 (4):610-618. doi:10.4161/rna.24201
4. Soper T, Mandin P, Majdalani N, Gottesman S, Woodson SA (2010) Positive regulation by small RNAs and the role of Hfq. Proceedings of the National Academy of Sciences 107 (21):9602-9607. doi:10.1073/pnas.1004435107
5. De Lay N, Schu DJ, Gottesman S (2013) Bacterial small RNA-based negative regulation: Hfq and its accomplices. The Journal of biological chemistry 288 (12):7996-8003. doi:10.1074/jbc.R112.441386
6. Jousselin A, Metzinger L, Felden B (2009) On the facultative requirement of the bacterial RNA chaperone, Hfq. Trends in microbiology 17 (9):399-405. doi:10.1016/j.tim.2009.06.003
7. Tsui HC, Leung HC, Winkler ME (1994) Characterization of broadly pleiotropic phenotypes caused by an hfq insertion mutation in Escherichia coli K-12. Molecular microbiology 13 (1):35-49
8. Wassarman KM, Repoila F, Rosenow C, Storz G, Gottesman S (2001) Identification of novel small RNAs using comparative genomics and microarrays. Genes & development 15 (13):1637-1651. doi:10.1101/gad.901001
9. Sittka A, Lucchini S, Papenfort K, Sharma CM, Rolle K, Binnewies TT, Hinton JC, Vogel J (2008) Deep sequencing analysis of small noncoding RNA and mRNA targets of the global post-transcriptional regulator, Hfq. PLoS genetics 4 (8):e1000163. doi:10.1371/journal.pgen.1000163
10. Zhang A, Wassarman KM, Ortega J, Steven AC, Storz G (2002) The Sm-like Hfq protein increases OxyS RNA interaction with target mRNAs. Molecular cell 9 (1):11-22
11. Sledjeski DD, Whitman C, Zhang A (2001) Hfq is necessary for regulation by the untranslated RNA DsrA. Journal of bacteriology 183 (6):1997-2005. doi:10.1128/jb.183.6.1997-2005.2001
12. Fantappie L, Metruccio MM, Seib KL, Oriente F, Cartocci E, Ferlicca F, Giuliani MM, Scarlato V, Delany I (2009) The RNA chaperone Hfq is involved in stress response and virulence in Neisseria meningitidis and is a pleiotropic regulator of protein expression. Infection and immunity 77 (5):1842-1853. doi:10.1128/iai.01216-08
13. Lenz DH, Mok KC, Lilley BN, Kulkarni RV, Wingreen NS, Bassler BL (2004) The small RNA chaperone Hfq and multiple small RNAs control quorum sensing in Vibrio harveyi and Vibrio cholerae. Cell 118 (1):69-82. doi:10.1016/j.cell.2004.06.009
14. Mika F, Hengge R (2013) Small Regulatory RNAs in the Control of Motility and Biofilm Formation in E. coli and Salmonella. International journal of molecular sciences 14 (3):4560-4579. doi:10.3390/ijms14034560
15. Chao Y, Vogel J (2010) The role of Hfq in bacterial pathogens. Current opinion in microbiology 13 (1):24-33. doi:10.1016/j.mib.2010.01.001
16. Schumacher MA, Pearson RF, Moller T, Valentin-Hansen P, Brennan RG (2002) Structures of the pleiotropic translational regulator Hfq and an Hfq-RNA complex: a bacterial Sm-like protein. The EMBO journal 21 (13):3546-3556. doi:10.1093/emboj/cdf322
17. Mura C, Randolph PS, Patterson J, Cozen AE (2013) Archaeal and eukaryotic homologs of Hfq: A structural and evolutionary perspective on Sm function. RNA Biol 10 (4):636-651. doi:10.4161/rna.24538
18. Sauer E, Weichenrieder O (2011) Structural basis for RNA 3'-end recognition by Hfq. Proceedings of the National Academy of Sciences of the United States of America 108 (32):13065-13070. doi:10.1073/pnas.1103420108
19. Link TM, Valentin-Hansen P, Brennan RG (2009) Structure of Escherichia coli Hfq bound to polyriboadenylate RNA. Proceedings of the National Academy of Sciences of the United States of America 106 (46):19292-19297. doi:10.1073/pnas.0908744106
20. Horstmann N, Orans J, Valentin-Hansen P, Shelburne SA, 3rd, Brennan RG (2012) Structural mechanism of Staphylococcus aureus Hfq binding to an RNA A-tract. Nucleic acids research 40 (21):11023-11035. doi:10.1093/nar/gks809
21. Sun X, Wartell RM (2006) *Escherichia coli* Hfq binds A_{18} and DsrA domain II with similar 2:1 Hfq_6/RNA stoichiometry using different surface sites. Biochemistry 45 (15):4875-4887. doi:10.1021/bi0523613
22. Panja S, Schu DJ, Woodson SA (2013) Conserved arginines on the rim of Hfq catalyze base pair formation and exchange. Nucleic Acids Res 41 (15):7536-7546. doi:10.1093/nar/gkt521
23. Stanek KA, Patterson-West J, Randolph PS, Mura C (2017) Crystal structure and RNA-binding properties of an Hfq homolog from the deep-branching Aquificae: conservation of the lateral RNA-binding mode. Acta crystallographica Section D, Structural biology 73 (Pt 4):294-315. doi:10.1107/s2059798317000031

bioRxiv preprint doi: https://doi.org/10.1101/150672; this version posted June 16, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

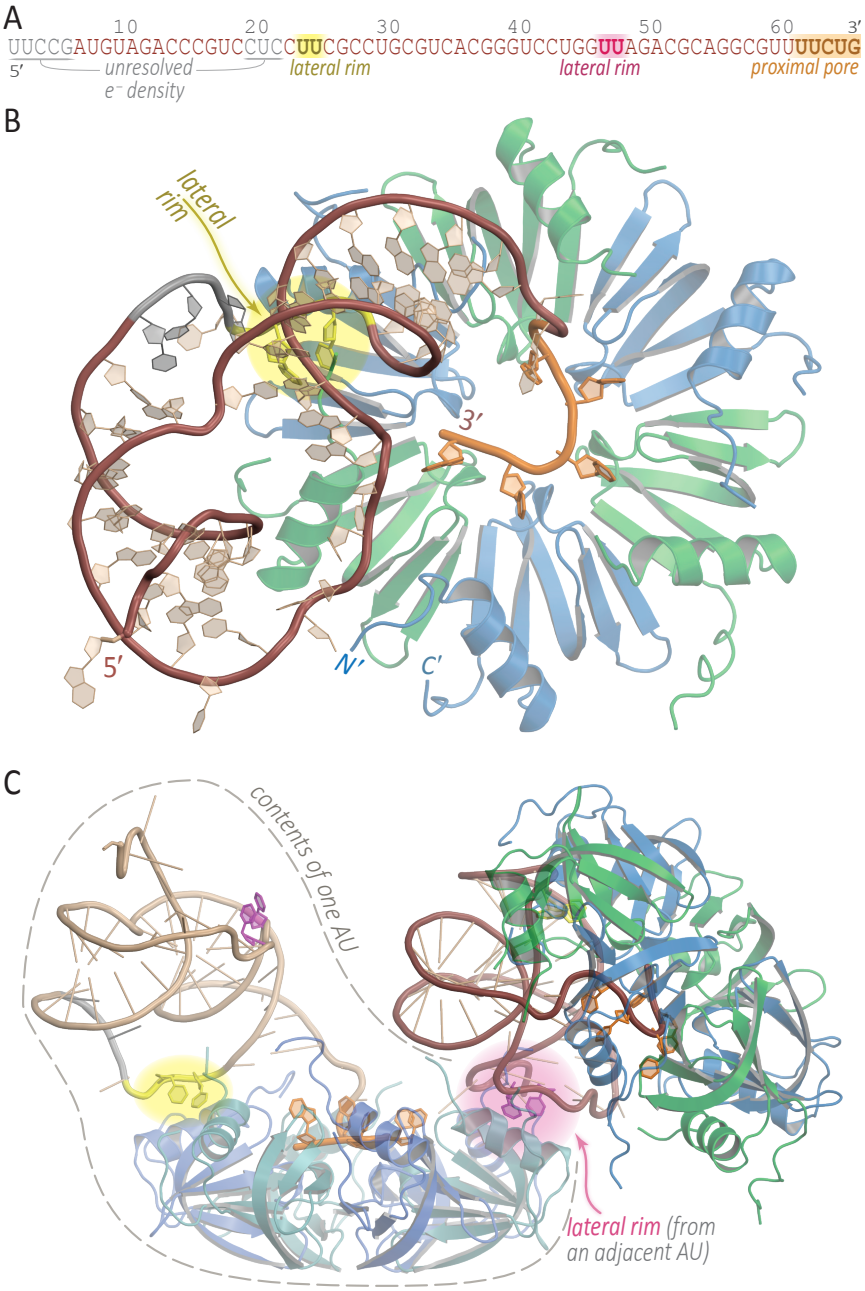
24. Schu DJ, Zhang A, Gottesman S, Storz G (2015) Alternative Hfq-sRNA interaction modes dictate alternative mRNA recognition. The EMBO journal 34 (20):2557-2573. doi:10.15252/embj.201591569
25. Dimastrogiovanni D, Frohlich KS, Bandyra KJ, Bruce HA, Hohensee S, Vogel J, Luisi BF (2014) Recognition of the small regulatory RNA RydC by the bacterial Hfq protein. Elife 3. doi:10.7554/eLife.05375
26. Schröder GF (2015) Hybrid methods for macromolecular structure determination: experiment with expectations. Curr Opin Struct Biol 31:20-27. doi:10.1016/j.sbi.2015.02.016
27. Schlundt A, Tants JN, Sattler M (2017) Integrated structural biology to unravel molecular mechanisms of protein-RNA recognition. Methods 118-119:119-136. doi:10.1016/j.ymeth.2017.03.015
28. Sherwood D, Cooper JB, Press OU (2011) Crystals, X-rays and Proteins: Comprehensive Protein Crystallography. Oxford University Press,
29. Cavanagh J, Fairbrother WJ, Palmer AG, Skelton NJ, Rance M (2010) Protein NMR Spectroscopy: Principles and Practice. Elsevier Science,
30. Bai X-c, McMullan G, Scheres SHW How cryo-EM is revolutionizing structural biology. Trends in Biochemical Sciences 40 (1):49-57. doi:10.1016/j.tibs.2014.10.005
31. Cheng Y, Grigorieff N, Penczek PA, Walz T (2015) A Primer to Single-Particle Cryo-Electron Microscopy. Cell 161 (3):438-449. doi:10.1016/j.cell.2015.03.050
32. Schluenzen F, Tocilj A, Zarivach R, Harms J, Gluehmann M, Janell D, Bashan A, Bartels H, Agmon I, Franceschi F, Yonath A (2000) Structure of functionally activated small ribosomal subunit at 3.3 angstroms resolution. Cell 102 (5):615-623
33. Carter AP, Clemons WM, Brodersen DE, Morgan-Warren RJ, Wimberly BT, Ramakrishnan V (2000) Functional insights from the structure of the 30S ribosomal subunit and its interactions with antibiotics. Nature 407 (6802):340-348
34. Ban N, Nissen P, Hansen J, Moore PB, Steitz TA (2000) The complete atomic structure of the large ribosomal subunit at 2.4 A resolution. Science (New York, NY) 289 (5481):905-920
35. Cohen SB, Graham ME, Lovrecz GO, Bache N, Robinson PJ, Reddel RR (2007) Protein composition of catalytically active human telomerase from immortal cells. Science (New York, NY) 315 (5820):1850-1853. doi:10.1126/science.1138596
36. Nguyen THD, Galej WP, Bai X-c, Savva CG, Newman AJ, Scheres SHW, Nagai K (2015) The architecture of the spliceosomal U4/U6•U5 tri-snRNP. Nature 523 (7558):47-52
37. Agafonov DE, Kastner B, Dybkov O, Hofele RV, Liu WT, Urlaub H, Luhrmann R, Stark H (2016) Molecular architecture of the human U4/U6.U5 tri-snRNP. Science (New York, NY) 351 (6280):1416-1420. doi:10.1126/science.aad2085
38. Tangprasertchai NS, Zhang X, Ding Y, Tham K, Rohs R, Haworth IS, Qin PZ (2015) An Integrated Spin-Labeling/Computational-Modeling Approach for Mapping Global Structures of Nucleic Acids. Methods Enzymol 564:427-453. doi:10.1016/bs.mie.2015.07.007
39. Putnam CD, Hammel M, Hura GL, Tainer JA (2007) X-ray solution scattering (SAXS) combined with crystallography and computation: defining accurate macromolecular structures, conformations and assemblies in solution. Quarterly reviews of biophysics 40 (3):191-285. doi:10.1017/s0033583507004635
40. Yadav DK, Lukavsky PJ (2016) NMR solution structure determination of large RNA-protein complexes. Progress in nuclear magnetic resonance spectroscopy 97:57-81. doi:10.1016/j.pnmrs.2016.10.001
41. Lu P, Bai X-c, Ma D, Xie T, Yan C, Sun L, Yang G, Zhao Y, Zhou R, Scheres SHW, Shi Y (2014) Three-dimensional structure of human γ-secretase. Nature 512 (7513):166-170. doi:10.1038/nature13567
42. Bartesaghi A, Merk A, Banerjee S, Matthies D, Wu X, Milne JL, Subramaniam S (2015) 2.2 A resolution cryo-EM structure of beta-galactosidase in complex with a cell-permeant inhibitor. Science (New York, NY) 348 (6239):1147-1151. doi:10.1126/science.aab1576
43. Ferre-D'Amare AR, Zhou K, Doudna JA (1998) A general module for RNA crystallization. Journal of molecular biology 279 (3):621-631. doi:10.1006/jmbi.1998.1789
44. Ferre-D'Amare AR, Doudna JA (2001) Methods to crystallize RNA. Curr Protoc Nucleic Acid Chem Chapter 7:Unit 7 6. doi:10.1002/0471142700.nc0706s00
45. Evans P (2006) Scaling and assessment of data quality. Acta Crystallogr D Biol Crystallogr 62 (Pt 1):72-82. doi:10.1107/S0907444905036693
46. Eddy SR (2014) Computational analysis of conserved RNA secondary structure in transcriptomes and genomes. Annu Rev Biophys 43:433-456. doi:10.1146/annurev-biophys-051013-022950
47. Ke A, Doudna JA (2004) Crystallization of RNA and RNA-protein complexes. Methods 34 (3):408-414. doi:10.1016/j.ymeth.2004.03.027

bioRxiv preprint doi: https://doi.org/10.1101/150672; this version posted June 16, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

48. Sawaya MR (2007) Characterizing a crystal from an initial native dataset. Methods Mol Biol 364:95-120. doi:10.1385/1-59745-266-1:95
49. Patterson J, Mura C (2013) Rapid colorimetric assays to qualitatively distinguish RNA and DNA in biomolecular samples. J Vis Exp (72):e50225. doi:10.3791/50225
50. Sukhodolets MV, Garges S (2003) Interaction of Escherichia coli RNA polymerase with the ribosomal protein S1 and the Sm-like ATPase Hfq. Biochemistry 42 (26):8022-8034. doi:10.1021/bi020638i
51. Møller T, Franch T, Højrup P, Keene DR, Bächinger HP, Brennan RG, Valentin-Hansen P (2002) Hfq: A bacterial Sm-like protein that mediates RNA-RNA interaction. Mol Cell 9 (1):23-30
52. Nikulin A, Stolboushkina E, Perederina A, Vassilieva I, Blaesi U, Moll I, Kachalova G, Yokoyama S, Vassylyev D, Garber M, Nikonov S (2005) Structure of Pseudomonas aeruginosa Hfq protein. Acta crystallographica Section D, Biological crystallography 61 (Pt 2):141-146. doi:10.1107/s0907444904030008
53. Kadowaki MA, Iulek J, Barbosa JA, Pedrosa Fde O, de Souza EM, Chubatsu LS, Monteiro RA, de Oliveira MA, Steffens MB (2012) Structural characterization of the RNA chaperone Hfq from the nitrogen-fixing bacterium Herbaspirillum seropedicae SmR1. Biochimica et biophysica acta 1824 (2):359-365. doi:10.1016/j.bbapap.2011.11.002
54. Obregon KA, Hoch CT, Sukhodolets MV (2015) Sm-like protein Hfq: Composition of the native complex, modifications, and interactions. Biochimica et biophysica acta 1854 (8):950-966. doi:10.1016/j.bbapap.2015.03.016
55. Milligan JF, Groebe DR, Witherell GW, Uhlenbeck OC (1987) Oligoribonucleotide synthesis using T7 RNA polymerase and synthetic DNA templates. Nucleic acids research 15 (21):8783-8798
56. Price SR, Ito N, Oubridge C, Avis JM, Nagai K (1995) Crystallization of RNA-protein complexes. I. Methods for the large-scale preparation of RNA suitable for crystallographic studies. Journal of molecular biology 249 (2):398-408
57. Ferre-D'Amare AR, Doudna JA (1996) Use of cis- and trans-ribozymes to remove 5' and 3' heterogeneities from milligrams of in vitro transcribed RNA. Nucleic acids research 24 (5):977-978
58. Beckert B, Masquida B (2011) Synthesis of RNA by In Vitro Transcription. In: Nielsen H (ed) RNA: Methods and Protocols. Humana Press, Totowa, NJ, pp 29-41. doi:10.1007/978-1-59745-248-9_3
59. McPherson A, Gavira JA (2014) Introduction to protein crystallization. Acta Crystallographica Section F, Structural Biology Communications 70 (Pt 1):2-20. doi:10.1107/S2053230X13033141
60. McPherson A (1999) Crystallization of Biological Macromolecules. Cold Spring Harbor Laboratory Press,
61. Obayashi E, Oubridge C, Krummel DP, Nagai K (2007) Crystallization of RNA-Protein Complexes. In: Walker JM, Doublié S (eds) Macromolecular Crystallography Protocols: Volume 1, Preparation and Crystallization of Macromolecules. Humana Press, Totowa, NJ, pp 259-276. doi:10.1007/978-1-59745-209-0_13
62. Raghunathan K, Harris PT, Arvidson DN (2010) Trial by fire: are the crystals macromolecules? Acta Crystallographica Section F: Structural Biology and Crystallization Communications 66 (Pt 5):615-620. doi:10.1107/S1744309110012078
63. Salt or Protein Crystal? Hampton Research. https://hamptonresearch.com/documents/growth_101/20.pdf.
64. Folichon M, Arluison V, Pellegrini O, Huntzinger E, Regnier P, Hajnsdorf E (2003) The poly(A) binding protein Hfq protects RNA from RNase E and exoribonucleolytic degradation. Nucleic acids research 31 (24):7302-7310
65. Mura C, Cascio D, Sawaya MR, Eisenberg DS (2001) The crystal structure of a heptameric archaeal Sm protein: Implications for the eukaryotic snRNP core. Proc Natl Acad Sci U S A 98 (10):5532-5537. doi:10.1073/pnas.091102298
66. Mura C, Phillips M, Kozhukhovsky A, Eisenberg D (2003) Structure and assembly of an augmented Sm-like archaeal protein 14-mer. Proceedings of the National Academy of Sciences of the United States of America 100 (8):4539-4544. doi:10.1073/pnas.0538042100

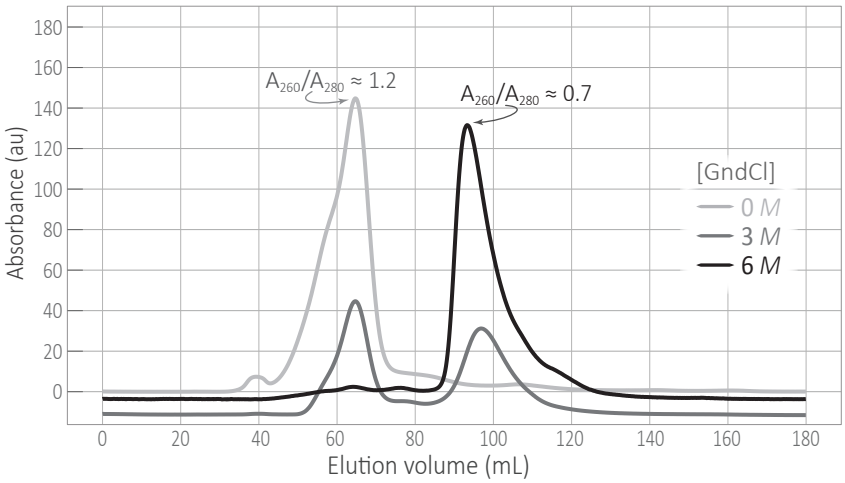
bioRxiv preprint doi: https://doi.org/10.1101/150672; this version posted June 16, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Stanek & Mura, **Figure 1**. Crystal structure of *E. coli* Hfq bound to a RydC sRNA.



bioRxiv preprint doi: https://doi.org/10.1101/150672; this version posted June 16, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Stanek & Mura, **Figure 2**. Size-exclusion chromatography of Hfq samples: Impact of GndCl on elution profiles and co-purifying nucleic acid content.



bioRxiv preprint doi: https://doi.org/10.1101/150672; this version posted June 16, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Table 1: A comprehensive list of Hfq structures in the PDB, including co-crystal structures with nucleotides and RNAs.

PDB ID (year)	Space-group	d_{min} (Å)	Solvent content (by vol)	Macromolecular complex [species]; other notes	Crystallization information (format, precipitants, other solution conditions/notes)	Citation
1KQ1 (2002)	$P2_1$	1.55	35.9%	$(Hfq)_6$ hexamer [*Staphylococcus aureus*]	HDVD[1] at 298 K; pH 4.6; $(NH_4)_2SO_4$, NaOAc	[1]
1KQ2 (2002)	$C222_1$	2.71	43.3%	$(Hfq)_6$•r(AU_5G) RNA, bound at proximal site; [*S. aureus*]	HDVD at 298 K; pH 7.5; HEPES, PEG-550, $MgCl_2$, KCl	[1]
1HK9 (2003)	$P6_1$	2.15	33.0%	$(Hfq)_6$ hexamer [*Escherichia coli*]	SDVD at 293 K; pH 4.6; 25% PEG-4000, 0.2 M NH_4-OAc, 0.2 M NaOAc	[2]
1U1S (2005)	$P2_12_12_1$	1.6	50.0%	$(Hfq)_6$ hexamer [*Pseudomonas aeruginosa*]	HDVD at 295 K; pH 8.5; 200 mM NH_4Cl, 12% PEG-4000, 50 mM Tris-HCl, 5 mM $CdCl_2$	[3]
1U1T (2005)	$P2_12_12_1$	1.9	51.7%	$(Hfq)_6$ hexamer [*P. aeruginosa*]	HDVD at 295 K; pH 6.5; 100 mM MES, 0.6 M $(NH_4)_2SO_4$, 1 M Li_2SO_4	[3]
2QTX (2007)	$P2_1$	2.5	45.3%	$(Hfq)_6$ hexamer [*Methanococcus jannaschii*]	SDVD at 277 K; pH 8.5; 0.1 M Tris, 0.2 M NH_4OAc, 25% PEG-3350	[4]
3GIB (2009)	$P2_12_12$	2.4	45.0%	$(Hfq)_6$•r$(A)_9$ RNA, bound to distal site; [*E. coli*]	HDVD at 298 K; pH 9.5; 0.1 M CHES, 40% v/v MPD	[5]
3HFN (2009)	$P3$	2.3	42.7%	$(Hfq)_6$ hexamer [*Anabaena PCC 7120*]	SDVD at 277 K; pH 3.5; 0.1 citric acid, 2 M $(NH_4)_2SO_4$	[6]
3HFO (2009)	$F222$	1.3	36.1%	$(Hfq)_6$ hexamer [*Synechocystis sp. PCC 6803*]	SDVD at 292 K; pH 7; 60% Tacsimate[2]	[6]
3INZ (2010)	$P2_12_12_1$	1.7	41.2%	$(Hfq)_6$ hexamer, H57T mutant; [*P. aeruginosa*]	VD at 295 K; pH 8.5; 50 mM NaCl, 100 mM NH_4Cl, 7.5% PEG-MME 550, 50 mM TrisCl, 10 mM $CdCl_2$	[7]
3M4G (2010)	$P1$	2.05	41%	$(Hfq)_6$ hexamer, H57A mutant; [*P. aeruginosa*]	HDVD at 293 K; pH 8.5; 50 mM NaCl, 100 mM NH_4Cl, 7.5% PEG-MME 550, 50 mM TrisCl, 10 mM $ZnCl_2$	[7]
2Y90 (2011)	$P6$	2.25	30%	$(Hfq)_6$ hexamer [*E. coli*]	HDVD at 293 K; pH 8; 0.1 M Tris, 1.6 M $(NH_4)_2SO_4$	[8]
2YHT (2011)	$P1$	2.9	30%	$(Hfq)_6$ hexamer [*E. coli*]	SDVD at 293 K; pH 5.4; 0.1 M Na^+citrate, 30% PEG-3350	[8]
2YLB (2011)	$P6_1$	1.15	39%	$(Hfq)_6$ hexamer [*Salmonella typhimurium*]	pH 7; 0.1 M HEPES, 0.5% Jeffamine, 1.1 M malonate	[9]
2YLC (2011)	$P6$	1.3	40%	$(Hfq)_6$•r$(U)_6$ RNA, bound at proximal pore; [*S. typhimurium*]	pH 8; 0.2 M NaSCN, 20% PEG-3350	[9]
3AHU (2011)	$F222$	2.2	42.0%	$(Hfq)_6$•r$(AG)_3A$ RNA (SELEX-derived aptamer), bound to distal site; [*Bacillus subtilis*]	HDVD at 293 K; pH 6.5; 0.2 M MES, 1.8 M $(NH_4)_2SO_4$, 0.01 M $CoCl_2$	[10]
3HSB (2011)	$I422$	2.2	39.6%	$(Hfq)_6$•r$(AG)_3A$ RNA (SELEX-derived aptamer), bound to distal site; [*B. subtilis*]	HDVD at 293 K; pH 6.5; 0.1 M MES, 1.8 M $(NH_4)_2SO_4$, 0.015 M $CoCl_2$	[10]
3QHS (2011)	$P1$	2.85	37.7%	$(Hfq)_6$ hexamer; full-length protein; [*E. coli*]	HDVD at 277 K; pH 6.5; 0.1 M Bis-Tris, 30% v/v PEG-MME 550, 0.05 M $CaCl_2$	[11]
3RER (2011)	$P1$	1.7	43.7%	$(Hfq)_6$•r(AU_6A) RNA•ADP [*E. coli*]	HDVD at 283 K; pH 6.2; 0.1 M cacodylate, 100 mM NaCl, 12% w/v PEG-8000	[12]

bioRxiv preprint doi: https://doi.org/10.1101/150672; this version posted June 16, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

PDB ID (year)	Space-group	d_{min} (Å)	Solvent content (by vol)	Macromolecular complex [species]; other notes	Crystallization information (format, precipitants, other solution conditions/notes)	Citation
3RES (2011)	*I*2	2.0	42.6%	$(Hfq)_6$•ADP [*E. coli*]	HDVD at 283 K; pH 4.2; 200 mM NH_4OAc, 100 mM NaOAc, 22% w/v PEG-4000	[12]
3SB2 (2011)	$P2_1$	2.63	43.0%	$(Hfq)_6$ hexamer [*Herbaspirillum seropedicae*]	SDVD at 291 K; pH 7.0; 0.1 M PCB (Na-propionate, Na-cacodylate, Bis-tris propane), 25% w/v PEG-1500	[13]
3QO3 (2012)	*C*2	2.15	48.3%	$(Hfq)_6$•ATP [*E. coli*]	SDVD at 295 K; pH 7.5; 0.1 M HEPES, 10% w/v PEG-8000, 8% v/v ethylene glycol	[14]
3QSU (2012)	*P*3	2.2	35.9%	$(Hfq)_6$•$r(A)_7$ RNA, bound to distal site; [*S. aureus*]	HDVD at 298 K; pH 6.5; 0.1 M Na-cacodylate, 12% v/v MPD, 0.2 M $Zn(OAc)_2$, 0.1 M KCl	[15]
3QUI (2013)	$P2_12_12_1$	1.93	37.9%	$(Hfq)_6$•{ADP, AMP-PNP}[3] [*P. aeruginosa*]	HDVD at 295 K; pH 8; 0.2 M $(NH_4)_2SO_4$, 0.2 M NaCl, 50 mM TrisCl	[16]
3VU3 (2013)	*I*222	2.85	58.9%	$(Hfq)_6$•Catalase HPII[4] [*E. coli*]	HDVD at 293 K; pH 9; 0.1 M TrisCl, 0.18 M NaCl, 10% w/v PEG-4000	[17]
4HT8 (2013)	*C*2	1.9	43.0%	$(Hfq)_6$•$r(A)_7$ RNA [*E. coli*]	HDVD at 283 K; pH 7.9; 200 mM NH_4OAc, 100 mM Tris, 26% v/v isopropanol	[18]
4HT9 (2013)	*I*2	1.8	33.0%	$(Hfq)_6$•$r(A)_7$•$r(AU_6A)$ RNAs [*E. coli*]	HDVD at 283 K; pH 6.2; 0.1 M cacodylate, 0.1 M NaCl, 12% w/v PEG-8000	[18]
4J5Y (2013)	$P2_12_12_1$	2.1	44.8%	$(Hfq)_6$•ATP [*P. aeruginosa*]	HDVD at 295 K; pH 8; 0.2 M $(NH_4)_2SO_4$, 0.2 M NaCl, 50 mM TrisCl	[16]
4J6W (2013)	$P2_12_12_1$	1.8	46.3%	$(Hfq)_6$•CTP [*P. aeruginosa*]	HDVD at 295 K; pH 8; 0.2 M $(NH_4)_2SO_4$, 0.2 M NaCl, 50 mM TrisCl	[16]
4J6X (2013)	$P2_12_12_1$	2.22	44.9%	$(Hfq)_6$•UTP [*P. aeruginosa*]	HDVD at 295 K; pH 8; 0.2 M $(NH_4)_2SO_4$, 0.2 M NaCl, 50 mM TrisCl	[16]
4J6Y (2013)	$P2_12_12_1$	2.14	44.1%	$(Hfq)_6$; GTP not found in density; [*P. aeruginosa*]	HDVD at 295 K; pH 8; 0.2 M $(NH_4)_2SO_4$, 0.2 M NaCl, 50 mM TrisCl, 4.75 mM GTP	[16]
4JLI (2014)	*H*3	1.79	35.7%	$(Hfq)_6$ hexamer, F42W mutant; [*E. coli*]	HDVD; pH 8.0-9.0; 0.1 M Tris, 22-28% PEG-3350, 26-32% isopropanol	[19]
4JRI (2014)	$P2_1$	1.83	38.0%	$(Hfq)_6$ hexamer, F39W mutant; [*E. coli*]	HDVD; pH 8.0-9.0; 0.1 M Tris, 22-28% PEG-3350, 26-32% isopropanol	[19]
4JRK (2014)	$P22_12_1$	1.89	39.9%	$(Hfq)_6$ hexamer, F11W mutant; [*E. coli*]	HDVD; pH 8.0-9.0; 0.1 M Tris, 22-28% PEG-3350, 26-32% isopropanol	[19]
4JUV (2014)	$P2_1$	2.19	40.2%	$(Hfq)_6$ hexamer, Y25W mutant; [*E. coli*]	HDVD at 295 K; pH 8.0-9.0; 0.1 M Tris, 22-28% PEG-3350, 26-32% isopropanol	[19]
4MMK (2014)	$P2_1$	2.16	40.3%	$(Hfq)_6$ hexamer, Q8A mutant; [*P. aeruginosa*]	HDVD at 303 K; pH 6.5; 50 mM TrisCl, 7% w/v PEG-2000 MME, 2% v/v MPD, 20 µM $ZnCl_2$	[20]
4MML (2014)	*P*6	1.8	36.5%	$(Hfq)_6$ hexamer, D40A mutant; [*P. aeruginosa*]	HDVD at 303 K; pH 6.5; 50 mM TrisCl, 7% w/v PEG-2000 MME, 2% v/v MPD, 20 µM $ZnCl_2$	[20]
4NL2 (2014)	$P2_122_1$	2.6	43.1%	$(Hfq)_6$ hexamer [*Listeria monocytogenes*]	HDVD at 298 K; pH 7.5; 0.1 M HEPES, 40% 1,2-propanediol	[21]
4NL3 (2014)	*C*2	3.1	48.7%	$(Hfq)_6$•$r(U)_6$ RNA, bound at proximal pore; [*L. monocytogenes*]	HDVD at 298 K; pH 7.5; 0.1 M HEPES, 40% 1,2-propanediol	[21]
4NOY (2014)	$P2_122_1$	2.8	43.3%	$(Hfq)_6$ hexamer, F43W mutant; [*L. monocytogenes*]	HDVD at 298 K; pH 7.5; 0.1 M HEPES, 40% 1,2-propanediol	[21]
4PNO (2014)	*P*6	0.97	33.4%	$(Hfq)_6$•$r(U)_6$ RNA, bound at proximal pore; [*E. coli*]	SDVD at 293 K; pH 8.0; 0.1 M HEPES/NaOH, 12% w/v PEG-3350, 0.25M KSCN	[22]

bioRxiv preprint doi: https://doi.org/10.1101/150672; this version posted June 16, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

PDB ID (year)	Space-group	d_{min} (Å)	Solvent content (by vol)	Macromolecular complex [species]; other notes	Crystallization information (format, precipitants, other solution conditions/notes)	Citation
4V2S (2014)	$P2_12_12_1$	3.48	55%	$(Hfq)_6$•RydC sRNA (65 nt) [*E. coli* Hfq, *S. enterica* RydC]	SDVD; pH 6.5; 0.2 M tri-sodium citrate, 0.1 M Na-cacodylate, 15% isopropanol	[23]
4QVC (2015)	$P2_12_12_1$	1.99	49.3%	$(Hfq)_6$•r(AUAACUA) RNA [*E. coli*]	HDVD at 281 K; pH 5.5; 0.1 M citrate, 12% w/v PEG-4000	[24]
4QVD (2015)	$P2_12_12_1$	1.97	49.5%	$(Hfq)_6$•r(AACUAAA) RNA [*E. coli*]	HDVD at 281 K; pH 7.2; 0.1 M HEPES, 16% w/v MPEG-5000[5]	[24]
4RCB (2015)	*P*6	1.63	38.3%	$(Hfq)_6$ hexamer [*E. coli*]	SDVD; 1.6 M $(NH_4)_2SO_4$, 0.5 M LiCl	[25]
4RCC (2015)	*P*6	1.98	38.3%	$(Hfq)_6$ hexamer [*E. coli*]	SDVD; pH 8.5; 0.1 M TrisCl, 1.5 M $(NH_4)_2SO_4$, 15% v/v glycerol	[25]
4Y91 (2015)	$P2_12_12_1$	2.66	39.0%	$(Hfq)_6$•r$(U)_6$ RNA, bound at proximal pore; [*Thermotoga maritima*]	HDVD at 291 K; pH 8.5; tri-potassium citrate, 30% w/v PEG-3350	n/a
4X9C (2016)	$P2_12_12_1$	1.4	41.7%	$(Hfq)_6$ hexamer [*M. jannaschii*]	HDVD at 296 K; pH 8; 0.1 M TrisCl, 50% v/v PEG-200	[26]
4X9D (2016)	$P2_12_12_1$	1.5	41.9%	$(Hfq)_6$•UMP [*M. jannaschii*]	HDVD at 296 K; pH 8; 0.1 M TrisCl, 50% v/v PEG-200	[26]
5DY9 (2016)	$P2_1$	1.6	34.9%	$(Hfq)_6$•AMP (Y68T mutant of Hfq); [*M. jannaschii*]	HDVD at 296 K; pH 8; 0.1 M TrisCl, 50% v/v PEG-200	[26]
5I21 (2016)	*P*6	1.55	37.0%	$(Hfq)_6$ hexamer, Y55W mutant; [*P. aeruginosa*]	HDVD at 303 K; pH 6.5; 50 mM TrisCl, 7% w/v PEG-2000 MME, 2% v/v MPD	n/a
5SZD (2017)	*P*1	1.49	40.2%	$(Hfq)_6$ hexamer [*Aquifex aeolicus*]	SDVD at 291 K; pH 5.5; 0.1 M Na-cacodylate, 5% w/v PEG-8000, 35% v/v MPD, 0.1 M $[Co(NH_3)_6]Cl_3$	[27]
5SZE (2017)	*P*6	1.5	46.1%	$(Hfq)_6$•r$(U)_6$ RNA, bound at lateral rim; [*A. aeolicus*]	SDVD at 291 K; pH 5.5; 0.1 M Na-cacodylate, 5% w/v PEG-8000, 35% v/v MPD, 1.0 M GndCl	[27]

Notes

[1] 'HDVD' and 'SDVD' stand for the hanging-drop and sitting-drop vapor diffusion methods, respectively.

[2] Tacsimate is "a mixture of titrated organic acid salts" that contains 1.8 M malonic acid, 0.25 M ammonium citrate tribasic, 0.12 M succinic acid, 0.3 M DL-malic acid, 0.4 M sodium acetate trihydrate, 0.5 M sodium formate, and 0.16 M ammonium tartrate dibasic (for more information, see http://hamptonresearch.com/documents/product/hr000175_what_is_tacsimate_new.pdf).

[3] This structure contains molecules of both ADP and the non-hydrolyzable ATP analog AMP-PNP.

[4] In this serendipitous co-crystal structure of *E. coli* Hfq and catalase HPII, an Hfq hexamer was found to bind each subunit of a HPII tetramer.

[5] MPEG, an acronym for methoxypolyethylene glycol (also known as PEG monomethyl ether), has a covalent formula of $CH_3(OCH_2CH_2)_nOH$, versus $H(OCH_2CH_2)_nOH$ for simple PEGs.

bioRxiv preprint doi: https://doi.org/10.1101/150672; this version posted June 16, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

References (Note: The following are specific to Table 1 only; can be merged with main text for final production?)

[1] Schumacher MA, RF Pearson, T Moller, P Valentin-Hansen, and RG Brennan. Structures of the pleiotropic translational regulator Hfq and an Hfq-RNA complex: A bacterial Sm-like protein. *EMBO Journal*. 2002; 21: 3546-56.

[2] Sauter C, J Basquin, and D Suck. Sm-like proteins in Eubacteria: The crystal structure of the Hfq protein from *Escherichia coli*. *Nucleic Acids Research*. 2003; 31: 4091-8.

[3] Nikulin A, E Stolboushkina, A Perederina, I Vassilieva, U Blaesi, I Moll, . . . S Nikonov. Structure of *Pseudomonas aeruginosa* Hfq protein. *Acta Crystallographica, Section D: Biological Crystallography*. 2005; 61: 141-6.

[4] Nielsen JS, A Boggild, CB Andersen, G Nielsen, A Boysen, DE Brodersen, and P Valentin-Hansen. An Hfq-like protein in archaea: Crystal structure and functional characterization of the Sm protein from *Methanococcus jannaschii*. *RNA*. 2007; 13: 2213-23.

[5] Link TM, P Valentin-Hansen, and RG Brennan. Structure of *Escherichia coli* Hfq bound to polyriboadenylate RNA. *Proceedings of the National Academy of Sciences of the United States of America*. 2009; 106: 19292-7.

[6] Bøggild A, M Overgaard, P Valentin-Hansen, and DE Brodersen. Cyanobacteria contain a structural homologue of the Hfq protein with altered RNA-binding properties. *FEBS Journal*. 2009; 276: 3904-15.

[7] Moskaleva O, B Melnik, A Gabdulkhakov, M Garber, S Nikonov, E Stolboushkina, and A Nikulin. The structures of mutant forms of Hfq from *Pseudomonas aeruginosa* reveal the importance of the conserved His57 for the protein hexamer organization. *Acta Crystallographica, Section F: Structural Biology and Crystallization Communications*. 2010; 66: 760-4.

[8] Bonnefond L, P Schellenberger, J Basquin, G Demangeat, C Ritzenthaler, R Chênevert, . . . C Sauter. Exploiting Protein Engineering and Crystal Polymorphism for Successful X-ray Structure Determination. *Crystal Growth & Design*. 2011; 11: 4334-4343.

[9] Sauer E, and O Weichenrieder. Structural basis for RNA 3′-end recognition by Hfq. *Proceedings of the National Academy of Sciences of the United States of America*. 2011; 108: 13065-70.

[10] Someya T, S Baba, M Fujimoto, G Kawai, T Kumasaka, and K Nakamura. Crystal structure of Hfq from *Bacillus subtilis* in complex with SELEX-derived RNA aptamer: Insight into RNA-binding properties of bacterial Hfq. *Nucleic Acids Research*. 2012; 40: 1856-1867.

[11] Beich-Frandsen M, B Vecerek, B Sjoblom, U Blasi, and K Djinovic-Carugo. Structural analysis of full-length Hfq from *Escherichia coli*. *Acta Crystallographica, Section F: Structural Biology and Crystallization Communications*. 2011; 67: 536-40.

[12] Wang W, L Wang, Y Zou, J Zhang, Q Gong, J Wu, and Y Shi. Cooperation of *Escherichia coli* Hfq hexamers in DsrA binding. *Genes & Development*. 2011; 25: 2106-17.

[13] Kadowaki MA, J Iulek, JA Barbosa, O Pedrosa Fde, EM de Souza, LS Chubatsu, . . . MB Steffens. Structural characterization of the RNA chaperone Hfq from the nitrogen-fixing bacterium *Herbaspirillum seropedicae* SmR1. *Biochimica et Biophysica Acta*. 2012; 1824: 359-65.

[14] Hammerle H, M Beich-Frandsen, B Vecerek, L Rajkowitsch, O Carugo, K Djinovic-Carugo, and U Blasi. Structural and biochemical studies on ATP binding and hydrolysis by the *Escherichia coli* RNA chaperone Hfq. *PLoS ONE*. 2012; 7: e50892.

[15] Horstmann N, J Orans, P Valentin-Hansen, SA Shelburne, 3rd, and RG Brennan. Structural mechanism of *Staphylococcus aureus* Hfq binding to an RNA A-tract. *Nucleic Acids Research*. 2012; 40: 11023-35.

[16] Murina V, N Lekontseva, and A Nikulin. Hfq binds ribonucleotides in three different RNA-binding sites. *Acta Crystallographica, Section D: Biological Crystallography*. 2013; 69: 1504-13.

[17] Yonekura K, M Watanabe, Y Kageyama, K Hirata, M Yamamoto, and S Maki-Yonekura. Post-transcriptional regulator Hfq binds catalase HPII: Crystal structure of the complex. *PLoS ONE*. 2013; 8: e78216.

[18] Wang W, L Wang, J Wu, Q Gong, and Y Shi. Hfq-bridged ternary complex is important for translation activation of *rpoS* by DsrA. *Nucleic Acids Research*. 2013; 41: 5938-48.

[19] Robinson KE, J Orans, AR Kovach, TM Link, and RG Brennan. Mapping Hfq-RNA interaction surfaces using tryptophan fluorescence quenching. *Nucleic Acids Research*. 2014; 42: 2736-49.

[20] Murina VN, BS Melnik, VV Filimonov, M Uhlein, MS Weiss, U Muller, and AD Nikulin. Effect of conserved intersubunit amino acid substitutions on Hfq protein structure and stability. *Biochemistry (Biokhimiia)*. 2014; 79: 469-77.

[21] Kovach AR, KE Hoff, JT Canty, J Orans, and RG Brennan. Recognition of U-rich RNA by Hfq from the Gram-positive pathogen *Listeria monocytogenes*. *RNA*. 2014; 20: 1548-59.

[22] Schulz EC, and O Barabas. Structure of an *Escherichia coli* Hfq:RNA complex at 0.97 Å resolution. *Acta Crystallographica, Section F: Structural Biology Communications*. 2014; 70: 1492-7.

[23] Dimastrogiovanni D, KS Frohlich, KJ Bandyra, HA Bruce, S Hohensee, J Vogel, and BF Luisi. Recognition of the small regulatory RNA RydC by the bacterial Hfq protein. *eLife*. 2014; 3.

[24] Wang L, W Wang, F Li, J Zhang, J Wu, Q Gong, and Y Shi. Structural insights into the recognition of the internal A-rich linker from OxyS sRNA by *Escherichia coli* Hfq. *Nucleic Acids Research*. 2015; 43: 2400-11.

bioRxiv preprint doi: https://doi.org/10.1101/150672; this version posted June 16, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

[25] Ji-Yong FS-QSY-LSC-YWP-QS. Limited Proteolysis Improves *E. coli* Hfq Crystal Structure Resolution. *Chinese Journal of Biochemistry and Molecular Biol*. 2015; 31: 1102-1108.

[26] Nikulin A, A Mikhailina, N Lekontseva, V Balobanov, E Nikonova, and S Tishchenko. Characterization of RNA-binding properties of the archaeal Hfq-like protein from *Methanococcus jannaschii*. *Journal of Biomolecular Structure & Dynamics*. 2017; 35: 1615-1628.

[27] Stanek KA, J Patterson-West, PS Randolph, and C Mura. Crystal structure and RNA-binding properties of an Hfq homolog from the deep-branching Aquificae: conservation of the lateral RNA-binding mode. *Acta Crystallographica, Section D: Structural Biology*. 2017; 73: 294-315.