

Lead-seq: transcriptome-wide structure probing *in vivo* using lead(II) ions

Christian Twittenhoff^{1,†}, Vivian B. Brandenburg^{1,†}, Francesco Righetti^{1,†}, Aaron M. Nuss², Axel Mosig³, Petra Dersch^{2,4} and Franz Narberhaus^{1,*}

¹Microbial Biology, Ruhr University Bochum, 44780 Bochum, Germany, ²Department of Molecular Infection Biology, Helmholtz Centre for Infection Research, 381214 Braunschweig, Germany, ³Department of Biophysics, Ruhr University Bochum, 44780 Bochum, Germany and ⁴Institute of Infectiology, Center for Molecular Biology of Inflammation, University of Münster, 48149 Münster, Germany

Received November 22, 2019; Revised April 08, 2020; Editorial Decision May 04, 2020; Accepted May 06, 2020

ABSTRACT

The dynamic conformation of RNA molecules within living cells is key to their function. Recent advances in probing the RNA structurome *in vivo*, including the use of SHAPE (Selective 2'-Hydroxyl Acylation analyzed by Primer Extension) or kethoxal reagents or DMS (dimethyl sulfate), provided unprecedented insights into the architecture of RNA molecules in the living cell. Here, we report the establishment of lead probing in a global RNA structuromics approach. In order to elucidate the transcriptome-wide RNA landscape in the enteric pathogen *Yersinia pseudotuberculosis*, we combined lead(II) acetate-mediated cleavage of single-stranded RNA regions with high-throughput sequencing. This new approach, termed 'Lead-seq', provides structural information independent of base identity. We show that the method recapitulates secondary structures of tRNAs, RNase P RNA, tmRNA, 16S rRNA and the *rpsT* 5'-untranslated region, and that it reveals global structural features of mRNAs. The application of Lead-seq to *Y. pseudotuberculosis* cells grown at two different temperatures unveiled the first temperature-responsive *in vivo* RNA structurome of a bacterial pathogen. The translation of candidate genes derived from this approach was confirmed to be temperature regulated. Overall, this study establishes Lead-seq as complementary approach to interrogate intracellular RNA structures on a global scale.

INTRODUCTION

RNA is a versatile biomolecule with diverse functions, ranging from carrying genetic information to regulatory as well as catalytic activities. Interactions between non-

adjacent nucleotides (nt) establish secondary structures, whereas long-range interactions within the RNA molecule (e.g. kissing loops, tetraloop:receptor interactions or pseudoknots) and inter-molecular interactions (e.g. RNA–RNA or RNA–protein interactions) contribute to the formation of a three-dimensional (3D) structure. Structured regions in RNAs often have an impact on transcription, transcript maturation and degradation as well as translation, and thereby modulate gene expression. Non-coding RNAs (ncRNAs), such as tRNAs, rRNAs, sRNAs and ribozymes, typically fold into stable defined secondary structures. In contrast, many mRNAs harbor regulatory structural motifs that undergo dynamic conformational alterations in response to physical or chemical stimuli. Two known *cis*-encoded RNA regulators are ligand-binding riboswitches (1,2) and RNA thermometers (RNATs) (3). RNATs are frequently located in the 5'-UTR (5'-untranslated region) of bacterial mRNAs, regulating translation of heat shock- or virulence-related genes in response to temperature (4,5). The thermosensor adopts a thermo-labile secondary structure that prevents binding of the ribosome to the Shine-Dalgarno sequence at low temperatures. Upon increasing temperature, the RNAT gradually melts, allowing for ribosome binding and translation initiation (3). The reversibility of the zipper-like mechanism precisely adjusts translation efficiency to the cellular demand (6).

Since the structure and function of RNA molecules are interconnected, the secondary and tertiary architectures of RNAs are of great interest. Several methods are available to inspect folded RNAs. Secondary structures can be predicted computationally (7) by comparative sequence alignment or by predicting the thermodynamically most favorable structure (8). Experimentally, RNA structures can be solved by biophysical methods, such as nuclear magnetic resonance spectroscopy, X-ray crystallography, or cryo-electron microscopy (9–12). A widely used experimental approach is the probing of RNA structures

*To whom correspondence should be addressed. Tel: +49 234 322 3100; Fax: +49 234 321 4620; Email: franz.narberhaus@rub.de

†The authors wish it to be known that, in their opinion, the first three authors should be regarded as Joint First Authors.

in solution with the help of enzymes or chemicals (13). Recently, powerful global or transcriptome-wide structure probing methods have been established. These couple RNA structure probing techniques to next generation sequencing and uncover the RNA structure of any given organism, even at different physiological conditions (14–16). The probes used in these approaches are either nucleases that digest the ribose-phosphate backbone, or chemicals that modify the RNA bases or ribose moiety. Due to their biological and physical properties, both enzymatic and chemical probes have their advantages and limitations (17,18). Enzymatic probes cleave nt either in a sequence-specific (e.g. RNase T1 that preferentially cleaves at unpaired guanines) or in a sequence-unspecific (e.g. Nuclease S1) manner (19). RNase V1 is the only probe that permits mapping of double-stranded nt (20). Due to their bulky structure, RNases may not be able to reach all single- or double-stranded regions and are not able to penetrate cell membranes, which limits their application to *in vitro* approaches (19). In contrast, chemical probes are suitable for *in vivo* probing due to their small size. All presently available RNA modification probes exclusively react with the bases or the backbone of single-stranded regions. For example, DMS (dimethyl sulfate) alkylates the Watson–Crick face of adenine and cytosine residues leading to termination of reverse transcription (21–24). Complementary to DMS-mediated probing, small carbodiimides form adducts with exposed Gs and Us of RNA structures. While the use of the carbodiimide CMCT (*N*-cyclohexy-*N'*(2-morpholinoethyl)carbodiimide metho-*p*-toluene sulfonate) is restricted to *in vitro* applications (19,25), recently, the EDC (1-ethyl-3-(3-dimethylaminopropyl)-carbodiimide) reagent was established for probing of RNA molecules in intact cells (26,27). A most recent report describes the use of N₃-kethoxal for labeling of single-stranded guanine bases in live cells (28). SHAPE (selective 2'-hydroxyl acylation analyzed by primer extension) reagents are less base-specific (29). They acylate the 2'-hydroxy group of ribose in all four nt, if they are single-stranded, and are applicable under cellular as well as cell-free conditions (30–33).

Divalent metal cations, such as lead(II), are another option for RNA structure probing (34). Lead cleavage of single-stranded nt has been used to determine RNA structures *in vitro* (35,36) and *in vivo* (37–39). Recently, lead probing was applied to resolve temperature-sensitive regions in an RNAT that mediates translational regulation of a cyanobacterial heat shock gene (6). Lead induces RNA cleavage by two distinct mechanisms. The first mechanism is rather strong and highly specific. It occurs near or at metal ion binding sites but is rarely observed (40,41). The second, weaker cleavage activity is more frequently detected and rather unspecific. Here, lead ions preferentially attack bulges, loops, and other single-stranded RNA regions, but do not cleave stacked nt or those involved in higher-order interactions (42–44). Both cleavage events are based on the same mode of action. Lead ions interact with the 2'-OH group of the ribose moiety to induce a nucleophilic attack of the 2'-oxygen on the vicinal phosphodiester, which results in 2',3'-cyclic phosphorus and 5'-hydroxyl groups as cleavage products (45,46).

To interrogate the dynamic RNA structure of the extraordinarily temperature-responsive human pathogen *Yersinia pseudotuberculosis*, we recently applied the parallel analysis of RNA structure with temperature elevation (PARTE) approach (47). Nuclease-based probing at 25°C, 37°C and 42°C mimicked environmental, host body and heat shock temperatures, respectively (48). Since this procedure relies on the structure-specific nucleases S1 and V1 (cleaving single- and double-stranded nt, respectively), it is limited to *in vitro* applications. Here, we describe a new method named Lead-seq that is able to chart bacterial RNA structures *in vivo*. Lead(II) acetate-induced RNA cleavage was followed by cDNA synthesis and next-generation sequencing, allowing a transcriptome-wide view on the structural RNA-landscape in *Y. pseudotuberculosis*.

MATERIALS AND METHODS

Strains and plasmids

All utilized bacterial strains, oligonucleotides and plasmids are summarized in Supplementary Tables S1–3, respectively. Correctness of nucleotide sequences was confirmed by automated sequencing (Eurofins Genomics, Martinsried, Germany). Enzymes for cloning were obtained from Thermo Scientific (St. Leon-Rot, Germany). Polymerase chain reaction (PCR), DNA manipulations, DNA restrictions and transformations were performed as described before (49).

Plasmid construction

Plasmids pBO7300 and pBO7301 were constructed as follows: the 5'-UTR of *groES* (YPK_3349) or *clpB* (YPK_3823) including 30 nt of the respective coding region (CDR) were PCR-amplified using primer pairs YPK_3349_fw/YPK_3349_rv or YPK_3823_fw/YPK_3823_rv, respectively. Afterward, the resulting fragments were ligated into pBAD2-*bgab*-His via NheI and EcoRI restriction sites, resulting in reporter plasmids pBO7300 or pBO7301. Repressing and derepressing point mutations were introduced by site-direct mutagenesis.

Determination of ideal lead acetate concentration

To achieve single-hit kinetics for *in vivo* RNA probing, we treated *Escherichia coli* cells (Supplementary Figure S1) with different concentrations of lead(II) acetate at 25°C and 37°C. All lead(II) acetate solutions were prepared freshly. A 1 M lead(II) acetate solution was mixed with 2 ml prewarmed (to 25°C/37°C) 4× LB medium (4 % (w/v) NaCl, 4 % (w/v) tryptone, 2 % (w/v) yeast extract) and distilled water to a final volume of 8 ml. Lead solutions of different concentrations (such as 175, 350 or 525 mM), corresponding to a final concentration of 50, 100 or 150 mM, respectively, were prepared. To determine appropriate lead(II) acetate concentrations for lead-mediated *in vivo* probing, cultures of *E. coli* DH5α (20 ml) were grown at 25°C and 37°C in LB medium (1 % (w/v) NaCl, 1 % (w/v) tryptone, 0.5 % (w/v) yeast extract) to an OD₆₀₀ = 0.5. The

lead solutions were added to the culture and cells were incubated with shaking for 7 min. Lead induced RNA cleavage was stopped by addition of 10 ml cold 500 mM ethylenediaminetetraacetic acid (EDTA) and cultures were placed on ice. Cells were pelleted, resuspended in 250 μ l cold resuspension buffer (10 mM Tris-HCl, 100 mM NaCl, 1 mM EDTA, pH = 8.0) and 250 μ l lysis buffer (50 mM Tris-HCl, 8% (w/v) sucrose, 0.5% (w/v) Triton X-100, 10 mM EDTA, 4 mg/ml lysozyme, pH = 8.0) was added. After incubation at 65°C for 90 s, RNA was obtained from samples by four times extraction with acid phenol and, subsequently, three times extraction with chloroform/isoamylalcohol (24:1). After ethanol precipitation, RNA pellets were resuspended in 50 μ l ddH₂O. RNA concentration was measured with a NanoDrop spectrophotometer ND-1000 (peqlab, Erlangen, Germany). First, total RNA was analyzed by agarose gel electrophoresis (Supplementary Figure S1 and degradation of 23S as well as 16S rRNA was determined by densitometry (AlphaEaseFC program). Second, to determine the extent of lead cleavage-induced degradation degree of the *groES* 5'-UTR:*gfp* mRNA, we performed primer extension (primer *gfp_rev_seq*; Supplementary Table S2). Cultures of *E. coli* DH5 α (20 ml), harboring plasmid pBO2267 (*Salmonella enterica* serovar Typhimurium M556 translational 5'-UTR *groES:gfp*) were grown as described above. For primer extension, 20 μ g RNA from each sample were mixed with 2.5 μ l dNTPs (4 mM), 2 μ l 5' labeled (³²P) Primer (*gfp_rev_seq*) as well as RNase inhibitor (RiboLock RI; Thermo Scientific) and heated to 80°C for 2 min. Subsequently, the sample was slowly cooled to 50°C followed by addition of 4 μ l 5 \times first strand buffer, 1 μ l DTT (100 μ M), 1 μ l RNase inhibitor (RiboLock RI; Thermo Fisher Scientific) and 1 μ l SuperScript III reverse transcriptase (Invitrogen). After incubation at 52°C for 60 min, the reaction was stopped by addition of 2.7 μ l NaOH (1 M) and heating to 70°C for 15 min. The samples were ethanol precipitated and reaction products were analyzed on an 8% polyacrylamide gel. The DNA sequencing reaction was performed with the same primer using the Thermo Sequenase cycle sequencing kit (USB) and plasmid pBO2267 (Supplementary Table S3) as template. Degradation of the full-length *groES* 5'-UTR:*gfp* mRNA was determined by densitometry (AlphaEaseFC program). The two experiments allowed to determine the appropriate concentrations of lead(II) acetate to equalize the extent of partial RNA digestion at 25 and 37°C.

***In vivo* probing of the *Y. pseudotuberculosis* transcriptome**

Cultures of *Y. pseudotuberculosis* YPIII (Supplementary Table S1) were grown at 25°C and 37°C in LB medium to an OD₆₀₀ = 0.5. Immediately, a 525 mM (final concentration 150 mM) and a 245 mM (final concentration 70 mM) lead acetate solution (2 ml 4 \times LB + 6 ml lead(II) acetate in distilled water) was added to the culture grown at 25 and 37°C, respectively. The cells were incubated under continuous shaking for 7 min, the reaction was stopped by addition of 10 ml cold 500 mM EDTA and cultures were placed on ice. Samples treated the same way but omitting lead(II) acetate treatment (2 ml 4 \times LB + 6 ml distilled water) served as internal control. RNA isolation was carried out by apply-

ing a hot phenol extraction protocol (49). Briefly, harvested cells were resuspended in 250 μ l resuspension buffer and 250 μ l lysis buffer prior to incubation at 65°C for 90 s. Afterward, 500 μ l prewarmed (65°C), water saturated phenol was added, samples were vortexed intensively and incubated at 65°C for 3 min. Tubes were frozen in liquid nitrogen for at least 30 s and subjected to centrifugation (13,000 rpm, 10 min, room temperature). The withdrawn aqueous phase was again extracted four times with phenol, followed by addition of 300 μ l chloroform:isoamylalcohol (24:1) were added and vigorous vortexing for 30 s. After a centrifugation step (15682 g, 10 min, room temperature), the aqueous phase was transferred into a new tube and the chloroform extraction was repeated. Finally, the RNA was precipitated by addition of 1/10 volume sodium acetate (pH 4.5) and 2.5 volume ice-cold 95% ethanol and incubation at -20°C for 1 h. DNA was digested using the TURBO DNase (Ambion) according to the manufacturer's instructions, purified with phenol:chloroform:isoamyl alcohol and RNA concentrations were measured using a NanoDrop spectrophotometer ND-1000 (peqlab). The RNA quality was verified using the Agilent RNA 6000 Nano Kit on the Agilent 2100 Bioanalyzer (Agilent Technologies).

Library preparation

First, rRNA was depleted from 8 μ g of total RNA using the Ribo-Zero rRNA removal Kit (Gram-Negative Bacteria; Illumina). Subsequently, the RNA (1.5 μ g) was fragmented by sonication, using a Covaris Adaptive Focused Acoustics device (Covaris), to a median fragment size of 200 nt. Resulting fragments were phosphorylated at the 5'-end via T4 polynucleotide kinase (Thermo Fisher Scientific). Afterward, fragments were ligated to sample-specific 5'- and 3'-adapter oligonucleotides (Supplementary Table S4) using T4 RNA ligase (Thermo Fisher Scientific). Next, RNA samples were subjected to reverse transcription (RT) using SuperScript III reverse transcriptase (Invitrogen) as well as a primer complementary to the 3'-adapter. To enrich for correctly ligated and reverse transcribed fragments, cDNA was amplified by 15 cycles of polymerase chain reaction using PCR primers identical to corresponding Illumina adapters (Supplementary Table S4). The PCR products were separated on a 2% agarose gel, fragments (150–500 bp) were cut out and purified using the QIAquick gel extraction kit (QIAGEN). The Illumina cluster station was used to perform Cluster generation. Single-end sequencing on the HiSeq2500 and Genome Analyzer II \times followed a standard protocol. The fluorescent images were processed to sequences and transformed to FastQ format using the Genome Analyzer Pipeline Analysis software 1.8.2 (Illumina).

Sequencing data analysis and lead score calculation

To evaluate the quality of the sequencing output, we used the program FastQC (Babraham Bioinformatics). 5'- and 3'-adapter sequences were removed with cutadapt (50). All sequenced libraries were mapped to the *Y. pseudotuberculosis* transcriptome with bowtie (version 0.12.8) (51). The transcriptome was prepared according to the genome

(NC_010465.1) and a transcriptome annotation published earlier (48).

The -1 position upstream of a cDNA read corresponds to the structural state (potential lead cleavage site) of the particular nt. After mapping, the number of reads ending on that position were counted for each nt of the transcriptome, using the previously published Perl script 'sam2tab.pl' (52). The resulting raw counts (rc) were then normalized to normed counts (nc) via

$$nc = \frac{rc + 5}{\text{read depth}}.$$

The normalization to the read depth accounts for differences between the subsamples upstream in the experimental pipeline, as e.g. in RNA isolation and sequencing depth (53). Overestimation of high fold-changes when observing sites with very few raw counts were avoided by the addition of five pseudocounts (54). The lead score was calculated by comparing lead-treated and untreated samples. It is defined as

$$\text{lead score} = \log_2 \frac{nc_{\text{lead}} + m}{nc_{\text{control}} + m},$$

where nc_{lead} and nc_{control} are the normed reverse transcription stops in the lead(II) acetate and the control library, respectively. The factor m represents the 80th percentile of the set of all normed counts from nc_{lead} and nc_{control} . It accounts for the overall level of normed counts in both subsamples and is included to additionally control for the amplitude of background signals in the experiment. Statistics (P -values of unpaired two-sided Student's t -test (P), Kolmogorov-Smirnov test (D), Spearman's as well as Pearson correlation coefficient) were computed using python scipy (python2.7, scipy1.1.0).

Visualization of lead scores

To enhance comprehensibility of visualized lead scores, scores < 0.5 were set to zero when plotting explicit scores. Plotting was realized with the python package matplotlib v2.2.2. Varna (v3-93) (55) was used to visualize lead scores mapped to secondary structure.

Testing of reliability of lead scores

Location and secondary structure of tRNAs were identified with tRNAScan-SE (56). Identical sequences as well as transcripts without lead scores were filtered out, resulting in a test dataset of 32 tRNAs. Since lead scores result from counting 5'-ends of mapped reads, and we restricted the read length to a minimum size of 20 nt during read mapping, lead scores are missing for the last 20 nt of each mapped transcript. Consequently, those nt were excluded from the test dataset, and the final test dataset consisted of 1429 nt (sum of ~ 45 nt from each of the 32 tRNAs), of which 764 are in a paired and 665 are in an unpaired state.

To test the reliability of lead scores, we calculated the Positive Predictive Value (PPV) as well as the sensitivity for each tRNA, regarding scores above different thresholds. The PPV is defined as the proportion of unpaired nt with a lead score (true positives) in all nt with a lead score. It ranges

between 0 and 1, with 1 meaning that all nt above the threshold are unpaired in the reference structure. The sensitivity can assume values between 0 and 1 as well. It is defined as the proportion of true positives in all nt in an unpaired state in the reference structure, and reflects how many of the actual unpaired nt can be found via the lead score (57).

Secondary structure prediction with lead scores

For the data-guided structure prediction, we used the soft-constraint folding based approach by Washietl *et al.* (58), implemented in the programs RNAfold and RNApvmmin of the ViennaRNA package v2.3.5 (59). To use the lead scores for structure prediction, they were first rescaled, setting the 96th percentile to a lead score of 2. Then the top 4% of the scores were set to a rescaled value of 2. All rescaled scores < 1 were excluded from the structure prediction. The resulting values were passed to RNApvmmin, with the ratio of the weighting factors tau and sigma set to 0.6. The obtained perturbation energies are then further processed by RNAfold, where the lead scores - driven structure is predicted. The implementation of lead scores into RNA structure prediction was evaluated using the same test dataset of 32 tRNAs as in the testing of the reliability of scores. Comparisons between secondary structures were based on the coarse-grained structure and calculated with RNAdistance.

Identification of RNA thermometer candidates

Following our previous work on the *Y. pseudotuberculosis*' *in vitro* structurome (48), we applied the Lead-Seq method at different temperatures (25°C + 37°C) to identify new RNAT candidates. For this purpose, Δ scores were calculated for each nt by subtracting lead scores at 37°C from lead scores at 25°C for each position. For each transcript, the top 70% Δ scores of all positive scores as well as the bottom 70% of all negative Δ scores were used to form the set Δs_{adj} . Each value of Δs_{adj} was mapped back to its position in the transcript. A sliding window of 31 nt length was moved across the entire transcript, forming the set $\Delta s_{\text{adj}}(\text{local})$. After each movement of the window, the Mann-Whitney-U test was performed for the two groups $\Delta s_{\text{adj}}(\text{local})$ inside the window and Δs_{adj} of the entire transcript, using scipy.stats.mannwhitneyu. Afterward, the corresponding p-value was \log_{10} -converted. If the average Δ score above the window was smaller than the average Δ score above the entire transcript, the $\log_{10}(P\text{-value})$ was additionally multiplied with -1 . If < 21 nt were included in $\Delta s_{\text{adj}}(\text{local})$, the output of the test was set to 0.

β -galactosidase activity assay

For the β -galactosidase activity assays, *E. coli* DH5 α cells carrying the 5'-UTR:*bgaB* fusion plasmids were grown overnight in LB with ampicillin at 25°C. Before being inoculated with an overnight culture (OD₆₀₀ = 0.2), LB media supplemented with ampicillin was pre-warmed to 25°C. After growth to an OD₆₀₀ = 0.5, transcription was induced with 0.01% w/v L-arabinose. The culture was split up and shifted to pre-warmed 100 ml flasks (temperatures indicated

in the respective figure). Subsequently, the cultures were incubated for 30 min and 200 μ l samples were taken for the β -galactosidase assay. The β -galactosidase assay was carried out as described previously (48). Mean standard deviations were calculated from at least three biological replicates with three technical replicates each.

Alignment-based structure comparison of RNAT sequences and secondary structure visualization

Enterobacterial genome sequences were retrieved from the NCBI database (<https://www.ncbi.nlm.nih.gov/genome/microbes/>). The local sequence and structure alignment was carried out using the LocARNA program (60,61) with default settings. The consensus structure was computed using RNAalifold (59) with default settings and the respective LocARNA-derived alignment file.

RESULTS

Lead(II) acetate maps the *in vivo* RNA structure of a bacterium

The goal of this study was to establish the use of the lead(II) ions as a structure-specific probe in a transcriptome-wide structure probing approach. It was previously demonstrated that lead(II) acetate is suitable for the identification of single-stranded RNA regions *in vitro* and *in vivo* (37). To assess the influence of temperature on lead(II) acetate-mediated RNA cleavage, we first adjusted the lead(II) acetate concentrations, such that a similar amount of RNA is cleaved at 25 and at 37°C. It has been reported that single-hit kinetics are achieved when around 10–20 % of the RNA is cleaved (52). In pilot experiments with *E. coli*, we grew cells to mid-logarithmic growth phase at 25 or 37°C and added lead(II) acetate to obtain different final concentrations of lead(II) ions. Isolated total RNA from these lead-treated cultures was run on agarose gels and inspected for intact 16S and 23S rRNAs by densitometry (Supplementary Figure S1A). Additionally, we analyzed cleavage of a single plasmid-expressed mRNA (*groES* 5'-UTR:*gfp* mRNA) by primer extension and densitometry (Supplementary Figure S1B). At 25°C, treatment of cells with 150 mM lead(II) acetate led to degradation of 17% of the rRNA and 22% of the *groES* mRNA. At 37°C, about 22 % of the rRNA and 27 % of the mRNA were degraded when cells were treated with 70 mM lead(II) acetate. Based on these results, 150 mM and 70 mM lead(II) acetate were deemed suitable to approach single-hit kinetics and equal lead(II) reactivity at 25 and 37°C, respectively.

To map the entire RNA structure of living *Y. pseudotuberculosis* cells, we coupled lead(II) ion-based structure probing (37,45) with next-generation sequencing (Lead-seq, Figure 1A). RNA was isolated from lead(II) acetate-treated *Y. pseudotuberculosis* cells that were grown at 25 and 37°C. After rRNA depletion, the remaining RNA was subjected to random fragmentation. 5'-adapters were ligated to the 5'-phosphate of RNA fragments, which were generated by T4 polynucleotide kinase treatment (T4 PNK). The 5'-adapter comprised a 6-nt barcode at the 3'-end of the adapter sequence (Supplementary Table S4). An adapter ligated to

the 3'-end of the RNA fragments served as primer binding site for RT. Subsequently, cDNA products were amplified by PCR using primers complementary to the 5'- and 3'-adapter sequences. The cDNA libraries were subjected to high-throughput sequencing generating 8.8–10.5 Mio cDNA reads for each library (Supplementary Table S5). The data analysis pipeline comprised trimming of the adapter sequences and mapping of the reads to the *Y. pseudotuberculosis* YPIII transcriptome. To ensure unambiguous mapping of the sequenced reads, we only considered reads with a size >20 nt. As only the 5'-ends of mapped reads carry structural information (i.e. correspond to unpaired or lead-cleaved nt), the size-filtering leads to a 'blind 3'-end' corresponding to the last 20 nt of each transcript. RT run-off events were counted in the libraries generated with RNA isolated from cells treated with or without (control) lead(II) acetate. Finally, we computed the lead score for each nt of the transcriptome. The lead score is the logarithmic ratio of RT run-off counts in the lead(II) acetate-treated library over RT run-off counts in the control library. To examine whether Lead-seq possesses a cleavage bias toward certain nt, we counted the number of adenines, cytosines, guanines or uracils that were cleaved. We found almost equal cleavage of all four nt at 25 and 37°C (Figure 1B). To explore the transcriptome's overall folding behavior, we inspected the distribution of the lead scores of all 3 million nt of the transcriptome at both tested temperature conditions (Figure 1C). We found a slight but significant increase of lead scores at the elevated temperature indicating melting of the transcriptome from 25 to 37°C (mean 25°C = 0.13; mean 37°C = 0.23; $P = 1e-13$).

Lead-seq is capable of mapping well-characterized RNA structures

To assess the accuracy of the Lead-seq approach, we examined the lead score profiles of well-studied and stable RNA molecules with conserved secondary structures. The profiles of tRNA^{Asp} measured at 25 and 37°C largely reflected the expected secondary structure (Figure 2). Most of the enhanced lead reactivity corresponded to unpaired nt in the D and anticodon loops or in single-stranded nucleotides between two stems (₈UA₉), not only in tRNA^{Asp} but also in other tRNAs (Supplementary Figure S2). As previously mentioned, the last 20 nt of each transcript are excluded from the analysis due to technical reasons, resulting in a lack of lead scores in the 3'-CCA region and the T ψ C loop (Figure 2A and B; Supplementary Figure S2). Several paired nt featured lead scores as well, e.g. ₆GG₇ in tRNA^{Asp} at 25°C (Figure 2B). Enhanced accessibility of nt in the first position of a stem has previously been reported for other probing agents like SHAPE (62). Moreover, the cleavage specificity and efficiency of lead(II) ions has been reported to depend on the structural context of the flanking base-pairs (63). Cleavage of paired nt might suggest a dynamic tRNA structure ensemble comprising mature, premature, or misfolded tRNAs. Enhanced cleavage by lead(II) ion seems to be corresponding to physical exposure on the outside of the 3D-structure of tRNA^{Asp} (Figure 2C, (64)). The lead scores of the D stems of tRNA^{Asp} as well as tRNA^{His} indicated an unfolding at higher temperature (Figure 2A and B; Supple-

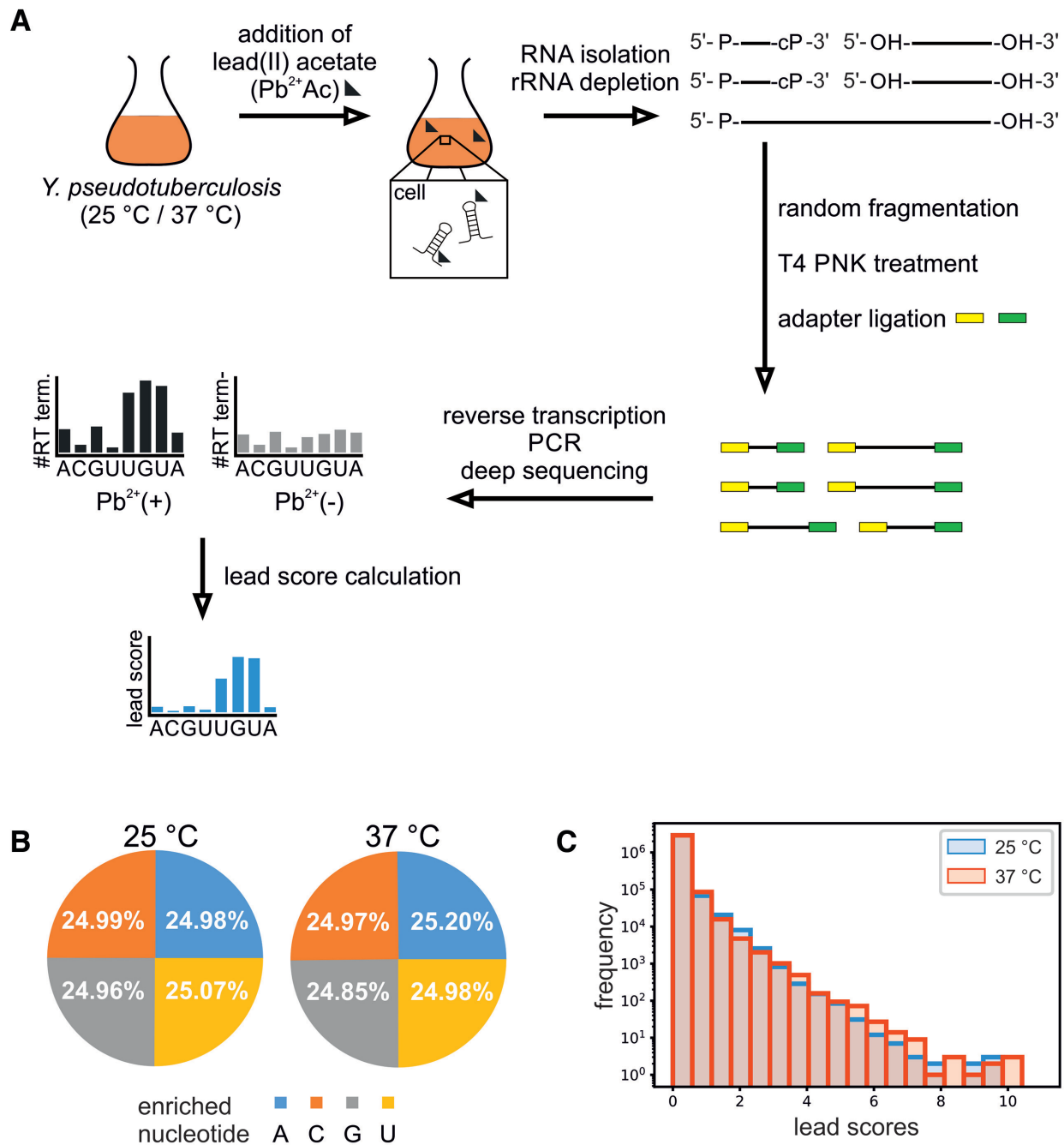


Figure 1. Mapping lead cleavage sites by high-throughput sequencing. (A) RNA from lead(II) acetate-treated *Yersinia pseudotuberculosis* cells, grown at 25 or 37°C. Lead(II) ion-mediated attack of single-stranded nucleotides generates 2',3'-cyclic phosphate (cP) and 5'-hydroxyl groups as cleavage products (45). After RNA isolation and rRNA depletion, total RNA was randomly fragmented and treated with T4 PNK. The RNA fragments were ligated to 5'—(yellow) and 3'—(green) adapters and subjected to RT. The cDNA products were PCR-amplified prior to Illumina high-throughput sequencing. Cleavage events per nt were counted in the lead-treated and control library. The lead score was calculated as the log ratio between lead-induced and spontaneous cleavage events. (B) Pie charts displaying the ratio of lead-cleavage preference toward adenines (As), cytosines (Cs), guanines (Gs) and uraciles (Us) at 25 and 37°C. Moreover, pie charts representing these ratios corrected for the genomic abundance of each nt are shown. (C) Distribution of lead scores at 25°C and 37°C ($n = 3009815$). Over all sampled nt, the mean lead score is significantly higher at 37°C than at 25°C (mean score at 25°C = 0.13, mean score at 37°C = 0.23, $P = 1.0e-13$, $D = 0.71$).

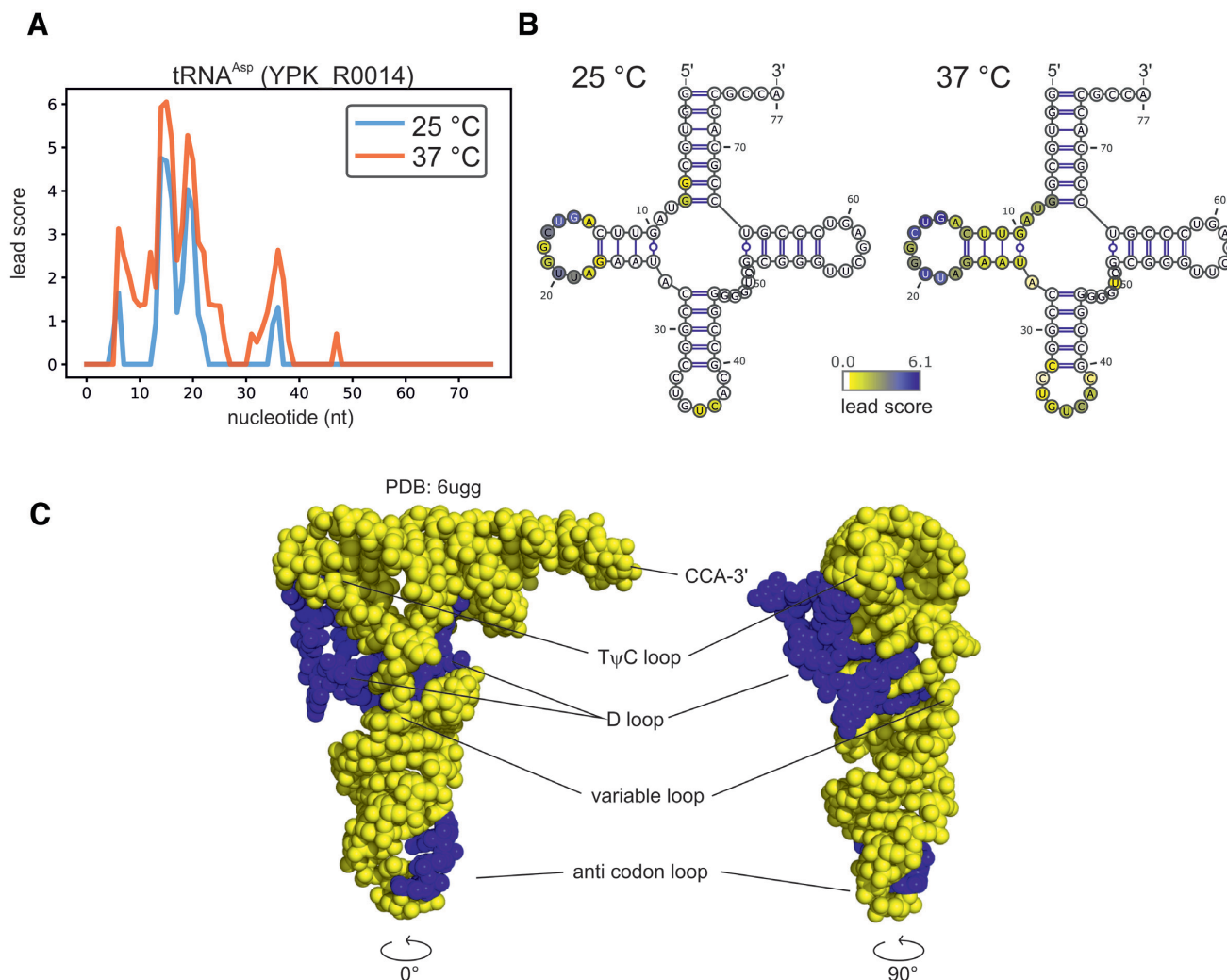


Figure 2. Lead score profile of tRNA^{His} matches the expected structure. (A) Lead score profiles of tRNA^{His} (YPK_R0043) obtained at 25 and 37°C. A positive lead score indicates that the nt is in an unstructured conformation. (B) The expected secondary structure of tRNA^{His} (taken from tRNAScan-SE (56)) was superimposed with lead scores obtained at 25 and 37°C. Lead score values are marked according to the indicated color code. (C) 3D representation of tRNA^{His} with lead-accessible nucleotides colored in blue.

mentary Figure S2), which is in line with the findings from temperature-dependent SHAPE probing of tRNA^{Asp} from *E. coli* (65).

We tested the performance of the Lead-seq approach using a reference dataset including 32 tRNAs present on the *Y. pseudotuberculosis* YPIII chromosome (for construction of the test dataset: see 'Materials and Methods' section). With this test dataset, we investigated whether the calculation of lead scores enhances the predictive power of structure predictions. Higher lead scores should provide more reliable information about the pairing state of a nt. To verify this, we applied different minimum thresholds, and excluded all nt with a lead score below these thresholds from the dataset. We then calculated the PPV and sensitivity for the remaining nt in the test dataset. Likewise, we calculated PPV and sensitivity for the raw counts (RT run-off events) from control- (untreated) and treatment- (lead(II) acetate-treated) experiments. This test was performed for the raw counts and lead scores at 25°C (Figure 3A), as well as 37°C

(Figure 3B). At both temperatures, the lead scores improved the detection of single-stranded regions in tRNA structures. In general, we found that unpaired nt exhibited significantly higher lead scores at both temperatures (Figure 3C). Thus, Lead-seq is able to map single-stranded nt.

In order to test whether Lead-Seq is capable of mapping structures of longer RNA molecules, we inspected the lead score profile of the RNase P RNA (Figure 4A). The secondary structure of the catalytic RNA of RNase P is known to be conserved and consists of 18 stem-loop structures as well as several tertiary interactions (66,67). The Lead-seq profiles of the RNase P RNA measured at 25°C and 37°C were well correlated (Pearson correlation = 0.7) consistent with the expected thermostability of the molecule (Figure 4A). Overall, the lead scores recapitulated the proposed secondary structure of the *Yersinia* RNase P RNA, and many high lead scores coincided with cleavage sites from previous lead(II) acetate-based structural investigations (Figure 4B) (36–37,68).

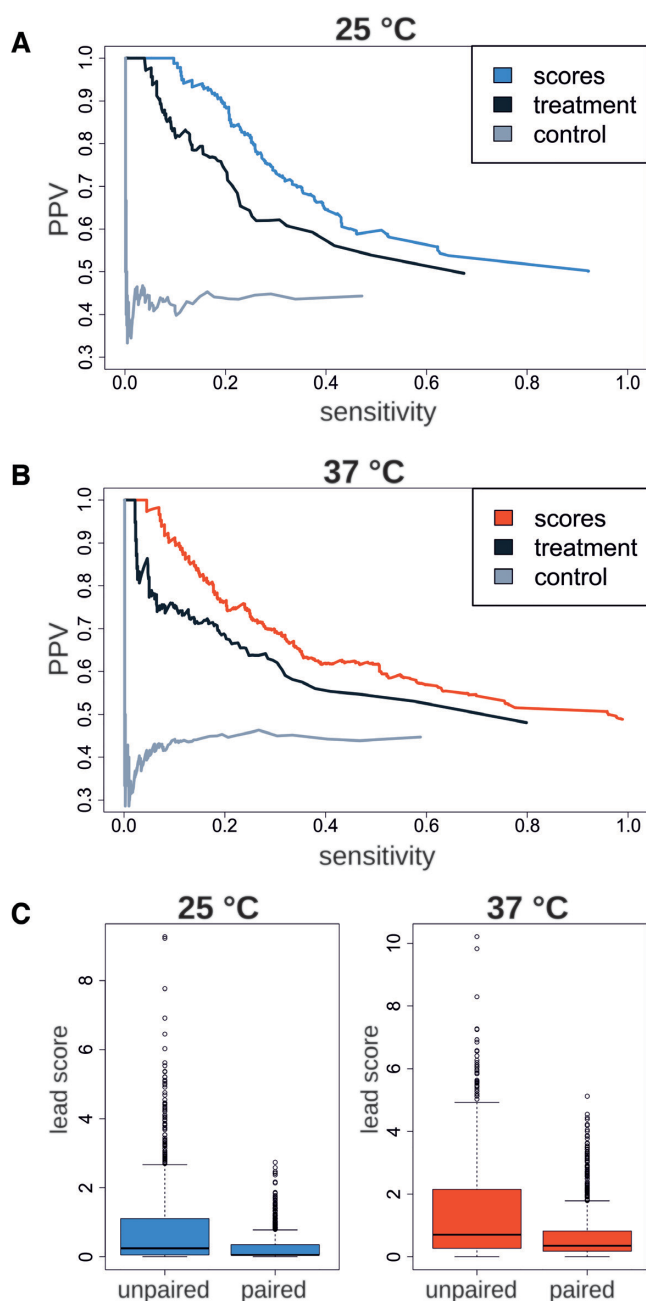


Figure 3. Lead-seq performance on a tRNA reference structural dataset. The structure of 32 tRNAs was used to assess the reliability of lead scores, including 764 paired and 665 unpaired nt. After excluding scores smaller than a certain threshold from the dataset, PPV and sensitivity were calculated from lead scores at 25°C (A) and 37°C (B). This was repeated using different thresholds. The distribution of lead scores at unpaired and paired sites at 25°C (mean unpaired = 0.84, mean paired = 0.26, $P < 2.2 \times 10^{-16}$, $D = 0.22$) and 37°C (mean unpaired = 1.41, mean paired = 0.69, $P < 2.2 \times 10^{-16}$, $D = 0.22$) is illustrated as boxplots (C).

Additionally, we inspected the lead score profiles of the tmRNA at 25 and 37°C (Supplementary Figure S3A). The tmRNA rescues stalled ribosomes on defective mRNAs with the helper protein SmpB in a process called *trans*-translation (69). Overall, the obtained lead scores fit the expected secondary structure (70) (Supplementary Figure

S3B). Nt₁₆UU₁₇ (as well as nt₃₃₃GAC₃₃₅, which are not visible in our approach) are supposed to be involved in SmpB binding as revealed by standard *in vivo* lead probing of tmRNA from wild type and Δ *smpB* *E. coli* strains (71). Our finding that ₁₆UU₁₇ are cleaved by lead at both temperatures suggests that the majority of tmRNAs in *Y. pseudotuberculosis* is not bound by SmpB under the tested conditions. Based on the data evaluation for tRNAs as well as larger RNA molecules, we conclude that Lead-seq is a suitable method to map RNA structures *in vivo*.

Lead-seq complements presently available RNA structure probing techniques

We inspected the 5'-UTR of *rpsT* encoding the ribosomal protein S20, which was previously analyzed in *E. coli* by SHAPE probing and is supposed to harbor a structured RNA motif with unknown function (72). With the exception of only one nt, which is positioned in an internal loop, the sequences of this region are identical in *Y. pseudotuberculosis* and *E. coli* (Supplementary Figure S4A). The RNA motif constitutes a single hairpin, in which single-stranded nt are prone to lead cleavage as well as SHAPE modification (Supplementary Figure S4B). The reliability of both methods seems comparable for the small RNA structure, and some overlap of enhanced lead- and SHAPE-reactivity can be found.

To further validate Lead-seq by other probing methods, we extended our analysis to larger RNAs and compared the results from the 16S rRNA of structure probing experiments with DMS (73) and SHAPE (74) with the lead scores derived from our experiments (Figure 5). Consistent with an intricate structure within the 30S ribosome, only few cuts/modifications were found by all three methods. A 3D representation of the 16S rRNA alone or the 30S ribosome shows that the lead cleavage occurs at surface-exposed nucleotides (Supplementary Movies S1 and 2). As in the 5'-UTR of *rpsT*, some intersection between lead scores and the reactivities from other methods occur. Strikingly, several unpaired sites are highly reactive only toward one of the used chemical probes. In the future, the combination of several probes in one study should be suitable to yield more comprehensive structural information.

Lead-seq data improve secondary-structure prediction

The structures of all tRNAs were predicted with and without guidance by lead scores. The distance between the resulting MFE structures and the reference structure was determined based on the coarse-grained structures (Figure 6A). The initial prediction without incorporation of Lead scores was correct for about half of all tested tRNAs (Figure 6B). For 15 of the 32 tRNAs, the classic approach predicted MFE structures not in accordance with the reference structures. For nine tRNAs—and thus 60% of all initially incorrectly predicted tRNAs—the inclusion of our experimental Lead-seq data improved the prediction. Strikingly, the Lead scores had a correcting influence on the MFE structure when the initial prediction was rather inaccurate, while the structure often remained unchanged when the initial structure was already close to the reference structure. In

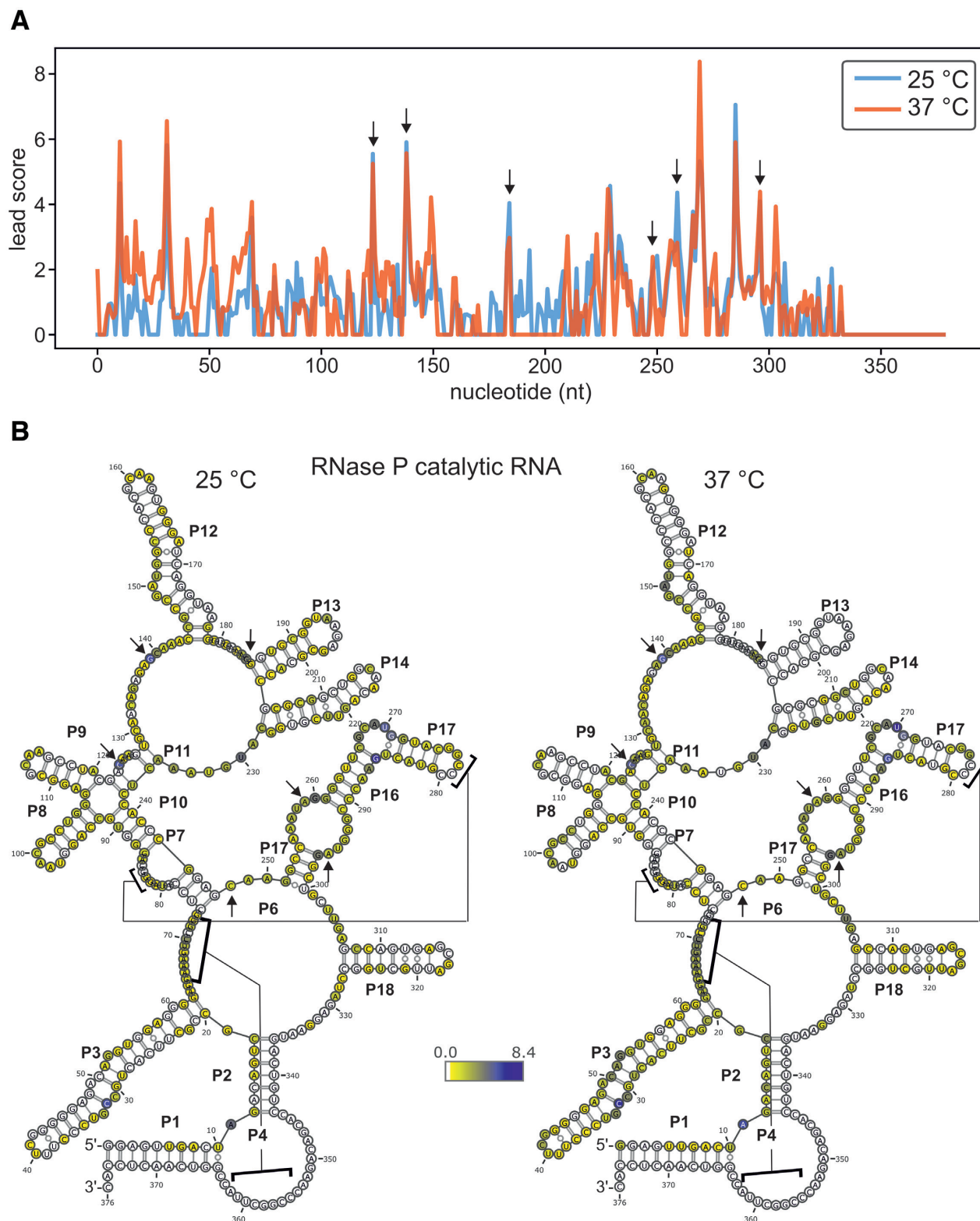


Figure 4. Lead score profiles of RNase P RNA matches the validated structure. (A) Lead score profiles of M1 RNA obtained at 25 and 37°C. (B) The lead scores of the M1 RNA (25 and 37°C) were overlaid with its expected secondary structure (105). Nt with positive lead score values, indicating a single-stranded conformation, are indicated in color. Proposed pseudoknots are indicated. Black arrows indicate prominent sites of lead(II) cleavage that were also observed by (36,38,68).

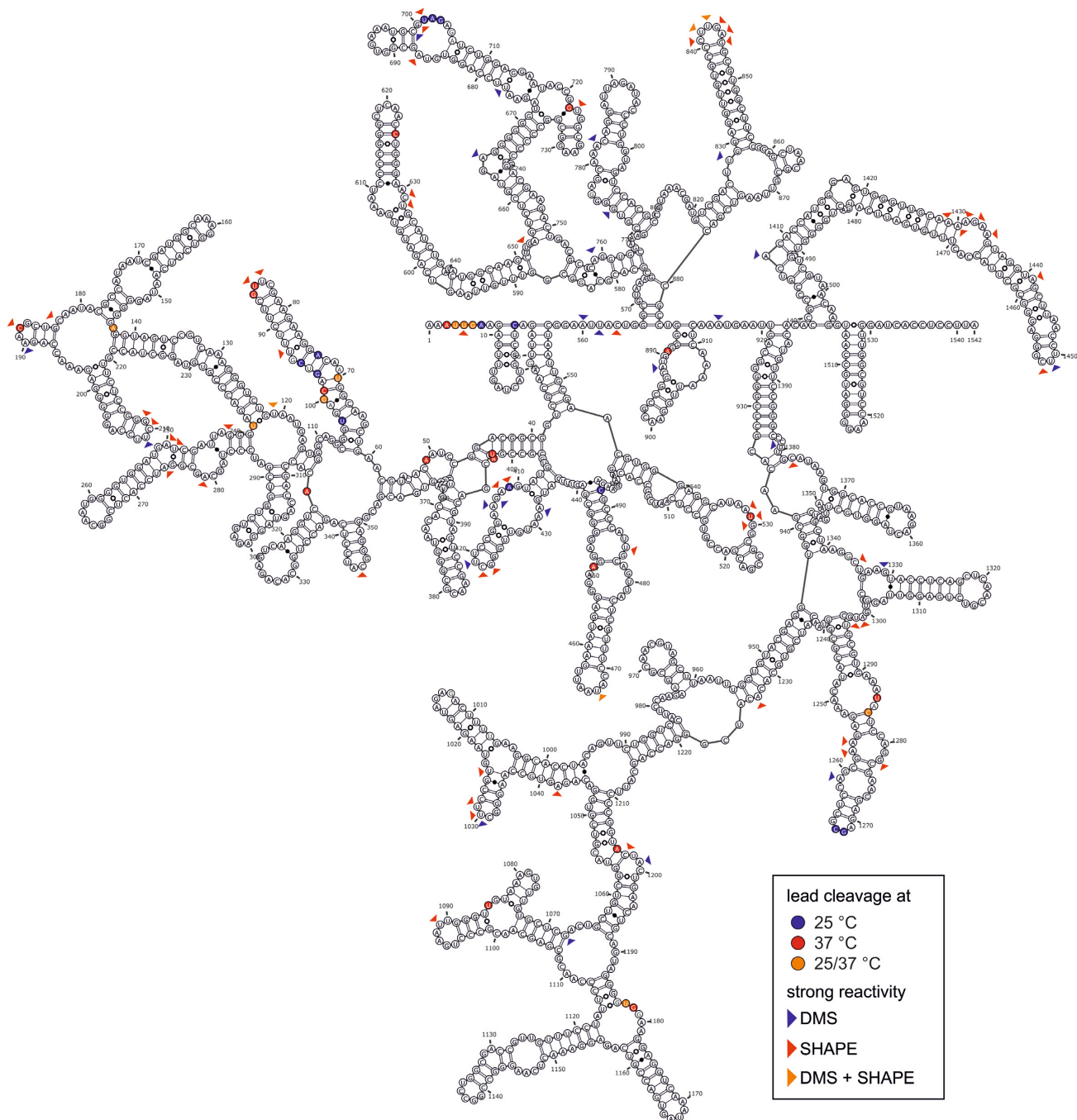


Figure 5. Lead-seq can compete with other global structure probing approaches. Secondary structure of the 16S rRNA. Nucleotides which are cleaved by lead(II) ions at 25°C, 37°C or at both temperatures are marked in blue, red or orange, respectively. Structural information from transcriptome-wide *in vivo* studies of 16S RNA using either DMS (106) or SHAPE reagents (74) as probe are indicated by blue, red or orange arrow heads, respectively.

total, the predictions of 28% of the tRNAs were positively influenced by the lead scores (Figure 6C).

Metagene analysis of the *Y. pseudotuberculosis* *in vivo* structure

Although all mRNA structures are distinct (or in other words, have their own structural ‘footprint’ (72)), metagene analysis can provide interesting insights into overall struc-

tural features. We used our Lead-seq dataset to examine the temperature-responsiveness of the coding RNA structure (mRNAs possessing a 5′-UTR > 14 nt, $n = 1553$) at 25 and 37°C. On average, mRNAs seem to be less structured at 37°C compared to 25°C (mean 25°C = 0.126, mean 37°C = 0.238, $P = 6.3e-11$), suggesting that coding RNAs tend to unfold at higher temperatures (Supplementary Figure S5A). To assess the differential melting behavior within mRNAs, we compared average lead scores of 5′-UTRs (Supplemen-

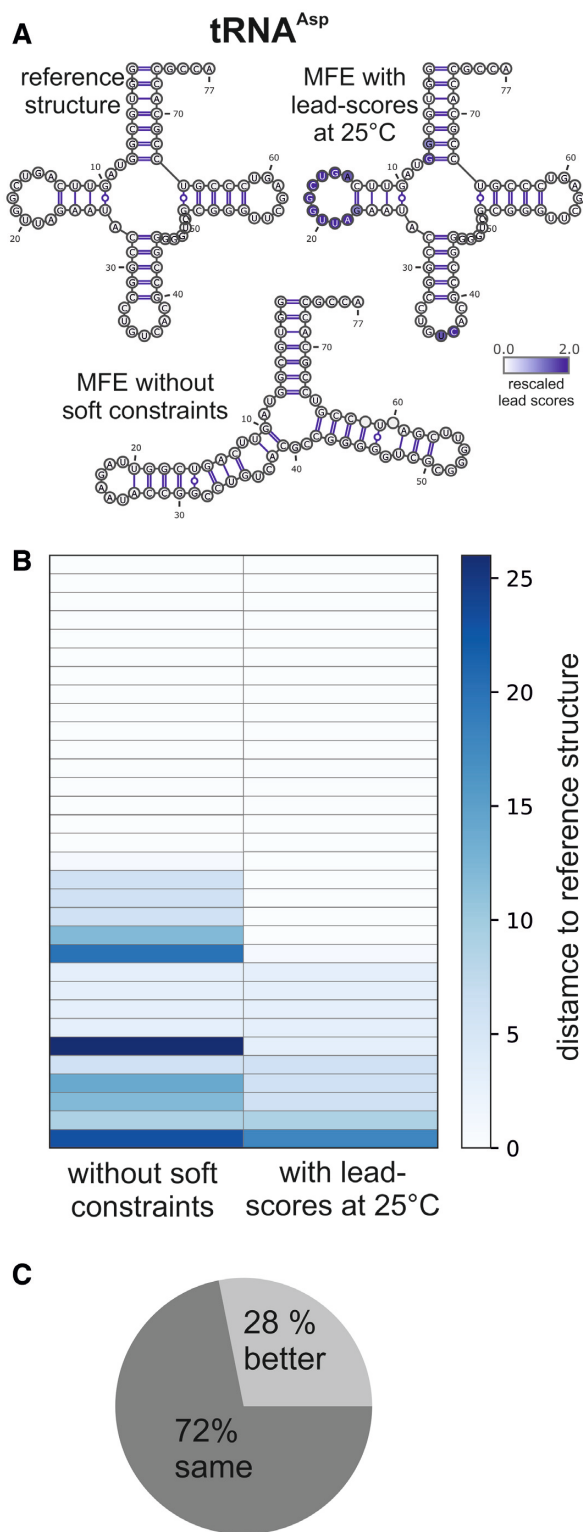


Figure 6. Comparison between RNA structure predictions with and without incorporation of Lead-seq data. (A) Examples of minimum free energy structures (MFE structures). The MFE structure of tRNA^{Asp} (center) is not in accordance to the reference structure (left). This could be improved by using Lead scores as soft constraints (right). (B) Distance to the reference structure of MFE structures of 32 tRNAs predicted without (left) and with (right) incorporation of Lead-seq data, based on coarse grained representations of the structures. (C) Fraction of tRNAs whose predicted MFE structure could be improved by Lead scores.

tary Figure S5B) as well as CDRs (Supplementary Figure S5C) and found the majority of both to melt from 25°C to 37°C. Transcript-wise examination of the 5'-UTR's and CDR's respective Δ score (25–37°C) revealed only a moderate correlation (Supplementary Figure S5D, Spearman's correlation = 0.68). Moreover, the mean Δ score of 5'-UTRs (mean Δ score = -0.133) is more negative than the corresponding value of CDRs (mean Δ score = -0.110), suggesting stronger melting of 5'-UTRs (Supplementary Figure S5E, $P = 1.1e-07$).

Next, we had a closer look at structural features across coding transcript regions by analyzing an average mRNA lead score profile (Figure 7A and B). All 1016 mRNAs with 5'-UTRs longer than 50 nt were aligned at the translation start codon and the average lead score of the last 48 nt of the 5'-UTRs as well as the first 48 nt (16 codons) of the CDR was calculated. We found 5'-UTRs to be more structured than CDRs at 25°C (mean 5'-UTR = 0.145, mean CDR = 0.163, $P = 5.6e-16$) (Figure 7A). At 37°C, we found nearly equal average lead scores for 5'-UTRs and CDRs (mean 5'-UTR = 0.265, mean CDR = 0.257, $P = 0.003$) (Figure 7B). In line with the results described above (Supplementary Figure S5A–E), this finding suggests that 5'-UTRs unfold at 37°C to a larger extent than CDRs. We also observed a local maximum of unstructuredness at the translation start codon at both temperatures (Figure 7A and B), which coincides with findings from the *in vivo* probing of the *E. coli* transcriptome (72). In line with the *E. coli* study but in contrast to the meta-analysis from our *in vitro* probing approach (48), we did not find a local minimum around the Shine–Dalgarno (SD) sequence (-10 ± 4 nt from the start codon) (Figure 7A and B).

Finally, we analyzed the structuredness of single nt within base triplets. Whether a three-nt periodicity in CDRs exists or not is controversial (72,75). To address this issue in *Yersinia*, we computed the average lead score of the first, second and third nt within the first 16 base triplets of all CDRs as well as the last 16 base triplets of all 5'-UTRs. At both temperatures, the first nt of the triplet was slightly less structured within CDRs (P 25°C = 0.01, P 37°C = 0.009) (Figure 7C and D), while this differential periodicity was absent in 5'-UTRs.

Discovery and validation of new RNA thermometers (RNATs)

Yersinia pseudotuberculosis adapts to virulence conditions by taking advantage of the temperature-controlled thermodynamic characteristics of RNA molecules (48,76). Since our study provides information on RNA structures at two temperatures relevant in this context (25°C as proxy for environmental conditions and 37°C mimicking virulence conditions), one should be able to extract known and potentially new RNATs from our datasets. This was indeed the case. We used a sliding-window approach to identify local structural rearrangements of protein-coding transcripts (Figure 8A). This method identifies transcript regions with significant local melting events compared to the general unfolding of the whole transcript. Following earlier studies from our lab, we additionally factored in the difference between average lead scores at 37 and 25°C (Δ score) by im-

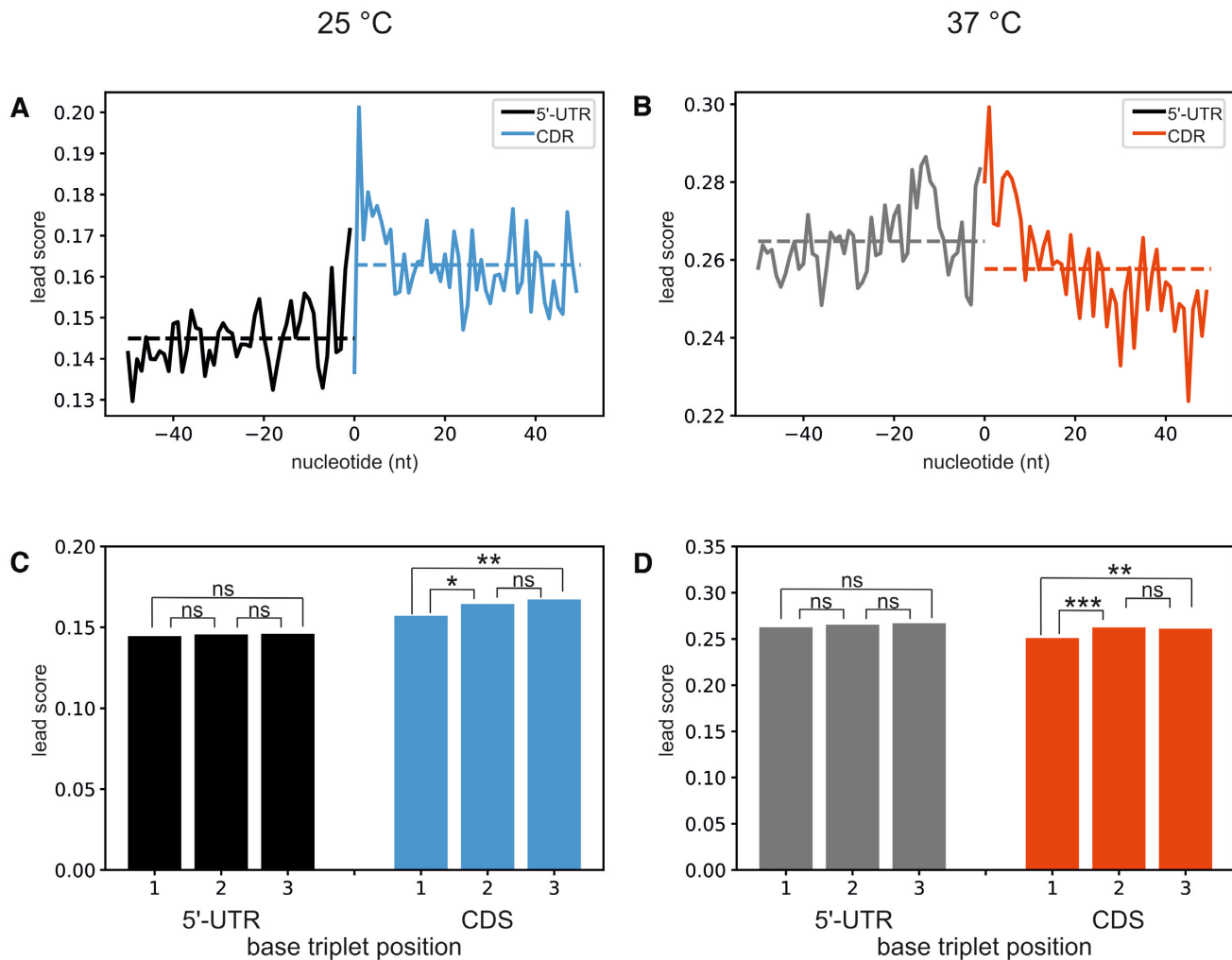


Figure 7. Structural analysis of the average lead profile across and within protein-coding transcripts. Average lead score for each nt across the last 50 nt of the 5'-UTR and the first 50 nt of the CDR at 25°C (A), and at 37°C (B). Average lead score for the first, second, and third base of each codon in the 5' UTR (last 48 nt) and the CDRs (first 48 nt) at 25°C (C), and 37°C (D). Statistical significance determined by unpaired two-sided *t*-test (asterisks abbreviation: ns $P > 0.05$; * $P < 0.05$; ** $P < 0.01$; *** $P < 0.001$; **** $P < 0.0001$). (A and B) $n = 1016$, (C and D) $n = 16\ 256$

plementing a sign-conversion. RNA regions with a strongly negative $p_{\log, \text{sign}}$ can be interpreted as regions undergoing temperature-dependent conformational changes. mRNAs exhibiting a $p_{\log, \text{sign}} < -3$ around the SD region (-10 ± 4 from the start codon) were considered promising RNAT candidates (Figure 8B). We found that the SD region upstream of *ailA*, *trxA*, *sodB*, *pepN* and *cysK-2* showed significant unfolding from 25 to 37°C (Table 1). The respective 5'-UTRs are known RNATs, which were identified and experimentally validated in our previous PARS-based (parallel analysis of RNA structure) study (48). Moreover, we identified several other potential RNA-based thermosensors that melt at elevated temperature, such as the 5'-UTRs upstream of *groES* and *clpB* both coding for heat shock chaperones (Table 1; Figure 9A and B). To corroborate that Lead-seq is capable of detecting novel RNATs, we experimentally measured temperature-dependent translation mediated by the *groES* or *clpB* 5'-UTRs. The respective 5'-UTRs were translationally fused to *bgab*, a gene coding for a heat-stable β -galactosidase (Figure 8C). In this well-established plasmid-

based reporter gene system (48), transcription is driven by the arabinose-inducible P_{BAD} promoter. Reporter gene activity in *E. coli* was several-fold induced after a shift from 25 to 37°C (Figure 8D). This temperature-controlled activity was comparable to the *Y. pseudotuberculosis kata* 5'-UTR, a known RNA-based thermoregulator (48).

It is interesting to note that the newly identified thermosensors appear to be conserved structural elements. The *groES* 5'-UTR of enterobacteria is composed of two hairpins containing conserved as well as covariant nt (Supplementary Figure S6A). The consensus structure masks the SD region and AUG codon by base-pairing interactions (Supplementary Figure S6B) and is known to mediate translational control and differential expression of the *groESL* operon in *Salmonella* (77). A high sequence conservation is also evident in the *clpB* 5'-UTR (Supplementary Figure S6C). In the consensus structure, SD region and start codon are involved in a base-pairing interaction (Supplementary Figure S6D). Superimposing the consensus structures of both new RNATs with the respective lead scores at

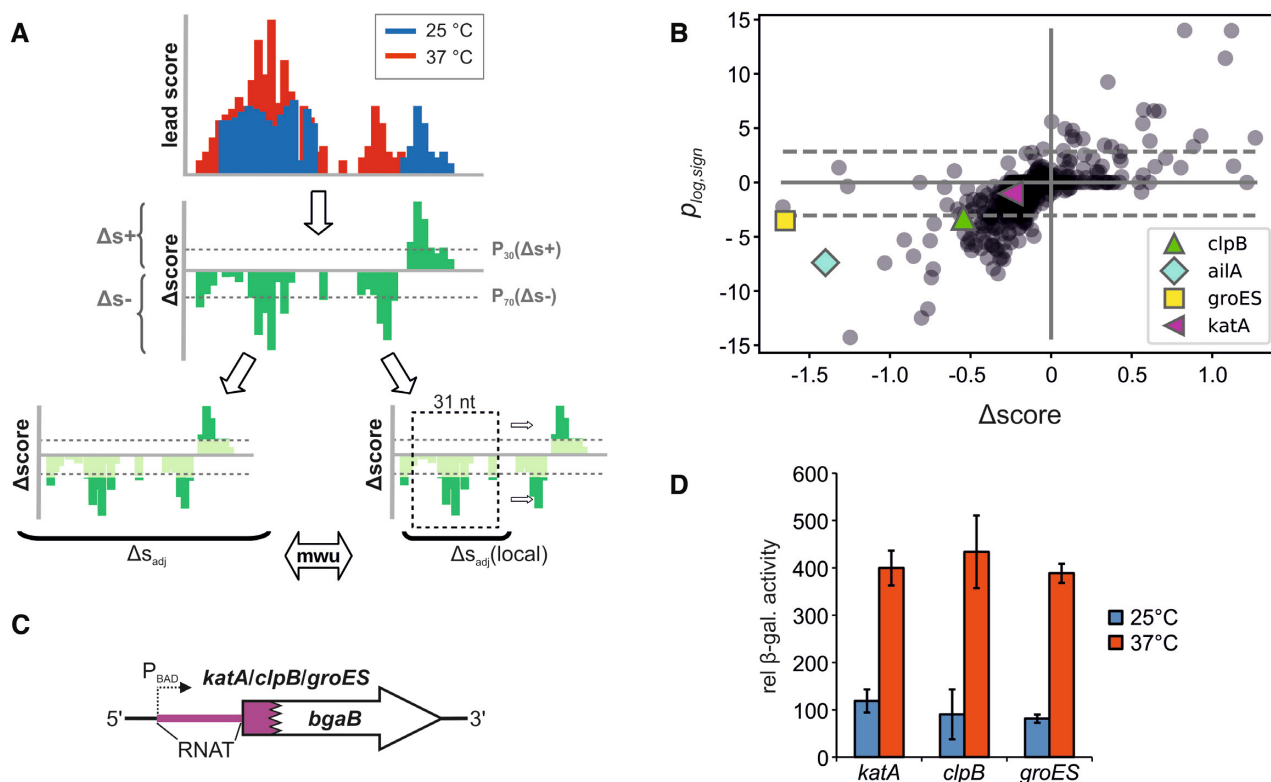


Figure 8. Detection of RNATs via lead scores and their experimental validation. (A) Schematic representation of the intra-transcript comparison method. First, the scores from experiments at 25°C (blue) and 37°C (red) of a transcript (top) form the basis for the calculation of Δscores at the single nt level (middle). From the set of all Δscores of a transcript, the percentiles P₃₀(Δs+) and P₇₀(Δs-) are calculated indicated by dashed lines. Score Δs_{adj} is based on all Δscores of the transcript being greater than P₃₀(Δs+) or smaller than P₇₀(Δs-) (bottom left). Next, a window is moved over the transcript, and Δs_{adj} of nt inside the window form Δs_{adj}(local) (bottom right). Scores Δs_{adj} and Δs_{adj}(local) are subsequently compared using the Mann–Whitney–U rank test (mwu). (B) Relation between general unfolding (Δscore) and local unfolding (p_{log.sign}) at Shine-Dalgarno regions sampled at 25 and 37°C. Significance levels of p_{log.sign} < -3 or p_{log.sign} > 3 indicate a significant lower or higher structuredness at 37°C than at 25°C, respectively, and are indicated by dashed lines. n = 902. (C) Schematic representation of reporter gene constructs used to study translational thermoregulation of potential thermosensors. The 5'-UTRs upstream of *clpB* and *groES* were translationally fused to *bgaB* encoding a heat-stable β-galactosidase, while transcription is under control of an arabinose-inducible promoter (P_{BAD}). (D) *Escherichia coli* DH5α cells harboring the corresponding plasmid (technical triplicates per each construct) were grown to an OD₆₀₀ = 0.5 at 25°C. Afterward, transcription was induced with 0.01% (w/v) L-arabinose and the cultures were split immediately: One half remained at 25°C while the other was transferred into pre-warmed flasks at 37°C. After 30 min incubation, samples were taken for subsequent β-galactosidase assay. The displayed results represent the mean activities from five independent experiments (biological replicates). Mean standard deviations are indicated as error bars.

25 and 37°C revealed higher lead scores for the *groES* 5'-UTR at 37°C compared to 25°C consistent with the thermolability of both hairpins (Figure 9A). In particular, the SD region is resistant to lead(II) acetate-mediated cleavage at 25°C but highly accessible at 37°C. Temperature-induced melting is also evident in the *clpB* 5'-UTR (Figure 9B). At 25°C, the 5'-UTR constitutes a stable structure as only few nucleotides are susceptible to lead cleavage. The attacked nucleotides are located within single-strand loop regions within the secondary structure model. At 37°C, the accessibility to lead(II) ions increased. To further explore the functionality of the newly discovered thermosensors, we designed stabilizing (called REP) as well as destabilizing point mutations (called DEREP) that were supposed to either clamp or loosen the structure of the 5'-UTRs (Figure 9A and B). The obtained variants were tested for temperature-dependent translation control using the *bgaB* reporter gene system (Figure 9C). As expected, the REP mutations led to full repression of reporter gene activity indicating a stable RNA structure that impairs translation initiation both

at 25 and 37°C. In contrast, the destabilized variants produced two- to three-fold higher BgaB activity relative to the wildtype 5'-UTRs already at 25°C. This result confirms the identity of these RNA elements as labile RNAT structures, whose destabilization by strategic mutations immediately facilitates ribosome access even at the low temperature. Overall, these results show that the Lead-seq approach can uncover dynamic RNA structures when applied at two different temperatures.

DISCUSSION

Transcriptome-wide RNA structure probing has been developed only recently and various methods have been established for probing *in vitro* and *in vivo*. Mapping of RNA structures in living cells is limited by the ability of the probing agent to penetrate the cell wall or membrane, which in turn is determined by the probe's physical properties such as size or charge. Furthermore, the electrostatic properties of the RNA environment determine the reactivity of chem-

Table 1. Lead and PARS scores of selected RNAT candidates

locus ^a	gene name ^a	Lead-seq ^b		PARS ^c	Thermal control ^d	Reference
		$\Delta p_{\log, \text{sign}}$	Δscore			
YPK_1268	<i>ailA</i>	-7.39	-1.40	0.52	Y	(48)
YPK_1429	<i>cysK-2</i>	-3.44	-0.27	0.11	Y	(48)
YPK_1863	<i>sodB</i>	-4.78	-0.45	-0.22	Y	(48)
YPK_2645	<i>pepN</i>	-6.27	-0.49	-0.62	Y	(48)
YPK_3349	<i>clpB</i>	-3.26	-0.54	-0.29	Y	This study
YPK_3823	<i>groES</i>	-3.56	-1.66	-0.45	Y	This study
YPK_4035	<i>trxA</i>	-3.16	-0.19	-0.85	s:Y l:N	(48)

^aLocus ID and gene name of the downstream gene potentially thermoregulated by the putative RNAT.

^bAverage values of the SD region (-14 to -6 from the start codon) calculated from lead scores at 25 and 37°C.

^cDifference in the PARS scores of the SD region from 37 to 25°C. PARS scores were taken from Righetti *et al.*, (48).

^dExperimental confirmation of temperature-dependent translational regulation in reporter gene studies from this study or studying the *in vitro* transcriptome of *Y. pseudotuberculosis* (48). (Y) indicates thermal control, (N) no thermal control, (s) and (l) short and long variants of the 5'-UTR, respectively.

ical probes (32,78). Like DMS or SHAPE reagents, lead(II) acetate is a suitable reagent for RNA structure determination in living cells. Hitherto, this probe was successfully used to determine the structure of single RNA molecules (37,38) and to identify metal ion binding sites within RNA (79).

In this proof-of-concept study we took advantage of the ability of lead(II) acetate to pass bacterial membranes and to attack ribonucleotides without sequence bias (Figure 1B). The parallel analysis of RNA from untreated cells identifies background signals derived from spontaneous RT stops, e.g. caused by stable RNA secondary structure, covalent modification and RNA degradation or processing (80). Such correction against untreated samples has commonly been used, for example in Mod-seq (81) or Structure-Seq2 (82). We generated lead scores by calculating the logarithmic ratios between RT run-offs in the lead treatment library and the control library. The magnitude of the lead score corresponds to the enhanced accessibility of nt, indicating a higher probability for a nt to be in a single-stranded conformation. Computation of logarithmic ratios as quantitative representation for structuredness has also been used in previous approaches, such as PARS (53,54), FragSeq (83) and CIRS-seq (25).

In earlier studies, it was reported that lead(II) acetate is able to map complex RNA structures (44,46). Our data show that Lead-seq has the potential to interrogate the genome-wide RNA structure landscape within living cells. The lead score profiles of tRNAs (Figure 3), RNase P RNA (Figure 4) and tmRNA (Supplementary Figure S3) were consistent with their accepted secondary structures. DMS or SHAPE reagents are alternative probes to investigate higher-order structures of RNA molecules (84). These compounds were used for comparison of *in vitro* and *in vivo* structures of single RNAs (85,86) as well as on a genome-wide scale (72,87–89). Although the folding conditions in dilute buffer or in a crowded intracellular environment differ, genome-wide SHAPE approaches barely found noticeable differences in mRNA structures between *in vivo* and *in vitro* data from mouse embryonic stem cells (87) and *E. coli* (72). Future comparative studies are needed to learn whether this is also true for other organisms. Since lead is applicable to both, cellular RNA after extraction and refolding as well as RNA within living cells, the Lead-seq ap-

proach has great potential to address this question in bacteria.

It was previously discussed that combining multiple structure probing techniques will provide the most definitive structure of certain RNA molecules (90). Thus, Lead-seq in combination with SHAPE and/or other methods might eliminate structural information gaps. An example for the additive value of two approaches is the mapping of the *rpsT* leader region by Lead-seq in *Y. pseudotuberculosis* (this study) and by SHAPE in *E. coli* (72) (Supplementary Figure S4). The comparison of Lead-seq, DMS (73) and SHAPE probing (74) on the well-known structure of the 16S rRNA (Figure 5) reveals complementary but also idiosyncratic patterns indicative of the different resolution power of these methods.

The final goal of all RNA structuromics methods is to generate detailed information on the folding state of cellular RNAs. At present, all experimental techniques are limited to the identification of single-stranded residues. These results provide valuable information that can be utilized to improve the quality of predictions by classical RNA secondary structure prediction algorithms (91). We show that this is also possible with Lead-seq data by combining secondary structure prediction with lead scores via the method developed by Washietl *et al.* (58). The prediction of 28% of the tested tRNA could be improved (Figure 6) and the benefit was particularly strong for those RNAs that initially showed a higher distance to the reference structure.

In addition to the analysis of selected individual RNAs to validate the approach, we used the Lead-seq method to investigate global structural features of coding transcripts. We found the CDR to be less structured (higher lead scores) than the upstream 5'-UTRs at 25°C (Figure 7A). The same trend was reported for mRNAs from metazoan (92), HIV-1 virus (93), rice (94) and human (54). Consistent with the results from the metagene analysis of the *Y. pseudotuberculosis in vitro* structurome (48) as well as the *in vivo* structurome of *E. coli* (72), 5'-UTRs and CDRs seem to be equally structured at 37°C (Figure 7B). Consequently, 5'-UTRs undergo stronger unfolding from 25 to 37°C than CDRs. Similar observations were made by DMS-based *in vivo* probing of *Oryza sativa* L. (rice) seedlings grown at 22 or 42°C (94). We observed a slight but significant structural pattern

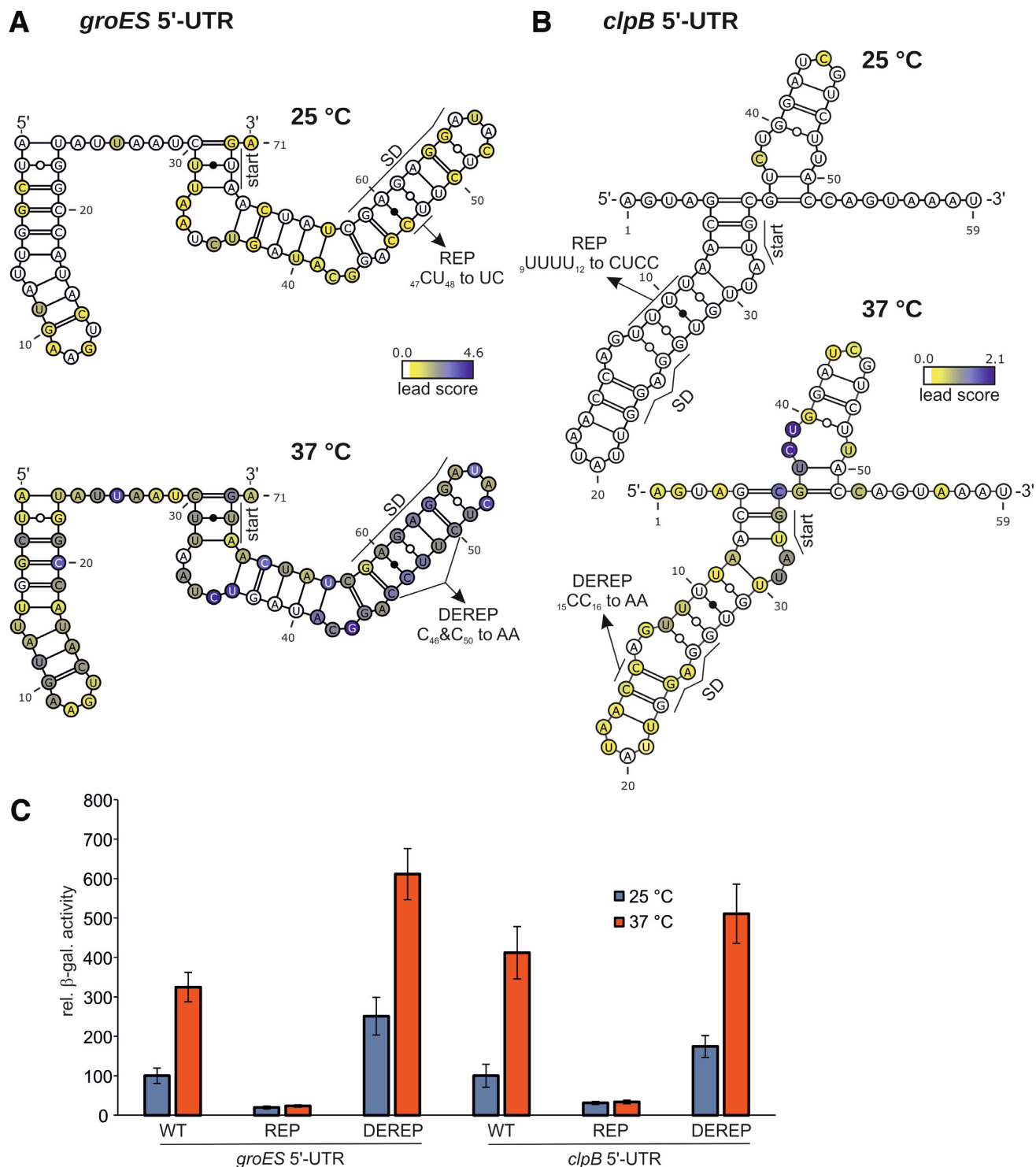


Figure 9. Lead-seq data indicate melting of the *groES* and *clpB* 5'-UTR's secondary structures. Lead scores obtained at 25 or 37°C are depicted in the secondary consensus structures (see Supplementary Figure S6) of the enterobacterial *groES* 5'-UTR (A) and *clpB* 5'-UTR (B). The SD sequence as well as the translation start codon are marked and the lead scores are represented by the indicated color code. Stabilizing (REP) as well as destabilizing mutations (DEREP) are depicted within the structures colored with lead scores obtained at 25 or 37°C, respectively. (C) Point mutations within the 5'-UTRs of *groES* and *clpB* impair RNAT functionality. WT and mutated 5'-UTRs (REP and DERE variants. For details see panel A and B) were translationally fused to the *bgaB* reporter gene and BgaB activity was measured as described in Figure 6. The displayed results represent the mean activities from three independent experiments (biological replicates). Mean standard deviations are indicated as error bars.

within codons, such that on average the third position in a codon was less structured (i.e. more prone to lead cleavage) than the preceding ones (Figure 7B). A similar structural bias in codons was reported in *Arabidopsis thaliana* (95), yeast (53), *E. coli* (75) and mouse (25). It is believed that this structural feature serves as a hidden code in RNA structure, which influences and regulates translation efficiency (25,95). Another explanation might be co-translational mRNA decay following the last translating ribosome, which can lead to a 3-nt periodicity pattern in the CDR (96).

Lead-seq applied to *Y. pseudotuberculosis* grown at two different temperatures allowed us to scrutinize its temperature-responsive *in vivo* structurome. We observed a general unfolding of mRNA structures (Supplementary Figure S5), in particular in 5'-UTRs. Thus, we attempted to localize structural melting processes, especially around the SD region (Figure 8). This strategy not only retrieved RNATs known from previous studies, but also uncovered novel thermosensors upstream of *groES* and *clpB* (Table 1). Both genes encode chaperones known to rescue proteins during the heat shock response (97–100). Strikingly, the homolog *groES* 5'-UTR found in *Salmonella enterica* also harbors a dynamic temperature-sensitive RNA structure. It drives differential translation of the *groESL* operon reducing the level of GroES protein over GroEL at low temperatures (77). Interestingly, many base pairs of the *groES*-RNAT of *Salmonella* and *Yersinia* are either covariant or conserved among the species, especially in the hairpin comprising the SD region (Figure 9). The alignment-based structure is further supported by the observed lead reactivity at 25°C, and it suggests a functional role of this structural element. The newly identified *clpB* RNAT also bears high sequence and structure identity among enterobacteria (Supplementary Figure S6). Structure and sequence conservation was also found in the *shuA/chuA* thermosensors from *Shigella dysenteriae*/*E. coli* regulating genes involved in iron acquisition (101). In contrast, less conserved leader regions with high structural similarity were found upstream of *cnfY* or *cnfI* from *Y. pseudotuberculosis* or uropathogenic *E. coli*, respectively (102).

Recent reports suggest a remarkable difference between the outcome of temperature-induced RNA melting in prokaryotes and eukaryotes. While unfolding of UTRs of bacterial mRNA facilitates ribosome loading and translation initiation, such melting rather induces mRNA degradation by the exosome in eukaryotic organisms (47,94). Lead-seq or other global structure probing techniques might be applied to address how temperature or other environmental parameters affect the RNA structurome in a variety of organisms. The potential of such methods is exemplified by a recently discovered thermoswitch controlling expression of a cold shock protein in *Listeria monocytogenes* (103).

The Lead-seq approach was able to map RNA structures well in line with previous findings. Thus, we believe that global lead probing has the potential to become a valuable method to map the dynamic RNA structuromes *in vitro* and *in vivo*. The established protocol is fairly straight-forward and direct since it does not rely on modification-induced reverse transcription termination of chemically modified nt. Therefore, it should be adaptable to various applica-

tions. For example, one should be able to use it for footprinting of RNA-binding proteins by comparison of the *in vivo* structuromes from wild type strains and strains lacking a certain RNA-binding protein (81). Another potential of global structure probing is the identification of ligand-binding sites in RNA molecules. Changes in SHAPE reactivity were observed near the ligand-binding site in the *thiM* TPP riboswitch aptamer domain in the presence and absence of TPP *in vitro* (85). In combination with ribosome profiling, Lead-seq should be able to correlate the effect of RNA structures and translation efficiency, as previously accomplished by other structure probing approaches (72,104).

DATA AVAILABILITY

Data is available at the Gene Expression Omnibus (GEO) database under accession number GSE140649 (<https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE140649>).

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

FUNDING

German Research Foundation [DFG NA 240/10-2]. Funding for open access charge: DFG [NA 240/10-2].
Conflict of interest statement. None declared.

REFERENCES

- Serganov, A. and Nudler, E. (2013) A decade of riboswitches. *Cell*, **152**, 17–24.
- Breaker, R.R. (2012) Riboswitches and the RNA world. *Cold Spring Harb. Perspect. Biol.*, **4**, a003566.
- Kortmann, J. and Narberhaus, F. (2012) Bacterial RNA thermometers: molecular zippers and switches. *Nat. Rev. Microbiol.*, **10**, 255–265.
- Loh, E., Righetti, F., Eichner, H., Twittenhoff, C. and Narberhaus, F. (2018) RNA thermometers in bacterial pathogens. *Microbiol. Spectr.*, **6**, RWR-0012–2017.
- Wei, Y. and Murphy, E.R. (2016) Temperature-dependent regulation of bacterial gene expression by RNA thermometers. In: Larramendy, M.L. and Soloneski, S. (eds). *Nucleic Acids—from Basic Aspects to Laboratory Tools Specific*. IntechOpen, London. pp. 157–181.
- Kortmann, J., Szodrok, S., Rinnenthal, J., Schwalbe, H. and Narberhaus, F. (2011) Translation on demand by a simple RNA-based thermosensor. *Nucleic Acids Res.*, **39**, 2855–2868.
- Wright, P.R., Mann, M. and Backofen, R. (2018) Structure and interaction prediction in prokaryotic RNA biology. *Microbiol. Spectr.*, **6**, RWR-0001–2017.
- Mathews, D.H., Moss, W.N. and Turner, D.H. (2010) Folding and finding RNA secondary structure. *Cold Spring Harb. Perspect. Biol.*, **2**, a003665.
- Fica, S.M. and Nagai, K. (2017) Cryo-electron microscopy snapshots of the spliceosome: Structural insights into a dynamic ribonucleoprotein machine. *Nat. Struct. Mol. Biol.*, **24**, 791–799.
- Fang, X., Stagno, J.R., Bhandari, Y.R., Zuo, X. and Wang, Y.X. (2015) Small-angle X-ray scattering: a bridge between RNA secondary structures and three-dimensional topological structures. *Curr. Opin. Struct. Biol.*, **30**, 147–160.
- Kennedy, S.D. (2016) NMR methods for characterization of RNA secondary structure. In: Turner, D.H. and Mathews, D.H. (eds). *RNA Structure Determination: Methods and Protocols*, Humana Press Inc., Totowa, pp. 253–264.

12. Lengyel, J., Hnath, E., Storms, M. and Wohlfarth, T. (2014) Towards an integrative structural biology approach: combining Cryo-TEM, X-ray crystallography, and NMR. *J. Struct. Funct. Genomics*, **15**, 117–124.
13. Huntzinger, E., Possedko, M., Winter, F., Moine, H., Ehresmann, C. and Romby, P. (2005) Probing RNA structures with enzymes and chemicals *in vitro* and *in vivo*. In: Hartmann, R.K., Bindereif, A., Schön, A. and Westhof, E. (eds). *Handb. RNA Biochem.* Wiley-VCH, Weinheim. pp. 151–171.
14. Ignatova, Z. and Narberhaus, F. (2017) Systematic probing of the bacterial RNA structure to reveal new functions. *Curr. Opin. Microbiol.*, **36**, 14–19.
15. Mortimer, S.A., Kidwell, M.A. and Doudna, J.A. (2014) Insights into RNA structure and function from genome-wide studies. *Nat. Rev. Genet.*, **15**, 469–479.
16. Strobel, E.J., Yu, A.M. and Lucks, J.B. (2018) High-throughput determination of RNA structures. *Nat. Rev. Genet.*, **19**, 615–634.
17. Silverman, I.M., Berkowitz, N.D., Gosai, S.J. and Gregory, B.D. (2016) Genome-wide approaches for RNA structure probing. In: Yeo, E.W. (ed). *Advances in Experimental Medicine and Biology*. Springer, Cham, pp. 29–59.
18. Ziehler, W.A. and Engelke, D.R. (2000) Probing RNA structure with chemical reagents and enzymes. *Curr. Protoc. Nucleic Acid Chem.*, **6.1.1–6.1.21**.
19. Ehresmann, C., Baudin, F., Mougél, M., Romby, P., Ebel, J.P. and Ehresmann, B. (1987) Probing the structure of RNAs in solution. *Nucleic Acids Res.*, **15**, 9109–9128.
20. Lowman, H.B. and Draper, D.E. (1986) On the recognition of helical RNA by cobra venom V1 nuclease. *J. Biol. Chem.*, **261**, 5396–5403.
21. Peattie, D.A. (1979) Direct chemical method for sequencing RNA. *Proc. Natl. Acad. Sci. U.S.A.*, **76**, 1760–1764.
22. Peattie, D.A. and Gilbert, W. (1980) Chemical probes for higher-order structure in RNA. *Proc. Natl. Acad. Sci. U.S.A.*, **77**, 4679–4682.
23. Inoue, T. and Cech, T.R. (1985) Secondary structure of the circular form of the *Tetrahymena* rRNA intervening sequence: a technique for RNA structure analysis using chemical probes and reverse transcriptase. *Proc. Natl. Acad. Sci. U.S.A.*, **82**, 648–652.
24. Lempereur, L., Nicoloso, M., Riehl, N., Ehresmann, C., Ehresmann, B. and Bachellerie, J.-P. (1985) Conformation of yeast 18S rRNA. Direct chemical probing of the 5' domain in ribosomal subunits and in deproteinized RNA by reverse transcriptase mapping of dimethyl sulfate-accessible sites. *Nucleic Acids Res.*, **13**, 8339–8357.
25. Incarnato, D., Neri, F., Anselmi, F. and Oliviero, S. (2014) Genome-wide profiling of mouse RNA secondary structures reveals key features of the mammalian transcriptome. *Genome Biol.*, **15**, 491.
26. Mitchell, D., Renda, A.J., Douds, C.A., Babitzke, P., Assmann, S.M. and Bevilacqua, P.C. (2019) *In vivo* RNA structural probing of uracil and guanine base-pairing by 1-ethyl-3-(3-dimethylaminopropyl)carbodiimide (EDC). *RNA*, **25**, 147–157.
27. Wang, P.Y., Sexton, A.N., Culligan, W.J. and Simon, M.D. (2019) Carbodiimide reagents for the chemical probing of RNA structure in cells. *RNA*, **25**, 135–146.
28. Wang, X., Gong, J., Chen, Y., Wu, T., Wang, F., Yang, S., Yuan, Y., Luo, G., Chen, K., Hu, L. *et al.* (2020) Keth-seq for transcriptome-wide RNA structure mapping. *Nat. Chem. Biol.*, **16**, 489–492.
29. Busan, S., Weidmann, C.A., Sengupta, A. and Weeks, K.M. (2019) Guidelines for SHAPE reagent choice and detection strategy for RNA structure probing studies. *Biochemistry*, **58**, 2655–2664.
30. Merino, E.J., Wilkinson, K.A., Coughlan, J.L. and Weeks, K.M. (2005) RNA structure analysis at single nucleotide resolution by selective 2'-hydroxyl acylation and primer extension (SHAPE). *J. Am. Chem. Soc.*, **127**, 4223–4231.
31. Wilkinson, K.A., Merino, E.J. and Weeks, K.M. (2006) Selective 2'-hydroxyl acylation analyzed by primer extension (SHAPE): quantitative RNA structure analysis at single nucleotide resolution. *Nat. Protoc.*, **1**, 1610–1616.
32. McGinnis, J.L., Dunkle, J.A., Cate, J.H.D. and Weeks, K.M. (2012) The mechanisms of RNA SHAPE chemistry. *J. Am. Chem. Soc.*, **134**, 6617–6624.
33. Spitale, R.C., Crisalli, P., Flynn, R.A., Torre, E.A., Kool, E.T. and Chang, H.Y. (2013) RNA SHAPE analysis in living cells. *Nat. Chem. Biol.*, **9**, 18–20.
34. Forconi, M. and Herschlag, D. (2009) Metal ion-based RNA cleavage as a structural probe. In: *Biophysical, Chemical, and Functional Probes of RNA Structure, Interactions and Folding: Part A*. pp. 91–106.
35. Malmgren, C., Wagner, E.G.H., Ehresmann, C., Ehresmann, B. and Romby, P. (1997) Antisense RNA control of plasmid R1 replication. *J. Biol. Chem.*, **272**, 12508–12512.
36. Ciesiolka, J., Hardt, W.-D., Schlegl, J., Erdmann, V.A. and Hartmann, R.K. (1994) Lead-ion-induced cleavage of RNase P RNA. *Eur. J. Biochem.*, **219**, 49–56.
37. Lindell, M., Romby, P. and Wagner, E.G.H. (2002) Lead(II) as a probe for investigating RNA structure *in vivo*. *RNA*, **8**, 534–541.
38. Lindell, M., Brännvall, M., Wagner, E.G.H. and Bra, M. (2005) Lead (II) cleavage analysis of RNase P RNA *in vivo*. *RNA*, **11**, 1348–1354.
39. Valverde, C., Lindell, M., Wagner, E.G.H. and Haas, D. (2004) A repeated GGA motif is critical for the activity and stability of the riboregulator RsmY of *Pseudomonas fluorescens*. *J. Biol. Chem.*, **279**, 25066–25074.
40. Brown, R.S., Dewan, J.C. and Klug, A. (1985) Crystallographic and biochemical investigation of the lead(II)-catalyzed hydrolysis of yeast phenylalanine tRNA. *Biochemistry*, **24**, 4785–4801.
41. Rubin, J.R. and Sundaralingam, M. (1983) Lead ion binding and RNA chain hydrolysis in phenylalanine tRNA. *J. Biomol. Struct. Dyn.*, **1**, 629–646.
42. Brunel, C., Romby, P., Westhof, E., Ehresmann, C. and Ehresmann, B. (1991) Three-dimensional model of *Escherichia coli* Ribosomal 5S RNA as deduced from structure probing in solution and computer modeling. *J. Mol. Biol.*, **221**, 293–308.
43. Ciesiolka, J., Lorenz, S. and Erdmann, V.A. (1992) Structural analysis of three prokaryotic 5S rRNA species and selected 5S rRNA-ribosomal-protein complexes by means of Pb(II)-induced hydrolysis. *Eur. J. Biochem.*, **204**, 575–581.
44. Gornicki, P., Baudin, F., Romby, P., Wiewiorowski, M., Kryzosiak, W., Ebel, J.P., Ehresmann, C. and Ehresmann, B. (1989) Use of lead(II) to probe the structure of large RNA's. Conformation of the 3' terminal domain of *E. coli* 16S rRNA and its involvement in building the tRNA binding sites. *J. Biomol. Struct. Dyn.*, **6**, 971–984.
45. Kirsebom, L.A. and Ciesiolka, J. (2005) Pb²⁺-induced cleavage of RNA. In: Hartman, R.K., Bindereif, A., Schön, A. and Westhof, E. (eds). *Handbook of RNA Biochemistry*. Wiley-VCH, Weinheim. pp. 214–228.
46. Brown, R.S., Hingerty, B.E., Dewan, J.C. and Klug, A. (1983) Pb(II)-catalysed cleavage of the sugar-phosphate backbone of yeast tRNA^{Phe}-implications for lead toxicity and self-splicing RNA. *Nature*, **303**, 543–546.
47. Wan, Y., Qu, K., Ouyang, Z., Kertesz, M., Li, J., Tibshirani, R., Makino, D.L., Nutter, R.C., Segal, E. and Chang, H.Y. (2012) Genome-wide measurement of RNA folding energies. *Mol. Cell*, **48**, 169–181.
48. Righetti, F., Nuss, A.M., Twittenhoff, C., Beele, S., Urban, K., Will, S., Bernhart, S.H., Stadler, P.F., Dersch, P. and Narberhaus, F. (2016) Temperature-responsive *in vitro* RNA structure of *Yersinia pseudotuberculosis*. *Proc. Natl. Acad. Sci. U.S.A.*, **113**, 7237–7242.
49. Sambrook, J. and W Russell, D. (2001) In: *Molecular Cloning: a Laboratory Manual*. Cold Spring Harbor Lab Press, NY.
50. Martin, M. (2011) Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet J.*, **17**, 10–12.
51. Langmead, B., Trapnell, C., Pop, M. and Salzberg, S.L. (2009) Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol.*, **10**, R25.
52. Wan, Y., Qu, K., Ouyang, Z. and Chang, H.Y. (2013) Genome-wide mapping of RNA structure using nuclease digestion and high-throughput sequencing. *Nat. Protoc.*, **8**, 849–869.
53. Kertesz, M., Wan, Y., Mazor, E., Rinn, J.L., Nutter, R.C., Chang, H.Y. and Segal, E. (2010) Genome-wide measurement of RNA secondary structure in yeast. *Nature*, **467**, 103–107.
54. Wan, Y., Qu, K., Zhang, Q.C., Flynn, R.A., Manor, O., Ouyang, Z., Zhang, J., Spitale, R.C., Snyder, M.P., Segal, E. *et al.* (2014) Landscape and variation of RNA secondary structure across the human transcriptome. *Nature*, **505**, 706–709.
55. Darty, K., Denise, A. and Ponty, Y. (2009) VARNA: interactive drawing and editing of the RNA secondary structure. *Bioinformatics*, **25**, 1974–1975.

56. Lowe, T.M. and Chan, P.P. (2016) tRNAscan-SE On-line: integrating search and context for analysis of transfer RNA genes. *Nucleic Acids Res.*, **44**, W54–W57.
57. Xu, Z., Almudevar, A. and Mathews, D.H. (2012) Statistical evaluation of improvement in RNA secondary structure prediction. *Nucleic Acids Res.*, **40**, e26.
58. Washietl, S., Hofacker, I.L., Stadler, P.F. and Kellis, M. (2012) RNA folding with soft constraints: reconciliation of probing data and thermodynamic secondary structure prediction. *Nucleic Acids Res.*, **40**, 4261–4272.
59. Lorenz, R., Bernhart, S.H., zu Siederdisen, C., Tafer, H., Flamm, C., Stadler, P.F. and Hofacker, I.L. (2011) ViennaRNA Package 2.0. *Algorith. Mol. Biol.*, **6**, 26.
60. Smith, C., Heyne, S., Richter, A.S., Will, S. and Backofen, R. (2010) Freiburg RNA tools: a web server integrating IntaRNA, ExpaRNA and LocARNA. *Nucleic Acids Res.*, **38**, 373–377.
61. Will, S., Joshi, T., Hofacker, I.L., Stadler, P.F. and Backofen, R. (2012) LocARNA-P: accurate boundary prediction and improved detection of structural RNAs. *RNA*, **18**, 900–914.
62. Sükösd, Z., Swenson, M.S., Kjems, J. and Heitsch, C.E. (2013) Evaluating the accuracy of SHAPE-directed RNA secondary structure predictions. *Nucleic Acids Res.*, **41**, 2807–2816.
63. Ciesiołka, J., Michałowski, D., Wrzesiński, J., Krajewski, J. and Krzyżosiak, W.J. (1998) Patterns of cleavages induced by lead ions in defined RNA secondary structure motifs. *J. Mol. Biol.*, **275**, 211–220.
64. Chan, C.W., Badong, D., Rajan, R. and Mondragon, A. (2019) Crystal structures of an unmodified bacterial tRNA reveal intrinsic structural flexibility and plasticity as general properties of unbound tRNAs. *RNA*, **26**, 278–289.
65. Wilkinson, K.A., Merino, E.J. and Weeks, K.M. (2005) RNA SHAPE chemistry reveals nonhierarchical interactions dominate equilibrium structural transitions in tRNA(Asp) transcripts. *J. Am. Chem. Soc.*, **127**, 4659–4667.
66. Chen, J. and Pace, N. (1997) Identification of the universally conserved core of ribonuclease P RNA. *RNA*, **3**, 557–560.
67. Haas, E.S., Brown, J.W., Pitulle, C. and Pace, N.R. (1994) Further perspective on the catalytic core and secondary structure of ribonuclease P RNA. *Proc. Natl. Acad. Sci. U.S.A.*, **91**, 2527–2531.
68. Zito, K., Hüttenhofer, A. and Pace, N.R. (1993) Lead-catalyzed cleavage of ribonuclease P RNA as a probe for integrity of tertiary structure. *Nucleic Acids Res.*, **21**, 5916–5920.
69. Keiler, K.C. (2008) Biology of *trans*-translation. *Annu. Rev. Microbiol.*, **62**, 133–151.
70. Zwieb, C., Gorodkin, J., Knudsen, B., Burks, J. and Wower, J. (2003) The nuclear protein database (NPD): sub-nuclear localisation and functional annotation of the nuclear proteome. *Nucleic Acids Res.*, **31**, 446–447.
71. Ivanova, N., Lindell, M., Pavlov, M., Schiavone, L.H., Wagner, E.G.H. and Ehrenberg, M. (2007) Structure probing of tmRNA in distinct stages of *trans*-translation. *RNA*, **13**, 713–722.
72. Mustoe, A.M., Busan, S., Rice, G.M., Hajdin, C.E., Peterson, B.K., Ruda, V.M., Kubica, N., Nutiu, R., Baryza, J.L. and Weeks, K.M. (2018) Pervasive regulatory functions of mRNA structure revealed by high-resolution SHAPE probing. *Cell*, **173**, 181–195.
73. Mustoe, A.M., Lama, N.N., Irving, P.S., Olson, S.W. and Weeks, K.M. (2019) RNA base-pairing complexity in living cells visualized by correlated chemical probing. *Proc. Natl. Acad. Sci. U.S.A.*, **116**, 24574–24582.
74. McGinnis, J.L., Liu, Q., Lavender, C.A., Devaraj, A., McClory, S.P., Fredrick, K. and Weeks, K.M. (2015) In-cell SHAPE reveals that free 30S ribosome subunits. *Proc. Natl. Acad. Sci. U.S.A.*, **112**, 2425–2430.
75. Del Campo, C., Bartholomäus, A., Fedyunin, I. and Ignatova, Z. (2015) Secondary structure across the bacterial transcriptome reveals versatile roles in mRNA regulation and function. *PLoS Genet.*, **11**, e1005613.
76. Nuss, A.M., Heroven, A.K., Waldmann, B., Reinkensmeier, J., Jarek, M., Beckstette, M. and Dersch, P. (2015) Transcriptomic profiling of *Yersinia pseudotuberculosis* reveals reprogramming of the Crp regulon by temperature and uncovers Crp as a master regulator of small RNAs. *PLoS Genet.*, **11**, e1005087.
77. Cimdin, A., Roßmanith, J., Langklotz, S., Bandow, J.E. and Narberhaus, F. (2013) Differential control of *Salmonella* heat shock operons by structured mRNAs. *Mol. Microbiol.*, **89**, 715–731.
78. Lavery, R. and Pullman, A. (1984) A new theoretical index of biochemical reactivity combining steric and electrostatic factors. An application to yeast tRNAPhe. *Biophys. Chem.*, **19**, 171–181.
79. Streicher, B., von Ahsen, U. and Schroeder, R. (1993) Lead cleavage sites in the core structure of group I intron-RNA. *Nucleic Acids Res.*, **21**, 311–317.
80. Bevilacqua, P.C. and Assmann, S.M. (2018) Technique development for probing RNA structure *in vivo* and genome-wide. *Cold Spring Harb. Perspect. Biol.*, **10**, a032250.
81. Talkish, J., May, G., Lin, Y., Woolford, J.L. Jr and Moman, C.J. (2014) Mod-seq: high-throughput sequencing for chemical probing of RNA structure probing of RNA structure. *RNA*, **20**, 713–720.
82. Ritchey, L.E., Su, Z., Tang, Y., Tack, D.C., Assmann, S.M. and Bevilacqua, P.C. (2017) Structure-seq2: sensitive and accurate genome-wide profiling of RNA structure *in vivo*. *Nucleic Acids Res.*, **45**, e135.
83. Underwood, J.G., Uzirov, A. V., Katzman, S. and Onodera, C.S. (2010) FragSeq: transcriptome-wide RNA structure probing using high-throughput sequencing. *Nat. Methods*, **7**, 995–1001.
84. Mortimer, S.A. and Weeks, K.M. (2007) A fast-acting reagent for accurate analysis of RNA secondary and tertiary structure by SHAPE chemistry. *J. Am. Chem. Soc.*, **129**, 4144–4145.
85. Watters, K.E., Yu, A.M., Strobel, E.J., Settle, A.H. and Lucks, J.B. (2016) Characterizing RNA structures *in vitro* and *in vivo* with selective 2'-hydroxyl acylation analyzed by primer extension sequencing (SHAPE-Seq). *Methods*, **103**, 34–48.
86. Diaz-Toledano, R., Lozano, G. and Martinez-Salas, E. (2017) In-cell SHAPE uncovers dynamic interactions between the untranslated regions of the foot-and-mouth disease virus RNA. *Nucleic Acids Res.*, **45**, 1416–1432.
87. Spitale, R.C., Flynn, R.A., Zhang, Q.C., Crisalli, P., Lee, B., Jung, J.W., Kuchelmeister, H.Y., Batista, P.J., Torre, E.A., Kool, E.T. *et al.* (2015) Structural imprints *in vivo* decode RNA regulatory mechanisms. *Nature*, **519**, 486–490.
88. Rouskin, S., Zubradt, M., Washietl, S., Kellis, M. and Weissman, J.S. (2014) Genome-wide probing of RNA structure reveals active unfolding of mRNA structures *in vivo*. *Nature*, **505**, 701–705.
89. Smola, M.J., Calabrese, J.M. and Weeks, K.M. (2015) Detection of RNA-protein interactions in living cells with SHAPE. *Biochemistry*, **54**, 6867–6875.
90. Mitchell, D., Assmann, S.M. and Bevilacqua, P.C. (2019) Probing RNA structure *in vivo*. *Curr. Opin. Struct. Biol.*, **59**, 151–158.
91. Lorenz, R., Wolfinger, M.T., Tanzer, A. and Hofacker, I.L. (2016) Predicting RNA secondary structures from sequence and probing data. *Methods*, **103**, 86–98.
92. Li, F., Zheng, Q., Ryvkin, P., Dragomir, I., Desai, Y., Aiyer, S., Valladares, O., Yang, J., Bambina, S., Sabin, L.R. *et al.* (2012) Global analysis of RNA secondary structure in two metazoans. *Cell Rep.*, **1**, 69–82.
93. Watts, J.M., Dang, K.K., Gorelick, R.J., Leonard, C.W., Bess, J.W., Swanstrom, R., Burch, C.L. and Weeks, K.M. (2009) Architecture and secondary structure of an entire HIV-1 RNA genome. *Nature*, **460**, 711–716.
94. Su, Z., Tang, Y., Ritchey, L.E., Tack, D.C., Zhu, M., Bevilacqua, P.C. and Assmann, S.M. (2018) Genome-wide RNA structure reprogramming by acute heat shock globally regulates mRNA abundance. *Proc. Natl. Acad. Sci. U.S.A.*, **115**, 12170–12175.
95. Ding, Y., Tang, Y., Kwok, C.K., Zhang, Y., Bevilacqua, P.C. and Assmann, S.M. (2014) *In vivo* genome-wide profiling of RNA secondary structure reveals novel regulatory features. *Nature*, **505**, 696–700.
96. Pelechano, V., Wei, W. and Steinmetz, L.M. (2015) Widespread co-translational RNA decay reveals ribosome dynamics. *Cell*, **161**, 1400–1412.
97. Weibezahn, J., Tessarz, P., Schlieker, C., Zahn, R., Maglica, Z., Lee, S., Zentgraf, H., Weber-Ban, E.U., Dougan, D.A., Tsai, F.T.F. *et al.* (2004) Thermotolerance requires refolding of aggregated proteins by substrate translocation through the central pore of ClpB. *Cell*, **119**, 653–665.
98. Bukau, B., Weissman, J. and Horwich, A. (2006) Molecular chaperones and protein quality control. *Cell*, **125**, 443–451.

99. Hartl,F.U., Bracher,A. and Hayer-Hartl,M. (2011) Molecular chaperones in protein folding and proteostasis. *Nature*, **475**, 324–332.
100. Hayer-Hartl,M., Bracher,A. and Hartl,F.U. (2016) The GroEL-GroES chaperonin machine: a nano-cage for protein folding. *Trends Biochem. Sci.*, **41**, 62–76.
101. Kouse,A.B., Righetti,F., Kortmann,J., Narberhaus,F. and Murphy,E.R. (2013) RNA-Mediated thermoregulation of iron-acquisition genes in *Shigella dysenteriae* and pathogenic *Escherichia coli*. *PLoS One*, **8**, e63781.
102. Twittenhoff,C., Heroven,A.K., Mühlen,S., Dersch,P. and Narberhaus,F. (2020) An RNA thermometer dictates production of a secreted bacterial toxin. *PLoS Pathog.*, **16**, e1008184.
103. Ignatov,D., Vaitkevicius,K., Durand,S., Cahoon,L., Sandberg,S.S., Liu,X., Kallipolitis,B.H., Rydén,P., Freitag,N., Condon,C. *et al.* (2020) An mRNA-mRNA interaction couples expression of a virulence factor and its chaperone in *listeria monocytogenes*. *Cell Rep.*, **30**, 4027–4040.
104. Burkhardt,D.H., Rouskin,S., Zhang,Y., Li,G.W., Weissman,J.S. and Gross,C.A. (2017) Operon mRNAs are organized into ORF-centric structures that predict translation efficiency. *Elife*, **6**, e22037.
105. Brown,J.W., Haas,E.S., Gilbert,D.G. and Pace,N.R. (1994) The Ribonuclease P database. *Nucleic Acids Res.*, **17**, 3660–3662.
106. Mustoe,A.M., Lama,N.N., Irving,P.S., Olson,S.W. and Weeks,K.M. (2019) RNA base-pairing complexity in living cells visualized by correlated chemical probing. *Proc. Natl. Acad. Sci. U.S.A.*, **116**, 24574–24582.