




RESEARCH PAPER



A high-throughput search for intracellular factors that affect RNA folding identifies *E. coli* proteins PepA and YagL as RNA chaperones that promote RNA remodelling

Alejandra Matsuri Rojano-Nisimura ^a, Lucas G. Miller ^b, Aparna Anantharaman^b, Aaron T. Middleton^a, Elroi Kibret^a, Sung H. Jung^c, Rick Russell^a, and Lydia M. Contreras ^b

^aDepartment of Molecular Biosciences, The University of Texas at Austin, Austin, TX, USA; ^bMcKetta Department of Chemical Engineering, The University of Texas at Austin, Austin, TX, USA; ^cDepartment of Biomedical Engineering, The University of Texas at Austin, Austin, TX, USA

ABSTRACT

General RNA chaperones are RNA-binding proteins (RBPs) that interact transiently and non-specifically with RNA substrates and assist in their folding into their native state. In bacteria, these chaperones impact both coding and non-coding RNAs and are particularly important for large, structured RNAs which are prone to becoming kinetically trapped in misfolded states. Currently, due to the limited number of well-characterized examples and the lack of a consensus structural or sequence motif, it is difficult to identify general RNA chaperones in bacteria. Here, we adapted a previously published *in vivo* RNA regional accessibility probing assay to screen genome wide for intracellular factors in *E. coli* affecting RNA folding, among which we aimed to uncover novel RNA chaperones. Through this method, we identified eight proteins whose deletion gives changes in regional accessibility within the exogenously expressed *Tetrahymena* group I intron ribozyme. Furthermore, we purified and measured *in vitro* properties of two of these proteins, YagL and PepA, which were especially attractive as general chaperone candidates. We showed that both proteins bind RNA and that YagL accelerates native refolding of the ribozyme from a long-lived misfolded state. Further dissection of YagL showed that a putative helix-turn-helix (HTH) domain is responsible for most of its RNA-binding activity, but only the full protein shows chaperone activity. Altogether, this work expands the current repertoire of known general RNA chaperones in bacteria.

ARTICLE HISTORY

Revised 1 November 2024
Accepted 11 November 2024

KEYWORDS

RNA chaperones; RNA folding; RNA-binding; protein-RNA interactions; nucleic acid binding proteins



Introduction


The biological functions of many RNAs require them to fold into specific conformations. RNA folding is a complex, hierarchical process in which initial chain compaction and local structure formation often result in non-native structures or contacts that become fixed by tertiary contacts to generate misfolded conformations [1,2]. These misfolded states can be long-lived, with large free energy barriers for refolding to their corresponding native states [3]. General RNA chaperones are RNA-binding proteins (RBPs) that bind RNA transiently and with low specificity (affinities typically in the μM range) to resolve misfolded conformers by promoting local unfolding. Upon protein-RNA complex dissociation, the RNA has an additional opportunity to fold to its native, functional state [4,5].

Known examples of general RNA chaperones in *E. coli* include the cold-shock domain protein, CspA, and the StpA protein. CspA is a relatively small (7.4 kDa) protein and a member of the OB-fold superfamily [6]. OB-fold proteins share a characteristic five-stranded β -barrel structure with two positively charged regions surrounding an exposed aromatic

patch, allowing intermediate (μM range) nucleic-acid binding [7]. Specifically, CspA binds single-stranded RNA (ssRNA) and acts as an RNA chaperone during cold-shock (when intracellular CspA concentration is about 10^{-4} M) by resolving stem loop structures within nascent mRNAs and allowing for transcript elongation [8,9]. In the case of StpA, this protein has been shown to have annealing and strand displacement activity *in vitro* and to be capable of increasing the frequency and/or lifetime of local unfolding events within structured RNAs [10]. StpA is also a small (15.3 kDa) protein and is composed of two structural domains. Intermediate nucleic acid binding (μM range affinity) occurs via electrostatic interactions between the RNA and the positively charged C-terminal domain (CTD) of the protein [11]. Other proteins with RNA chaperone activity in *E. coli* include ribosomal proteins, with S12 being the best characterized example [12,13], and ATP-dependent general RNA remodelers from the DEAD-box helicase family, like SrmB, CsdA and RhlE (reviewed in [14]).

General RNA chaperones are both structurally and functionally diverse. Nevertheless, an emerging theme is that, in

CONTACT Lydia M. Contreras  lcontrer@che.utexas.edu  McKetta Department of Chemical Engineering, The University of Texas at Austin, 200 E Dean Keeton St., Austin, TX 78712, USA

 Supplemental data for this article can be accessed online at <https://doi.org/10.1080/15476286.2024.2429956>

© 2024 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited. The terms on which this article has been published allow the posting of the Accepted Manuscript in a repository by the author(s) or with their consent.

addition to their RNA chaperone functions, several of these proteins have specialized functions that involve binding to DNA substrates. For instance, CspA acts as a transcriptional enhancer by recognizing a single-stranded CCAAT motif and promoting the transcription of genes such as *hns* and *gyrA* [15]. Likewise, StpA, a paralog of the HN-S global regulator that participates in DNA condensation, binds curved DNA and has been proposed to serve as a molecular *hns* backup [16,17]. These findings suggest that additional DNA-binding proteins (DBPs) might also moonlight as RNA chaperones. However, identification of other DBPs (or proteins in general) capable of remodelling RNA substrates is limited by the lack of shared structural or sequence motifs; thereby making it infeasible to predict RNA chaperone functions using bioinformatics.

To identify cellular factors that promote RNA folding, we have adapted the *in vivo* RNA Structural Sensing System (iRS [3] assay previously developed by our group [18,19]. The iRS [3] uses sequence-specific, user-designed antisense RNA (asRNA) probes against an RNA of interest by coupling hybridization of the probe to activation of a downstream GFP fluorescence reporter sequence. As such, this assay allows for the evaluation of changes in local RNA structures and the identification of potential functional regions (such as those that participate in RNA-protein interactions) [20]. We previously used this iRS³ assay to profile the folding of the *Tetrahymena* group I intron ribozyme using a collection of sequence-specific asRNA probes and benchmarked our results against those obtained by *in vivo* DMS probing [18].

Here, we coupled the iRS³ assay to a Tn5 transposon library and identified genes that, when knocked out, give changes in the accessibility of a key region of the *Tetrahymena* group I ribozyme when it is expressed in *E. coli*. This ribozyme was selected due to its extensive structural characterization *in vitro*, which includes cryo-EM resolved structures of its native and non-native states and chemical probing structural maps [21–26], and its known retained activity when expressed in *E. coli*. Through this approach, we identified two genes, *yagL* and *pepA*, that may encode RNA chaperones. Supporting this hypothesis, we found that purified PepA protein, shown previously to interact functionally with DNA [27], can also bind single-stranded RNA (ssRNA) and double-stranded RNA (dsRNA). Furthermore, the purified YagL protein, which was previously uncharacterized, accelerates native refolding of a long-lived misfolded conformation of the *Tetrahymena* ribozyme and binds both ssRNA and dsRNA. Via homology modelling and deep learning, we identified a predicted helix-turn-helix (HTH) RNA-binding domain in the C-terminus region of YagL and showed, using protein truncations, that this predicted HTH region is primarily responsible for the RNA-binding activity of YagL. In contrast, both the N- and C-terminal domains are required for its chaperone role *in vitro*, suggesting additional functional roles for the N-terminus of YagL. Together, our results expand the repertoire of known RNA chaperone proteins in *E. coli* and introduce a methodology to investigate their influence on RNA

folding *in vivo*, which could be applicable to other RNA chaperones and other RNA molecules.

Materials and methods

Plasmids and strains

A detailed list of all strains, plasmids and primers used in this study can be found under supplementary Table S1

EZ-Tn5 transposition and library preparation

A DNA fragment containing tetR was PCR amplified from pACYC184 [28]. pCML375 was constructed by inserting this fragment into the EZ-Tn5-carrying plasmid pMOD<MCS>; (Lucigen). The transposon DNA was purified by digesting pCML375 using PvuII-HF (NEB), and transposomes were generated following the manufacturer's instructions. 1 U/ μ L of the EZ-Tn5 transposome was used for electroporation into *E. coli* MG1655 as previously described in [29].

After transformation, recovered cells were grown on eight large plates (150 \times 15 mm), containing Luria-Bertani (LB) medium (Benton-Dickenson and Company) supplemented with 10 μ g/mL tetracycline (Amresco). To ensure good library coverage, each plate was partitioned into 68 sections and 20–30 colonies were collected from each partition.

$$\frac{8 \text{ large plates}}{1} \times \frac{68 \text{ partitions}}{\text{large plate}} \times \frac{20 - 30 \text{ CFU}}{\text{partition}} \\ = 10,880 \sim 3x \text{ coverage of suspected genes in } E. coli$$

A previously published toehold-mediated reporter plasmid containing a tetracycline-inducible antisense RNA (asRNA) targeting the P3 and P4 domains of the *Tetrahymena* ribozyme (herein Probe 3) fused to GFP [18,30], as well as the ribozyme target transcript under a pBAD promoter (sequence included in Supplementary Table S2) was transformed into the isolated colonies by chemical transformation (this plasmid is referred to as pCML2533 in Supplementary Table S1). In short, cells from 20 to 30 colonies were treated with CaCl₂ to make them competent and were then transformed with the pCML2533 plasmid and plated in LB-agar supplemented with 10 μ g/mL tetracycline and 50 μ g/mL kanamycin (Amresco). From the resulting plates, library collections were assembled by collecting transformed colonies directly from the plates by resuspending them in \sim 500 μ L of fresh liquid LB media. Five different library collections were obtained, each containing recovered cells from \sim 14 plates.

Cells were subcultured to prepare 25% (v/v) glycerol stocks which were stored at -80°C .

Screening of the transposon library and FACS sorting

A 500 μ L glycerol cryo-stock aliquot was thawed completely and used to inoculate 20 mL of LB medium supplemented with 10 μ g/mL tetracycline and 50 μ g/mL kanamycin (Amresco). Cultures were incubated at 37°C and 120 rpm until they reached target OD_{600–0.15}. At that point, the *in vivo* reporter was induced by adding 800 μ L of 20% arabinose (final concentration 0.8%) and 20 μ L of 100 mg/mL

anhydrotetracycline (aTc) (final concentration 100 ng/ μ l). Prior to sorting, cultures were incubated for 2.5 h.

Cell sorting was performed using a FACSAria III cell sorter (Becton Dickinson) equipped with a blue solid-state laser (488 nm excitation) using a 100 μ m nozzle at a sorting rate of \sim 300 events/s and a sorting efficacy between 75% and 90%. Settings were adjusted, and the data was visualized using the FACSDiva software (BD and Company). Fluorescence scatter was used to sort cells into four different populations based on their GFP fluorescence intensity (GFP-A). The area of interest was determined for high fluorescence expressing cells by comparing the scatter distribution of the transposon library relative to wild type *E. coli* MG1655 cells.

Isolates from the region of interest corresponded to 1.1–1.7% of the total population. Cells were spread on agar plates containing LB with 10 μ g/mL tetracycline and 50 μ g/mL kanamycin (Amresco), and 377 different CFU were obtained. For storage, glycerol cryo-stocks were made for each isolate and were stored at -80°C .

Validation of isolated mutants after FACS by 96-well plate reader screens

To validate the outcome of the high-throughput screen, colonies were screened again using a Cytation3 plate reader in a 96-well plate. Colonies corresponding to the 377 isolates obtained after Fluorescence Activated Cell Sorting (FACS) were re-grown in 5 mL overnight cultures supplemented with 10 μ g/mL tetracycline and 50 μ g/mL kanamycin (Amresco). The next morning, 96-well polystyrene plates (Greiner) were filled, with 300 μ L of fresh LB broth and 0.5% kanamycin (Amresco) and inoculated with 3 μ L of overnight culture (6 wells per overnight, corresponding to three technical replicates for induced and uninduced conditions). After 2 h, 0.15 μ L of 100 mg/mL aTc (final concentration 100 ng/ μ l) were added to the wells. Six microlitre of 40% arabinose (final concentration 0.8%) was added to half of the wells, corresponding to the induced condition. After 5 h of incubation, final OD₆₀₀ measurement and fluorescence readings were collected. Sixty-four isolates were selected, using a fluorescence ratio log₂ fold-change cut-off of >0.5 and a p-value of <0.05 , for follow-up identification and characterization experiments. A schematic overview of the screening and validation process after FACS is shown in Supplementary Figure S1.

Whole-genome sequencing and identification of transposon insertions

Overnight cultures were started for each of the 64 library isolates using 5 mL of LB medium supplemented with standard amounts of tetracycline. These cultures were pooled together into eight different subcultures. To this purpose, 500 μ L of overnight culture (62.5 μ L per isolate) was used to seed 500 mL of fresh LB medium (1:1000 dilution). Cultures were grown to saturation and used to make glycerol cryo-stocks to be stored at -80°C .

For each subculture, genomic DNA (gDNA) was extracted using a QIAamp UCP DNA Micro Kit (QIAGEN) and following manufacturer's instructions. Total yields were \sim 6–8 μ g of genomic DNA per pool. gDNA was extracted from wild type *E. coli* MG1655 cells as a control. Genomic DNA was submitted to the GSAF core facility (UT Austin) for library prep and sequencing. The samples were analysed for quality check using a bioanalyzer (Agilent). DNA libraries were prepared using standard Illumina kits and were run using a MiSeq sequencer (Illumina) in a 2×250 paired-end scheme.

The following computational pipeline was used to identify transposon insertions: (i) quality control checks were performed using fastqc [31]. (ii) adapter sequences were trimmed using CUTADAPT [32]. (iii) the trimmed sequences in FASTQ format were used as inputs for the Transposon Insertion Finder (TIF) program [33]. As an additional quality check, the Bedtools' coverage utility [34] was used after trimming to assess the coverage depth and breadth of our pooling approach; obtaining a coverage \sim 12–60 \times for each sequenced pool. The 19-bp mosaic end (ME) sequence (5'-CTG TCT CTT ATA CAC ATC T - 3') recognized by the Tn5 transposase and its reverse complement were used as head and tail end sequence inputs for the TIF program. Additionally, the length of the resulting transposon site duplication (TSD) was set to 9 bp. The resulting FASTA file, containing all identified sequences flanking a transposon sorted by TSD, was used to map the transposon insertion start and end positions by BLAST search against the *E. coli* K-12 MG1655 reference genome.

In total, 31 different transposon insertions were mapped using this approach (Table 1).

Fluorescence measurements and regional accessibility mapping using the in vivo RNA structural sensing system (iRS³)

Fluorescence measurements were performed by adapting a previously published protocol [19]. In short, *E. coli* cells (*i.e.*, genomic mutants of potential chaperone genes identified by transposon screening, Table 1), were transformed with the toehold-mediated reporter plasmid carrying Probe 3 (see EZ-Tn5 transposition and library preparation section above). Additionally, strains were transformed in parallel with two previously published control probes [19]: 'p-O-iRS3GG-scramble', which encodes for a random 15 nt asRNA probe that does not have complementarity with the genome of *E. coli* and should thus remain in an RBS-sequestered conformation and generate only background, low-end GFP signal, and 'p-O-iRS3GG-open', which is based on the scramble probe but the cis-blocking region that sequesters the RBS is mutated such that the probe is always in a free, open conformation and should generate the max, high-end of GFP signal. These control probes were used in every experiment to validate the induction of the probes upon chemical addition and to assess the fluorescence range of the experiment. No major variations were observed in the fluorescence of these control probes between experimental runs (Supplementary Figure S4).

Table 1. List of genes identified through whole-genome sequencing (WGS). Thirty-one different transposon insertions were identified after cell sorting and sequencing. The sequencing pool number corresponds to the number of sample(s) in which the insertion was mapped out of the eight different pools of isolates for which gDNA was extracted. Insertion start and insertion end are the bp positions from the *E. coli* reference genome (*Escherichia coli* K-12 MG1655, accession number U00096) at which the transposon start and end sequence was mapped. Gene descriptions were obtained from GenBank.

Sequencing pool#	Insertion Start	Insertion End	Gene	Description
1	414911	414919	<i>sbpC</i>	ATP dependent, structure specific DNA nuclease
1	4291955	4291963	<i>nrfE</i>	cytochrome c-type biogenesis protein, NrfE
1	4485073	4485081	<i>pepA</i>	aminopeptidase
2	581419	581427	<i>nohD</i>	DNA-packaging protein NU1 homolog
2	581419	581427	<i>tfaD</i>	Protein TfaD
2	3169527	3169535	<i>ygiW</i>	BOF family protein YgiW
2	3993162	3993170	<i>cyaA</i>	adenylate cyclase
2	4616796	4616804	<i>yjiW</i>	putative glycyl-radical enzyme activating enzyme
2	4616796	4616804	<i>yjiJ</i>	DUF3029 domain-containing protein YjiJ
3, 5	85626	85634	<i>ilvI</i>	acetolactate synthase/acetohydroxybutanoate synthase, catalytic subunit
3,5	551513	551521	<i>purK</i>	N5-carboxyaminoimidazole ribonucleotide synthase
3	3845314	3845322	<i>adeD</i>	adenine deaminase
3, 8	3981176	3981184	<i>yifK</i>	putative transporter YifK
4	293654	293662	<i>yagL</i>	CP4-6 prophage; resolvase-like catalytic domain-containing protein YagL
4, 7	417166	417174	<i>phoB</i>	DNA-binding transcriptional dual regulator PhoB
4	571209	571217	<i>ybcL</i>	DLP12 prophage; periplasmic protein YbcL
4, 7, 8	3131712	3131720	<i>yghQ</i>	putative transport protein YghQ
4, 5	3537690	3537698	<i>yhgF</i>	putative RNA-binding protein YhgF
4	4384804	4384812	<i>yjeM</i>	putative transporter YjeM
5	3958302	3958310	<i>ilvC</i>	ketol-acid reductoisomerase (NADP(+))
6	290000	290008	<i>argF</i>	CP4-6 prophage; ornithine carbamoyltransferase
6	3063153	3063161	<i>ygfD</i>	methylmalonyl-CoA mutase-interacting GTPase
6	3570659	3570667	<i>glgX</i>	limit dextrin alpha-1,6-glucohydrolase
6	4276489	4276497	<i>yjxC</i>	c-di-GMP phosphodiesterase PdeC
5, 6, 7	3408468	3408476	<i>panF</i>	pantothenate:Na(+) symporter
7	3689953	3689961	<i>bcsC</i>	endo-1,4-D-glucanase
7	4310415	4310423	<i>yjCW</i>	D-allose ABC transporter ATP binding subunit
7	4368784	4368792	<i>fxsA</i>	protein FxsA
7	4507089	4507097	<i>yjhE</i>	KpLE2 phage-like element; putative membrane protein (pseudogene)
8	3732908	3732916	<i>xyIG</i>	xylose ABC transporter ATP binding subunit
8	3870472	3870480	<i>yidT</i>	putative D-galactonate transporter

While the fluorescence of the controls was not replicated, the results from those runs were not considered.

Transformed strains harbouring the reporter plasmids were grown overnight at 37°C and 120 rpm. Fifty microlitre of overnight culture was used to inoculate 5 mL of fresh LB medium (1:100 dilution) supplemented with 50 µg/mL of kanamycin (Amresco). Two tubes with fresh medium were seeded per overnight culture. Cultures were allowed to grow until they reached target OD₆₀₀ ~0.15. At that point, 5 µL of anhydrotetracycline (aTc) (final concentration 100 ng/µL) was added to the cultures to induce expression of the asRNA probe. This condition represents the target-uninduced control (herein after referred to as uninduced for simplicity). Additionally, 200 µL of 20% arabinose (final concentration 0.8%) was added to one of the culture pairs for induction of the target RNA (ribozyme) (induced condition).

GFP fluorescence was measured 2.5 h post-induction on a FACSCalibur flow cytometer (Becton Dickinson) equipped with a 488 nm argon laser and a 530 nm FL1 logarithmic amplifier. Sample data were collected using the CellQuest Pro software (BD and Company) with a user-defined gate. Fluorescence measurements were collected from ~250,000 cells per sample and analysed using Microsoft Excel (Microsoft for Windows). The medians of uninduced (asRNA probe only) and induced (asRNA probe + ribozyme) samples were used to calculate the fluorescence relative ratio of each biological replicate tested.

Protein purification

stpA was cloned into the *Bam*HI and *Sac*I sites of pET-21a (+) via Gibson assembly using pCYFP [35] as a starting backbone and through the addition of homology arms to the *stpA* sequence by PCR. The resulting plasmid (pCML2868) was used to purify StpA-H6 (containing six additional His residues at the C terminus) via Ni-NTA pull downs as described in [36]. The same cloning approach was used to generate plasmids for the purification of PepA-H6 (pCML2870) and YagL-H6 (pCML3388). Successful protein expression and band sizes were confirmed by Coomassie blue staining (Supplementary Figure S2). After purification, protein concentrations were determined via Bradford assay using bovine serum albumin (BSA) as a standard. We confirmed that PepA was functionally active after purification by performing a peptidase activity assay (see Supplementary Methods). The activity of our purified PepA was determined to be ~73.1 nKat/mg of protein (Supplementary Figure S3). Previous determinations using this method have found the activity of PepA to be 30–150 nKat/mg of protein [37].

Gibson assembly was also used for the generation of YagL domain truncations [38]. Briefly, primers were designed to amplify the YagL coding region from pCML3388 corresponding to the HTH domain (187–232) or the full-length protein without this region (1–186) with homology arms for insertion between StyI and XhoI sites of pET21a (+). These inserts were ligated into the pET21a (+) backbone to form YagL_HTH (pCML3739) and YagL_dHTH (pCML3725). Protein expression, purification, and quality checks were performed as described above. Protein purity

was further verified by LC-MS/MS. Samples were digested with trypsin, desalted, then run on the Dionex LC and Orbitrap Fusion 2 for LC-MS/MS with a 2-h run time. Raw data files were analysed using the Proteome Discoverer version 2.5 and Scaffold version 5. The intended his-tagged purified proteins were the most abundant proteins detected via mass spectrometry. No proteins were detected within 10-fold of the Normalized Quantitative Value for each purified protein sample.

Northern blotting analysis

To measure the steady state levels of the iRS [3] transcript, RNA was extracted from cells expressing the iRS [3] reporter 3 h post-induction using the Direct-zol™ RNA MiniPrep kit (Zymo). After extraction, the RNA was subjected to a Northern Blot analysis as described in [39,40]. The iRS [3] transcript was blotted using probe I (5'-GCCCATTAACATCACC - 3'). For the *in vivo* assays, the iRS [3] transcript is expressed from a pLtetO promoter.

In vitro RNA-Protein binding assays

Filter binding assays were performed by adapting previously published protocols [10,41,42]. Two 5'-end dephosphorylated oligonucleotides corresponding to a 21-mer sequence (5'-AUGUGGAAAAUCUCUAGCAGU-3', herein '21 R+'), and the complementary sequence (5'-CUGCUAGAGAUUUUCCACAU-3', herein "21 R-"), previously published in [10], were custom synthesized by Genelink. An additional short RNA hairpin (5'-GCTCTAGAGCATTATGTTTCAGATAAGG-3', herein 'hairpin'), previously published in [10] was custom synthesized by IDT. The P [31] end-labelled oligonucleotides were prepared as described in [43]. Binding reactions were carried out by incubating 50 fmol of labelled RNA oligonucleotide with increasing concentrations of the respective protein (0–10 μM) for 10 min at room temperature in the binding buffer (75 mM Tris-HCl pH 7.5, 0.4 mM spermidine, 0.1 mM MgCl₂, 50 mM NaCl, 0.5 mM EDTA, 0.25 mM DTT and 6% glycerol; 60 μL total reaction volume).

After incubation, 50 μL of the binding mixture reactions were applied onto a Bio-Dot apparatus (BioRad) assembled with membranes pre-equilibrated with binding buffer without salts (top: nitrocellulose membrane (0.45 μm pore size; BioRad); bottom: Nylon N+ membrane (0.2 μm pore size; Amersham/Cytiva) and washed twice with 100 μL binding buffer. The membranes were placed on Whatman chromatography paper (Amersham/Cytiva) to air dry for ~10 min and then covered with saran wrap. The membranes were exposed to a phosphor screen (GE Healthcare) for 1 h. Following exposure, the phosphor screen was imaged using a Typhoon Phosphorimager.

Kinetic chaperone activity assays

The activity of candidate chaperone proteins was evaluated using a previously published *in vitro* two-step catalytic activity assay [44,45]. In this assay, catalytic activity is used as a readout to monitor native state folding of the *Tetrahymena* ribozyme.

Materials were prepared as described in [46] and measurements were performed with slight modifications. Specifically, the ribozyme (0.15 μM) was incubated into 20 mM Na-MOPS (pH

7.0) and 5 mM MgCl₂ for 6 min at 25°C to give a population of predominantly misfolded ribozyme and then placed on ice. Folding reactions were initiated by the addition of 6 μL of purified enzyme (*i.e.*, StpA, YagL, PepA, etc.) at different concentrations (or the equivalent volume of chaperone storage buffer [20 mM Tris-Cl (pH 7.5), 500 mM KCl, 1 mM EDTA, 0.2 mM DTT, and 50% glycerol (vol/vol); 6 μL]). Refolding was monitored at 25°C. At various time points, reaction aliquots (2 μL) were transferred to a folding quench solution [2 μL of 95 mM MgCl₂, 1 mM guanosine, 1 mg/mL proteinase K and 20 mM Na-MOPS (pH 7.0)] to stop the folding reaction and create the necessary conditions for the subsequent substrate cleavage reaction. Quenched aliquots were kept on ice until the cleavage reactions were initiated.

During the catalytic step, substrate cleavage reactions were initiated by adding 1 μL of trace 5'-[31]p-labelled substrate oligonucleotide (~20,000 dpm/μL) to the folding reaction aliquots. This substrate (5'-CCCUCUA₅-3', abbreviated rSA₅) mimics the 5'-splice-site junction and is cleaved by the ribozyme to give a shorter radiolabeled product (5'-CCCUCU-3') (47). Cleavage reactions were stopped after 2 min with 2-fold excess of an EDTA-containing gel-loading solution [72% formamide (vol/vol), 100 mM EDTA, 0.4 mg/ml xylene cyanol, and 0.4 mg/ml bromophenol blue]. The fraction of cleaved rSA₅ was determined by running the reaction products on a 20% denaturing polyacrylamide gel. Measurement of this fraction for the reaction aliquots from various folding times was used to monitor the formation of the native ribozyme as a function of folding time [47]. Substrate cleavage reactions were performed for 2 min, which allows for essentially complete cleavage of the substrate bound to the native ribozyme and insignificant cleavage of the substrate bound to the misfolded ribozyme. Thus, under these conditions, the fraction of the substrate that is cleaved provides a good measure of the fraction of the ribozyme that is in the native state.

Results

High-throughput screen uncovers pool of candidate proteins that may assist RNA folding

Given the heterogeneous nature of proteins that function as general RNA chaperones, we performed a high-throughput screen to identify proteins that affect RNA folding. We anticipated that these proteins would likely include nucleic acid-binding proteins (and specifically RBPs) with general RNA chaperone activity. For the screen, we used *in vivo* regional RNA structure probing, which provides a framework for studying RNA molecules in their cellular context. The *in vivo* RNA Structural Sensing System (iRS [3]) identifies structurally accessible regions by probing 9–16 nt regions of a target transcript with user-designed, complementary asRNA probes [30]. Successful binding and hybridization of the asRNA probe to the target RNA region disrupts an adjacent cis-repressed hairpin structure that sequesters the ribosome-binding site (RBS), thus, preventing translation of the green fluorescence protein (GFP) reporter (schematic shown in Figure 1(A), bottom). Importantly, this method has been successfully applied in different contexts to interrogate functional sites and conformational arrangements within different types of RNA molecules [20,48] and to capture the effects of

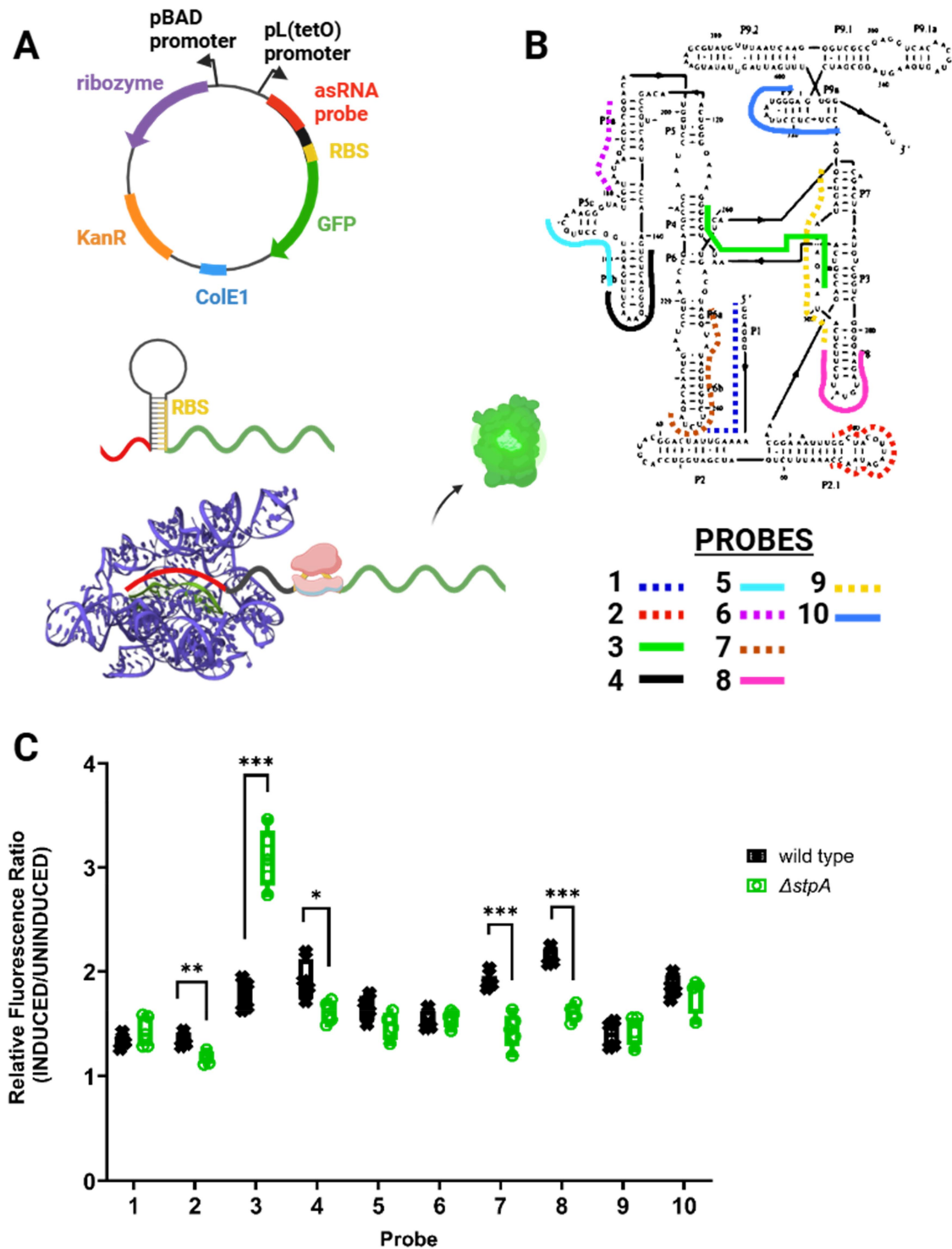


Figure 1. Accessibility mapping captures structural rearrangements of the *tetrahymena* group I intron ribozyme. (A) plasmid map and schematic representation of asRNA probing methodology used to target the ribozyme. The iRS [3] reporter plasmid contains a pBAD promoter followed by the group I intron ribozyme sequence and GFP fused to a user-designed asRNA probe under a TetR-regulated, pL(tetO) promoter. In the iRS [3] approach, asRNA binding alleviates sequestration of the ribosome binding site (RBS) region and allows translation of the GFP reporter. (B) Schematic of the target regions for each asRNA, designed previously (18) to probe the regional accessibility of the ribozyme. Probe 3 (shown in green) is the asRNA probe complementary to both the P3 and P4 domains of the ribozyme. (C) iRS [3] fluorescence shifts corresponding to 10 probes targeting unique regions within the ribozyme expressed in *E. coli* wild type BW25113 and $\Delta stpA$ strains. For each probe, fluorescence ratios were calculated by dividing induced (asRNA probe + ribozyme) by target- uninduced (asRNA probe only) median fluorescence values. Relative fluorescence ratios for each probe are shown in the graph and represent the measured fluorescence for at least 4 independent biological replicates. Asterisks denote statistically significant differences between the relative fluorescence of the iRS [3] system when expressed in the $\Delta stpA$ strain relative to the wild-type parent strain (unpaired t-test; *p-value < 0.05, **p-value < 0.01, ***p-value < 0.001).

protein interactions *in vivo* [20]. Thus, we anticipated that this probing method could detect changes in the folding pathway of a structured target RNA upon disrupting genes encoding RNA chaperones (loss of function) that influence its folding. Specifically, we used the *Tetrahymena* group I intron ribozyme (herein referred to as ribozyme), since the structure and folding of this molecule have been extensively characterized by different *in vivo* and *in vitro* methods in the past [18,21–23]. The ribozyme has a compact native structure, and it tends to misfold into a well-defined, native-like conformation in the absence of RNA chaperones [21]. The native and misfolded structures of the ribozyme have recently been solved using cryo-EM, providing a high-resolution view of structural differences to guide our regional accessibility profiling [24–26].

We first profiled changes in the regional accessibility of the ribozyme upon deletion of the known RNA chaperone StpA, which has been shown to promote annealing and strand displacement of model RNA substrates *in vitro* [10]. We used 10 sequence-specific asRNA probes [18] that target different regions of the ribozyme to profile changes in regional accessibility that could indicate chaperone-mediated structural rearrangements. Two additional control probes representative of the high and low ends of fluorescence were used in every experiment to validate probe induction and fluorescence detection (Supplementary Figure S4) [19]. The fluorescence signal for the 10 asRNA probes was measured when expressed in wildtype *E. coli* and compared to the signal measured in a single-deletion mutant *E. coli* Δ stpA (obtained and verified by genomic PCRs from the Keio collection, [49]). The specific regions of the ribozyme targeted by each probe are shown in Figure 1(B). For each probe, we compared the fluorescence ratio for the Δ stpA strain [the fluorescence value upon induction of the ribozyme relative to an uninduced (probe-only) control] with the analogous ratio for the wild-type strain (Figure 1(C)). Notably, upon deletion of *stpA*, we observed a roughly 75% increase in fluorescence for Probe 3 (p-value <0.001) which targets the P3 and P4 domains of the ribozyme, indicating an StpA-related change in accessibility within this region. These results agree with our previously published data showing that Probe 3 can capture accessibility changes in a well-characterized ribozyme variant that lacks important tertiary contacts and increases solvent exposure of the core [18,22]. The increase in Probe 3 accessibility in the absence of StpA likely results from (1) an increased population of folded conformation(s) with exposure of the complementary nucleotides and/or (2) an increased population of less stable conformations, with transient accessibility occurring during local unfolding events. This region of the ribozyme, which includes the P3 helix, displays increased exposure to footprinting reagents in the known misfolded conformation, and P3 is required to unwind and rewind during refolding from this misfolded conformation to the native state. [21,50]. In these scenarios, during the folding of the ribozyme, we expect the formation of the P3 helix to be the rate-limiting step, favouring hybridization of the asRNA probes targeting this region [51]. Additionally, another well-studied chaperone, CYT-19, accelerates this transition *in vitro* and thereby reduces exposure of this region [52]. We also

observed minor, but significant, drops in the fluorescence signal of Probes 7 and 8 (~25% reductions; p-value <0.001) which target primarily domains P6b and P8, respectively. These helices are on the surface of the native ribozyme structure and may be decreased in accessibility in some non-native conformations. Similarly, more subtle changes (<15% drops, p-value <0.05) at the L2.1 and L5b loops, targeted by Probe 2 and Probe 4, respectively, could be explained by changes in the accessibility of these regions. Importantly, confirmation of the ability to capture established RNA conformational changes upon deletion of the known RNA chaperone StpA served as a proof of concept for a larger screen, motivating the use of this accessibility probing approach to screen for additional general chaperone proteins that contribute to the folding of the ribozyme.

Thus, we designed a Tn5-transposon library and incorporated the iRS [3] system expressing Probe 3 to screen for genes encoding candidate RNA-chaperone proteins. The library was prepared to ensure the coverage of 3× the number of suspected genes in *E. coli* (see Materials and Methods). As illustrated in Figure 2, transposed strains harbouring the reporter Probe 3 plasmid were sorted based on their higher levels of fluorescence relative to wild type *E. coli* using fluorescence-activated cell sorting (FACS). Like when probing in the presence and absence of the StpA chaperone (albeit we did not detect StpA in our screen, see Discussion for possible reasons), we attributed high Probe 3 fluorescence in transposed strains to changes in the folding dynamics of the ribozyme when expressed *in vivo* due to the disruption of a gene that influences RNA folding. After fluorescence sorting, the isolated mutants were subjected to a second screen to validate their fluorescence signal (Supplementary Figure S1, log₂ fold-change cut-off of >0.5). From this second screen, high-fluorescence mutants were pooled, and transposon insertions were identified through whole-genome sequencing (WGS) (see Materials and Methods). Using this approach, 31 unique transposon insertions were mapped within different genes (distributed all throughout the coding sequence of genes; names and insertion coordinates are listed in Table 1) and were identified as potential RNA chaperone candidates. The identified candidate genes encode proteins that participate in a wide variety of cellular processes including cAMP biosynthesis, transmembrane transport, plasmid recombination, amino acid metabolism, and response to metal ions, among others. If these candidates represent true chaperones, it is likely that RNA chaperoning is a ‘moonlighting’ function like for previously described RNA chaperone proteins [4]

RNA accessibility profiling of the tetrahymena group I intron ribozyme in single-gene knockout strains identifies *PepA* and *YagL* as putative RNA chaperones

To further validate candidate genes and rule out multiple transposon insertions in our pooled strains, we performed further accessibility measurements on single-gene knockouts of our candidate chaperone-encoding genes. Specifically, we sought to validate changes in Probe 3 fluorescence (reflecting changes in the regional accessibility at the P3/P4 region of the ribozyme) by transforming the iRS [3] reporter carrying Probe 3 into single-gene knockout strains of the identified candidate genes (Table 1). To conduct these experiments, we cured out the KanR antibiotic marker from

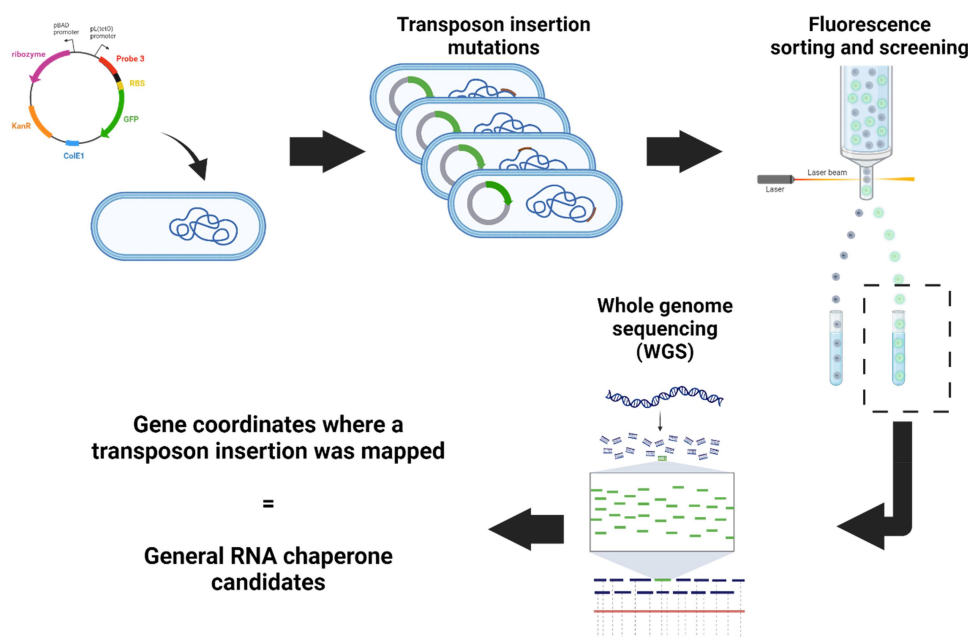


Figure 2. Methodology to uncover candidate proteins that affect RNA folding in vivo. A library of *E. coli* MG1655 cells with random transposon insertions of the TetR cassette was generated by adapting previously published procedures (77). The plasmid containing the iRS [3] system with the probe 3 asRNA was transformed into the library. Cells were sorted based on their GFP expression. High-fluorescence isolates were prepped for whole-genome sequencing, allowing for transposon insertion mapping. Image created with BioRender.com.

single-gene knockout strains in the Keio collection for 16/31 of our candidate genes using FLP recombination and confirmed the removal by colony PCR (Supplementary Table S1). Removing the antibiotic marker ensured that the strains were compatible with the accessibility reporter plasmid, which contains a kanamycin resistance cassette, and reduced the potential metabolic burden of using multiple antibiotics. Of the remaining 15 candidate genes, we were not able to obtain knockout strains for 10

of them (*nrfE*, *nohD*, *tfaD*, *yjjW*, *purK*, *adeD*, *ygfD*, *yjcW*, *xylG*, and *yidT*), and we were not successful at curing out the KanR cassette for five Keio collection strains (*yjll*, *ilvC*, *glgX*, *yjcC* and *panF*).

We validated the expected fluorescence increases, indicating enhanced accessibility for the region targeted by Probe 3, for 8/16 tested strains relative to the wild-type strain (*sbcC*, *pepA*, *yglW*, *yagL*, *argF*, *yifK*, *fxsA* and *yjhE*) (p-value < 0.05; Figure 3). Of these eight genes, we were particularly intrigued

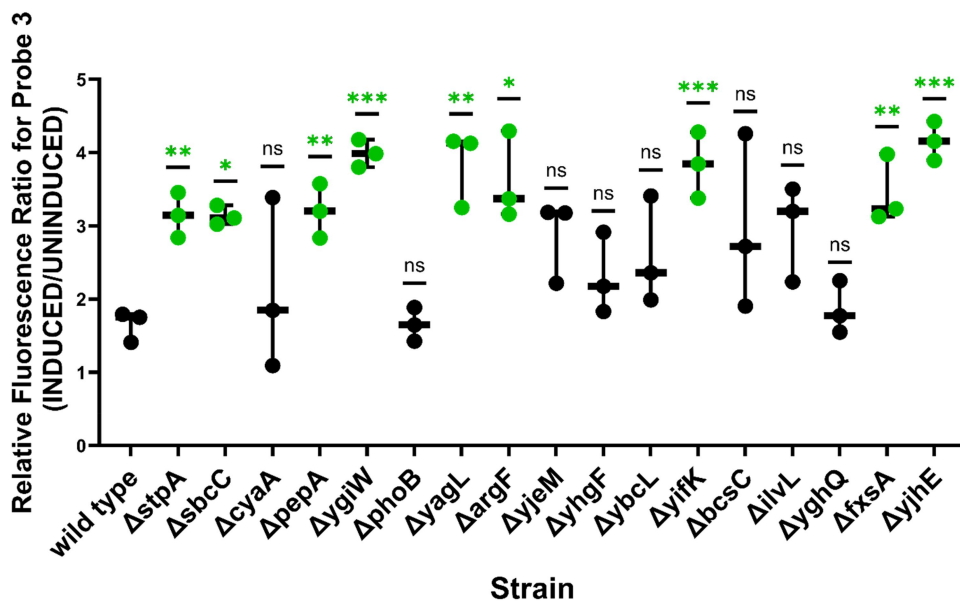


Figure 3. Fluorescence of iRS [3]-targeting of ribozyme region P3/P4 on single-gene knockouts validated 8/16 cellular factors. Fluorescence ratios were calculated by dividing induced (probe 3 asRNA + ribozyme) by target-uninduced (probe 3 asRNA only) median fluorescence values. Individual data points shown in the graph represent the paired induced/uninduced ratio obtained for three independent biological replicates. Asterisks denote significant difference between the fluorescence ratio of probe 3 when expressed on the single-knockout strain relative to when expressed on wildtype *E. coli* BW25113 (unpaired t-test; *p-value < 0.05, **p-value < 0.01, ***p-value < 0.001). *E. coli* Δ stpA was included as a positive control for these experiments.

by *pepA* and *yagL*. The aminopeptidase A protein, PepA, was previously shown to be a multifunctional protein with DNA-binding capabilities [53,54]. Similarly, the *yagL* gene was predicted to encode a DNA-binding recombinase. Because previously characterized general RNA chaperones in *E. coli* (CspA, StpA) are DNA-binding proteins that moonlight as RNA remodellers, we pursued investigations of the roles of PepA and YagL in RNA folding. We excluded SbcCD subunit C (SbcC) from further studies because *in vitro* reconstitution of this complex for biochemical confirmation studies presented a challenge given that this protein forms a complex with subunit D (SbcD) for its DNA nuclease function [55]. Additionally, we chose not to prioritize the uncharacterized genes *yifK*, *fxsA* and *yjhE* for further investigation. For *cyaA*, *bcsC*, and *ilvL* great variation between biological replicates in our fluorescence assay prevented us from validating the high-throughput results. These could be interesting candidates for future work evaluating chaperone activity using a different methodology.

To better understand the influence of PepA and YagL on the folding of the ribozyme, we evaluated the accessibility of additional regions using the full set of asRNA probes for this target [18,56]; targeted regions are listed in Table 2]. As shown in Figure 4(A), when using additional ribozyme-specific asRNA probes on a $\Delta pepA$ strain expressing the *Tetrahymena* ribozyme, we observe a similar fluorescence pattern to that generated on the $\Delta stpA$ strain (results shown in Figure 1). Specifically, Probe 3 (targeting the P3/P4 domain) shows significantly higher fluorescence in the absence of the PepA protein (~150% increase, p-value <0.001). We also observed modest but significant drops in the fluorescence of Probe 7 and Probe 8 [~20% (p-value <0.05) and ~40% (p-value <0.01) reductions] which target the P6b helix and the adjacent tetraloop-receptor (J6a/6b) and the L8 loop, respectively. A minor increase in fluorescence was also detected for the regions targeted by Probe 6 (~30% increase, p-value <0.001). Since the fluorescence changes are not uniform for the different probes, we infer that the increase in fluorescence observed when using Probe 3 is due at least in part to structural changes that result in increased accessibility in the region targeted by this probe. This was further validated via Northern blotting by probing transcript levels of each asRNA using a GFP-specific probe, which ruled out intracellular probe concentration as a contributing factor to the observed changes in fluorescence (Supplementary Figure S5).

When we measured the fluorescence of the ribozyme-specific asRNA probes on the $\Delta yagL$ strain expressing the ribozyme (results shown in Figure 4(B)), we obtained a similar fluorescence change for Probe 3 (targeting the P3 and P4 domains) to that observed for the $\Delta stpA$ and $\Delta pepA$ strains, relative to the wild-type strain (~98% increase, p-value <0.01). Additionally, we observed unique changes in fluorescence, relative to the wild-type strain, for Probe 5 and Probe 10 (targeting the P5c hairpin and the 3'

end/splice site region, respectively). For Probe 5, we observed a ~38% increase in GFP reporter fluorescence (p-value <0.001). This probe targets P5c, part of the P5abc domain, which stabilizes the catalytic core of the ribozyme by forming tertiary contacts and forms rapidly

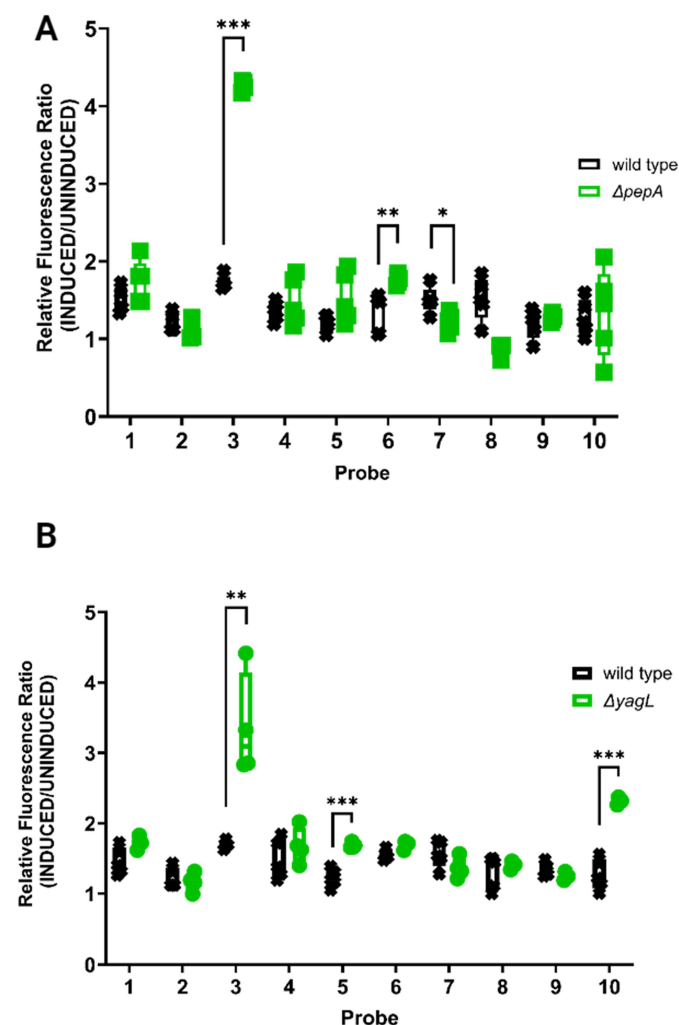


Figure 4. Protein-mediated changes on the accessibility profile of the *Tetrahymena* ribozyme suggest RNA chaperone role for PepA and YagL. (A) Fluorescence of iRS [3]-targeting of the *Tetrahymena* ribozyme when expressed in *E. coli* wild type BW25113 and $\Delta pepA$. For each probe, fluorescence ratios were calculated by dividing paired induced (asRNA probe + ribozyme) by target-uninduced (asRNA probe only) median fluorescence values. Individual data points shown in the graph ('□'. $\Delta pepA$ and 'x' -wild type) represent the obtained values for independent biological quadruplets. (B) Ribozyme accessibility profile captured by the iRS [3] assay when expressed in *E. coli* wild type and $\Delta yagL$. For each probe, fluorescence ratios were calculated by dividing induced (asRNA probe + ribozyme) by uninduced (asRNA probe only) median fluorescence values for independent biological quadruplets (depicted as 'O' - $\Delta yagL$ and 'x' -wild type in the graph). Asterisks denote significant difference in the fluorescence ratio for observed in the mutant strain ($\Delta pepA$ or $\Delta yagL$) relative to that of the wild-type strain (unpaired t-test; *p-value < 0.05, **p-value < 0.01, ***p-value < 0.001).

Table 2. Regions targeted by the 10 asRNA probes used to profile the *tetrahymena* gl intron ribozyme. Ten asRNA probes previously published in sowa et al. (2015) (18) were used to measure changes in regional accessibility for specific regions along the ribozyme transcript.

Probe 1	P1 (5' end post-splicing)	Probe 5	P5c (catalytic activity)	Probe 9	P9.2 (guides folding of the core)
Probe 2	L2.1 (tertiary contact)	Probe 6	P5-a (A-rich bulge)	Probe 10	P10 (3' end/splice site)
Probe 3	P3-J3/4 (3' splicing/catalytic region)	Probe 7	P6b (tetraloop receptor)		
Probe 4	L5b (catalytic activity)	Probe 8	L8 (intron processing)		

during folding of the ribozyme [57]. Similarly, Probe 10 shows an ~82% increase in fluorescence (p-value < 0.001). Probe 10 targets the 3' end and the P9 domain of the ribozyme, which participate in peripheral interactions and may be increased in accessibility in the misfolded conformation and during refolding to the native state [24–26,58]. We also confirmed that fluorescence shifts are not due to differences in transcript levels of the accessibility probes across strains via Northern blotting analysis using a GFP-specific radiolabeled probes (Supplementary Figure S5). We note that the Northern blots showed a significant reduction in transcript levels for the iRS [3] probe in the $\Delta yagL$ strain. However, we concluded that the observed increases in fluorescence signal from the accessibility probing assay are likely due to changes in the abilities of the probes to hybridize to their target regions; since a reduction in probe levels would only reduce the magnitude of the observed changes, no significant reductions in fluorescence were observed for any of the probed regions.

Together, these results isolated PepA and YagL as two proteins that are likely capable of remodelling RNA *in vivo*, as determined by changes in accessibility patterns of key regions within the ribozyme upon their deletion. We did not perform further tests on other genes identified via the high throughput screen, but those genes represent promising candidates for future work.

YagL accelerates native refolding of the tetrahymena group I intron ribozyme in vitro from a misfolded state

To investigate how PepA and YagL affect the folding of ribozyme, we tested the ability of these proteins to accelerate native refolding from the long-lived misfolded conformation. To measure refolding, we used a two-step, or ‘discontinuous’, folding assay in which catalytic activity is used to monitor native ribozyme folding (Figure 5(A)). In this experiment, the ribozyme is pre-incubated with Mg^{2+} to generate the misfolded conformation. Then, during the first assay step, the ribozyme is allowed to refold to the native

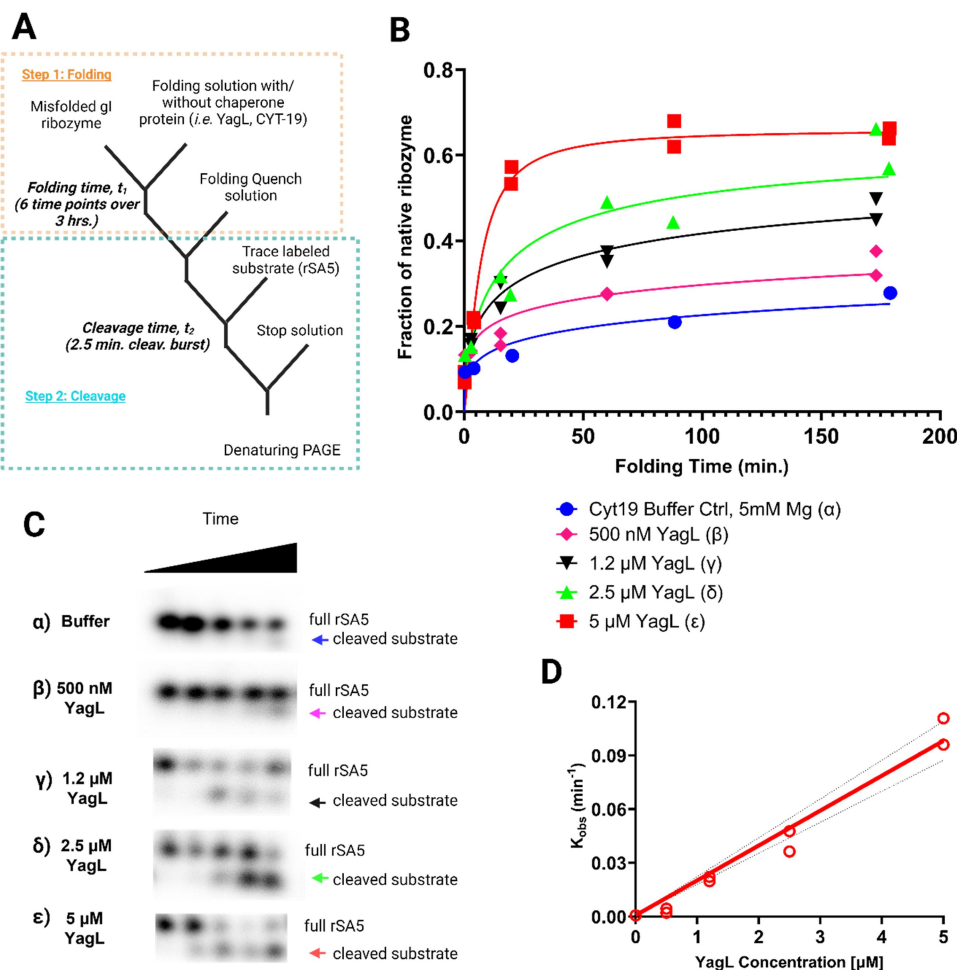


Figure 5. Acceleration of native ribozyme folding by YagL. (A) The *Tetrahymena* gl intron ribozyme was pre-incubated into a misfolded state and then added into folding reactions without YagL (a), with 500 nM YagL (b), 1.2 μ M YagL (c), 2.5 μ M YagL (d), or 5 μ M YagL (e). Reactions were stopped at different times, after which radiolabeled rSA5 (which mimics the 5'-splice-site junction cleaved by the ribozyme) was added to perform substrate cleavage reactions. After quenching, reactions were stopped and analysed by denaturing PAGE to quantify product formation. (B) The fraction of cleaved substrate was quantified and used to generate plots of the fraction of native ribozyme as a function of folding time. From these plots, average rate constants from two independent determinations of native state formation were 0.0008 min^{-1} (a), 0.0032 min^{-1} (b), 0.0210 min^{-1} (c), 0.0421 min^{-1} (d), and 0.100 min^{-1} (e). The faster reactions gave end points of 0.66, and end points for the slower reactions were forced to this value. Data points for individual determinations are provided on graph, curves represent the best fit results. (C) Gel images used to quantify cleavage product formation (these images were obtained for one out of the two independent determinations performed; raw and processed images for all replicates can be found in the supplementary information- appendix 3 for this publication). The cleaved substrate, indicated by the arrows, increases with folding time and with YagL concentration. (D) Observed rate constants from (B) plotted against YagL concentration. Dotted black lines denote the 95% confidence intervals. From this linear fitting, the second-order rate constant K_{cat}/K_M was determined to be $\sim 1.9 \times 10^4 \text{M}^{-1} \text{min}^{-1}$ in the presence of 5 mM Mg^{2+} .

state in the presence of various concentrations of chaperone protein (or no chaperone protein as a control), with reaction aliquots stopped and removed at various times. In the second step, the fraction of native ribozyme is determined for each of these time points by measuring the fraction of an added oligonucleotide substrate that is rapidly cleaved by the native ribozyme [21,59] [44].

Using this assay, we found that purified YagL from *E. coli* accelerates native refolding of the ribozyme in a concentration-dependent manner (Figure 5(B-C)). In the absence of YagL, refolding occurred slowly, on the timescale of hours, consistent with previous work [21,59]. We plotted the observed rate constants against YagL concentration, obtaining a k_{cat}/K_M value of $1.9 \pm 0.1 \times 10^4 M^{-1} min^{-1}$ (Figure 5(D)). This value is comparable to that for the established CYT-19 RNA chaperone under similar experimental conditions [60]. The data also revealed the possibility of upward curvature in the concentration dependence, which might indicate the cooperative involvement of multiple protein molecules. Importantly, in this assay, YagL was inactivated by proteinase K after the folding step and before the substrate cleavage step. Thus, the experiment demonstrates that YagL functions as a chaperone in this folding reaction, via increasing the rate constant for ribozyme refolding to the native state, not via direct modulation of the catalytic activity of the ribozyme.

In contrast, purified PepA did not accelerate ribozyme refolding compared to the control reaction lacking PepA (Supplementary Figure S6). We did not evaluate whether PepA plays other chaperone roles such as destabilization of local structures within the ribozyme. Future work should further investigate the functional role and relevant conditions in which PepA interacts productively with RNA substrates.

PepA and YagL are RNA-binding proteins that bind single-stranded and double-stranded RNA

Previously characterized general chaperones in *E. coli* (such as CspA and StpA) bind RNA with modest affinities in the μM range. Further, mutations that increase the RNA-binding affinity of StpA resulted in reduced chaperone activity (as measured by the ability of StpA to promote cis-splicing of the *td* intron) [10], suggesting that there may be an optimal affinity range for RNA chaperone activity.

Although DNA-binding activity has been shown for PepA [27,61] and predicted for YagL [62], [63] their ability to directly bind RNA was unknown. Therefore, we used nitrocellulose filter binding to measure the binding of these proteins to three 5'-[31P]-labelled short RNA oligonucleotides that were used previously to evaluate StpA-RNA binding (10). Specifically, we used two random 21-mer RNA sequences ('21 R+' and its complementary sequence "21 R-") to evaluate binding to ssRNA, and we used a short hairpin loop (herein referred to as 'hairpin' for simplicity) to evaluate binding to a dsRNA. As shown in Figure 6(A-C), PepA bound both 21 R+ and 21 R- with affinities in the low μM range ($2.0 \pm 0.6 \mu M$ and $0.6 \pm 0.3 \mu M$, respectively), demonstrating that this

protein is capable of binding ssRNA. Further, PepA bound to the hairpin oligonucleotide with an estimated K_D value of $0.6 \pm 0.4 \mu M$, suggesting that it can bind both ssRNA and dsRNA without an apparent substrate preference. Likewise, YagL bound RNA in the low μM range, with a detectable preference for dsRNA. YagL bound both 21 R+ and 21 R- with similar affinities ($4.5 \pm 0.3 \mu M$ and $5.0 \pm 1.2 \mu M$, respectively), and it bound the hairpin twofold more tightly ($2.4 \pm 0.3 \mu M$; Figure 7(C)). For the ssRNAs, the binding curves displayed hints of sigmoidal behaviour, raising the possibility of cooperative binding, but the fitted Hill coefficients were close to one, and for simplicity we fit these curves using a simple, hyperbolic-binding equation (Figure 7(C)).

These results suggest that PepA and YagL modestly interact with RNA (μM range), like the general *E. coli* RNA chaperones StpA and CspA. For contrast, StpA was shown to bind short ssRNA with a K_D value of ~ 580 nM in filter-binding assays and to bind structured RNAs with lower affinity (ranging from 12.3 to 24.7 μM depending on the RNA substrate) in isothermal titration calorimetry (ITC) measurements (10, 59). Similarly, CspA has been reported to interact with its natural partner ACB (anti-cold box) with a K_D value of $\sim 12 \mu M$ [8].

A structurally conserved helix-turn-helix (HTH) domain mediates YagL interactions with nucleic acids

While PepA is a moonlighting protein whose multiple functions have been documented in the literature [27,64] and whose crystal structure has been solved [65], little was known about the structure and function of YagL. After identifying YagL as a novel RNA chaperone in *E. coli* and confirming its ability to both accelerate refolding of the ribozyme and interact with ssRNA and dsRNA substrates *in vitro*, we wanted to further investigate the structural features of YagL that could explain its RNA chaperone role.

YagL was an uncharacterized, 27.2 kDa protein predicted to function as a DNA recombinase [62]. To determine an initial protein structural model for YagL and to identify structurally similar proteins, we subjected its full sequence to homology modelling via Phyre2. Using this approach, YagL shows structural similarity to resolvase proteins (Supplementary Information - Appendix 1). In parallel, we used deep learning protein structure predictions to generate a model for YagL via AlphaFold [66]. Notably, this approach (in addition to homology modelling predictions via Phyre2) led to the identification of a helix-turn-helix (HTH) domain in the C-terminus of YagL (Supplementary Figure S7, panels A&B). The HTH domain found in YagL is predominantly composed of highly conserved, positively charged residues that could mediate charge-charge interactions with RNA substrates (Supplementary Figure S7, panels D&F). This binding mode is supported by additional molecular docking simulations that identified residues like Lys, Arg, His, and Trp in the HTH domain of YagL as DNA-interfacing residues that could also mediate RNA binding (Supplementary Figure S8, Supplementary Information - Appendix 2). In these simulations, 6 DNA ligands (Supplementary Table S3) collected from homologous protein structures identified in the Phyre2

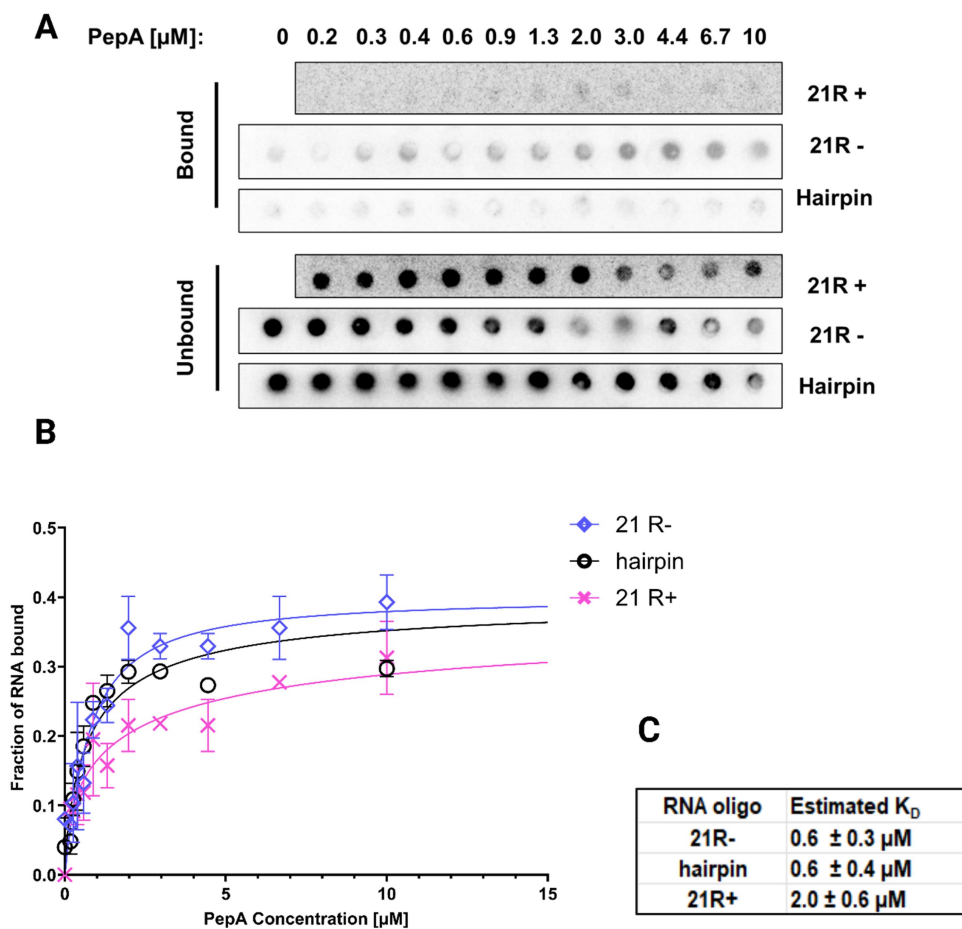


Figure 6. PepA binds both ssRNA and dsRNA. Filter binding assays were performed to test binding of PepA to two single-stranded 21mer RNA sequences: '21 R+' (5'-AUGUGGAAAUCUCUAGCAGU-3') and '21 R-' (5'-CUGCUAGAGAUUUUCCACAU-3'). A short hairpin loop oligo (5'-GCTCTAGAGCATTATGTTTCAGATAAGG-3') was used to evaluate binding to small structured RNAs. (A) Representative membrane images of the bound and unbound signals. Filter binding experiments were conducted in experimental duplicates for each reaction. (B) Non-linear fits were used to generate binding curves for PepA and the three RNA substrates ('21+', '21-', and 'hairpin'). Error bars represent the variation between the experimental replicates. (C) Estimated K_D values for each substrate.

analysis were used as a proxy to investigate RNA binding (as many of the amino acid residues responsible for binding these biomolecules overlap [67–69]).

HTH domains are becoming increasingly identified in multi-functional proteins that bind to nucleic acid substrates (e.g., transcription factors and proteins of the La domain family which serve an RNA chaperone role in eukaryotes [70]). Thus, we decided to evaluate if this predicted HTH domain in YagL was responsible for its RNA-binding and RNA chaperone roles by creating two YagL protein truncations; YagL-dHTH, which included the N-terminal domain of the protein minus the predicted HTH domain, and YagL-HTH, which comprised the predicted C-terminal HTH domain (Figure 8(A), detailed sequences, are included in Supplementary Table S4).

We performed filter-binding assay experiments using the purified YagL-dHTH and YagL-HTH protein truncations. As shown in Figure 8 panels B&C, the predicted HTH domain alone was able to bind RNA with estimated affinities between 3 and 6 μM depending on the RNA substrate used. However, unlike the full YagL protein, YagL-HTH displayed a reversed binding preference relative to the full-length YagL, with twofold tighter binding for the ssRNAs. Additionally, there was pronounced upward curvature in

the binding curves for all the RNAs, giving Hill coefficients of approximately two and indicating cooperative binding. For the YagL-dHTH protein, minimal oligonucleotide binding was observed at any tested protein concentration (Figure 8 panels D&E), suggesting that the C-terminus region of YagL, encoding the predicted HTH domain, is responsible for most of the RNA-binding activity of YagL.

To evaluate if the predicted HTH domain was also responsible for the RNA chaperone activity of YagL, its ability to accelerate native refolding of the ribozyme was assessed using the two-step folding assay described above. As shown in Figure 9, only the full YagL protein was capable of significantly accelerating the refolding of the ribozyme. Thus, we conclude that while the C-terminus HTH domain of YagL is sufficient to bind RNA substrates, the full protein is needed for its RNA chaperone function.

Discussion

The diverse functions of RNAs depend on their folding into precise, native structures [71]. General RNA chaperones allow folding to occur in biologically relevant time scales by

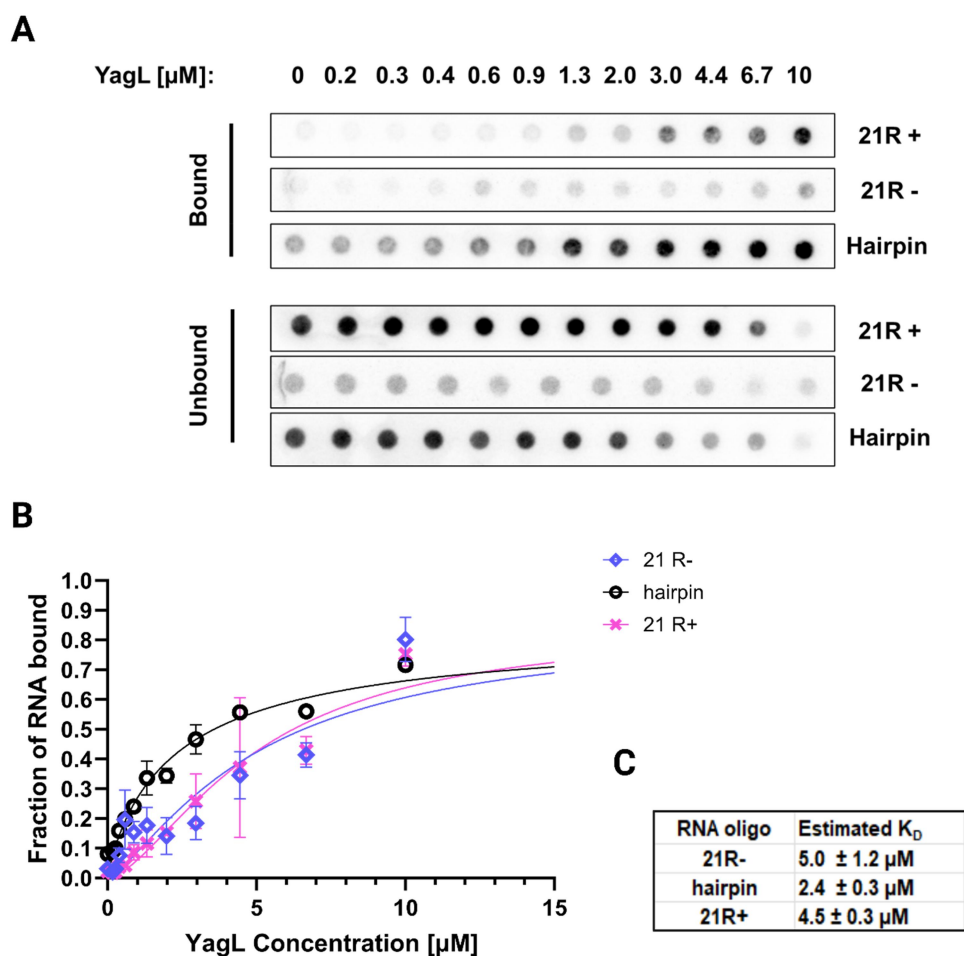


Figure 7. YagL binds preferentially to dsRNA. filter binding assays were performed to test binding of YagL to two single-stranded 21mer RNA sequences: '21 R+' (5'-AUGUGGAAAUCUCUAGCAGU-3') and '21 R-' (5'-CUGCUAGAGAUUUUCCACAU-3'). A short hairpin loop oligo (5'-GCTCTAGAGCATTATGTTTCAGATAAGG-3') was used to evaluate binding to small structured RNAs. (A) Representative membrane images of the bound and unbound signals. Filter binding experiments were conducted in experimental duplicates for each reaction. (B) Non-linear fits were used to generate binding curves for YagL and the three RNA substrates ('21+', '21-', and 'hairpin'). Error bars represent the variation between the experimental replicates. (C) Estimated K_D values for each substrate.

interacting transiently with RNA substrates and resolving non-native intermediate structures [72,73]. However, the lack of a shared structural or sequence motif has limited their discovery and characterization. Here, we repurposed a regional accessibility assay (iRS [3]) to identify 31 candidate RNA chaperone proteins (Table 1), and we validated *in vivo* effects upon gene deletion for eight of them (*sbcC*, *pepA*, *yglW*, *yagL*, *argF*, *yifK*, *fxsA* and *yjhE*) (Figure 3). We further purified and characterized two of them, YagL and PepA, demonstrating RNA-binding activity for both proteins (Figure 6 and Figure 7) and *bona fide* RNA chaperone activity for YagL (Figure 5). For both the *in vivo* and *in vitro* probing of chaperone activity, we used the well-characterized *Tetrahymena* group I intron ribozyme as an example of a highly structured RNA that is prone to misfolding and benefits from general RNA chaperone activity. Thus, we anticipate that the chaperone action of these proteins will extend to other structured RNAs that require assistance in their folding.

The eight candidate RNA chaperone proteins identified in this study (*sbcC*, *pepA*, *yglW*, *yagL*, *argF*, *yifK*, *fxsA* and *yjhE*) add to the growing list of proteins with demonstrated (or

hypothesized) chaperone activity. The large number of proteins with general chaperone activity may reflect that this activity can arise simply from preferential protein binding and thereby stabilization of unfolded or partially unfolded RNA intermediates. Consistent with this idea, RNA chaperone proteins frequently have other functions and act secondarily as chaperones, a role termed 'moonlighting' [74]. Indeed, among the eight chaperone candidates for which we validated ribozyme accessibility increases *in vivo*, we identified several genes encoding predicted and/or known DNA-binding proteins (i.e. *pepA*, *sbcC*, *yagL*), and proteins whose functions include DNA-binding have been shown previously to be good candidates for moonlighting as general RNA chaperones [75,76].

It is notable that many known RNA chaperone proteins were not detected in our high-throughput screen. Most generally, lack of detection of a chaperone protein could result from functional redundancy with respect to ribozyme accessibility – i.e. the two proteins facilitate the same structural transitions that give changes in accessibility, and even with one gene deleted, the activity conferred by the other protein is present at a saturating level. Note that such a result would not

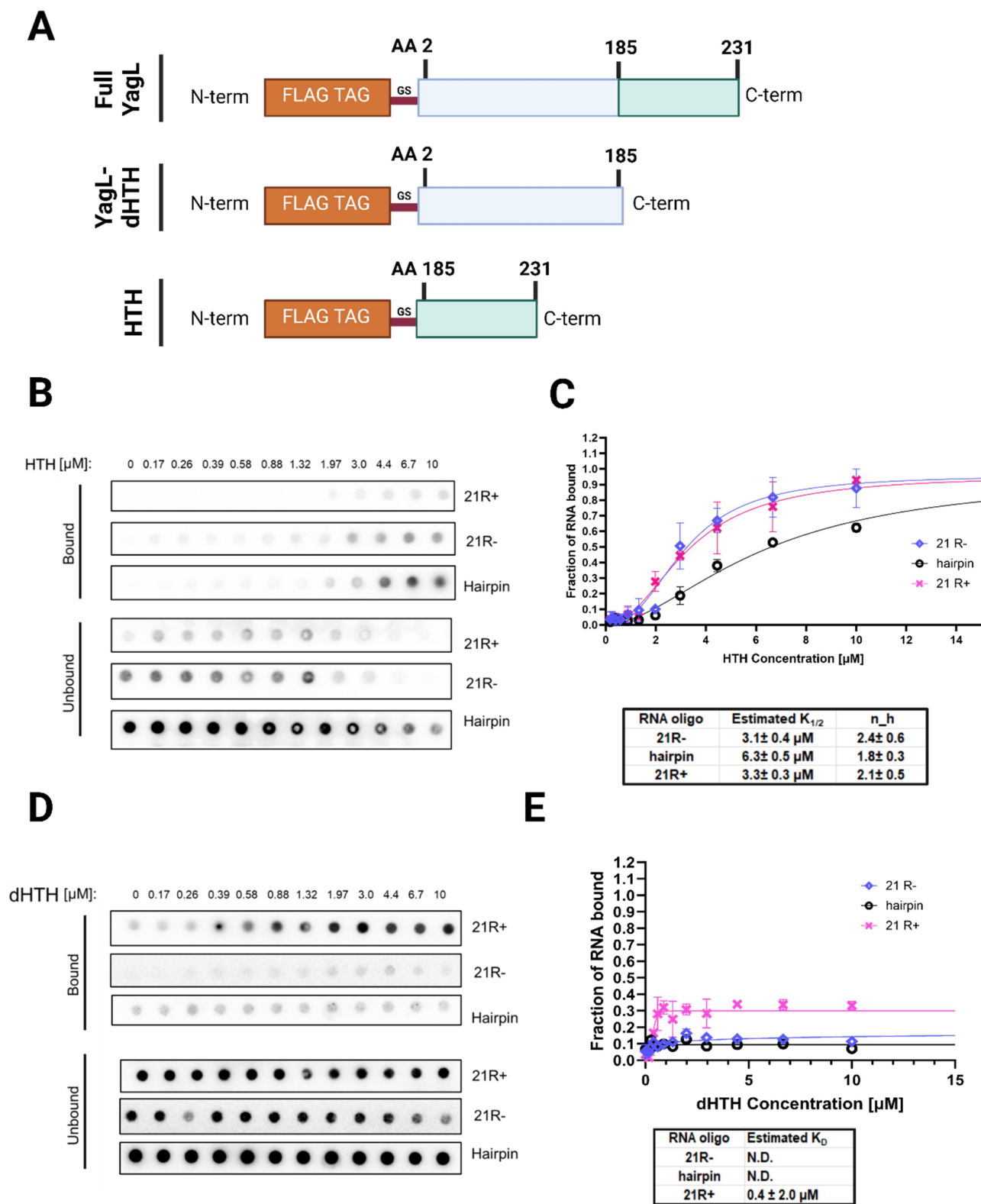


Figure 8. The predicted HTH domain of YagL is responsible for its RNA-binding activity. (A) schematic visualization of the two YagL protein truncations constructed to determine their contribution to RNA-binding. YagL-dHTH includes the N-terminus domain of the protein minus a predicted HTH domain located at the C-terminus of the protein. HTH consists of the last 46 amino acids of YagL. Nucleotide sequences of these protein truncations are included in supplementary table S4. (B&D) filter binding assays were performed to evaluate binding of the HTH and dHTH protein truncations to two single-stranded 21mer RNA sequences, '21 R+' and '21 R-', and a small, structured RNA, 'hairpin'. (C&E) the bound and unbound fractions were quantified to generate binding curves for the HTH and the dHTH truncated protein respectively. For the HTH truncation, sigmoidal curves were obtained, suggesting cooperative binding of ~ 2 functional units (hill coefficients: 2.4 ± 0.6 (21 R-), 1.8 ± 0.3 (hairpin), and 2.1 ± 0.5 (21 R+)). Image created with Biorender.com.

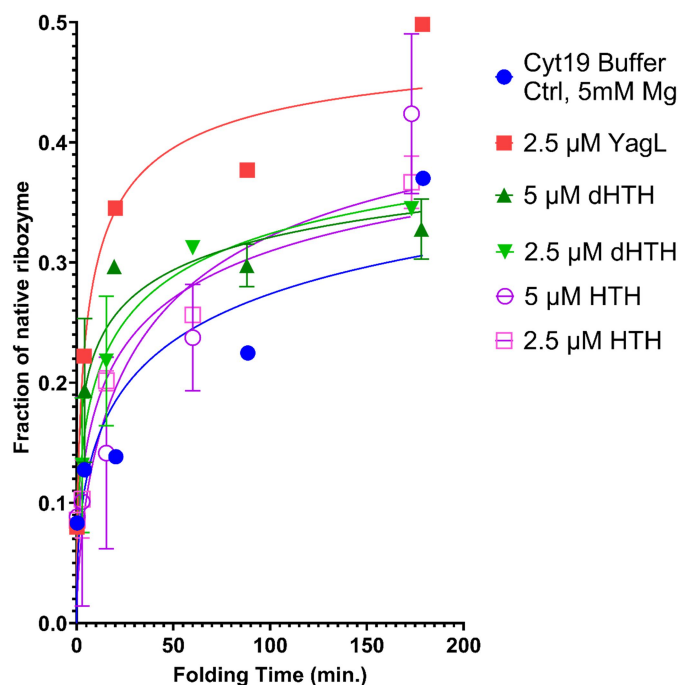


Figure 9. The full YagL protein is required for it to accelerate re-folding of the *Tetrahymena* ribozyme. A two-step catalytic activity assay was performed by pre-incubating the ribozyme into a misfolded state and then adding different concentrations of the ‘dHTH’ or the ‘HTH’ protein truncations. The full length YagL protein was used as a positive control. Reactions were stopped at different times, after which radiolabeled rS45 was added to perform substrate cleavage reactions. The fraction of cleaved product was quantified and used to generate plots of the fraction of native ribozyme as a function of folding time. Error bars represent the variation between experimental duplicates. Only the full YagL protein was capable of significantly accelerating re-folding of the ribozyme, as evidenced by the rapid increase in fraction of native ribozyme at earlier time points (red) compared to the buffer only control (blue).

necessarily indicate that the two proteins are functionally redundant for folding of all RNAs, just the one being probed. It is also possible that a given protein could go undetected under one set of growth conditions while being readily detectable under other growth conditions due to changes in protein level and/or the range or levels of RNA substrates requiring chaperone activity. Additionally, chaperone proteins are unlikely to be detected using our experimental workflow if they make a large contribution to cellular fitness. This is due to sub-culturing and pooling steps preceding sequencing that could cause some mutants to be lost, favouring mutations that are less detrimental to the organism. This might explain why we did not detect the known chaperone StpA in the high-throughput assay, despite finding that an individual deletion of *stpA* resulted in accessibility changes that were similar to those observed upon deletion of *pepA* or *yagL*. Changes in the pooling approach in combination with the implementation of new sequencing approaches, such as long-read sequence, represent good alternatives to overcome current detection limitations.

Our biochemical analysis of YagL and PepA revealed properties that strongly support the hypothesis of general RNA chaperone function for these two proteins. Both proteins bind ssRNA with affinities in the low μM range (Figure 6 and Figure 7) ssRNA-binding activity is expected to be a near-universal feature of RNA chaperones because the ability of chaperones to

accelerate folding transitions requires, by definition, that they bind preferentially and thereby stabilize unfolded intermediates, which are likely to include regions of ssRNA. YagL and PepA also bind dsRNA with similar affinity. This activity may contribute to chaperone activity by trapping helical segments of structured RNA as they transiently unfold from tertiary contacts [77]. Alternatively, or in addition, dsRNA-binding activity may reflect ‘collateral’ effects of the additional functions of these proteins, as PepA binds DNA in Xer site-specific recombination and transcription regulation, and YagL is predicted to function as a resolvase from homology modelling. The relatively modest affinity for RNA appears to be an emerging general property of RNA chaperone proteins, which may reflect the necessity to form a complex that is sufficiently stable to accelerate RNA unfolding but sufficiently transient that it allows for efficient release and subsequent folding of the RNA. Thus, while ATP-dependent chaperones (DEAD-box helicase proteins) use their ATPase activity to cycle between states that have high or low affinity for ssRNA [14,78,79] ATP-independent chaperones may need to ‘thread the needle’ by having intermediate RNA affinity.

Our biochemical experiments also show directly that YagL possesses robust chaperone activity, as it strongly accelerates re-folding of the misfolded *Tetrahymena* ribozyme. This chaperone activity requires the full-length YagL protein, despite the RNA binding and multimerization activities being primarily dependent on the HTH domain (Figure 8). Interestingly, PepA does not possess detectable activity for accelerating this refolding process, yet it does impact accessibility of regions of the ribozyme *in vivo* (those complementary to probes 3, 6, 7 and 8, Figure 4(A)). Likewise, StpA lacks detectable activity for *Tetrahymena* ribozyme refolding (Supplementary Figure S9) despite affecting accessibility of the ribozyme *in vivo* (Figure 1) and previous demonstrations of chaperone activity for another group I intron and model RNA substrates [10,80]. Together, these results highlight that different chaperone proteins have apparently achieved some degree of specialization to the types of RNA folding transitions that they accelerate. This specialization may reflect differences in properties such as RNA binding affinity, specificity, and ability to multimerize; and it probably contributes to our finding that several different chaperone proteins function in *in vivo* folding of a complex, structured RNA like the *Tetrahymena* ribozyme.

Our findings provide initial insights into the chaperone activities of YagL and PepA. In the future work, it will be interesting to further investigate their mechanisms of RNA recognition and to elucidate the repertoire of native substrates for these two proteins, as well as for the other candidates that emerged from the high-throughput screen. In addition, our work illustrates how the iRS [3] assay could be applied to study *in vivo* folding and chaperone activity in the context of other structured RNAs.

Acknowledgments

We would like to thank Jessica Podnar and Dennis Wylie (GSAF core of The University of Texas at Austin) for their help in planning the sequencing strategy for the transposon library. We also thank Richard Salinas (Microscopy and Flow Cytometry core of the University of Texas at Austin) for his continuous advice during the flow cytometry experiments. Authors

would also like to acknowledge Dr Maria Person, Michelle Gadush, and the Mass Spectrometry Facility at the University of Texas at Austin for helping in performing Mass Spectrometry Protein Identification experiments. We acknowledge Dr Jeff Barrick for kindly providing us with the Keio collection deletion strains, Dr Phanourios Tamamis for providing advice and guidance regarding the molecular docking experiments, and Dr Mia Mihailovic for insightful discussions and advice on the iRS³ experiments. The authors would like to acknowledge BioRender.com for assistance with figure creation. Molecular graphics and analyses performed with UCSF ChimeraX, developed by the Resource for Biocomputing, Visualization, and Informatics at the University of California, San Francisco, with support from the National Institutes of Health R01-GM129325 and the Office of Cyber Infrastructure and Computational Biology, National Institute of Allergy and Infectious Diseases.

Funding

This work was supported by the Welch Foundation [F-1756]; NIH [R35 Grant GM131777 to R.R. and R01 Grant GM135495 to L.M.C.]; and the Fulbright Program [Fulbright García-Robles to A.M.R.N.]. Funding for open access charge: Welch Foundation [F-1756].

Data availability statement

All outputs from Phyre2 simulations and HADDOCK outputs are available in Figshare, at 10.6084/m9.figshare.21899445.

Whole-Genome Sequencing data for the strains used in this study is available from the NCBI Sequence Read Archive via BioProject accession number PRJNA923723.

Author contributions

Designed Research: A.M.R.N., L.G.M., L.M.C., R.R.; Performed experiments: A.M.R.N., L.G.M., A.A., A.T.M., S.H.J., E.K.; Analyzed data: A.M.R.N., L.G.M.; Wrote paper: A.M.R.N., L.G.M., L.M.C., R.R.

ORCID

Alejandra Matsuri Rojano-Nisimura  <http://orcid.org/0009-0006-7539-590X>

Lucas G. Miller  <http://orcid.org/0009-0008-0575-6030>

Lydia M. Contreras  <http://orcid.org/0000-0001-5010-5511>

References

- Woodson SA. Recent insights on RNA folding mechanisms from catalytic RNA. *Cell Mol Life Sci.* 2000;57(5):796–808. doi: 10.1007/s000180050042
- Russell R, Millett IS, Tate MW, et al. Rapid compaction during RNA folding. *Proc Natl Acad Sci U S A.* 2002;99(7):4266–4271. doi: 10.1073/pnas.072589599
- Russell R, Zhuang X, Babcock HP, et al. Exploring the folding landscape of a structured RNA. *Proc Natl Acad Sci USA.* 2002;99(1):155–160. doi: 10.1073/pnas.221593598
- Herschlag D. RNA chaperones and the RNA folding problem. *J Biol Chem.* 1995;270(36):20871–20874. doi: 10.1074/jbc.270.36.20871
- Chakrabarti S, Hyeon C, Ye X, et al. Molecular chaperones maximize the native state yield on biological times by driving substrates out of equilibrium. *Proc Natl Acad Sci U S A.* 2017;114(51):E10919–E10927. doi: 10.1073/pnas.1712962114
- Schindelin H, Jiang W, Inouye M, et al. Crystal structure of CspA, the major cold shock protein of *Escherichia coli*. *Proc Natl Acad Sci U S A.* 1994;91(11):5119–5123. doi: 10.1073/pnas.91.11.5119
- Theobald DL, Mitton-Fry RM, Wuttke DS. Nucleic acid recognition by ob-fold proteins. *Annu. Rev. Biophys. Biomol. Struct.* 2003;32(1):115–133. doi: 10.1146/annurev.biophys.32.110601.142506
- Rennella E, Sára T, Juen M, et al. RNA binding and chaperone activity of the *E. coli* cold-shock protein CspA. *Nucleic Acids Res.* 2017;45(7):4255–4268. doi: 10.1093/nar/gkx044
- Bae W, Jones PG, Inouye M. CspA, the major cold shock protein of *Escherichia coli*, negatively regulates its own gene expression. *J Bacteriol.* 1997;179(22):7081–7088. doi: 10.1128/jb.179.22.7081-7088.1997
- Mayer O, Rajkowitz L, Lorenz C, et al. RNA chaperone activity and RNA-binding properties of the *E. coli* protein StpA. *Nucleic Acids Res.* 2007;35(4):1257–1269. doi: 10.1093/nar/gkl1143
- Cusick ME, Belfort M. Domain structure and RNA annealing activity of the *Escherichia coli* regulatory protein StpA. *Mol Microbiol.* 1998;28(4):847–857. doi: 10.1046/j.1365-2958.1998.00848.x
- Coetzee T, Herschlag D, Belfort M. *Escherichia coli* proteins, including ribosomal protein S12, facilitate in vitro splicing of phage T4 introns by acting as RNA chaperones. *Genes Dev.* 1994;8(13):1575–1588. doi: 10.1101/gad.8.13.1575
- Semrad K, Green R, Schroeder R. RNA chaperone activity of large ribosomal subunit proteins from *Escherichia coli*. *RNA.* 2004;10(12):1855–1860. doi: 10.1261/rna.7121704
- Jarmoskaite I, Russell R. RNA helicase proteins as chaperones and remodelers. *Annu Rev Biochem.* 2014;83(1):697–725. doi: 10.1146/annurev-biochem-060713-035546
- Brandi A, Pon CL, Gualerzi CO. Interaction of the main cold shock protein CS7.4 (CspA) of *Escherichia coli* with the promoter region of *hns*. *Biochimie.* 1994;76(10–11):1090–1098. doi: 10.1016/0300-9084(94)90035-3
- Müller CM, Dobrindt U, Nagy G, et al. Role of histone-like proteins H-NS and StpA in expression of virulence determinants of uropathogenic *Escherichia coli*. *J Bacteriol.* 2006;188(15):5428–5438. doi: 10.1128/JB.01956-05
- Uyar E, Kurokawa K, Yoshimura M, et al. Differential binding profiles of StpA in wild-type and *hns* mutant cells: a comparative analysis of cooperative partners by chromatin immunoprecipitation-microarray analysis. *J Bacteriol.* 2009;191(7):2388–2391. doi: 10.1128/JB.01594-08
- Sowa SW, Vazquez-Anderson J, Clark CA, et al. Exploiting post-transcriptional regulation to probe RNA structures in vivo via fluorescence. *Nucleic Acids Res.* 2015;43(2):e13. doi: 10.1093/nar/gku1191
- Leistra AN, Mihailovic MK, Contreras LM. Fluorescence-based methods for characterizing RNA interactions in vivo. *Methods Mol Biol.* 2018;1737:129–164. doi: 10.1007/978-1-4939-7634-8_9
- Vazquez-Anderson J, Mihailovic MK, Baldrige KC, et al. Optimization of a novel biophysical model using large scale in vivo antisense hybridization data displays improved prediction capabilities of structurally accessible RNA regions. *Nucleic Acids Res.* 2017;45(9):5523–5538. doi: 10.1093/nar/gkx115
- Russell R, Herschlag D. Probing the folding landscape of the tetrahymena ribozyme: commitment to form the native conformation is late in the folding pathway. *J Mol Biol.* 2001;308(5):839–851. doi: 10.1006/jmbi.2001.4751
- Das R, Kwok LW, Millett IS, et al. The fastest global events in RNA folding: electrostatic relaxation and tertiary collapse of the Tetrahymena ribozyme. *J Mol Biol.* 2003;332(2):311–319. doi: 10.1016/s0022-2836(03)00854-4
- Wan Y, Suh H, Russell R, et al. Multiple unfolding events during native folding of the tetrahymena group I ribozyme. *J Mol Biol.* 2010;400(5):1067–1077. doi: 10.1016/j.jmb.2010.06.010
- Su Z, Zhang K, Kappel K, et al. Cryo-em structures of full-length Tetrahymena ribozyme at 3.1 Å resolution. *Nature.* 2021;596(7873):603–607. doi: 10.1038/s41586-021-03803-w
- Li S, Palo MZ, Pintilie G, et al. Topological crossing in the misfolded Tetrahymena ribozyme resolved by cryo-em. *Proc Natl Acad Sci U S A.* [2022 Sep 13];119(37):e2209146119. doi: 10.1073/pnas.2209146119
- Bonilla SL, Vicens Q, Kieft JS. Cryo-em reveals an entangled kinetic trap in the folding of a catalytic RNA. *Sci Adv.* 2022;8(34):eabq4144. doi: 10.1126/sciadv.abq4144

- [27] Charlier D, Kholti A, Huysveld N, et al. Mutational analysis of *Escherichia coli* PepA, a multifunctional dna-binding aminopeptidase. *J Mol Biol.* 2000;302(2):411–426. doi: [10.1006/jmbi.2000.4067](https://doi.org/10.1006/jmbi.2000.4067)
- [28] Rose RE. The nucleotide sequence of pACYC184. *Nucleic Acids Res.* 1988;16(1):355. doi: [10.1093/nar/16.1.355](https://doi.org/10.1093/nar/16.1.355)
- [29] Goryshin IY, Jendrisak J, Hoffman LM, et al. Insertional transposon mutagenesis by electroporation of released Tn5 transposition complexes. *Nat Biotechnol.* 2000;18(1):97–100. doi: [10.1038/72017](https://doi.org/10.1038/72017)
- [30] Ekdahl AM, Rojano-Nisimura AM, Contreras LM. Engineering toehold-mediated switches for native RNA detection and regulation in bacteria. *J Mol Biol.* 2022;434(18):167689. doi: [10.1016/j.jmb.2022.167689](https://doi.org/10.1016/j.jmb.2022.167689)
- [31] Andrews S. FastQC: a quality control tool for high throughput sequence data [Online]. 2011. Available online at <http://www.bioinformatics.babraham.ac.uk/projects/fastqc/>
- [32] Martin M. CUTADAPT removes adapter sequences from high-throughput sequencing reads. *EMBnet J.* 2011;17(1):10–12. doi: [10.14806/ej.17.1.200](https://doi.org/10.14806/ej.17.1.200)
- [33] Nakagome M, Solovieva E, Takahashi A, et al. Transposon insertion finder (TIF): a novel program for detection of de novo transpositions of transposable elements. *BMC Bioinformatics.* 2014 [cited 2014 Mar 14];15(1):71. doi: [10.1186/1471-2105-15-71](https://doi.org/10.1186/1471-2105-15-71)
- [34] Quinlan AR, Hall IM. Bedtools: a flexible suite of utilities for comparing genomic features. *Bioinformatics.* 2010;26(6):841–842. doi: [10.1093/bioinformatics/btq033](https://doi.org/10.1093/bioinformatics/btq033)
- [35] Rojano-Nisimura AM, Haning K, Janovsky J, et al. Codon selection affects recruitment of ribosome-associating factors during translation. *ACS Synth Biol.* 2020;9(2):329–342. doi: [10.1021/acsynbio.9b00344](https://doi.org/10.1021/acsynbio.9b00344)
- [36] Dubey AK, Baker CS, Romeo T, et al. RNA sequence and secondary structure participate in high-affinity CsrA-rna interaction. *RNA.* 2005;11(10):1579–1587. doi: [10.1261/rna.2990205](https://doi.org/10.1261/rna.2990205)
- [37] Stressler T, Tanzer C, Ewert J, et al. Simple purification method for a recombinantly expressed native his-tag-free aminopeptidase a from *Lactobacillus delbrueckii*. *Protein Expr Purif.* 2017;131:7–15. doi: [10.1016/j.pep.2016.10.010](https://doi.org/10.1016/j.pep.2016.10.010)
- [38] Gibson DG, Young L, Chuang RY, et al. Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nat Methods.* 2009;6(5):343–345. doi: [10.1038/nmeth.1318](https://doi.org/10.1038/nmeth.1318)
- [39] Mihailovic MK, Ekdahl AM, Chen A, et al. Uncovering transcriptional regulators and targets of sRNAs using an integrative data-mining approach: H-NS-Regulated RseX as a case study. *Front Cell Infect Microbiol.* 2021 [cited 2021 Jul 13];11:696533. doi: [10.3389/fcimb.2021.696533](https://doi.org/10.3389/fcimb.2021.696533)
- [40] Zhang F, Ramsay ES, Woodson SA. In vivo facilitation of *Tetrahymena* group I intron splicing in *Escherichia coli* pre-ribosomal RNA. *RNA.* 1995;1(3):284–292.
- [41] Rio DC. Filter-binding assay for analysis of RNA-protein interactions. *Cold Spring Harb Protoc.* 2012 [cited 2012 Oct 1];2012(10):1078–1081. doi: [10.1101/pdb.prot071449](https://doi.org/10.1101/pdb.prot071449)
- [42] Altschuler SE, Lewis KA, Wuttke DS. Practical strategies for the evaluation of high-affinity protein/nucleic acid interactions. *J Nucleic Acids Investig.* 2013;4(1):19–28. doi: [10.4081/jnai.2013.e3](https://doi.org/10.4081/jnai.2013.e3)
- [43] Gonzalez-Rivera JC, Orr AA, Engels SM, et al. Computational evolution of an RNA-binding protein towards enhanced oxidized-RNA binding. *Comput Struct Biotechnol J.* 2019 [cited 2019 Dec 27];18:137–152. doi: [10.1016/j.csbj.2019.12.003](https://doi.org/10.1016/j.csbj.2019.12.003)
- [44] Potratz JP, Russell R. RNA catalysis as a probe for chaperone activity of dead-box helicases. *Methods Enzymol.* 2012;511:111–130. doi: [10.1016/B978-0-12-396546-2.00005-X](https://doi.org/10.1016/B978-0-12-396546-2.00005-X)
- [45] Gracia B, Russell R. RNA catalytic activity as a probe of chaperone-mediated RNA folding. *Methods Mol Biol.* 2014;1086:225–237. doi: [10.1007/978-1-62703-667-2_13](https://doi.org/10.1007/978-1-62703-667-2_13)
- [46] Bhaskaran H, Russell R. Kinetic redistribution of native and misfolded RNAs by a dead-box chaperone. *Nature.* 2007;449(7165):1014–1018. doi: [10.1038/nature06235](https://doi.org/10.1038/nature06235)
- [47] Russell R, Herschlag D. New pathways in folding of the tetrahymena group I RNA enzyme. *J Mol Biol.* 1999;291(5):1155–1167. doi: [10.1006/jmbi.1999.3026](https://doi.org/10.1006/jmbi.1999.3026)
- [48] Leistra AN, Amador P, Buvanendiran A, et al. Rational modular RNA engineering based on in vivo profiling of structural accessibility. *ACS Synth Biol.* 2017;6(12):2228–2240. doi: [10.1021/acsynbio.7b00185](https://doi.org/10.1021/acsynbio.7b00185)
- [49] Baba T, Ara T, Hasegawa M, et al. Construction of *Escherichia coli* K-12 in-frame, single-gene knockout mutants: the Keio collection. *Mol Syst Biol.* 2006;2(1):2006.0008. doi: [10.1038/msb4100050](https://doi.org/10.1038/msb4100050)
- [50] Mitchell D 3rd, Jarmoskaite I, Seval N, et al. The long-range P3 helix of the tetrahymena ribozyme is disrupted during folding between the native and misfolded conformations. *J Mol Biol.* 2013;425(15):2670–2686. doi: [10.1016/j.jmb.2013.05.008](https://doi.org/10.1016/j.jmb.2013.05.008)
- [51] Zarrinkar PP, Williamson JR. Kinetic intermediates in RNA folding. *Science.* [1994 Aug 12];265(5174):918–924. doi: [10.1126/science.8052848](https://doi.org/10.1126/science.8052848)
- [52] Mohr S, Stryker JM, Lambowitz AM. A dead-box protein functions as an atp-dependent RNA chaperone in group I intron splicing. *Cell.* 2002;109(6):769–779. doi: [10.1016/S0092-8674\(02\)00771-7](https://doi.org/10.1016/S0092-8674(02)00771-7)
- [53] Nguyen Le Minh P, Nadal M, Charlier D. The trigger enzyme PepA (aminopeptidase A) of *Escherichia coli*, a transcriptional repressor that generates positive supercoiling. *FEBS Lett.* 2016;590(12):1816–1825. doi: [10.1002/1873-3468.12224](https://doi.org/10.1002/1873-3468.12224)
- [54] Commichau FM, Stülke J. Trigger enzymes: bifunctional proteins active in metabolism and in controlling gene expression. *Mol Microbiol.* 2008;67(4):692–702. doi: [10.1111/j.1365-2958.2007.06071.x](https://doi.org/10.1111/j.1365-2958.2007.06071.x)
- [55] Connelly JC, de Leau Es, Leach DR. DNA cleavage and degradation by the SbcCD protein complex from *Escherichia coli*. *Nucleic Acids Res.* 1999;27(4):1039–1046. doi: [10.1093/nar/27.4.1039](https://doi.org/10.1093/nar/27.4.1039)
- [56] Mihailovic MK, Vazquez-Anderson J, Li Y, et al. High-throughput in vivo mapping of RNA accessible interfaces to identify functional sRNA binding sites. *Nat Commun.* 2018 [cited 2018 Oct 4];9(1):4084. doi: [10.1038/s41467-018-06207-z](https://doi.org/10.1038/s41467-018-06207-z)
- [57] Zheng M, Wu M, Tinoco I Jr. Formation of a GNRA tetraloop in P5abc can disrupt an interdomain interaction in the tetrahymena group I ribozyme. *Proc Natl Acad Sci U S A.* 2001;98(7):3695–3700. doi: [10.1073/pnas.051608598](https://doi.org/10.1073/pnas.051608598)
- [58] Zarrinkar PP, Williamson JR. The P9.1-P9.2 peripheral extension helps guide folding of the tetrahymena ribozyme. *Nucleic Acids Res.* [1996 Mar 1];24(5):854–858. doi: [10.1093/nar/24.5.854](https://doi.org/10.1093/nar/24.5.854)
- [59] Russell R, Das R, Suh H, et al. The paradoxical behavior of a highly structured misfolded intermediate in RNA folding. *J Mol Biol.* 2006;363(2):531–544. doi: [10.1016/j.jmb.2006.08.024](https://doi.org/10.1016/j.jmb.2006.08.024)
- [60] Tijerina P, Bhaskaran H, Russell R. Nonspecific binding to structured RNA and preferential unwinding of an exposed helix by the CYT-19 protein, a dead-box RNA chaperone. *Proc Natl Acad Sci USA.* 2006;103(45):16698–16703. doi: [10.1073/pnas.0603127103](https://doi.org/10.1073/pnas.0603127103)
- [61] Charlier D, Hassanzadeh G, Kholti A, et al. Involved in pyrimidine regulation of the *Escherichia coli* carbamoylphosphate synthetase operon encodes a sequence-specific DNA-binding protein identical to XerB and PepA, also required for resolution of ColEI multimers. *J Mol Biol.* 1995;250(4):392–406. doi: [10.1006/jmbi.1995.0385](https://doi.org/10.1006/jmbi.1995.0385)
- [62] Huntley RP, Sawford T, Mutowo-Muullenet P, et al. The GOA database: gene ontology annotation updates for 2015. *Nucleic Acids Res.* 2015 Jan [cited 2014 Nov 6];43(Database issue):D1057–1063. doi: [10.1093/nar/gku1113](https://doi.org/10.1093/nar/gku1113)
- [63] Hohmann KF, Blümler A, Heckel A, et al. The RNA chaperone StpA enables fast RNA refolding by destabilization of mutually exclusive base pairs within competing secondary structure elements. *Nucleic Acids Res.* 2021;49(19):11337–11349. doi: [10.1093/nar/gkab876](https://doi.org/10.1093/nar/gkab876)
- [64] Khan I, Chen Y, Dong T, et al. Genome-scale identification and characterization of moonlighting proteins. *Biol Direct.* [2014 Dec

- 11];9(1):30. doi: [10.1186/s13062-014-0030-9](https://doi.org/10.1186/s13062-014-0030-9) PMID: 25497125; PMCID: PMC4307903.
- [65] Sträter N, Sherratt DJ, Colloms SD. X-ray structure of aminopeptidase a from *Escherichia coli* and a model for the nucleoprotein complex in *xer* site-specific recombination. *Embo J.* 1999;18(16):4513–4522. doi: [10.1093/emboj/18.16.4513](https://doi.org/10.1093/emboj/18.16.4513)
- [66] Jumper J, Evans R, Pritzel A, et al. Highly accurate protein structure prediction with AlphaFold. *Nature.* 2021;596(7873):583–589. doi: [10.1038/s41586-021-03819-2](https://doi.org/10.1038/s41586-021-03819-2)
- [67] Bartas M, Červeň J, Guziurová S, et al. Amino acid composition in various types of nucleic acid-binding proteins. *Int J Mol Sci.* 2021 [cited 2021 Jan 18];22(2):922. doi: [10.3390/ijms22020922](https://doi.org/10.3390/ijms22020922)
- [68] Terribilini M, Lee JH, Yan C, et al. Prediction of RNA binding sites in proteins from amino acid sequence. *RNA.* 2006;12(8):1450–1462. doi: [10.1261/rna.2197306](https://doi.org/10.1261/rna.2197306)
- [69] Zhang J, Ma Z, Kurgan L. Comprehensive review and empirical analysis of hallmarks of DNA-, RNA- and protein-binding residues in protein chains [published correction appears in *brief bioinform. Brief Bioinform.* 2019;20(4):1250–1268. 2020 Sep 25];21(5):1856]. doi: [10.1093/bib/bbx168](https://doi.org/10.1093/bib/bbx168)
- [70] Aravind L, Anantharaman V, Balaji S, et al. The many faces of the helix-turn-helix domain: transcription regulation and beyond. *FEMS Microbiol Rev.* 2005;29(2):231–262. doi: [10.1016/j.femsre.2004.12.008](https://doi.org/10.1016/j.femsre.2004.12.008)
- [71] Breaker RR, Joyce GF. The expanding view of RNA and DNA function. *Chem Biol.* 2014;21(9):1059–1065. doi: [10.1016/j.chembiol.2014.07.008](https://doi.org/10.1016/j.chembiol.2014.07.008)
- [72] Schroeder R, Barta A, Semrad K. Strategies for RNA folding and assembly. *Nat Rev Mol Cell Biol.* 2004;5(11):908–919. doi: [10.1038/nrm1497](https://doi.org/10.1038/nrm1497)
- [73] Russell R. RNA misfolding and the action of chaperones. *Front Biosci.* 2008;2008 Jan 1;13(13):1–20. Published doi:[10.2741/2557](https://doi.org/10.2741/2557)
- [74] Albihlal WS, Gerber AP. Unconventional RNA-binding proteins: an uncharted zone in RNA biology. *FEBS Lett.* 2018;592(17):2917–2931. doi: [10.1002/1873-3468.13161](https://doi.org/10.1002/1873-3468.13161)
- [75] Cassidy LA, Maher LJ 3rd. Having it both ways: transcription factors that bind DNA and RNA. *Nucleic Acids Res.* 2002;30(19):4118–4126. doi: [10.1093/nar/gkf512](https://doi.org/10.1093/nar/gkf512)
- [76] Hudson WH, Ortlund EA. The structure, function and evolution of proteins that bind DNA and RNA. *Nat Rev Mol Cell Biol.* 2014;15(11):749–760. doi: [10.1038/nrm3884](https://doi.org/10.1038/nrm3884)
- [77] Pan C, Potratz JP, Cannon B, et al. Dead-box helicase proteins disrupt RNA tertiary structure through helix capture. *PLoS Biol.* 2014 [cited 2014 Oct 28];12(10):e1001981. doi: [10.1371/journal.pbio.1001981](https://doi.org/10.1371/journal.pbio.1001981)
- [78] Hilbert M, Karow AR, Klostermeier D. The mechanism of atp-dependent RNA unwinding by DEAD box proteins. *Biol Chem.* 2009;390(12):1237–1250. doi: [10.1515/BC.2009.135](https://doi.org/10.1515/BC.2009.135)
- [79] Hoffman LM, Jendrisak JJ, Meis RJ, et al. Transposome insertional mutagenesis and direct sequencing of microbial genomes. *Genetica.* 2000;108(1):19–24. doi: [10.1023/a:1004083307819](https://doi.org/10.1023/a:1004083307819)
- [80] Waldsich C, Grossberger R, Schroeder R. RNA chaperone StpA loosens interactions of the tertiary structure in the td group I intron in vivo. *Genes Dev.* 2002;16(17):2300–2312. doi: [10.1101/gad.231302](https://doi.org/10.1101/gad.231302)