Identification of conserved RNA regulatory switches in living cells using RNA secondary structure ensemble mapping and covariation analysis

Ivana Borovská [1,2,9], Chundan Zhang[1,9], Sarah-Luisa J. Dülk [1,3,4], Edoardo Morandi[1], Marta F. S. Cardoso [1], Billal M. Bourkia[1], Daphne A. L. van den Homberg[5], Michael T. Wolfinger [6,7,8], Willem A. Velema [5] & Danny Incarnato [1] ✉

RNA molecules can populate ensembles of alternative structural conformations; however, comprehensively mapping RNA conformational landscapes within living cells presents notable challenges and has, as such, so far remained elusive. Here, we generate transcriptome-scale maps of RNA secondary structure ensembles in both *Escherichia coli* and human cells, uncovering features of structurally heterogeneous regions. By combining ensemble deconvolution and covariation analyses, we report the discovery of several bacterial RNA thermometers in the 5′ untranslated regions (UTRs) of the *cspG, cspI, cpxP* and *lpxP* mRNAs of *Escherichia coli*. We mechanistically characterize how these thermometers switch structure in response to cold shock and reveal the CspE chaperone-mediated regulation of *lpxP*. Furthermore, we introduce a method for the transcriptome-scale mapping of 5′ UTR structures in eukaryotes and leverage it to uncover RNA structural switches regulating the differential usage of open reading frames in the 5′ UTRs of the *CKS2* and *TXNL4A* mRNAs in HEK293 cells. Collectively, this work reveals the complexity of RNA structural dynamics in living cells and provides a resource to accelerate the discovery of regulatory RNA switches.

RNA molecules are pivotal orchestrators of virtually every cellular process, functioning as genetic information carriers and master regulators of gene expression. These roles are intricately intertwined with the ability of RNA molecules to fold into complex structures. Recent strides in the fields of chemical biology and transcriptomics have allowed for concurrent examination of the secondary structures of thousands of transcripts in a single experiment[1]. Although the theoretical folding space of RNA molecules is vast[2] and many RNAs have been reported to populate an ensemble of alternative base-pairing states[3–8], most transcriptome-wide studies have focused on determining a single conformation for each transcript[9–13], limiting the predictive power of such models and hindering our understanding of fundamental regulatory mechanisms. To overcome this limitation, a number of methods[3,4,6,7,14–17], either dependent on or independent of thermodynamics[18], have been devised to deconvolute RNA structural ensembles from chemical probing data. These methods have revealed a growing repertoire of dynamic RNA structural ensembles, typically populating a small number of conformations, within both viral and human RNAs, with crucial regulatory roles[1,19,20]. The SARS-CoV-2 frameshifting element has been reported by multiple studies to

sample a large number of conformations[4,5,21–23], one of which involves a ~1.1-kb-long long-range interaction that appears to be essential for efficient ribosomal frameshifting[5]. The 7SK noncoding RNA that controls P-TEFb, a regulator of the transcriptional elongation, populates an ensemble composed of a P-TEFb-bound and a P-TEFb-unbound conformation, whose relative stoichiometries are dependent on cell transcription and proliferation[6]. The human telomerase RNA component adopts two alternative conformations in cells, the minor of which is characterized by misfolding of the CR4/5 and template/pseudoknot domains and, therefore, is incapable of binding to TERT[8].

We previously introduced DRACO[4], an algorithm capable of deconvolving RNA structure ensembles from chemical probing data read out through mutational profiling[14,24,25] (MaP). In MaP experiments, RNA molecules are first incubated in the presence of a chemical probe, which induces covalent modifications on the RNA at the level of unpaired (or structurally flexible) nucleotides. Sites of chemical modification are then recorded as complementary DNA (cDNA) mutations during reverse transcription (RT) and decoded by high-throughput sequencing. By analyzing comutation patterns in sequencing reads, DRACO can estimate the number of conformations populating the structure ensemble for a given RNA, as well as reconstruct their structures and estimate their relative stoichiometries. Furthermore, unlike the majority of the available ensemble deconvolution methods, which present substantial computational overheads, DRACO is optimized for fast computations, thus enabling transcriptome-wide analyses.

In this study, we perform a transcriptome-scale exploration of RNA secondary structure ensembles in living cells and introduce a generalized framework for the identification of functional regulatory RNA structural switches, by combining DRACO-mediated ensemble deconvolution of transcriptome-wide MaP data and prioritization of functional structures using an automated evolutionary conservation assessment, termed DeConStruct. By mapping structurally heterogeneous regions across the entire *Escherichia coli* transcriptome and their dynamics in response to cold adaptation, we demonstrate that this framework can effectively identify RNA thermometers. Then, by developing a method for transcriptome-wide mapping of 5′ untranslated regions (UTRs) of eukaryotic cells, here named 5′UTR-MaP, we explore the secondary structure ensembles across the 5′ UTRome of human cells, further showing that our framework can identify RNA structural switches regulating open reading frame (ORF) usage. Altogether, our study unveils the complexity of the RNA secondary structure ensemble landscape of living cells.
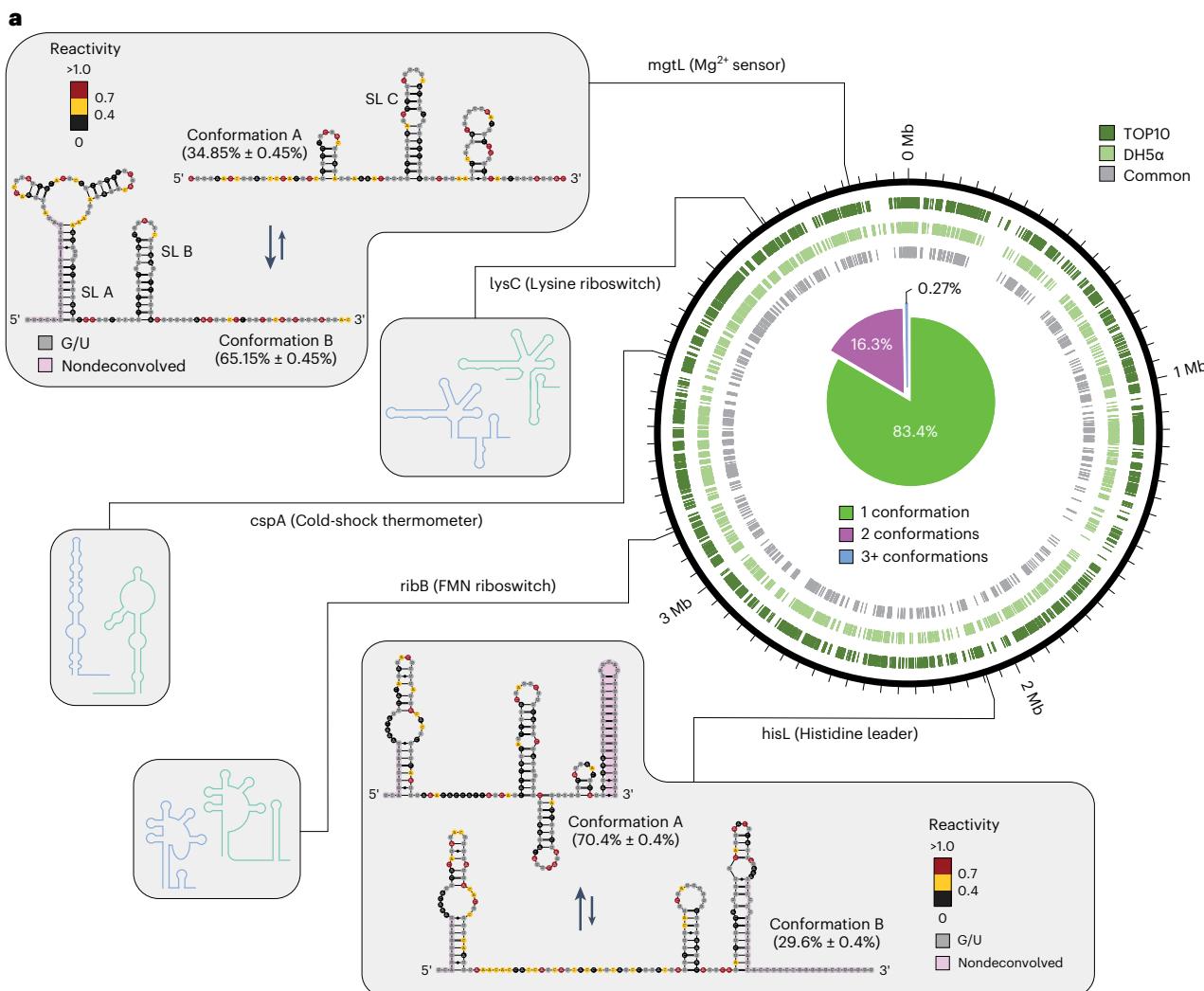
Results

Transcriptome-wide mapping of bacterial RNA structural ensembles identifies general features of structurally heterogeneous RNAs

To comprehensively profile RNA structural ensembles across the *E. coli* transcriptome, we performed in vivo dimethyl sulfate (DMS) probing on exponentially growing DH5α and TOP10 cells at 37 °C. Ribosomal RNA (rRNA)-depleted samples were subjected to DMS-MaPseq analysis, yielding approximately 1 billion paired-end reads for each experiment. Additionally, we generated libraries from total RNA obtained from both in vivo probed cells and ex vivo probed RNA following deproteinization, thereby facilitating the derivation of optimized folding parameters. Analysis of mutation distributions showed an enrichment for mutations on A and C, constituting on average 56.15% ± 0.55% and 34.15% ± 0.55% of all mutations, respectively (Supplementary Fig. 1a), as expected from DMS preference for modifying A and C. Area under the receiver operator characteristic curve analysis of 16S and 23S rRNAs confirmed the enrichment of DMS-induced mutations on unpaired bases of rRNA reference structures (area under the curve: 0.873–0.914; Supplementary Fig. 1b). DH5α and TOP10 bulk (ensemble average) DMS reactivities were highly correlated (r = 0.94, Pearson correlation coefficient; Supplementary Fig. 1c,d). We achieved a minimum sequencing depth

of 5,000× for roughly two thirds (62–66%) of the bases in the *E. coli* expressed transcriptome (transcripts per million mapped reads ≥10; Supplementary Fig. 1e), a coverage threshold we previously showed to be sufficient to ensure robust ensemble deconvolution by DRACO[4]. It is important to point out that such a coverage is only theoretical. Indeed, as DRACO analyzes the transcriptome in sliding windows, only considering reads harboring at least two DMS-induced mutations falling within the window's range, the effective number of reads used for the ensemble deconvolution is substantially lower than the theoretical sequencing depth for that window. To determine whether this sequencing depth would be sufficient to efficiently capture RNA secondary structure ensembles in our data, we selected four known riboswitches (namely, the *ribB* FMN riboswitch[26], the *mgtL* Mg²⁺ sensor[27], the *thiM* TPP riboswitch[28] and the *lysC* lysine riboswitch[29]) located within RNAs spanning different expression levels. We then randomly downsampled the reads mapping to these RNAs and analyzed each sample with DRACO. Notably, DRACO was able to identify multiple conformations in nearly 100% of the cases with an overall effective median read depth of ~1,460×, with the lysine riboswitch being detected across all subsamples at an effective median read depth of just ~416× (Supplementary Fig. 2). To be conservative, we decided to set the minimum effective read depth threshold for DRACO to 2,000× for downstream analyses (Methods). DRACO-mediated ensemble deconvolution of the *E. coli* transcriptome revealed that, among regions populating an equivalent number of conformations in both strains (Supplementary Table 1), encompassing 1,040,669 bases (accounting for over two thirds of the analyzed bases), approximately 16.6% populated two or more conformations (Fig. 1a). When also including regions populating two or more conformations but not necessarily the same number across both strains, we could recover an additional ~1.5% of structurally heterogeneous regions. The regions identified by DRACO included several known riboswitches, such as the *mgtL* Mg²⁺ sensor[27] and the *hisL* histidine leader[30], as well as the *cspA* RNA thermometer[31]. Under the used growth conditions, the cognate ligands for these RNA switches are expected to be abundant. Accordingly, the Mg²⁺ sensor predominantly adopted the conformation encompassing stem loops (SLs) A and B (conformation B: 65.15% ± 0.45%; Fig. 1a, top inset), which is favored at high Mg²⁺ concentrations. Similarly, the histidine leader predominantly favored the attenuated conformation (conformation A: 70.4% ± 0.4%; Fig. 1a, bottom inset), which facilitates complete leader peptide translation in the presence of high histidyl-tRNA levels. Altogether, these results confirm that our approach is indeed suited for the identification of regulatory RNA structural switches from transcriptome-wide chemical probing data.

To elucidate the features distinguishing regions populating a single conformation (hereafter referred to as '1 regions') from those exhibiting two or more conformations (hereafter referred to as '2+ regions'), we first examined the possibility that the separation between these regions might result from differences in their information content. As DRACO relies on A/C comutation patterns to perform ensemble deconvolution, we wondered whether 2+ regions might be enriched in A/C bases and, thus, might possess higher information content, but no such enrichment was observed (A+C in 1 regions: 49.4%, A+C in 2+ regions: 49.1%; P = 0.99, one-tailed Wilcoxon rank-sum test). Subsequently, we performed partition function folding, either unconstrained or constrained by bulk DMS-MaPseq reactivities and optimized folding parameters (Methods and Supplementary Fig. 3), thereby deriving base-pairing probabilities across the entire *E. coli* transcriptome. We calculated median Shannon entropies, a measure of the structural disorder across each base of a transcript, for both 1 and 2+ regions. Unexpectedly, Shannon entropies were significantly higher for 1 regions than for 2+ regions in both unconstrained (P = 7.4 × 10⁻¹⁵, Wilcoxon rank-sum test; Fig. 1b) and experimentally constrained predictions (P = 5.6 × 10⁻³², Wilcoxon rank-sum test; Supplementary Fig. 4a). This result suggested that 2+ regions may predominantly

Fig. 1 | In vivo mapping of *E. coli* RNA structure ensembles. a, Schematic representation of the *E. coli* genome. Regions populating two or more conformations in TOP10 (dark green), DH5α (light green) or both (gray) are indicated. Examples of riboswitches and RNA thermometers known to populate two alternative conformations are shown. For the *mgtL* Mg²⁺ sensor and the *hisL* histidine leader, reconstructed reactivities for the two conformations were averaged across DH5α and TOP10 cells and overlaid on the known structures. Bases falling outside of the region deconvolved by DRACO are marked in pink. Inset, pie chart depicting the percentages of bases in the *E. coli* transcriptome populating one, two or three or more conformations. Only bases populating the same number of conformations in DH5α and TOP10 cells were considered. **b**, Distribution of median Shannon entropies (from unconstrained predictions)

in 1 versus 2+ regions. **c**, Distribution of median unpaired probabilities (from unconstrained predictions) in 1 versus 2+ regions. **d**, Distribution of median bulk (ensemble average) in vivo DMS reactivities in 1 versus 2+ regions. **e**, Distribution of Gini indices calculated on bulk (ensemble average) in vivo DMS reactivities in 1 versus 2+ regions. **f**, Distribution of G+C content in 1 versus 2+ regions. **g**, Distribution of median percentage sequence conservation, calculated on a set of ten Gram-negative bacterial genomes, in 1 versus 2+ regions. For all box plots, boxes span the 25th to the 75th percentile. The center represents the median. Outliers (values below the 25th percentile − 1.5× the IQR or above the 75th percentile + 1.5× the IQR) are not shown. *P* values were calculated using a two-sided Wilcoxon rank-sum test.

occupy well-defined structural states, while 1 regions may exhibit greater disorder and exist in numerous states with lower probability. In line with this hypothesis, we observed that the median probability of bases to be unpaired was significantly higher in 1 regions compared to 2+ regions in unconstrained ($P = 9.8 \times 10^{-44}$, Wilcoxon rank-sum test; Fig. 1c) and even more prominently in experimentally constrained ($P = 1.4 \times 10^{-117}$, Wilcoxon rank-sum test; Supplementary Fig. 4b) predictions. Accordingly, median DMS reactivities were significantly higher and Gini indices were significantly lower in 1 regions as compared to 2+ regions, indicating that 1 regions tend to be less structured than 2+ regions (median reactivity $P = 3.8 \times 10^{-105}$, Gini index $P = 8.4 \times 10^{-133}$, Wilcoxon rank-sum test; Fig. 1d,e). The propensity of 2+ regions toward increased structural orderliness appeared to be at least partly sequence driven, as these regions exhibited a significantly higher G+C content ($P = 1.5 \times 10^{-17}$; Fig. 1f), as well as lower folding free energies than expected for sequences of same dinucleotide composition ($P = 5.4 \times 10^{-13}$, Wilcoxon rank-sum test; Supplementary Fig. 4c), as compared to 1 regions. Furthermore, comparative sequence analysis of ten Gram-negative genomes revealed that 2+ regions are significantly more conserved than 1 regions ($P = 1.1 \times 10^{-50}$, Wilcoxon rank-sum test; Fig. 1g).

We next wondered whether the observed structural heterogeneity of 2+ regions might arise from alternative transcript isoforms generated by alternative promoters and/or terminators[32]. We tested whether 2+ regions were enriched within transcripts generated from alternative promoters but did not observe any enrichment (expected: 10.9%, observed: 9.7%; $P = 0.87$, one-tailed binomial test). Similarly, no enrichment was observed for genes harboring alternative terminators (expected: 4.3%, observed: 3.8%; $P = 0.75$, one-tailed binomial test). Furthermore, we ruled out the possibility that structural heterogeneity of 2+ regions might preferentially arise as a consequence of interactions with small RNAs (sRNAs), as no significant enrichment was observed for experimentally validated sRNA interactions[33] ($P = 0.91$, one-tailed binomial test). We next investigated two additional potential confounding factors: RNA translation and RNA decay. To evaluate the propensity of 2+ regions toward higher structural heterogeneity in absence of any contribution by cellular factors, we extracted RNA from exponentially growing DH5α and TOP10 cells and subjected it to in vitro refolding and DMS-MaPseq analysis. We achieved high correlations and coverage as we did for the in-cell datasets (Supplementary Fig. 5a,b). Overall, in vitro refolded RNAs showed a higher fraction of 2+ regions (~32.3%; Supplementary Fig. 5c and Supplementary Table 2) as compared to in vivo samples. These regions included known RNA switches such as the Mg^{2+} sensor, the molybdenum cofactor riboswitch[34] and the *cspA* RNA thermometer (Supplementary Fig. 5d–f). Notably, when focusing on bases populating a consistent number of conformations across both strains, common to both in vivo and in vitro datasets, we observed that over two thirds (67.3%) of those encompassing in vivo-identified 2+ regions appeared to be structurally heterogeneous under in vitro conditions as well, with 94.1% of them populating the same number of conformations in cell and in vitro and 5.9% showing increased structural heterogeneity under in vitro conditions (Supplementary Fig. 5g and Supplementary Table 3). It is also worth noticing that, upon reanalysis of published ribosome profiling data[35], we observed that regions exhibiting structural heterogeneity (2+) under in vitro conditions but not in cell showed significantly higher ribosome occupancy and translation efficiencies as compared to regions being heterogeneous both in vivo and in vitro, or only in vivo ($P = 5.1 \times 10^{-10}$ and 4.8×10^{-11}, Wilcoxon rank-sum test; Supplementary Fig. 5h). This is in line with previous findings[13,36] showing that translation can actively unfold RNA structures in cells, thus suggesting that it might be partly masking the actual structural heterogeneity of cellular RNAs.

We further polished this high-confidence set of 2+ regions by discarding those overlapping with experimentally determined RNase E cleavage sites[37], for which the observed heterogeneity might derive from the presence of decay fragments. Importantly, feature reanalysis for this subset did not affect the aforementioned differences between 1 and 2+ regions (Supplementary Fig. 6). Furthermore, as 1 regions were on average ~3-fold less covered than 2+ regions and, as such, might be polluted by putative 2+ regions that were not deconvolved by DRACO, we also analyzed the same set of features in 2+ regions as compared to a random set of transcriptome regions of matching size and observed the exact same trends (Supplementary Fig. 7).
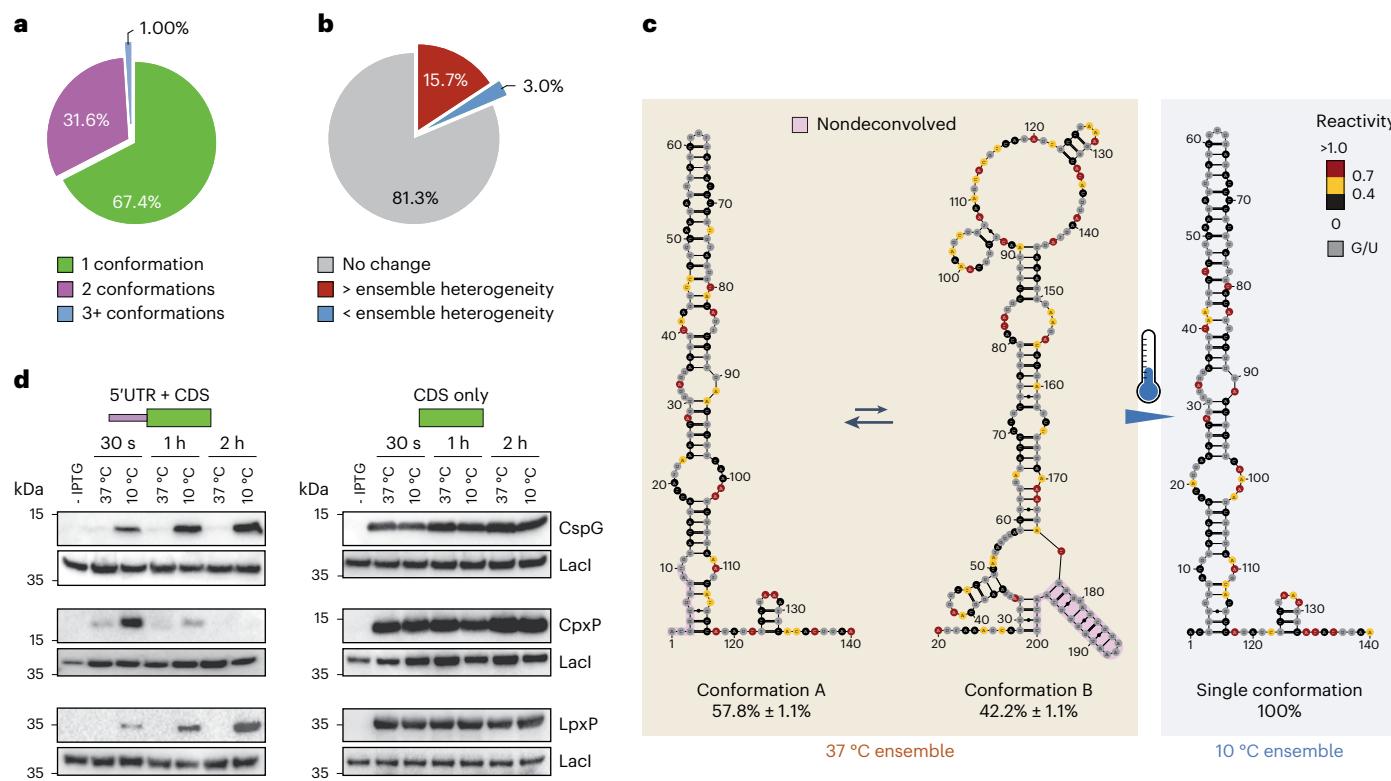
A generalized framework based on automatic covariation analysis accelerates the discovery of RNA regulatory switches

Sequence analysis of 2+ regions showed that they possess a substantially higher median percentage conservation, as compared to both 1 regions and random transcriptome regions (Fig. 1g and Supplementary Fig. 7f), thus suggesting that 2+ regions might be enriched for conserved RNA structural regulatory elements.

To evaluate this possibility, we implemented a framework, dubbed DeConStruct framework (Supplementary Fig. 8), which builds on top of the cm-builder method we previously introduced[22,38] but is greatly expanded to automatically build alignments of related sequences from a representative set of bacterial genomes, and validated it on known RNA switches (Supplementary Fig. 9). Firstly, we identified a total of 901 structurally heterogeneous regions whose DRACO-deconvolved reactivity profiles could be nonambiguously matched between DH5α and TOP10 cell in vivo data and generated experimentally informed RNA structure models. After discarding regions encompassing any known RNA structure element from RFAM and applying stringent alignment selection criteria (Methods), our framework identified 226 regions (~26.5%) for which at least one of the conformations showed robust covariation support, generally regarded as strong evidence of RNA structure functionality, as determined by R-scape[39] analysis. Taken together, our data hint at the existence of a previously unappreciated repertoire of functional regulatory RNA elements, likely including regulatory RNA switches, in bacterial RNAs. To facilitate the analysis and exploration of these regions, we further aggregated them into a browsable website (https://www.incarnatolab.com/datasets/Ensembles_Borovska_2025/). It is worth emphasizing that, because of a number of different factors (namely, inaccuracies in the thermodynamic model and in the resulting structure predictions, the used set of representative bacterial genomes and the stringent alignment selection criteria), this set likely represents an underestimate of the actual number of conserved, structurally heterogeneous regions in the *E. coli* transcriptome. Accordingly, approximately 40% of the regions showed at least one covarying base pair.

We next sought to exploit our framework to identify conserved bacterial RNA thermometers. We focused on cold shock response, as changes in RNA structure have been previously reported to be one of the hallmarks of cold adaptation in bacteria[35,40]. We therefore performed DMS-MaPseq analysis of exponentially growing *E. coli* cells shocked at 10 °C for 20 min. It has been previously shown that DMS reaction kinetics is slower at 10 °C than at 37 °C and, therefore, reaction times need to be increased to achieve similar modification rates[11,35]. We wondered whether this might cause artifacts as, during the timeframe of DMS modification, which would exceed the average half-life of *E. coli* mRNAs upon cold shock[41], the expressed transcriptome changes substantially. To this end, we shocked *E. coli* cells for 20 min at 10 °C, treated them with DMS for 2 or 30 min and compared their expression profiles to those of DMS-untreated cells by RNA-seq (Supplementary Fig. 10a). Gene expression analysis showed that, while DMS-untreated cells underwent the expected transcriptome changes between 2 and 30 min (for example, a robust upregulation of mRNAs encoding cold-induced cold shock proteins, Csps; $R^2 = 0.88$, Pearson correlation coefficient), no change occurred for cells treated with DMS ($R^2 = 0.99$, Pearson correlation coefficient). These data indicate that, despite slowed down reaction kinetics at 10 °C, addition of DMS nearly immediately blocks

Fig. 2 | Redistribution of RNA structure ensembles upon cold shock.
a, Pie chart depicting the percentages of bases in the *E. coli* transcriptome populating one, two or three or more conformations after cold shock. Only bases populating the same number of conformations in DH5α and TOP10 cells are considered. **b**, Pie chart depicting the percentages of bases in the *E. coli* transcriptome, for transcripts expressed both at 37 °C and at 10 °C, for which the ensemble heterogeneity increases (red), decreases (blue) or remains unchanged (gray) after cold shock. **c**, RNA secondary structure ensemble analysis of the

cspA RNA thermometer at 37 °C (left) and after cold shock (right). Reactivities were averaged across DH5α and TOP10 cells and overlaid on the predicted structures. Bases falling outside of the region deconvolved by DRACO are marked in pink. **d**, Western blot analysis of FLAG-tagged CspG, CpxP and LpxP expression at 37 °C and 10 °C, 30 min, 1 h and 2 h after IPTG induction and cold shock for constructs harboring both the 5′ UTR and the CDS (left) or only the CDS (right). LacI was used as the loading control. Analysis is representative of two independent biological replicates.

all cellular processes, including transcription and RNA decay, thus providing an instantaneous snapshot of the RNA structurome.

Cold-shocked cells showed well-correlated DMS reactivities ($r = 0.95$, Pearson correlation coefficient; Supplementary Fig. 10b) and formed a well-separated cluster with respect to exponentially growing cells at 37 °C in principal component analysis (Supplementary Fig. 10c), indicating the existence of substantial structural rearrangements at 10 °C. Surprisingly, DRACO-mediated ensemble deconvolution showed that, among regions populating an equivalent number of conformations in both strains (Supplementary Table 4), totaling 807,853 bases, approximately 32.6% populated two or more conformations, corresponding to a nearly twofold increase as compared to cells grown at 37 °C (Fig. 2a). When focusing solely on regions covered both at 37 °C and 10 °C, we observed that over fivefold more regions showed increased ensemble heterogeneity and, thus, populated a higher number of conformations at 10 °C than those with reduced ensemble heterogeneity (increased heterogeneity at 10 °C: ~15.7%, decreased heterogeneity at 10 °C: ~3%; Fig. 2b and Supplementary Table 5). While this increase in structural diversity upon cold shock might seem counterintuitive to traditional thermodynamic expectations, which predict fewer structural states at lower temperatures, we suggest that the decrease in temperature might reduce the entropy of 1 regions, promoting higher structuredness and fewer, more well-defined structural states.

Among the fewer regions that exhibited reduced ensemble heterogeneity upon cold shock, we observed the well-known *cspA* RNA thermometer[31]. Previous studies proposed that *cspA* can switch between a translationally incompetent conformation at 37 °C and a translationally competent conformation at 10 °C (refs. 31,35). However, such a model

cannot explain why, at 37 °C, CspA is among the top 10% expressed proteins in the *E. coli* proteome[42]. Accordingly, ensemble deconvolution analysis showed that, at 37 °C, the 5′ UTR of *cspA* populates two conformations, with the translationally competent conformation being the predominant one (conformation A: 57.8% ± 1.1%) and becoming the sole conformation upon cold shock (Fig. 2c). Interestingly, the two conformations of *cspA* could also be observed in our in vitro refolded dataset (Supplementary Fig. 5f), albeit with inverted stoichiometries (conformation A: 45.25% ± 0.55%, conformation B: 54.75% ± 0.55%), thus indicating that the cellular environment has a key role in determining conformation abundances in the ensemble. Ontology analysis of genes containing regions undergoing ensemble redistribution upon cold shock showed a significant enrichment for terms associated with response to temperature changes (response to heat $P = 2.3 \times 10^{-5}$, response to cold $P = 3.7 \times 10^{-4}$), stress response ($P = 1.2 \times 10^{-3}$), and pathways commonly modulated in response to cold shock, such as glycolysis ($P = 6.2 \times 10^{-7}$), fatty acid biosynthesis ($P = 1.5 \times 10^{-4}$), lipid biosynthesis and lipid metabolism ($P = 4.1 \times 10^{-6}$ and 3.4×10^{-3}) and protein folding and unfolding ($P = 7.7 \times 10^{-3}$ and 9.2×10^{-5}). Additionally, reanalysis of publicly available ribosome profiling data[35] indicated a moderate yet significant increase in translation efficiency for both genes encompassing regions of differential structural heterogeneity upon cold shock, 10 min after the temperature shift to 10 °C, as compared to 37 °C ($P = 1.1 \times 10^{-5}$, paired Wilcoxon rank-sum test; Supplementary Fig. 10d). This increase was not observed for genes encompassing regions whose RNA ensemble heterogeneity remained unchanged. Collectively, these data suggest the existence of a wide catalog of undiscovered RNA thermometers in bacteria.

To further explore this possibility, we selected three genes whose 5′ UTRs encompassed regions of increased structural heterogeneity upon cold shock and that displayed increased translation efficiency at 10 °C by ribosome profiling, as well as structural conservation, as revealed by our DeConStruct framework (namely, *cspG*, *cpxP* and *lpxP*). The *cspG* mRNA encodes for a Csp that belongs to the same family of CspA. *cspG* was previously shown to be robustly induced by cold shock at the transcriptional level[43,44]; however, to the best of our knowledge, its regulation at the translational level and the role of its 5′ UTR as an RNA thermometer, have not been reported. The *cpxP* mRNA encodes a periplasmic protein involved in sensing and mediating the adaptation to various cellular stresses that might result in protein misfolding[45,46], which is crucial for responses to heat and cold shock[47]. The *lpxP* mRNA encodes a palmitoleoyl transferase that catalyzes the palmitoylation of lipid A to maintain optimal outer membrane fluidity at low temperature[48]. We first tested whether the 5′ UTRs of these mRNAs had a role in regulating their translation upon cold shock by cloning these genes, with or without their 5′ UTRs, in IPTG-inducible constructs and by measuring their expression at 37 °C versus 10 °C. All three genes showed minimal to no expression at 37 °C but robust translation upon cold shock, while deletion of their 5′ UTRs abrogated their cold-mediated regulation (Fig. 2d). *cspG* and *lpxP* both showed increasing translation over a 2-h time course, while *cpxP* expression quickly increased within the first 30 min of cold shock and then rapidly decreased after 1 h.

We next wondered whether temperature shift alone would be sufficient to remodel the structure of the 5′ UTR of these genes. We, therefore, performed in vitro transcription of these mRNAs, either at 37 °C or 10 °C, followed by DMS-MaPseq analysis. Notably, both *cspG* and *cpxP* showed substantial structural rearrangements at 10 °C as compared to 37 °C (Extended Data Fig. 1a and Supplementary Fig. 11a). These temperature-induced RNA structural switches were supported by extensive covariation (Extended Data Fig. 1b and Supplementary Fig. 11b), underscoring their functional relevance. We further asked whether the regulation observed for the 5′ UTR of *cspG* was also shared by other cold-induced members of the Csp family. In particular, the 5′ UTR of *cspB* shares extensive sequence identity with the 5′ UTR of *cspG* (Supplementary Fig. 12a) while that of *cspI* diverges. *cspI* was previously shown to be subjected to translational control by its 5′ UTR but, just like for *cspG*, its putative role as an RNA thermometer was not investigated[49]. Intriguingly, all three 5′ UTRs showed a similar behavior, characterized by a large SL structure encompassing the entire 5′ UTR and part of the coding region, at 37 °C, which sequestered both the ribosome-binding site (RBS) and the start codon, and a register-shifted, shorter, SL structure involving the sole 5′ UTR at 10 °C, which left the RBS and the start codon available for translation initiation (Extended Data Fig. 1a and Supplementary Fig. 12b,c).

Analogously to the 5′ UTRs of Csp-encoding mRNAs, the 5′ UTR of *lpxP* showed two nearly equimolar conformations at 10 °C in cell (Fig. 3a), respectively characterized by an SL structure (conformation A) sequestering both RBS and start codon, which is also the single predominant conformation at 37 °C (Fig. 3b), and a register-shifted SL (SL$_{alt}$, conformation B) leaving both elements available for translation initiation. Both SL and SL$_{alt}$ showed extensive covariation support (Fig. 3c) and a covariance model (CM)-guided homology search revealed the existence of homologous structures in several other Gram-negative bacteria (Supplementary Fig. 13). We confirmed that conformation B corresponded to the translation-competent conformation by generating an SL$_{alt}$-stabilized mutant and validated its structure by targeted DMS-MaPseq analysis (Supplementary Fig. 14a). Stabilization of SL$_{alt}$ abrogated the cold-mediated regulation of *lpxP*, leading to its constitutive expression at both 37 °C and 10 °C (Fig. 3d). To further confirm that the observed regulation was indeed structure mediated, rather than being caused by other factors such as altered mRNA decay of the SL$_{alt}$-stabilized mutant, we adopted the PURE system[50],
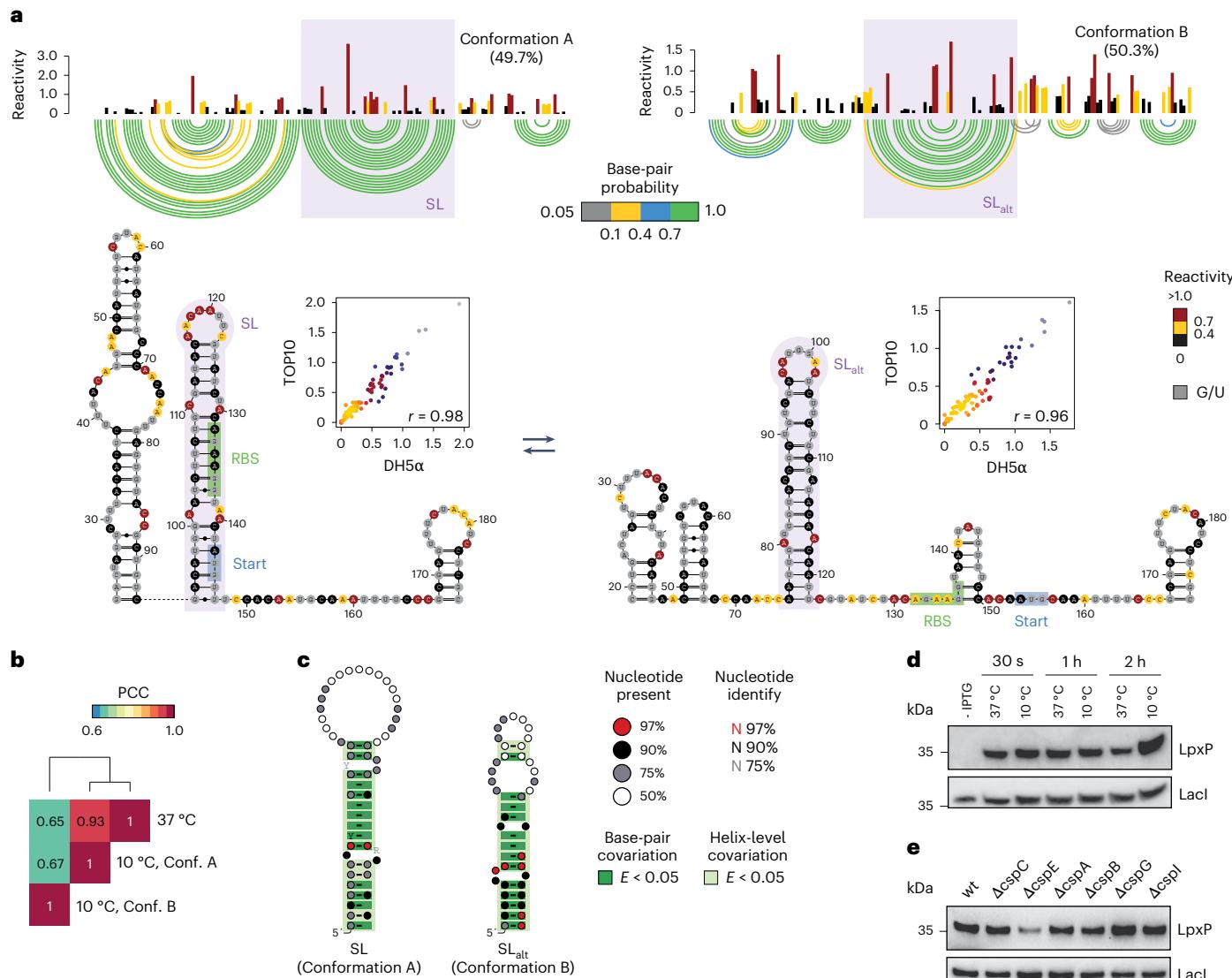
a reconstituted *E. coli* in vitro transcription–translation system. Expression of *lpxP* harboring the wild-type 5′ UTR could not be detected after 2 h at 37 °C, while the template with the coding sequence (CDS) only and the SL$_{alt}$-stabilized mutant were robustly translated (Supplementary Fig. 14b,c).

Unlike the 5′ UTRs of Csp-encoding mRNAs and of *cpxP*, however, the *lpxP* 5′ UTR showed no structure rearrangement between 37 °C and 10 °C under in vitro conditions, solely populating the translationally incompetent conformation at both temperatures (Supplementary Fig. 15). We hypothesized that the energy barrier between the translationally competent and translationally incompetent conformations might be so high that a switch from SL to SL$_{alt}$ could not happen spontaneously on a biologically relevant timescale; as such, it might be a chaperone-assisted process. Interestingly, reanalysis of published CspC and CspE CLIP-seq data from septicemic *E. coli*[51] showed binding of both proteins to the 5′ UTR of *lpxP*. Similarly, RNA immunoprecipitation sequencing analysis of Csp proteins in *Salmonella enterica* serovar Typhimurium[52], which we found to also carry a homologous structural switch (Supplementary Fig. 13), showed binding of CspE and, to a lesser extent, CspC to the 5′ UTR of *lpxP*, hinting at a highly conserved regulatory mechanism. To confirm the role of Csp proteins in regulating the structural switch in the *lpxP* 5′ UTR, we analyzed *lpxP* translation in *E. coli* Csp knockouts from the KEIO collection[53] and observed a ~50% reduction in LpxP expression in ΔCspE cells (Fig. 3e), which was not observed for the SL$_{alt}$-stabilized mutant (Supplementary Fig. 14d).

Transcriptome-scale ensemble deconvolution analysis of the human 5′UTRome identifies RNA structural switches regulating ORF usage

Armed with this new powerful framework, we next adopted it to identify regulatory RNA structural switches regulating the translation of human mRNAs. We decided to focus on 5′ UTRs, which are known to harbor a wide repertoire of translation regulatory elements[54]. However, as the human transcriptome is much larger than that of *E. coli* and ensemble deconvolution analysis is highly demanding in terms of sequencing depths, we could not resort to standard transcriptome-wide DMS-MaPseq analysis. Furthermore, because of their high G+C content and being positioned at the 5′-most end of mRNAs, 5′ UTRs tend to be underrepresented in traditional RNA-seq-like libraries. To address this, we sought to develop a method, here dubbed 5′UTR-MaP, which combines chemical probing and MaP analysis with the selective enrichment of 7-methylguanosine-capped RNA fragments (Extended Data Fig. 2). We applied 5′UTR-MaP to HEK293 cells treated with DMS in vivo at 37 °C, generating ~800 million reads across two biological replicate experiments, and achieved robust enrichment of mRNA 5′ ends (Fig. 4a and Supplementary Fig. 16a) as compared to standard DMS-MaPseq[25]. We further confirmed that 5′UTR-MaP can be also adopted for the analysis of SHAPE-treated samples, by generating libraries in HEK293 treated with 2-aminopyridine-3-carboxylic acid imidazolide[55] (2A3) (Supplementary Fig. 16b). 5′UTR-MaP successfully captured known RNA structure elements within human 5′UTRs[56,57] (Fig. 4b and Supplementary Fig. 16c). 5′UTR-MaP analysis of DMS-treated HEK293 cells exhibited high reproducibility (r = 0.94, Pearson correlation coefficient; Supplementary Fig. 16d) and successfully captured the enrichment for A/C mutations expected for DMS-treated RNAs (A: 41.25% ± 0.25%, C: 42.65% ± 0.95%; Supplementary Fig. 16e).

DRACO-mediated ensemble deconvolution of the HEK293 5′ UTRome revealed that, among regions populating an equivalent number of conformations in both replicate experiments (Supplementary Table 6), encompassing 148,461 bases across 2,240 5′ UTRs, approximately 7% populated two or more conformations, encompassing 183 5′ UTRs (Fig. 4c). When also including regions populating two or more conformations but not necessarily the same number across both experiments, we could recover an additional ~1% of structurally
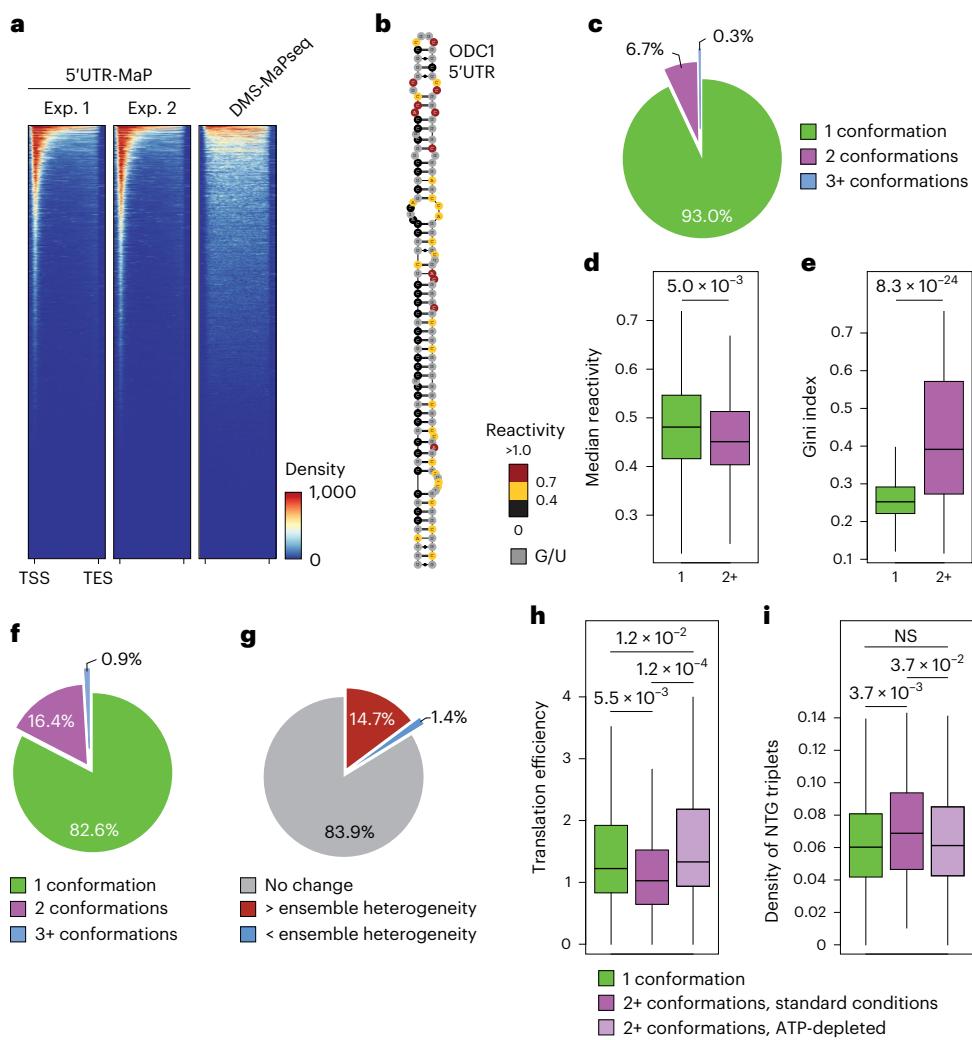
Fig. 3 | Characterization of the *lpxP* RNA thermometer. a, Secondary structure models for the two conformations of the *lpxP* 5′ UTR as identified by ensemble deconvolution analysis of cold-shocked bacteria, with overlaid in vivo DMS reactivities at 10 °C, along with reactivity profiles and base-pairing probabilities for both conformations. Reactivities are averaged across DH5α and TOP10 cells. The register-shifted SL and SL$_{alt}$ are highlighted in purple. Insets, scatter plot depicting the correlation of base reactivities for the deconvolved conformations across DH5α and TOP10 cells. **b**, Heat map of pairwise Pearson correlation coefficients (PCC) of normalized DMS reactivities across the two alternative conformations of the *lpxP* 5′ UTR at 10 °C and the sole 37 °C conformation. **c**, Structure models for SL and SL$_{alt}$ inferred by phylogenetic analysis. Base pairs showing significant covariation (as determined by R-scape) are boxed in dark green (*E* < 0.05). Helices showing helix-level covariation support (*E* < 0.05) are boxed in light green. **d**, Western blot analysis of SL$_{alt}$-stabilized mutant expression at 37 °C and 10 °C, 30 min, 1 h and 2 h after IPTG induction and cold shock. LacI was used as the loading control. Analysis is representative of two independent biological replicates. **e**, Western blot analysis of full-length (5′ UTR + CDS) FLAG-tagged LpxP expression at 10 °C, 1 h after IPTG induction, in wild-type or in Csp single-knockout clones from the KEIO collection. LacI was used as the loading control. Analysis is representative of two independent biological replicates.

heterogeneous regions. As observed for *E. coli* RNAs, 2+ regions of the HEK293 5′ UTRome showed significantly lower median reactivity and a higher Gini index as compared to 1 regions (median reactivity *P* = 5.0 × 10⁻³, Gini index *P* = 8.3 × 10⁻²⁴, Wilcoxon rank-sum test; Fig. 4d,e). A similar trend was also observed for Shannon entropies and G+C content, although the differences were not significant (Shannon entropy *P* = 0.09, G+C *P* = 0.14, Wilcoxon rank-sum test). Previous studies showed that 5′ UTR RNA structures can be actively unfolded by a number of adenosine triphosphate (ATP)-dependent RNA helicases[54,58], suggesting that our survey of structurally heterogeneous regions might represent an underestimate of the actual number of putative RNA structural switches in human 5′ UTRs. To investigate this possibility, we generated two replicate 5′UTR-MaP experiments in ATP-depleted HEK293 cells (Supplementary Fig. 17a). As expected, ATP depletion

resulted in an overall increase in 5′ UTR structuredness as compared to standard culture conditions, as demonstrated by a robust reduction in median reactivity and increase in Gini index for 5′ UTRs upon ATP depletion (Supplementary Fig. 17b,c). Furthermore, DRACO-mediated ensemble deconvolution revealed a robust increase in RNA structural heterogeneity in ATP-depleted cells as compared to cells grown under standard conditions. Of the regions populating an equivalent number of conformations in both replicate experiments in ATP-depleted cells (Supplementary Table 7), encompassing 153,283 bases across 2,546 5′ UTRs, approximately 17.3% populated two or more conformations, encompassing 511 5′ UTRs (Fig. 4f). When also including regions populating two or more conformations but not necessarily the same number across both experiments, we could recover an additional ~3.4% of structurally heterogeneous regions. When focusing solely on

Fig. 4 | In vivo mapping of HEK293 5′ UTR RNA structure ensembles. a, Heat map depicting the enrichment of reads at the 5′ end of transcripts in 5′UTR-MaP experiments, as compared to standard DMS-MaPseq[25]. TES, transcription end site. **b**, Known structure of SL1 in the 5′ UTR of the *ODC1* mRNA, as captured by 5′UTR-MaP[57]. **c**, Pie chart depicting the percentages of bases in the 5′ UTRome of HEK293 cultured under standard conditions, populating one, two or three or more conformations. Only bases populating the same number of conformations in both replicate experiments were considered. **d**, Distribution of median bulk (ensemble average) in vivo DMS reactivities in 1 versus 2+ regions. **e**, Distribution of Gini indices calculated on bulk (ensemble average) in vivo DMS reactivities in 1 versus 2+ regions. **f**, Pie chart depicting the percentages of bases in the 5′ UTRome of HEK293 upon ATP depletion, populating one, two or three or more conformations. Only bases populating the same number of conformations in both replicate experiments were considered. **g**, Pie chart depicting the percentages of bases in the 5′ UTRome of HEK293, for transcripts expressed under both standard and ATP-depleted conditions, for which the ensemble heterogeneity increases (red), decreases (blue) or remains unchanged (gray) upon ATP depletion. **h**, Distribution of translation efficiencies for genes whose 5′ UTRs encompass regions always populating 1 or 2+ conformations under standard or ATP-depleted conditions. NS< not significant. **i**, Distribution of densities of NTG triplets within regions populating 1 or 2+ conformations under standard or ATP-depleted conditions. For all box plots, boxes span the 25th to the 75th percentile. The center represents the median. Outliers (values below the 25th percentile − 1.5× the IQR or above the 75th percentile + 1.5× the IQR) are not shown. *P* values were calculated using a two-sided Wilcoxon rank-sum test.

regions covered both under standard and ATP-depleted conditions, we observed that over 10.5-fold more regions showed increased ensemble heterogeneity and, thus, populated a higher number of conformations upon ATP depletion than those with reduced ensemble heterogeneity (increased heterogeneity upon ATP depletion: ~14.7%, decreased heterogeneity upon ATP depletion: ~1.4%; Fig. 4g and Supplementary Table 8).

We next investigated whether differences in 5′ UTR structural heterogeneity were associated with different levels of translation. Reanalysis of published ribosome profiling data[59] revealed that 5′ UTRs populating a single conformation were significantly more efficiently translated as compared to those populating multiple conformations under standard conditions ($P = 5.5 \times 10^{-3}$, Wilcoxon rank-sum test) but slightly less efficiently translated than those whose heterogeneity

increased upon ATP depletion ($P = 1.2 \times 10^{-2}$, Wilcoxon rank-sum test; Fig. 4h). We wondered whether structurally heterogeneous 5′ UTRs might contain regulatory elements that would result into decreased translation efficiencies. Some of the better-characterized 5′ UTR elements that can lead to translation repression are G-quadruplexes[60–62] (G4s) and upstream ORFs[63–65] (uORFs). While G4-forming regions from publicly available rG4-seq datasets[66,67] did not show any significant enrichment within 2+ regions under standard conditions, we did observe a significantly higher density of NTG triplets as compared to both 1 regions ($P = 3.7 \times 10^{-3}$, Wilcoxon rank-sum test) and 2+ regions ($P = 3.7 \times 10^{-2}$, Wilcoxon rank-sum test) under ATP-depleted conditions (Fig. 4i), suggesting that coexisting alternative RNA secondary structures within these 5′ UTRs might have a role in regulating the usage of uORFs versus the main ORF.

To investigate this hypothesis, after generating DMS-MaPseq data for human rRNAs from total RNA and using it to derive optimized folding parameters as previously done for *E. coli* (Supplementary Fig. 18), we modeled experimentally informed RNA structures of 2+ regions under standard conditions. Once again, we leveraged our DeConStruct framework to prioritize putative functional RNA switches, encompassing 5′ UTR regions with a high density of NTG triplets. To facilitate the extraction of candidate homologous sequences from related genomes, we modified the framework to automatically build a database of related sequences by directly extracting (and degapping) the relevant portions from precomputed multiple genome alignments (that is, the multiz100way alignment of 100 species). While human 5′ UTR structures typically showed fewer covarying pairs than bacterial RNAs, likely because of the lower number of sequences in the analyzed alignments, they often showed helix-level covariations. We selected 5′ UTRs from two mRNAs: CDC28 protein kinase regulatory subunit 2 (*CKS2*) and thioredoxin-like 4A (*TXNL4A*). Both mRNAs formed two alternative conformations, confirmed by targeted DMS-MaPseq analysis (Fig. 5a and Extended Data Fig. 3a), for which the DeConStruct framework could identify notable covariation support to different extents (Fig. 5b and Extended Data Fig. 3b). For *CKS2*, we could nonambiguously pinpoint the uORF start codon by reanalyzing publicly available ribosome profiling data from lactimidomycin-treated HEK293 cells[68]. To simultaneously measure translation from both the main ORF and the uORF, we designed a vector carrying the 5′ UTR of *CKS2* plus the first 20 codons of the main ORF, encompassing the region identified by DRACO to be structurally heterogeneous, fused in frame to the sequence encoding EGFP, followed by the sequence encoding mCherry, positioned on the same frame as the uORF (Fig. 5c). As this alternative frame runs uninterrupted across the entire length of the EGFP, translation from the uORF results in the production of mCherry. We then designed mutations aimed at stabilizing either one of the two conformations sampled by *CKS2* 5′ UTR, taking care not to change the encoded amino acids on both frames. Fluorescence measurement across wild-type and mutant reporters showed that the stabilization of conformation A resulted into a reduction of translation of the main ORF while leaving translation from the uORF essentially unaltered. Oppositely, stabilization of conformation B did not affect translation of the main ORF but increased translation of the uORF. As the mCherry signal tends to be weaker than that of EGFP, we further confirmed these observations by replacing the mCherry with an HA tag and measuring the expression levels of the proteins translated from the two frames by western blotting (Fig. 5d). Western blot analysis confirmed the fluorescence results, showing an even stronger effect of the conformation-stabilizing mutations. For TXNL4A, instead, we could not easily pinpoint a predominant uORF from ribosome profiling data; therefore, we selected the top-scoring uORF after averaging out the results from two algorithms for the prediction of translation start sites[69–71] and designed our reporter by aligning its frame to the mCherry frame. Given the extremely high G+C content and predicted stability of conformation A, we could not easily design mutations to stabilize it, whereas we could design mutations stabilizing conformation B. Notably, stabilization of conformation B resulted in a robust increase of translation from the uORF, while leaving translation from the main ORF nearly unaltered (Extended Data Fig. 3c). Furthermore, we confirmed translation of the uORF by mutagenizing the uORF start codon from CUG to CCG, which reduced mCherry expression nearly twofold.

Discussion
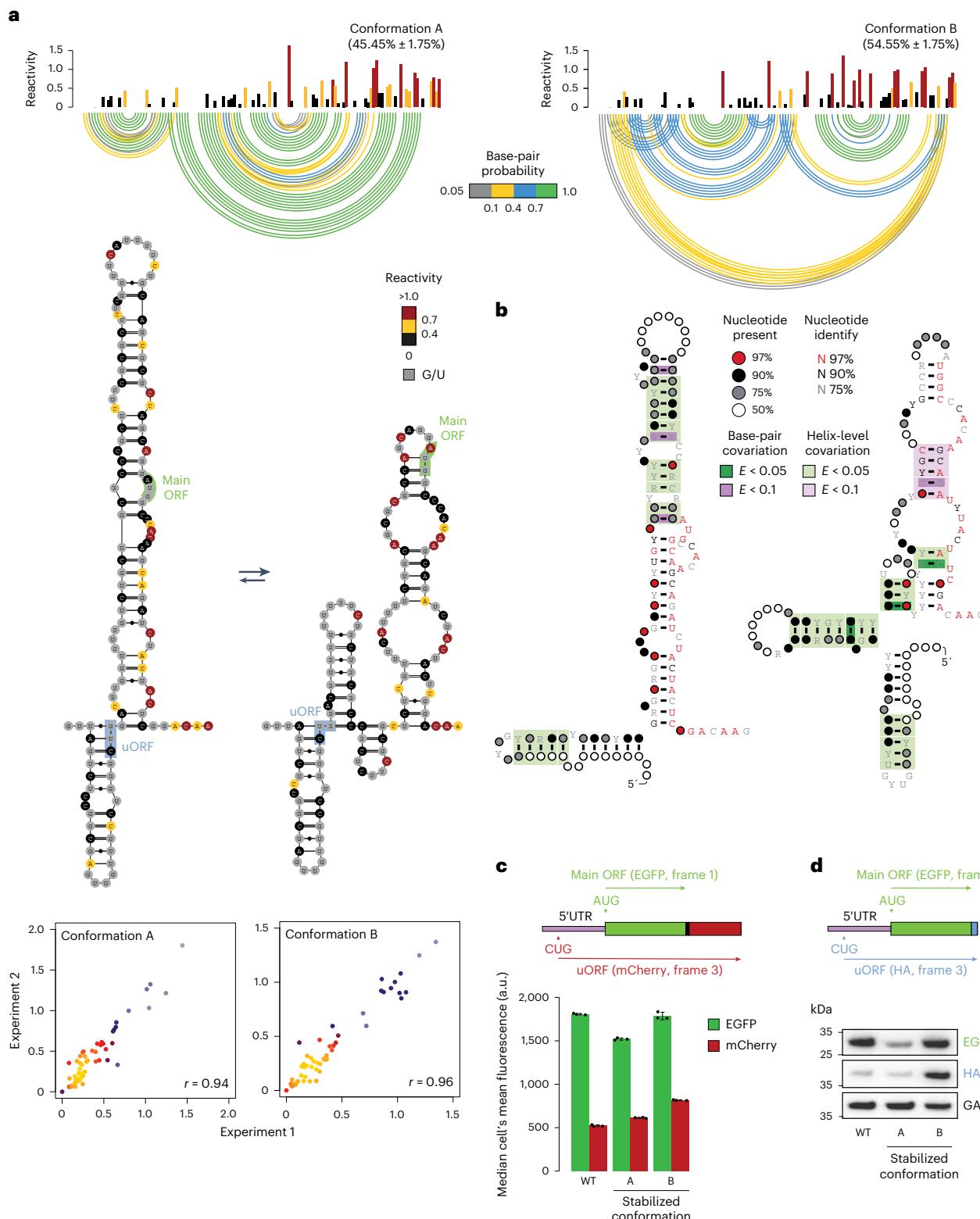
The ability of RNA molecules to undergo conformational changes in response to both internal and external signals makes it crucial to understand their structural dynamics to clarify their role in fine-tuning gene expression and their structure–function relationship at large.

In this study, we reported transcriptome-scale maps of RNA secondary structure ensembles in living cells and introduced a generalized framework, DeConStruct, which, by combining chemical probing-guided ensemble deconvolution and analysis of evolutionary conservation by covariation, accelerates the discovery of functional regulatory RNA structural switches that have so far remained largely elusive. By leveraging this framework in bacteria, we report the discovery of hundreds of candidate conserved RNA structural switches, which will provide an important resource for the identification of riboswitch classes and RNA thermometers. While characterizing the function and regulation of these putative switches constitutes a nontrivial challenge, reanalysis of a recently published dataset of RNA polymerase-pausing sites[72] showed a robust enrichment for these sites in 2+ regions as compared to 1 regions (42.8% for 1 regions versus 54.3% for 2+ regions; $P = 6.8 \times 10^{-16}$, one-tailed binomial test), suggesting that at least some of these sites might represent terminator–antiterminator transcriptional riboswitches. We further experimentally characterized a handful of candidates, demonstrating that our approach can recover both canonical, protein-independent thermometers, such as *cspG*, *cspI* and *cpxP*, and chaperone-dependent ones, such as *lpxP*. As the *lpxP* thermometer represents a true on–off temperature-controlled switch, oppositely to *cspA*, we can anticipate that it will have important applications in synthetic biology. Furthermore, we developed a method for the transcriptome-wide analysis of eukaryotic 5′ UTRs by chemical probing, 5′UTR-MaP, which, combined with the DeConStruct framework, enabled the discovery of RNA structural switches regulating translation of uORFs in human mRNAs, such as *CKS2* and *TXNL4A*.

Our data further challenge the traditional view of the cold shock response in bacteria, which has long been understood primarily in terms of RNA unfolding facilitated by the overexpression of Csps, or the increased structural rigidity of RNA because of lower temperatures. Instead, our findings suggest that the response is far more intricate, involving a complex redistribution of RNA structural ensembles. This indicates that cold shock induces a broader and more dynamic reorganization of RNA structures than previously thought, reflecting a sophisticated cellular adaptation to temperature changes. Furthermore, the large changes in relative stoichiometries observed for the *cspA* and *lpxP* ensembles upon temperature shift, which are substantially higher than one would typically expect from the temperature dependence from the Boltzmann distribution alone, reinforce the notion that the cellular environment has a crucial role in determining the composition of RNA ensembles in vivo.

Despite representing a crucial step toward a better understanding of the regulatory roles of RNA structures in living cells and the nuanced dynamics of RNA structure ensembles in response to environmental cues, the current study presents a number of limitations. Firstly, the use of DMS limits chemical probing to the interrogation of A and C, which might hamper the identification of small and A/C-poor structurally dynamic regions. This will hopefully be addressed in future studies by taking advantage of recent advances in chemical probing protocols and reagents[55,73] that allow querying all four nucleotides. Secondly, an implicit assumption of chemical probing-guided ensemble deconvolution analyses is that, to identify coexisting alternative structural states for an RNA, they need to interconvert at a rate that is slower than the timescale of the probing experiment. Indeed, transition barrier analysis for the alternative conformations identified in this study showed that, in general, they tend to be separated by high energy barriers (median at 37 °C: 8.6 kcal mol^{-1}, median at 10 °C: 13.5 kcal mol^{-1}; Supplementary Fig. 19a) and their interconversion requires the disruption or creation of, on average, more than 50% of the base pairs (52–57%; Supplementary Fig. 19b), thus suggesting that they are unlikely to spontaneously interconvert on a biologically relevant timescale. Rather, their interconversion in the cell is likely a chaperone-mediated process. Accordingly, 2+ regions tend to be significantly enriched for binding of the chaperones CspC and CspE as compared to 1 regions (CspC $P = 1.0 \times 10^{-25}$, CspE $P = 2.8 \times 10^{-34}$, one-tailed binomial test). Oppositely, as previously discussed, it is possible that 1 regions might represent an average of

Fig. 5 | Characterization of the *CKS2* uORF-regulating RNA structural switch. a, Secondary structure models for the two conformations of the *CKS2* 5′ UTR as identified by ensemble deconvolution from targeted DMS-MaPseq analysis, with overlaid in vivo DMS reactivities, along with reactivity profiles and base-pairing probabilities for both conformations. Reactivities are averaged across the two replicate experiments. The scatter plots depict the correlation of base reactivities for the deconvolved conformations across the two replicate experiments. **b**, Structure models for the two conformations of the *CKS2* 5′ UTR, inferred by phylogenetic analysis. Base pairs showing significant covariation (as determined by R-scape) are boxed in dark green

(*E* < 0.05) or purple (*E* < 0.1). Helices showing helix-level covariation support are boxed in light green (*E* < 0.05) or light purple (*E* < 0.1). **c**, Histogram depicting the median of cell's mean fluorescence in HEK293 cells expressing a dual-frame vector, harboring the 5′ UTR of *CKS2*, either wild type or mutagenized to stabilize either of the two conformations. Error bars represent the s.d. of three independent biological replicates. **d**, Western blot analysis of EGFP and HA expression in HEK293 cells expressing a dual-frame vector, harboring the 5′ UTR of *CKS2*, either wild type or mutagenized to stabilize either of the two conformations. GAPDH was used as the loading control. Analysis is representative of two independent biological replicates.

short-lived (that is, excited) states, interconverting at a faster rate, or might sample alternative conformations at stoichiometries too low to be detected by chemical probing-guided ensemble deconvolution analyses. Indeed, a well-known limitation of currently available ensemble deconvolution methods[3,4,6] is the impossibility to detect conformations with abundances below 5–10%.

Nevertheless, although the structural switches identified in this study might only represent a conservative estimate of the actual RNA structural diversity in living cells, our findings advance our understanding of gene expression regulation and will aid the development of innovative RNA-targeted therapeutic strategies[74–76].

Online content

Any methods, additional references, Nature Portfolio reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at https://doi.org/10.1038/s41587-025-02739-0.

References

1. Spitale, R. C. & Incarnato, D. Probing the dynamic RNA structurome and its functions. *Nat. Rev. Genet.* **24**, 178–196 (2023).
2. Zuker, M. & Sankoff, D. RNA secondary structures and their prediction. *Bull. Math. Biol.* **46**, 591–621 (1984).
3. Tomezsko, P. J. et al. Determination of RNA structural diversity and its role in HIV-1 RNA splicing. *Nature* **582**, 438–442 (2020).
4. Morandi, E. et al. Genome-scale deconvolution of RNA structure ensembles. *Nat. Methods* **18**, 249–252 (2021).
5. Lan, T. C. T. et al. Secondary structural ensembles of the SARS-CoV-2 RNA genome in infected cells. *Nat. Commun.* **13**, 1128 (2022).
6. Olson, S. W. et al. Discovery of a large-scale, cell-state-responsive allosteric switch in the 7SK RNA using DANCE-MaP. *Mol Cell* **82**, 1708–1723 (2022).
7. Yang, M. et al. In vivo single-molecule analysis reveals COOLAIR RNA structural diversity. *Nature* **609**, 394–399 (2022).
8. Forino, N. M. et al. Telomerase RNA structural heterogeneity in living human cells detected by DMS-MaPseq. *Nat. Commun.* **16**, 925 (2025).
9. Incarnato, D., Neri, F., Anselmi, F. & Oliviero, S. Genome-wide profiling of mouse RNA secondary structures reveals key features of the mammalian transcriptome. *Genome Biol.* **15**, 491 (2014).
10. Ding, Y. et al. In vivo genome-wide profiling of RNA secondary structure reveals novel regulatory features. *Nature* **505**, 696–700 (2014).
11. Rouskin, S., Zubradt, M., Washietl, S., Kellis, M. & Weissman, J. S. Genome-wide probing of RNA structure reveals active unfolding of mRNA structures in vivo. *Nature* **505**, 701–705 (2014).
12. Spitale, R. C. et al. Structural imprints in vivo decode RNA regulatory mechanisms. *Nature* **519**, 486–490 (2015).
13. Mustoe, A. M. et al. Pervasive regulatory functions of mRNA structure revealed by high-resolution SHAPE probing. *Cell* **173**, 181–195 (2018).
14. Homan, P. J. et al. Single-molecule correlated chemical probing of RNA. *Proc. Natl Acad. Sci. USA* **111**, 13858–13863 (2014).
15. Cordero, P. & Das, R. Rich RNA structure landscapes revealed by mutate-and-map analysis. *PLoS Comput. Biol.* **11**, e1004473 (2015).
16. Spasic, A., Assmann, S. M., Bevilacqua, P. C. & Mathews, D. H. Modeling RNA secondary structure folding ensembles using SHAPE mapping data. *Nucleic Acids Res.* **46**, 314–323 (2018).
17. Li, H. & Aviran, S. Statistical modeling of RNA structure profiling experiments enables parsimonious reconstruction of structure landscapes. *Nat. Commun.* **9**, 606 (2018).
18. Aviran, S. & Incarnato, D. Computational approaches for RNA structure ensemble deconvolution from structure probing data. *J. Mol. Biol.* **434**, 167635 (2022).
19. Bose, R., Saleem, I. & Mustoe, A. M. Causes, functions, and therapeutic possibilities of RNA secondary structure ensembles and alternative states. *Cell Chem Biol* **31**, 17–35 (2024).
20. Bonilla, S. L., Jones, A. N. & Incarnato, D. Structural and biophysical dissection of RNA conformational ensembles. *Curr. Opin. Struct. Biol.* **88**, 102908 (2024).
21. Ziv, O. et al. The short- and long-range RNA–RNA interactome of SARS-CoV-2. *Mol. Cell* **80**, 1067–1077 (2020).
22. Manfredonia, I. et al. Genome-wide mapping of SARS-CoV-2 RNA structures identifies therapeutically-relevant elements. *Nucleic Acids Res.* **48**, 12436–12452 (2020).
23. Schlick, T. et al. To knot or not to knot: multiple conformations of the SARS-CoV-2 frameshifting RNA element. *J. Am. Chem. Soc.* **143**, 11404–11422 (2021).
24. Siegfried, N. A., Busan, S., Rice, G. M., Nelson, J. A. E. & Weeks, K. M. RNA motif discovery by SHAPE and mutational profiling (SHAPE-MaP). *Nat. Methods* **11**, 959–965 (2014).
25. Zubradt, M. et al. DMS-MaPseq for genome-wide or targeted RNA structure probing in vivo. *Nat. Methods* **14**, 75–82 (2017).
26. Vitreschak, A. G., Rodionov, D. A., Mironov, A. A. & Gelfand, M. S. Regulation of riboflavin biosynthesis and transport genes in bacteria by transcriptional and translational attenuation. *Nucleic Acids Res.* **30**, 3141–3151 (2002).
27. Cromie, M. J., Shi, Y., Latifi, T. & Groisman, E. A. An RNA sensor for intracellular Mg^{2+}. *Cell* **125**, 71–84 (2006).
28. Winkler, W., Nahvi, A. & Breaker, R. R. Thiamine derivatives bind messenger RNAs directly to regulate bacterial gene expression. *Nature* **419**, 952–956 (2002).
29. Sudarsan, N., Wickiser, J. K., Nakamura, S., Ebert, M. S. & Breaker, R. R. An mRNA structure in bacteria that controls gene expression by binding lysine. *Genes Dev.* **17**, 2688–2697 (2003).
30. Chan, C. L. & Landick, R. The *Salmonella* Typhimurium his operon leader region contains an RNA hairpin-dependent transcription pause site. Mechanistic implications of the effect on pausing of altered RNA hairpins. *J. Biol. Chem.* **264**, 20796–20804 (1989).
31. Giuliodori, A. M. et al. The *cspA* mRNA is a thermosensor that modulates translation of the cold-shock protein CspA. *Mol. Cell* **37**, 21–33 (2010).
32. Conway, T. et al. Unprecedented high-resolution view of bacterial operon architecture revealed by RNA sequencing. *mBio* **5**, e01442-14 (2014).
33. Tierrafría, V. H. et al. RegulonDB 11.0: comprehensive high-throughput datasets on transcriptional regulation in *Escherichia coli* K-12. *Microb. Genom.* **8**, mgen000833 (2022).
34. Regulski, E. E. et al. A widespread riboswitch candidate that controls bacterial genes involved in molybdenum cofactor and tungsten cofactor metabolism. *Mol. Microbiol.* **68**, 918–932 (2008).
35. Zhang, Y. et al. A stress response that monitors and regulates mRNA structure is central to cold shock adaptation. *Mol. Cell* **70**, 274–286 (2018).
36. Beaudoin, J.-D. et al. Analyses of mRNA structure dynamics identify embryonic gene regulatory programs. *Nat. Struct. Mol. Biol.* **25**, 677–686 (2018).
37. Herzel, L., Stanley, J. A., Yao, C.-C. & Li, G.-W. Ubiquitous mRNA decay fragments in *E. coli* redefine the functional transcriptome. *Nucleic Acids Res.* **50**, 5029–5046 (2022).
38. Morandi, E., van Hemert, M. J. & Incarnato, D. SHAPE-guided RNA structure homology search and motif discovery. *Nat. Commun.* **13**, 1722 (2022).
39. Rivas, E., Clements, J. & Eddy, S. R. A statistical test for conserved RNA structure shows lack of evidence for structure in lncRNAs. *Nat. Methods* **14**, 45–48 (2017).
40. Zhang, Y. & Gross, C. A. Cold shock response in bacteria. *Annu. Rev. Genet.* **55**, 377–400 (2021).

41. Esquerré, T. et al. Genome-wide investigation of mRNA lifetime determinants in *Escherichia coli* cells cultured at different growth rates. *BMC Genomics* **16**, 275 (2015).

42. Schmidt, A. et al. The quantitative and condition-dependent *Escherichia coli* proteome. *Nat. Biotechnol.* **34**, 104–110 (2016).

43. Nakashima, K., Kanamaru, K., Mizuno, T. & Horikoshi, K. A novel member of the cspA family of genes that is induced by cold shock in *Escherichia coli*. *J. Bacteriol.* **178**, 2994–2997 (1996).

44. Etchegaray, J.-P. & Inouye, M. CspA, CspB, and CspG, major cold shock proteins of *Escherichia coli*, are induced at low temperature under conditions that completely block protein synthesis. *J. Bacteriol.* **181**, 1827–1830 (1999).

45. Ruiz, N. & Silhavy, T. J. Sensing external stress: watchdogs of the *Escherichia coli* cell envelope. *Curr. Opin. Microbiol.* **8**, 122–126 (2005).

46. MacRitchie, D. M., Buelow, D. R., Price, N. L. & Raivio, T. L. Two-component signaling and gram negative envelope stress response systems. *Adv. Exp. Med. Biol.* **631**, 80–110 (2008).

47. Phadtare, S. Recent developments in bacterial cold-shock response. *Curr. Issues Mol. Biol.* **6**, 125–136 (2004).

48. Carty, S. M., Sreekumar, K. R. & Raetz, C. R.Effect of cold shock on lipid A biosynthesis in *Escherichia coli*. Induction at 12 °C of an acyltransferase specific for palmitoleoyl-acyl carrier protein. *J. Biol. Chem.* **274**, 9677–9685 (1999).

49. Wang, N., Yamanaka, K. & Inouye, M. CspI, the ninth member of the CspA family of *Escherichia coli*, is induced upon cold shock. *J. Bacteriol.* **181**, 1603–1609 (1999).

50. Shimizu, Y. & Ueda, T. PURE technology. *Methods Mol. Biol.* **607**, 11–21 (2010).

51. Yair, Y. et al. Cellular RNA targets of cold shock proteins CspC and CspE and their importance for serum resistance in septicemic *Escherichia coli*. *mSystems* **7**, e0008622 (2022).

52. Michaux, C. et al. RNA target profiles direct the discovery of virulence functions for the cold-shock proteins CspC and CspE. *Proc. Natl Acad. Sci. USA* **114**, 6824–6829 (2017).

53. Baba, T. et al. Construction of *Escherichia coli* K-12 in-frame, single-gene knockout mutants: the KEIO collection. *Mol. Syst. Biol.* **2**, 2006.0008 (2006).

54. Leppek, K., Das, R. & Barna, M. Functional 5′ UTR mRNA structures in eukaryotic translation regulation and how to find them. *Nat. Rev. Mol. Cell Biol*. **19**, 158–174 (2018).

55. Marinus, T., Fessler, A. B., Ogle, C. A. & Incarnato, D. A novel SHAPE reagent enables the analysis of RNA structure in living cells with unprecedented accuracy. *Nucleic Acids Res.* **49**, e34 (2021).

56. Aziz, N. & Munro, H. N. Iron regulates ferritin mRNA translation through a segment of its 5′ untranslated region. *Proc. Natl Acad. Sci. USA* **84**, 8478–8482 (1987).

57. Manzella, J. M. & Blackshear, P. J. Regulation of rat ornithine decarboxylase mRNA translation by its 5′-untranslated region. *J. Biol. Chem.* **265**, 11817–11822 (1990).

58. Byeon, G. W. et al. Functional and structural basis of extreme conservation in vertebrate 5′ untranslated regions. *Nat. Genet.* **53**, 729–741 (2021).

59. Clamer, M. et al. Active ribosome profiling with RiboLace. *Cell Rep* **25**, 1097–1108 (2018).

60. Bugaut, A. & Balasubramanian, S. 5′-UTR RNA G-quadruplexes: translation regulation and targeting. *Nucleic Acids Res.* **40**, 4727–4741 (2012).

61. Endoh, T. & Sugimoto, N. Mechanical insights into ribosomal progression overcoming RNA G-quadruplex from periodical translation suppression in cells. *Sci. Rep.* **6**, 22719 (2016).

62. Balaratnam, S. et al. Investigating the NRAS 5′ UTR as a target for small molecules. *Cell Chem. Biol.* **30**, 643–657 (2023).

63. Morris, D. R. & Geballe, A. P. Upstream open reading frames as regulators of mRNA translation. *Mol. Cell. Biol.* **20**, 8635–8642 (2000).

64. Calvo, S. E., Pagliarini, D. J. & Mootha, V. K. Upstream open reading frames cause widespread reduction of protein expression and are polymorphic among humans. *Proc. Natl Acad. Sci. USA* **106**, 7507–7512 (2009).

65. Johnstone, T. G., Bazzini, A. A. & Giraldez, A. J. Upstream ORFs are prevalent translational repressors in vertebrates. *EMBO J.* **35**, 706–723 (2016).

66. Kwok, C. K., Marsico, G., Sahakyan, A. B., Chambers, V. S. & Balasubramanian, S. rG4-seq reveals widespread formation of G-quadruplex structures in the human transcriptome. *Nat. Methods* **13**, 841–844 (2016).

67. Zhao, J. et al. Enhanced transcriptome-wide RNA G-quadruplex sequencing for low RNA input samples with rG4-seq 2.0. *BMC Biol.* **20**, 257 (2022).

68. Lee, S. et al. Global mapping of translation initiation sites in mammalian cells at single-nucleotide resolution. *Proc. Natl Acad. Sci. USA* **109**, E2424 (2012).

69. Reuter, K., Biehl, A., Koch, L. & Helms, V. PreTIS: a tool to predict non-canonical 5′ UTR translational initiation sites in human and mouse. *PLoS Comput. Biol.* **12**, e1005170 (2016).

70. Gleason, A. C., Ghadge, G., Sonobe, Y. & Roos, R. P. Kozak similarity score algorithm identifies alternative translation initiation codons implicated in cancers. *Int. J. Mol. Sci.* **23**, 10564 (2022).

71. Gleason, A. C., Ghadge, G., Chen, J., Sonobe, Y. & Roos, R. P. Machine learning predicts translation initiation sites in neurologic diseases with nucleotide repeat expansions. *PLoS ONE* **17**, e0256411 (2022).

72. Yakhnin, A. V. et al. Robust regulation of transcription pausing in *Escherichia coli* by the ubiquitous elongation factor NusG. *Proc. Natl Acad. Sci. USA* **120**, e2221114120 (2023).

73. Mitchell, D., Cotter, J., Saleem, I. & Mustoe, A. M. Mutation signature filtering enables high-fidelity RNA structure probing at all four nucleobases with DMS. *Nucleic Acids Res.* **51**, 8744–8757 (2023).

74. Warner, K. D., Hajdin, C. E. & Weeks, K. M. Principles for targeting RNA with drug-like small molecules. *Nat. Rev. Drug Discov.* **17**, 547–558 (2018).

75. Childs-Disney, J. L. et al. Targeting RNA structures with small molecules. *Nat. Rev. Drug Discov.* **21**, 736–762 (2022).

76. Ellinger, E. et al. Riboswitches as therapeutic targets: promise of a new era of antibiotics. *Expert Opin. Ther. Targets* **27**, 433–445 (2023).

[1]Department of Molecular Genetics, Groningen Biomolecular Sciences and Biotechnology Institute (GBB), University of Groningen, Groningen, The Netherlands. [2]Institute of Molecular Physiology and Genetics, Centre of Biosciences, Slovak Academy of Sciences, Bratislava, Slovak Republic. [3]Department of Biology, Massachusetts Institute of Technology, Cambridge, MA, USA. [4]Broad Institute of MIT and Harvard, Cambridge, MA, USA. [5]Institute for Molecules and Materials, Radboud University Nijmegen, Nijmegen, The Netherlands. [6]Department of Theoretical Chemistry, University of Vienna, Vienna, Austria. [7]Research Group Bioinformatics and Computational Biology, Faculty of Computer Science, University of Vienna, Vienna, Austria. [8]RNA Forecast e.U., Vienna, Austria. [9]These authors contributed equally: Ivana Borovská, Chundan Zhang. ✉e-mail: d.incarnato@rug.nl

Methods

Strains, growth conditions and in vivo DMS probing

E. coli K-12 MG1655 derivative strains DH5α and TOP10 were streaked on Luria–Bertani (LB) plates and a single colony was picked, inoculated in 4 ml of LB broth and grown overnight at 37 °C with shaking. The day after, the culture was diluted to an optical density at 600 nm (OD_{600}) = 0.05 in 25 ml of LB broth and grown at 37 °C until $OD_{600} \approx 0.5$ (~2 h). For cold shock, 2 ml of this culture was mixed with 2 ml of LB broth prechilled to 0 °C in a water–ice slurry and then incubated at 10 °C for 20 min. For DMS (D186309, Merck) probing, DMS from a fresh 1:4 dilution in ethanol (~2.64 M) was added to the bacteria at a final concentration of 200 mM. Probing was conducted for 2 min at 37 °C or for 30 min at 10 °C (to achieve comparable modification efficiencies) with moderate shaking (800 rpm). Reactions were then quenched by addition of one volume of 1 M DTT, after which bacteria were collected by centrifugation at 17,000*g* for 1 min. The supernatant was discarded; the pellet was washed twice with 0.5 M DTT and then immediately subjected to RNA extraction.

Human HEK293 cells were cultured in high-glucose DMEM medium (L0104, Biowest), supplemented with 10% FBS (H1138, Merck), 25 U per ml penicillin and 25 µg ml^{-1} streptomycin at 37 °C and 5% CO_2. For ATP depletion experiments, cells were washed twice in PBS and then kept for 20 min in glucose-free DMEM (11966025, Thermo Fisher), supplemented with 10% FBS, 1 mM sodium pyruvate, 25 U per ml penicillin, 25 µg ml^{-1} streptomycin, 10 mM 2-deoxy-D-glucose (25972-M, Merck) and 10 mM sodium azide (71289, Merck) at 37 °C and 5% CO_2. DMS from a fresh 1:4 dilution in ethanol (~2.64 M) was directly added to the cells at a final concentration of 150 mM. Probing was conducted for 2 min at 37 °C. Reactions were then quenched by addition of one volume of 1 M DTT, after which cells were collected by centrifugation at 5,000*g* for 1 min. Supernatant was discarded and pellets were immediately lysed by direct addition of 1 ml of ice-cold TRIzol reagent (15596018, Thermo Fisher Scientific).

RNA extraction

For *E. coli*, cell pellets were resuspended in 62.5 µl of resuspension buffer (20 mM Tris-HCl pH 8.0, 80 mM NaCl and 10 mM EDTA pH 8.0), supplemented with 100 µg ml^{-1} final lysozyme (L6876, Merck) and 20 U of SUPERase•In RNase inhibitor (A2696, Thermo Fisher Scientific), by vigorous vortexing. Samples were incubated at room temperature for 1 min, followed by addition of 62.5 µl of lysis buffer (0.5% Tween-20, 0.4% sodium deoxycholate, 2 M NaCl and 10 mM EDTA). Samples were then inverted 5–10 times and incubated at room temperature for 2 min, followed by an additional 2 min on ice. Then, 1 ml of ice-cold TRIzol reagent was then added and samples were vigorously vortexed for 15 s.

Both bacterial and human samples were extracted as per manufacturer instructions. Residual genomic DNA (gDNA) was removed by digestion with TURBO DNase I (AM2239, Thermo Fisher Scientific) at 37 °C for 30 min.

DMS probing of bacterial in vitro refolded RNA

First, 10 µg of total RNA from exponentially growing *E. coli* was diluted in 89 µl of nuclease-free water, then heat-denatured at 95 °C for 2 min and immediately chilled on ice for 1 min. Next, 10 µl of ice-cold 10× folding buffer (250 mM HEPES pH 7.5 and 2 M KCl) were then added and samples were incubated at 37 °C for 15 min. Then, 1 µl of 500 mM $MgCl_2$ (prewarmed at 37 °C) was added and samples were incubated at 37 °C for 15 min to enable tertiary-structure formation. Probing was conducted by adding DMS at a final concentration of 200 mM and incubating the samples at 37 °C for 2 min. Reactions were then quenched by addition of 1 volume 1 M DTT, after which RNA was cleaned up on Monarch spin RNA cleanup columns (10 µg; T2030L, New England Biolabs) as per manufacturer instructions.

Extraction and DMS probing of bacterial native deproteinized rRNA

Native deproteinized *E. coli* rRNA was prepared as previously described[22]. Briefly, 2 ml of DH5α or TOP10 cells grown to $OD_{600} \approx 0.5$ were collected by centrifugation at 1,000*g* for 5 min (4 °C) and then resuspended in 1 ml of resuspension buffer (15 mM Tris-HCl pH 8.0, 450 mM sucrose and 8 mM EDTA), supplemented with 100 µg ml^{-1} final lysozyme. Samples were incubated at 22 °C for 5 min and then on ice for an additional 10 min, after which protoplasts were collected by centrifugation at 5,000*g* for 5 min (4 °C). The protoplast pellet was then resuspended in 120 µl of protoplast lysis buffer (50 mM HEPES pH 8.0, 200 mM NaCl, 5 mM $MgCl_2$ and 1.5% SDS), supplemented with 0.2 µg µl^{-1} proteinase K (P2308, Merck) and samples were incubated at 22 °C for 5 min, followed by 5 min on ice. SDS was precipitated by addition of 30 µl of SDS precipitation buffer (50 mM HEPES pH 8.0, 1 M potassium acetate and 5 mM $MgCl_2$), followed by centrifugation at 17,000*g* for 5 min (4 °C). Supernatant was extracted twice with phenol, chloroform and isoamyl alcohol (25:24:1), pre-equilibrated in RNA folding buffer (50 mM HEPES pH 8.0, 200 mM NaCl and 5 mM $MgCl_2$), and twice with chloroform. Deproteinized samples were then supplemented with 20 U of SUPERase•In RNase inhibitor equilibrated at 37 °C for 20 min. DMS from a 1:4 dilution in ethanol was added to a final concentration of 200 mM and samples were incubated at 37 °C for 2 min with shaking (800 rpm). Reactions were quenched by the addition of one volume of 1 M DTT and then cleaned up using Monarch spin RNA cleanup columns as per manufacturer instructions.

DMS probing of candidate bacterial RNA thermometers in vitro

T7 templates of *cspB*, *cspG*, *cspI*, *cpxP* and *lpxP*, including the 5′ UTR and CDS, were generated by PCR from DH5α gDNA using Q5 high-fidelity 2× master mix (M0492L, New England Biolabs). In vitro transcription reactions were performed using the HiScribe T7 high-yield RNA synthesis kit (E2040L, New England Biolabs) in 20 µl, using 1 µg of an equimolar pool of all templates. Reactions were incubated for 4 h at either 37 °C or 10 °C, after which RNA was probed by directly adding 200 mM final DMS to the reactions and incubating at 37 °C for 2 min or at 10 °C for 30 min. Reactions were then quenched by addition of one volume of 1 M DTT, after which RNA was cleaned up on Monarch spin RNA cleanup columns as per manufacturer instructions. Template DNA was then removed by digestion with TURBO DNase I (AM2239, Thermo Fisher Scientific) at 37 °C for 30 min and RNA samples were again cleaned up on Monarch spin RNA cleanup columns.

Bacterial DMS-MaPseq library preparation

DMS-MaPseq libraries were prepared as previously described[4], with minor changes. Before library preparation, highly abundant short RNA species, such as tRNAs, were depleted on Monarch spin RNA cleanup columns by loading a 1:1:1 mixture of total RNA in nuclease-free water, RNA-binding buffer and 100% ethanol. For transcriptome-wide DMS-MaPseq libraries, rRNA depletion was performed on 1.1 µg of total RNA using the RiboCop for bacteria kit (126, Lexogen), with two minor changes to the manufacturer's protocol; the denaturation temperature was increased to 95 °C and probe annealing temperature was lowered to 55 °C. Following rRNA depletion, RNA was cleaned up on Monarch spin RNA cleanup columns and eluted in 8 µl of nuclease-free water. For total RNA DMS-MaPseq libraries used for the optimization of folding parameters, 1 µg of total RNA was instead directly used as input for the subsequent step. RNA was supplemented with 2 µl of 100 µM random hexamers, 2 µl of deoxynucleoside triphosphates (dNTPs; 10 mM each) and 4 µl of 5× RT buffer (250 mM Tris-HCl pH 8.3, 375 mM KCl and 15 mM $MgCl_2$). Samples were then incubated at 94 °C for 5.5 min to simultaneously denature and fragment the RNA to a median size of 200 nt and immediately transferred to ice for 1 min. Samples were then supplemented with 1 µl of 0.1 M DTT, 20 U of SUPERase•In RNase inhibitor, 200 U of TGIRT-III enzyme (TGIRT50, InGex) and 25 ng µl^{-1}

actinomycin D (A1410, Merck) and incubated at 25 °C for 10 min, 57 °C for 1 h and 60 °C for 1 h. Addition of actinomycin D increased strand specificity by ~10%. TGIRT-III was degraded by adding 2 µg of proteinase K and incubating at 37 °C for 20 min. Proteinase K was inactivated by the addition of protease inhibitor cocktail (P8340, Merck). cDNA–RNA hybrids were then converted to double-stranded DNA (dsDNA) using the NEBNext Ultra II directional RNA second-strand synthesis module (E7550, New England Biolabs) by incubating at 16 °C for 1 h. dsDNA was cleaned up with 1.8 volumes of NucleoMag NGS cleanup and size select beads (744970, Macherey Nagel) and used as input for the NEBNext Ultra II DNA library prep kit for Illumina (E7645S, New England Biolabs) as per manufacturer instructions.

5′UTR-MaP library preparation

Before library preparation, ~1.5 µg of poly(A)+ RNA was enriched per sample using oligo d(T)$_{25}$ magnetic beads (S1419S, New England Biolabs). RNA was directly eluted from the beads by fragmentation in 4 mM MgCl$_2$ for 5.5 min at 94 °C and then cleaned up on Monarch spin RNA cleanup columns. Endogenous 5′-phosphate groups and 2′,3′-cyclic phosphates generated by chemical fragmentation were removed by treatment with 1 U of shrimp alkaline phosphatase (rSAP) (M0371L, New England Biolabs) in a final volume of 20 µl at 37 °C for 30 min, followed by cleanup on Monarch spin RNA cleanup columns. Decapping of 5′-capped RNA fragments was performed by treating the RNA with 5 U of Cap-Clip acid pyrophosphatase (C-CC15011H, CellScript) in a final volume of 20 µl at 37 °C for 1 h, followed by cleanup on Monarch spin RNA cleanup columns. Decapped RNA fragments were then ligated to an RNA adaptor (CUACACGACGCUCUUC-CGAUCU) harboring a 5′-biotin–TEG modification. Decapped RNA fragments and the RNA adaptor (1 µl of a 10 µM dilution) were first denatured by incubation at 70 °C for 5 min, after which the samples were snap-cooled on ice for 1 min. Samples were then supplemented with 30 U of high-concentration T4 RNA ligase 1 (single-stranded RNA ligase; M0437M, New England Biolabs) and ligation was performed in a final volume of 20 µl at 25 °C for 2 h in the presence of 12.5% PEG-8000 and 1 mM ATP. Then, 10 min before the end of the incubation, 20 µl of Dynabeads MyOne Streptavidin T1 beads (65601, Thermo Fisher Scientific) were aliquoted in a 2-ml tube, washed twice in 100 µl of 2× binding and wash buffer (10 mM Tris-HCl pH 7.5, 1 mM EDTA and 2 M NaCl) and then resuspended in 40 µl of the same buffer. RNA samples were then supplemented with 20 µl of nuclease-free water to dilute the PEG and then transferred to the washed beads. Samples were extensively vortexed and then incubated for 15 min at 22 °C in a thermomixer with constant shaking (1,000 rpm). Samples were then placed on the magnet, the supernatant was discarded and beads were washed twice with 500 µl of 1× binding and wash buffer by extensive vortexing. Two additional washes were then performed with 500 µl of nuclease-free water by incubating at 80 °C for 2 min. We found this step to be critical when preparing libraries from DMS-treated samples but not 2A3-treated samples, as DMS modifications introduce positive charges on the RNA that, because of the negative charge of the phosphate backbone, cause the RNA to aggregate. Heat denaturation at this stage allows washing away non-5′-cap-derived fragments. Ligated RNA fragments were eluted by incubating the beads in 50 µl of formamide elution buffer (95% formamide and 10 mM EDTA) at 95 °C for 3 min and then cleaned up on Monarch spin RNA cleanup columns. Eluted RNA fragments were ligated to a 5′-preadenylated and C3 spacer 3′-blocked DNA 3′ adaptor (rApp-AGATCGGAAGAGCACACGTCT-SpC3). RNA fragments and the adaptor (1 µl of a 10 µM dilution) were first denatured by incubation at 70 °C for 5 min, after which the samples were snap-cooled on ice for 1 min. Samples were then supplemented with 200 U of T4 RNA ligase 2, truncated KQ (M0373L, New England Biolabs), and ligation was performed in a final volume of 20 µl at 25 °C for 2 h in the presence of 12.5% PEG-8000. Samples were cleaned up on Monarch spin RNA cleanup columns. For 2A3-treated samples, RNA was eluted in 8 µl of

nuclease-free water, supplemented with 2 µl of 10 µM RT primer and 1 µl of 10 mM dNTPs; for DMS-treated samples, RNA was eluted in 9 µl of nuclease-free water, supplemented with 2 µl of 10 µM RT primer and 2 µl of 10 mM dNTPs. RNA was incubated at 70 °C for 5 min and then snap-cooled on ice for 1 min. RT reactions were performed in a final volume of 20 µl. For 2A3-treated samples reactions were supplemented with 4 µl of 5× RT buffer (250 mM Tris-HCl pH 8.0 and 375 mM KCl), 2 µl of DTT 0.1 M, 20 U of SUPERase•In RNase inhibitor, 200 U of Super-Script II RTase (18064022, Thermo Fisher Scientific) and 6 mM final MnCl$_2$ and incubated for 1.5 h at 42 °C, 10 min at 50 °C, 10 min at 55 °C, 10 min at 60 °C and 15 min at 75 °C. For DMS-treated samples, reactions were supplemented with 4 µl of 5× RT buffer (250 mM Tris-HCl pH 8.3, 375 mM KCl and 15 mM MgCl$_2$], 1 µl of 0.1 M DTT, 20 U of SUPERase•In RNase inhibitor and 200 U of TGIRT-III enzyme and incubated for 10 min at 42 °C, 1 h at 57 °C and 1 h at 60 °C. The TGIRT-III–RNA–cDNA complex was destroyed by the addition of 1 µl 10 M NaOH, followed by incubation at 95 °C for 3 min. Reactions were cleaned up on Monarch spin RNA cleanup columns, using one volume of RNA-binding buffer and one volume of 100% ethanol to only recover fragments ≥ 200 nt. Barcoding was performed by PCR using the NEBNext Ultra II Q5 master mix (M0544X, New England Biolabs) as per manufacturer instructions.

HEK293 total RNA DMS-MaPseq library prep

To mimic the same conditions used for the 5′UTR-MaP library preparation, 100 ng of total RNA per sample was fragmented in 4 mM MgCl$_2$ for 5.5 min at 94 °C and then cleaned up on Monarch spin RNA cleanup columns as per manufacturer instructions. The 2′,3′-cyclic phosphates generated by chemical fragmentation were removed by treatment with 1 U of rSAP in a final volume of 20 µl at 37 °C for 30 min, followed by heat inactivation of the enzyme at 70 °C for 5 min. Reactions were then supplemented with 20 U of T4 polynucleotide kinase (M0201L, New England Biolabs), 1 mM ATP and 5 mM DTT in a final volume of 50 µl and incubated at 37 °C for 1 h. The 5′-phosphorylated RNA fragments were then cleaned up on Monarch spin RNA cleanup columns and subjected to adaptor ligation, RT and PCR as detailed above.

Targeted DMS-MaPseq analysis of *CKS2* and *TXNL4A*

Targeted DMS-MaPseq analysis of *CKS2* and *TXNL4A* 5′ UTRs was performed using total RNA from HEK293 transfected for 24 h with the pEF6 vector carrying the wild-type 5′ UTR sequences as described below and probed with 150 mM DMS for 2 min at 37 °C. RT was carried out using gene-specific RT primers targeting the CDS of EGFP, harboring the reverse-complemented Illumina 3′ adaptor. Here, 3 µg of RNA was supplemented with 2 µl of 10 µM gene-specific RT primer and 2 µl of 10 mM dNTPs. RNA was incubated at 70 °C for 5 min and then snap-cooled on ice for 1 min. Samples were supplemented with 4 µl of 5× RT buffer (250 mM Tris-HCl pH 8.3, 375 mM KCl and 15 mM MgCl$_2$), 1 µl of 0.1 M DTT, 20 U of SUPERase•In RNase inhibitor and 200 U of TGIRT-III enzyme and incubated for 10 min at 50 °C, 1 h at 57 °C and 1 h at 60 °C. RT reactions were performed in a final volume of 20 µl. The TGIRT-III–RNA–cDNA complex was destroyed by the addition of 1 µl of 10 M NaOH, followed by incubation at 95 °C for 3 min. Reactions were cleaned up on Monarch spin RNA cleanup columns as per manufacturer instructions. Addition of the Illumina 5′ adaptor and barcoding were performed simultaneously by PCR, using 0.5 µM of i5 and i7 multiplexing primers, 0.025 µM of gene-specific forward primer harboring the Illumina 5′ adaptor and the NEBNext Ultra II Q5 master mix, as per manufacturer instructions.

Cloning of *cspG*, *cpxP* and *lpxP* constructs and mutagenesis of *lpxP*

Wild-type *cspG*, *cpxP* and *lpxP* FLAG-tagged, IPTG-inducible constructs, including the 5′ UTR and CDS, were prepared by amplifying the relevant regions from DH5α gDNA and cloning them in pET22b(+) vector (69744, Merck) between the XbaI and EcoRI sites. The exact

transcription start site (TSS) was determined from DMS-MaPseq coverage. Similarly, *cspG*, *cpxP* and *lpxP* FLAG-tagged, IPTG-inducible constructs, including the sole CDS, were cloned in pET22b(+) between the NdeI and EcoRI sites. For *cpxP*, as the identified candidate thermometer encompassed part of the CDS, the CDS was cloned starting at the third in-frame ATG codon by exploiting a naturally occurring, in-frame NdeI site. As RNA cotranscriptional folding can be influenced by the speed of the RNA polymerase, the vector's T7 promoter was replaced with a tac promoter. The SL_{alt}-stabilized *lpxP* 5′ UTR mutant was prepared using the Q5 site-directed mutagenesis kit (E0554S, New England Biolabs) as per manufacturer instructions. All cloning steps were performed in NEB 5α competent *E. coli* cells (C2987H, New England Biolabs). All vectors were verified by Sanger sequencing (Macrogen Europe). The sequences of primers used for cloning and mutagenesis are available in Supplementary Table 9. The vector containing the wild-type *lpxP* gene (inclusive of 5′ UTR) was deposited to Addgene (plasmid 212594).

Cloning and mutagenesis of *CKS2* and *TXNL4A* 5′ UTRs

Wild-type *CKS2* and *TXNL4A* 5′ UTRs were cloned in a modified pEF6 vector between the BamHI and EcoRI sites. Briefly, a sequence encoding EGFP (frame 1)–STOP–T2A–mCherry (frame 3) was assembled by PCR and cloned between the EcoRI and XbaI sites of the pEF6/V5-His vector (K961020, Thermo Fisher Scientific). *CKS2* was reverse-transcribed and amplified from HEK293 total RNA. *CKS2* mutants designed to stabilize conformations A or B were prepared using the Q5 site-directed mutagenesis kit as per manufacturer instructions. For *TXNL4A*, amplification proved much more challenging because of the extreme G+C content. Therefore, both wild-type and mutant stabilizing conformation B and the mutant disrupting the CUG start codon of the candidate uORF were prepared by PCR assembly of overlapping oligonucleotides using Q5 high-fidelity 2× master mix. As the candidate uORF of *TXNL4A* resided on frame 2, one nucleotide (G160) was deleted from a loop region at the end of the 5′ UTR of *TXNL4A* to align it to the mCherry frame. The sequences of primers used for cloning and mutagenesis are available in Supplementary Table 9.

Western blot analysis of bacterial protein expression at 37 °C versus 10 °C

Sanger-verified vectors were transformed in BL21(DE3) competent *E. coli* cells (C2627H, New England Biolabs). Two independent colonies were picked and inoculated in 3 ml of LB broth and grown overnight at 37 °C with shaking. The next day, bacteria were diluted to $OD_{600} \approx 0.05$ and grown until $OD_{600} \approx 0.3$. At this point, IPTG was added to a final concentration of 1 mM and cells were incubated with shaking at 37 °C for 30 min. Bacteria were then split into two separate aliquots, pelleted and resuspended in LB broth at 37 °C or 10 °C. Bacteria were then grown with shaking at 37 °C or 10 °C and 2-ml aliquots were collected after 30 min, 1 h or 2 h. Collected bacteria were pelleted and pellets were resuspended in 60 μl of lysis buffer (10 mM Tris-HCl pH 8.0, 10 mM EDTA and 0.1% Triton X-100), supplemented with 1 μg μl⁻¹ lysozyme and 1:100 dilution protease inhibitor cocktail. Samples were then subjected to ten cycles of sonication (5 s on, 5 s off) using a UP200St (Hielscher) ultrasonic processor. Protein concentrations were determined using a Pierce BCA protein assay kit (23225, Thermo Fisher Scientific) as per manufacturer instructions. Then, 30 μg of lysate was resolved on 10% SDS–PAGE gels, followed by transfer to nitrocellulose membrane using the iBlot 2 gel transfer system (IB21001, Thermo Fisher Scientific). Membranes were blocked by incubation for 1 h in 5% (w/v) nonfat dry milk (A0830, PanReac AppliChem ITW Reagents) in PBS, supplemented with 0.001% final Tween-20. Immunoblotting was performed using monoclonal anti-FLAG M2 antibody (F1804, Merck) or anti-LacI (9A5) universal antibody (EG1501, Kerafast) and Immobilon Forte western HRP substrate (WBLUF0100, Merck). PageRuler Plus (26619, Thermo Fisher Scientific) was used as the size standard.

Analysis of lpxP expression in *csp* knockouts

Wild-type and *csp*-knockout BW25113 *E. coli* cells from the KEIO collection[53] were first made competent using the Mix&Go! *E. coli* transformation kit (T3001, Zymo Research) and then transformed with the IPTG-inducible *lpxP* vector as described above. Two independent colonies were picked, inoculated in 3 ml of LB broth and grown overnight at 37 °C with shaking. The next day, bacteria were diluted to $OD_{600} \approx 0.05$ and grown until $OD_{600} \approx 0.5$. At this point, 1 ml of bacteria were directly mixed with 1 ml of ice-cold LB broth containing 0.02 mM IPTG and bacteria were incubated at 10 °C for 1 h with moderate shaking (800 rpm). Lysis and western blot analysis were conducted as described above. Knockout of *csp* genes was validated by PCR on gDNA from the individual clones.

In vitro transcription–translation using the PURE system

In vitro translation analysis of full-length wild-type and SL_{alt}-stabilized mutant or CDS-only *lpxP* was performed using the PURExpress in vitro protein synthesis kit (E6800S, New England Biolabs). Reactions were conducted in a final volume of 6.25 μl, using 2.5 μl of solution A, 1.875 μl of solution B, 0.1 μl of SUPERase•In RNase inhibitor and ~50 fmol of pET22b(+) template (harboring a T7 promoter instead of a tac promoter). Reactions were incubated at 37 °C for 1.5 h, then immediately mixed with 2× loading dye and resolved on a 12% polyacrylamide gel.

Western blot analysis of *CKS2* wild type and conformation-stabilizing mutants

Sanger-verified vectors were transfected in HEK293 cells. Briefly, on the first day, 800,000 cells were plated per well in a six-well plate precoated with 0.001% poly(L-lysine) (P8920, Merck). On the second day, 5 μg of plasmid DNA was transfected using 10 μl of Lipofectamine 2000 transfection reagent (11668019, Thermo Fisher Scientific) in 800 μl of Opti-MEM reduced-serum medium (51985034, Thermo Fisher Scientific). Then, 6 h after transfection, cells were supplemented with 1 ml of complete DMEM, supplemented with 20% FBS but without antibiotics. On the third day, cells were washed twice in PBS and then collected in radioimmunoprecipitation assay buffer (10 mM Tris-HCl pH 7.5, 150 mM NaCl, 0.1% SDS, 0.1% sodium deoxycholate and 1% Triton X-100), supplemented with protease inhibitor cocktail. After discarding membranes by centrifugation at 17,000g for 10 min (4 °C), protein concentrations were determined using a Pierce BCA protein assay kit as per manufacturer instructions. Then, 10 μg of lysate was resolved on 10% SDS–PAGE gels, followed by transfer to nitrocellulose membrane using the iBlot 2 gel transfer system. Membranes were blocked by incubation for 1 h in 5% (w/v) nonfat dry milk in PBS, supplemented with 0.001% final Tween-20. Immunoblotting was performed using anti-EGFP polyclonal antibody (CAB4211, Thermo Fisher Scientific), anti-HA tag polyclonal antibody (PA1-985, Thermo Fisher Scientific), anti-GAPDH monoclonal antibody (60004-1-Ig, Proteintech) and Immobilon Forte western HRP substrate. PageRuler Plus was used as the size standard.

Fluorescence microscopy analysis of *CKS2* and *TXNL4A* wild type and conformation-stabilizing mutants

For fluorescence microscopy analysis, on day one, 50,000 HEK293 cells were plated on 96-well flat-bottom plates precoated with 0.001% poly(L-lysine). On the second day, 75 ng of plasmid DNA was transfected using 0.625 μl of Lipofectamine 2000 transfection reagent in 50 μl of Opti-MEM reduced-serum medium. Then, 6 h after transfection, cells were supplemented with 100 μl of complete DMEM, supplemented with 20% FBS but without antibiotics. On the third day, cells were imaged with a Zeiss Observer Z1 widefield microscope and ×10 objective lens. The fluorescence signal per cell was quantified with Fiji[77] version 2.14.0/1.54g. The EGFP channel was used to identify particles by signal thresholding.

Processing of bacterial DMS-MaPseq data

Following sequencing, paired-end reads were clipped of sequencing adaptors using Cutadapt[78] version 4.4 (parameters: -A AGATCGGAAG -a AGATCGGAAG -m 100:100 -O 1) and merged using PEAR[79] version 0.9.11 (parameters: -n 100 -q 20 -u 0 -e -y 10G -z). Merged reads were then combined with R1 and the reverse-complemented R2 for read pairs that could not be merged. Next, a comprehensive annotation of *E. coli* transcriptional units with experimentally determined TSSs was built by aggregating 5′ UTR information from RegulonDB[33] and transcriptional units from EcoCyc[80] and the corresponding sequences were extracted from the *E. coli* str. K-12 substr. MG1655 genome (GenBank U00096.3). For the analysis of known riboswitches, a reference was built including the sole riboswitch regions ± 50 nt. Reads were then mapped to this reference using the rf-map tool of RNA Framework[81] version 2.8.3 and Bowtie2 (ref. 82) version 2.3.5.1 after clipping terminal bases with Phred quality < 20, discarding reads containing internal Ns and trimming the six 5′-most bases to account for possible mispriming artifacts (parameters: -b2 -cq5 20 -ctn -cmn 0 -cl 50 -mp --very-sensitive-local --nofw -b5 6). Alignments in SAM format were sorted and converted to BAM format using SAMtools[83] version 1.15.1. BAM alignments were then processed using RNA Framework's rf-count to generate both RC files (containing per-base mutations and coverage) and MM files (containing a map of mutated positions per read). Aligned reads spanning less than 100 nt of a transcript and reads having more than 10% mutated bases or fewer than two mutations were discarded. Insertions, ambiguously aligned deletions and deletions longer than 1 nt were ignored. Mutations were considered only if both the mutated base and the two surrounding bases had Phred quality > 20; consecutive mutations falling within 3 nt of each other were ignored (parameters: -m -mm -wl 2000 -ds 100 -es -na -ni -md 1 -dc 3 -me 0.1). DMS-MaPseq data from total RNA, used for the calibration of folding parameters described below, was analyzed with minor changes to the above protocol. Briefly, reads were mapped to a reference composed of only the 16S and 23S rRNA sequences. The minimum length spanned by reads was decreased to 90 nt (as total RNA DMS-MaPseq experiments were sequenced as single-read 100 bp, in contrast to rRNA-depleted DMS-MaPseq experiments that were sequenced as paired-end 150 bp) and reads harboring <2 mutations were also retained (parameters: -m -ds 90 -es -na -ni -dc 3 -ow -me 0.1 -md 1).

Processing of human 5′UTR-MaP data

Following sequencing, paired-end reads were clipped of sequencing adaptors using Cutadapt (parameters: -A AGATCGGAAG -a AGATCG-GAAG -m 75:75 -O 1) and merged using PEAR (parameters: -n 75 -q 20 -u 0 -e -y 10G -z). Merged reads were then combined with R1 and the reverse-complemented R2 for read pairs that could not be merged. Reads were then mapped to the MANE version 1.2 reference (plus the 18S and 28S rRNA sequences) using the rf-map tool of RNA Framework and Bowtie2 after clipping terminal bases with Phred quality < 20 and discarding reads containing internal Ns (parameters: -b2 -cq5 20 -ctn -cmn 0 -mp --very-sensitive-local -bnr). Alignments in SAM format were sorted and converted to BAM format using SAMtools. BAM alignments were then processed using RNA Framework's rf-count to generate both RC files (containing per-base mutations and coverage) and MM files (containing a map of mutated positions per read). Aligned reads spanning less than 90 nt of a transcript and reads having more than 10% mutated bases or fewer than two mutations were discarded. Insertions, ambiguously aligned deletions and deletions longer than 1 nt were ignored. Mutations were considered only if both the mutated base and the two surrounding bases had Phred quality > 20 and consecutive mutations falling within 3 nt of each other were ignored (parameters: -m -mm -wl 2000 -ds 100 -es -na -ni -md 1 -dc 3 -me 0.1). DMS-MaPseq data from total RNA, used for the calibration of folding parameters described below, were analyzed with minor changes to the above protocol. Briefly, no merging was performed as samples were sequenced

as single reads and reads were mapped to a reference composed of only the 18S and 28S rRNA sequences.

Optimization of folding parameters

RC files from total RNA DMS-MaPseq experiments were processed using RNA Framework's rf-norm to obtain normalized reactivity profiles (parameters: -sm 4 -nm 3 -rb AC -mm 1 -n 1,000). For *E. coli* secondary structure modeling, optimal slope (4.8) and intercept (−0.8) values were identified through jackknifing by simultaneously optimizing folding of 16S and 23S rRNAs over both in vivo and ex vivo deproteinized DMS-MaPseq data from both DH5α and TOP10 cells using RNA Framework's rf-jackknife (parameters: -rp -md 600 -x -m), ViennaRNA[84] version 2.5.1 and the modified Fowlkes–Mallows index[5].

For human secondary structure modeling, optimal slope (4.6) and intercept (−2) values were identified by simultaneously optimizing folding of 28S rRNA over both replicate experiments.

Ensemble deconvolution analysis

Ensemble deconvolution was performed using the DRACO algorithm[4]. Briefly, DRACO slides a window of a user-defined length along each transcript, retaining only those reads falling entirely within the window's boundaries. For each window, a graph is then generated by exploiting the comutation information so that, basically, each mutation in a read represents a vertex and two bases observed to comutate within the same read are connected by an edge. The normalized Laplacian of the graph's adjacency matrix is then subjected to eigen decomposition and eigengap analysis to identify the number of coexisting RNA conformations making up the ensemble. This number is then used to perform a soft partitioning of the graph (graph-cut) to reconstruct the individual reactivity profiles of the different conformations and their relative stoichiometries. In its original implementation, this graph-cut step involved randomly initializing the weight of each vertex for each conformation N times (with $N = 50$), followed by selection of the set of weights yielding the lowest normalized graph-cut score. This initial set of weights was then iteratively altered by a factor $\varepsilon = \frac{1}{2C}$, where C is the number of conformations making up the ensemble, until the normalized graph-cut score was minimized. As this procedure was performed only once, the risk was that the identified set of weights would represent only a local minimum of the graph-cut score rather than the true minimum, potentially leading to inconsistent conformation reconstruction across consecutive DRACO runs. Furthermore, the value of ε was typically too large to enable the accurate reweighting of the vertices (for instance, with $C = 2$ and $\varepsilon = 0.25$). To address these issues, we introduced the following improvements in the DRACO algorithm (available as version 1.2 from the repository https://github.com/dincarnato/draco/): (1) the number of random initializations N was increased to 500 (adjustable through the --softClusteringInits parameter); (2) the weight factor ε was lowered to 0.005 (adjustable through the --softClusteringWeightModule parameter); and (3) the entire graph-cut procedure is now repeated multiple times (adjustable through the --softClusteringIters parameter), to ensure convergence toward the true normalized graph-cut score minimum. Before running DRACO, the MM files generated by rf-count were preprocessed using the filterMM utility (available from the repository https://github.com/dincarnato/labtools) to discard reads having <2 A/C mutated bases and regions of extremely high coverage were randomly downsampled to achieve a maximum per-base coverage of 500,000×.

For *E. coli*, DRACO analysis was performed with a window size of 100 nt, slid in 5-nt increments, requiring a minimum base coverage of 2,000× and a minimum of 2,000 reads after filtering to perform the eigen deconvolution and repeating the graph-cut procedure 30 times (parameters: --absWinLen 100 --absWinOffset 5 --minBaseCoverage 2000 --minFilteredReads 2000 --minPermutations 10 --maxPermutations 50 --firstEigengapShift 0.95 --lookaheadEigengaps 1 --softClusteringIters 30 --softClusteringInits 500 --softClusteringWeightModule 0.005).

For HEK293, the window size was reduced to 90 nt and slid in 1-nt increments (parameters: --absWinLen 90 --absWinOffset 1) to account for the smaller library insert size, whereas all other parameters were left unchanged.

Evaluation of sequencing depth's effect on the ensemble deconvolution of known riboswitches

To evaluate the ability of DRACO to detect known riboswitches from in vivo probing data, we used data from TOP10 bacteria and selected four riboswitches belonging to mRNAs having different expression levels in our dataset. A reference was built including only the riboswitch ± 50 nt and reads were preprocessed and mapped as detailed in the previous paragraphs. The resulting MM files were then randomly subsampled using the extract function of the rf-mmtools utility of RNA Framework by setting the value of the -rs parameter to 2, 4, 8, 10, 40 or 80 to subsample 1/2, 1/4, 1/8, 1/10, 1/40 or 1/80 of the reads mapping to each riboswitch, respectively. A total of 20 random subsamplings were performed for each. The resulting MM files were then subjected to DRACO analysis (parameters: --absWinLen 100 --absWinOffset 1 --minBaseCoverage 2000 --minFilteredReads 2000 --minPermutations 10 --maxPermutations 50 --firstEigengapShift 0.95 --lookaheadEigengaps 1 --softClusteringIters 30 --softClusteringInits 500 --softClusteringWeightModule 0.005). If at least one window overlapping the riboswitch was found to populate >1 conformation, the riboswitch was considered detected.

Comparison of DH5α versus TOP10 strains, 37 °C versus 10 °C and standard versus ATP-depleted conditions

Correlation between experiments (related to Supplementary Figs. 1c,d, 5a, 10b and 16d) was calculated on the raw mutation frequencies of A/C bases in transcriptional units for which ≥50% of A/C bases had coverage ≥ 10,000× after removing outliers (raw reactivity > 0.1). The number of conformations populated by each base in the covered transcriptome (related to Figs. 1a, 2a and 4c,f) was determined by parsing DRACO's JSON-formatted output files. As DRACO uses a sliding window approach, consecutive overlapping windows might be found to populate different numbers of conformations; in such cases, overlapping bases were assigned the highest number of conformations. Windows populating different numbers of conformations between 37 °C and 10 °C (related to Fig. 2b) or between standard and ATP-depleted conditions (related to Fig. 4g) were identified as follows. First, windows populating one or two or more conformations were extracted from DRACO's JSON-formatted output files into BED format and overlapping windows were merged using the mergeBed tool of BEDTools[85] version 2.31.0. Any portion of the windows populating one conformation overlapping with the windows populating two or more conformations was removed using BEDTools' subtractBed. Then, windows populating one or two or more conformations common to both DH5α and TOP10 at either 37 °C or 10 °C or both replicate experiments in HEK293 cultured under standard or ATP-depleted conditions were identified by intersecting the corresponding sets from both experiments using BEDTools intersectBed. Only windows populating the same number of conformations in both DH5α and TOP10 or in both HEK293 replicate experiments were retained. Lastly, common windows populating two or more conformations in both DH5α and TOP10 at 37 °C or in both HEK293 replicate experiments under standard conditions and one conformation in both DH5α and TOP10 at 10 °C or in both HEK293 replicate experiments under ATP-depleted conditions (less than ensemble heterogeneity) or vice versa (greater than ensemble heterogeneity), as well as regions populating the same number of conformations in both strains or replicate experiments at both temperatures or culture conditions (no change), were identified by intersecting the windows set determined in the previous step using BEDTools intersectBed. Window coordinates were then intersected with gene coordinates to identify which genes contained windows showing differential ensemble heterogeneity between 37 °C and 10 °C or between standard and ATP-depleted conditions.

Translation efficiency analysis

Ribosome profiling and RNA-seq data for *E. coli* cells at 37 °C or shocked at 10 °C for 10 min were obtained from a previous study[35] (GSE103421). Reads were aligned to the same transcriptome reference used for DMS-MaPseq analysis, using RNA Framework's rf-map and Bowtie[86] version 1.3.1, allowing a maximum of to mapping positions (parameters: -ca3 CTGTAGGCACCATCAA -bnr -ow -bm 2 -bc 32000 -ba). After discarding all reads mapping to the rRNA operons, read counts for protein-coding genes containing windows showing differential heterogeneity between 37 °C and 10 °C as described above were calculated by intersecting CDS coordinates in BED format with the relevant BAM files using BEDTools intersectBed (parameter: -c). Only windows ≥ 50 nt (half of the window size used for DRACO analysis) were considered. For both Ribo-seq and RNA-seq data, per-gene reads per kilobase per million mapped reads (RPKMs) were calculated as follows:

$$RPKM = \frac{C}{NL} \times 1{,}000{,}000$$

where C is the read count on the gene, N is the total number of reads mapped in the experiment and L is the length of the gene in kilobases. Translation efficiency for each gene (related to Fig. 3e) was then calculated as follows:

$$\text{Translation efficiency} = \frac{RPKM_{Ribo-seq} + 0.1}{RPKM_{RNA-seq} + 0.1}$$

where 0.1 is a pseudo count added to avoid division by zero. Only genes expressed at ≥1 RPKM at both 37 °C and 10 °C were considered.

For HEK293, ribosome profiling and RNA-seq data were obtained from two previous studies[59,87] (GSE112353 and GSE228010). Reads were first aligned to a reference including rRNAs, tRNAs and small nucleolar RNAs, using RNA Framework's rf-map and Bowtie2 version 2.3.5.1. Unmapped reads were then aligned to the same transcriptome reference used for 5'UTR-MaP analysis and read counts for protein-coding genes containing windows showing differential heterogeneity between standard and ATP-depleted conditions described above were calculated by intersecting CDS coordinates in BED format with the relevant BAM files. Only windows ≥ 45 nt (half of the window size used for DRACO analysis) were considered. Only genes expressed at ≥1 RPKM both under standard and ATP-depleted conditions were considered.

Comparison of regions populating one versus two or more conformations

Eight features were evaluated for regions populating one versus two or more conformations: A+C content, G+C content, median Shannon entropy, median unpaired probability, median reactivity, Gini index, median percentage conservation and Z score (related to Figs. 1b–g and 4d,e and Supplementary Figs. 4, 6 and 7). For all analyses, only regions spanning at least half of the window size used for DRACO analysis (50 nt for *E. coli*, 45 nt for HEK293) were included. Furthermore, all 2+ regions were retained for this analysis (whether or not they populated the same number of conformations in both strains or replicate experiments), provided that they populated two or more conformations in both strains (or replicate experiments). First, bulk reactivity profiles for both DH5α and TOP10 grown at 37 °C or HEK293 grown under standard conditions were obtained by normalizing the respective RC files as described above using RNA Framework's rf-norm (parameters: -sm 4 -nm 3 -rb AC -mm 1 -n 1,000) and the resulting normalized XML reactivity files were combined using RNA Framework's rf-combine. From these XML files, reactivity data for regions populating one or two or more conformations were extracted and used to calculate the median reactivity and Gini index distributions. Combined XML files were then passed to RNA Framework's rf-fold to compute base-paring

probabilities and Shannon entropies (parameters: -sl 4.8 -in -0.8 -md 600 -dp -sh for *E. coli* or -sl 4.6 -in -2 -md 600 -dp -sh for HEK293). Unpaired probabilities per base were calculated as follows:

$$1 - \sum_{j=i}^{J} p(i,j)$$

where $p(i,j)$ is the base-pairing probability between nucleotides i and j over all possible J partners. For unconstrained predictions, the same parameters were used with the addition of the -i parameter to ignore experimental reactivities. Distributions of folding free energy Z scores were calculated on the nucleotide sequences corresponding to the regions populating one or two or more conformations in the absence of any constraint using ViennaRNA. For Z-score calculation, the sequence of each region was shuffled 100 times while preserving dinucleotide frequencies and the corresponding folding free energies were predicted using RNAfold. The Z score for each region was then calculated as follows:

$$Z = \frac{\triangle G - \mu}{\sigma}$$

where ΔG is the folding free energy for the original sequence, while μ and σ are the average and s.d., respectively, of the folding free energies across the 100 shuffled sequences.

The same 1 and 2+ regions, as defined above, were used for all analyses, including translation efficiency (related to Fig. 4h,i and Supplementary Figs. 5h and 10d) and gene ontology analyses. Gene ontology was performed using DAVID[88].

Sequence-level conservation analysis

To evaluate the conservation of regions populating one versus two or more conformations, a multiple-sequence alignment was computed using Mugsy[89] version 1.2.3 and ten Gram-negative bacteria genomes: *E. coli* str. K-12 substr. MG1655 (GenBank U00096.3), *S. enterica* subsp. enterica serovar Typhimurium str. LT2 (GenBank AE006468.2), *Shigella flexneri* 2a str. 2457T (GenBank AE014073.1), *Klebsiella pneumoniae* subsp. *pneumoniae* HS11286 (GenBank CP003200.1), *Yersinia pestis* CO92 (GenBank AL590842.1), *Enterobacter* sp. 638 (GenBank CP000653.1), *Serratia marcescens* strain KS10 (GenBank CP027798.1), *Pectobacterium carotovorum* strain WPP14 (GenBank CP027798.1), *Shigella dysenteriae* strain SWHEFF_49 (GenBank CP055055.1) and *Enterobacter cloacae* isolate 1382 (GenBank OW968328.1). The resulting alignment was parsed to calculate the percentage conservation at each position with respect to the *E. coli* genome.

Reactivity profile reconstruction and structure modeling for high-confidence regions

High-confidence structurally heterogeneous regions for which the deconvolved reactivity profiles could be nonambiguously matched between DH5α and TOP10 or between HEK293 replicate experiments (average correlation of reactivity profiles ≥ 0.65) were extracted using RNA Framework's rf-json2rc by including 20 extra bases on either side of the structure (parameters: -ec 1,000 -mom 0.35 -e 20 -cf 0.1 -i 0.1 -mcm 0.65 -mcr 0.65). The tool processes DRACO's JSON-formatted output files from two experiments, aggregating those regions showing sufficient agreement between the deconvolved reactivity profiles across the two experiments and yielding two RC files containing the per-base coverage and mutations across the different conformations reconstructed by DRACO for the analyzed RNAs. The resulting RC files were then processed using RNA Framework's rf-norm to yield normalized reactivity profiles (parameters: -sm 4 -nm 3 -rb AC -mm 1 -n 100). Structure modeling was performed using the consensusFold utility (available from the repository https://github.com/dincarnato/labtools), which leverages RNAalifold[90] to aggregate multiple reactivity profiles into a consensus secondary structure (parameters: -sl

4.8 -in -0.8 -md 600 for *E. coli* or -sl 4.6 -in -2 -md 600 for HEK293). For the modeling of secondary structures under cold shock conditions an additional parameter (-t 10) was specified to set the folding temperature to 10 °C.

Normalization of 5′UTR-MaP reactivity data

As 5′UTR-MaP selectively enriches 5′ UTR regions, which are intrinsically highly structured because of their high G+C content, traditional gene-level normalization of reactivities would lead to notable biases because of the low number of highly reactive bases. To address this issue, we adopted an experiment-level normalization approach. Briefly, bases covered across all experiments were sorted and values greater than 1.5× the interquartile range (IQR) + the 75th percentile were removed. After excluding these outliers, the next 10% of remaining bases common to all experiments were averaged, yielding an experiment-level normalization factor. We implemented this approach in the rf-normfactor tool of RNA Framework (parameters: -sm 4 -nm 3 -rb AC -mc 1,000). The resulting normalization factors were then passed to the rf-norm tool using the -nf parameter.

Covariation analysis

To evaluate the conservation of the identified *E. coli* structures, we implemented the evolutionary conservation analysis module of the DeConStruct framework, built on top of the cm-builder pipeline (available from the repository https://github.com/dincarnato/labtools) we previously introduced[4,22] (which exploits Infernal[91] version 1.1.3 and R-scape[39,92] version 2.0.0.q), to be able to handle full bacterial genomes rather than just individual transcripts. For each predicted structure (filtering out those with a known match in Rfam[93]) a CM was first built using Infernal's cmbuild and the sole *E. coli* sequence. The CM was then used to search a database of 7,598 representative archaeal and bacterial genomes (and associated plasmids when present) from RefSeq to iteratively identify putative homologs. In its original implementation, cm-builder used an *E*-value-based approach to search in the database. This approach had two main limitations. Firstly, the *E* value for the identified matches was dependent on the size of the searched database, potentially leading to different results with different database sizes. Secondly, it required the calibration of the CMs using Infernal's cmcalibrate module, a computationally intensive task, which is not easily scalable to hundreds of candidates. To address these issues, we implemented a bit-score-based search. Briefly, to trick Infernal into thinking that a CM had been calibrated, a fake set of ECMLC, ECMGC, ECMLI and ECMGI field values was introduced into the CM. These fields are only used to determine the *E* value of a database search but they do not affect the bit score. Then, a decoy database was built by randomly extracting and reversing ~10% of the sequences from the original genome database. Infernal's cmsearch was then used to search the CM against the decoy database. A noise threshold N was defined by taking the highest possible bit score returned by this search and by rounding it up to the nearest multiple of 5. If $N < 20$, then N was set to 20. The search was then repeated against the original database, retaining only those matches having bit score > N. Matches having <50% canonical base pairs and truncated hits covering <75% of the structure were discarded. The resulting set of candidate homologs was then realigned against the original CM using Infernal's cmalign. The whole procedure was repeated a maximum of three times. At each iteration, N was increased by 10 and the alignment of candidate homologs was analyzed using R-scape's average product correction (APC)-corrected G-test statistics and a relaxed E-value threshold of 0.1 (to account for those structures falling within coding regions for which sequence variation might be 'constrained' by the underlying amino acid sequence). If the number of significantly covarying base pairs dropped with respect to the previous iteration (except for the first iteration), the procedure was stopped. The final alignment was then polished by discarding sequences with a length that was significantly different

from the majority of the sequences in the alignment. This was achieved by converting sequence lengths to Z scores and discarding sequences with abs(Z score) > 2 and length difference > 10% with respect to the average sequence length in the alignment (implemented in the stockholmPolish tool available from the repository https://github.com/dincarnato/labtools). To further select only high-confidence alignments, we performed a stringent filtering by selecting alignments matching three criteria: (1) ≥25% of the helices showing helix-level covariation (R-scape's Lancaster aggregated E value < 0.05); (2) ≥12.5% of the base pairs showing covariation (R-scape APC-corrected G-test statistic E value < 0.1); and (3) ≥5 base pairs showing covariation.

For human RNA structures, sequences of candidate structural homologs were directly extracted from multiz100way MAF files (https://hgdownload.soe.ucsc.edu/goldenPath/hg38/multiz-100way/maf/) using the mafsInRegion tool of the kentUtils (available from the repository https://github.com/ENCODE-DCC/kentUtils) after lifting the identified structurally heterogeneous regions from transcriptome-level to genome-level coordinates using the transcriptome2genome tool (available from the repository https://github.com/dincarnato/labtools). Extracted MAF blocks were concatenated, gaps were removed and the resulting set of sequences was used as database for the cm-builder tool. This part of the analysis was implemented in the dbFromMAF tool (available as part of the DeConStruct pipeline from the repository https://github.com/dincarnato/papers). As this set of sequences represents a higher-confidence set as compared to the set of complete bacterial genomes used for the analysis of $E.$ $coli$ structures, two parameters were relaxed for the construction of alignments; at each iteration, the bit-score noise threshold (N) was increased by 5 (rather than 10) and matches having <35% (rather than 50%) canonical base pairs were discarded. No polishing was performed on the output alignments and filtering was relaxed by selecting all structures having at least three covarying base pairs (R-scape APC-corrected G-test statistic E value < 0.1) and two covarying helices (R-scape's Lancaster aggregated E value < 0.05).

Design of conformation-stabilizing mutants

The mutant stabilizing the SL$_{alt}$ conformation of $lpxP$ was designed manually by introducing point mutations in the 5′ half of the stem but taking care not to touch any nucleotide in the surroundings of the RBS residing on the 3′ half of the stem. Mutants stabilizing the different conformations of $CKS2$ and $TXNL4A$ were automatically designed using the rf-mutate tool of RNA Framework. For this purpose, the program was modified to enable specifying a target structure. For example, to stabilize conformation A, this was provided as the target structure, while conformation B was provided as the wild-type structure, so that the program would design mutations minimizing the probability of forming conformation B while simultaneously maximizing the probability of forming conformation A (with a maximum tolerated base-pair distance of 25%). Mutations were designed in such a way that the underlying amino acid sequences of both the uORF and the main ORF were preserved.

Evaluation of energy barriers and fraction changed base pairs between conformations

Transition barriers were estimated on the set of structures predicted from structurally heterogeneous regions whose DRACO-deconvolved reactivity profiles could be nonambiguously matched between DH5α and TOP10 cells as described above. Estimation was performed using DrFindpath, a component of the DrTransformer package[94]. DrFindpath uses the Findpath heuristic[95], which is implemented in the ViennaRNA library. The fraction of changed base pairs between conformations was calculated as follows:

$$F = 1 - \frac{C}{c_1 + c_2 + C}$$

where C is the number of base pairs common to both conformations and c_1 and c_2 are the numbers of base pairs unique to either conformation.

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability

Sequencing data were deposited to the Gene Expression Omnibus database under accession number GSE247244. Raw MM files for analysis with DRACO are available from Zenodo (https://doi.org/10.5281/zenodo.10357457)[96]. Additional processed files (including the browsable set of conserved structurally heterogeneous regions) are available online (https://www.incarnatolab.com/datasets/Ensembles_Borovska_2025.php). Source data are provided with this paper.

Code availability

The source codes of DRACO version 1.2, the cm-builder, filterMM, consensusFold, transcriptome2genome and stockholmPolish utilities and the DeConStruct framework are freely available from GitHub under the GPLv3 license (https://github.com/dincarnato/draco, https://github.com/dincarnato/labtools and https://github.com/dincarnato/papers).

References

77. Schindelin, J. et al. Fiji: an open-source platform for biological-image analysis. *Nat. Methods* **9**, 676–682 (2012).
78. Martin, M. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet J.* **17**, 10–12 (2011).
79. Zhang, J., Kobert, K., Flouri, T. & Stamatakis, A. PEAR: a fast and accurate Illumina Paired-End reAd mergeR. *Bioinformatics* **30**, 614–620 (2014).
80. Karp, P. D. et al. The EcoCyc database. *EcoSal Plus* **6**, eesp00022023 (2023).
81. Incarnato, D., Morandi, E., Simon, L. M. & Oliviero, S. RNA Framework: an all-in-one toolkit for the analysis of RNA structures and post-transcriptional modifications. *Nucleic Acids Res.* **46**, e97 (2018).
82. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9**, 357–359 (2012).
83. Li, H. et al. The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078–2079 (2009).
84. Lorenz, R. et al. ViennaRNA package 2.0. *Algorithms Mol. Biol.* **6**, 26 (2011).
85. Quinlan, A. R. & Hall, I. M. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**, 841–842 (2010).
86. Langmead, B., Trapnell, C., Pop, M. & Salzberg, S. L. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol.* **10**, R25 (2009).
87. Kim, D. et al. Pan-KRAS inhibitor disables oncogenic signalling and tumour growth. *Nature* **619**, 160–166 (2023).
88. Sherman, B. T. et al. DAVID: a web server for functional enrichment analysis and functional annotation of gene lists (2021 update). *Nucleic Acids Res.* **50**, W216–W221 (2022).
89. Angiuoli, S. V. & Salzberg, S. L. Mugsy: fast multiple alignment of closely related whole genomes. *Bioinformatics* **27**, 334–342 (2011).
90. Bernhart, S. H., Hofacker, I. L., Will, S., Gruber, A. R. & Stadler, P. F. RNAalifold: improved consensus structure prediction for RNA alignments. *BMC Bioinformatics* **9**, 474 (2008).
91. Nawrocki, E. P. & Eddy, S. R. Infernal 1.1: 100-fold faster RNA homology searches. *Bioinformatics* **29**, 2933–2935 (2013).
92. Rivas, E. RNA covariation at helix-level resolution for the identification of evolutionarily conserved RNA structure. *PLoS Comput. Biol.* **19**, e1011262 (2023).

93. Kalvari, I. et al. Rfam 14: expanded coverage of metagenomic, viral and microRNA families. *Nucleic Acids Res.* **49**, D192–D200 (2021).

94. Badelt, S., Lorenz, R. & Hofacker, I. L. DrTransformer: heuristic cotranscriptional RNA folding using the nearest neighbor energy model. *Bioinformatics* **39**, btad034 (2023).

95. Flamm, C., Hofacker, I. L., Maurer-Stroh, S., Stadler, P. F. & Zehl, M. Design of multistable RNA molecules. *RNA* **7**, 254–265 (2001).

96. Borovska, I. et. Al. Identification of conserved RNA regulatory switches in living cells using RNA secondary structural ensemble mapping and covariation analysis. Datasets. *Zenodo* https://doi.org/10.5281/zenodo.10357457 (2025).

Author contributions

I.B. and D.I. designed the experiments. I.B., C.Z., S.-L.J.D., M.F.S.C. and D.A.L.v.d.H. performed the experiments. E.M. and D.I. developed the DRACO algorithm. E.M., M.T.W. and D.I. performed data analysis and structure modeling. I.B., C.Z. and D.I. wrote the manuscript with input from all authors. D.I. supervised the research.

Competing interests

The authors declare no competing interests.

Additional information
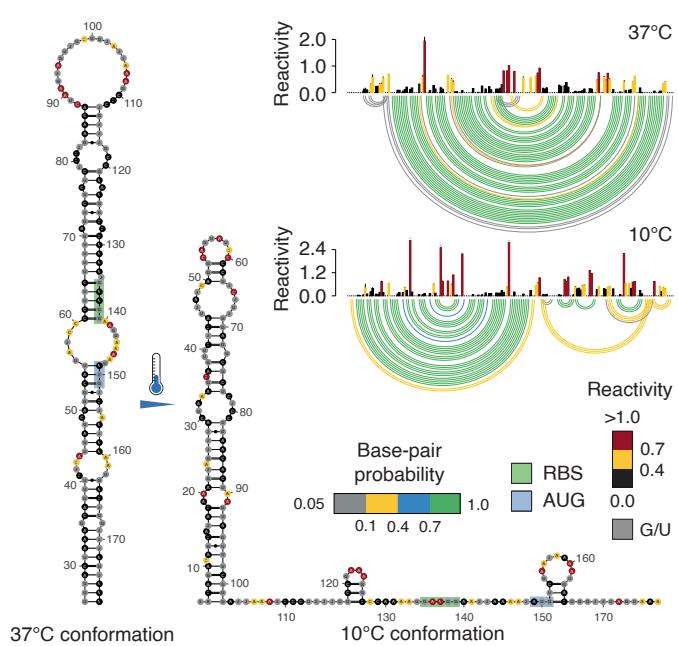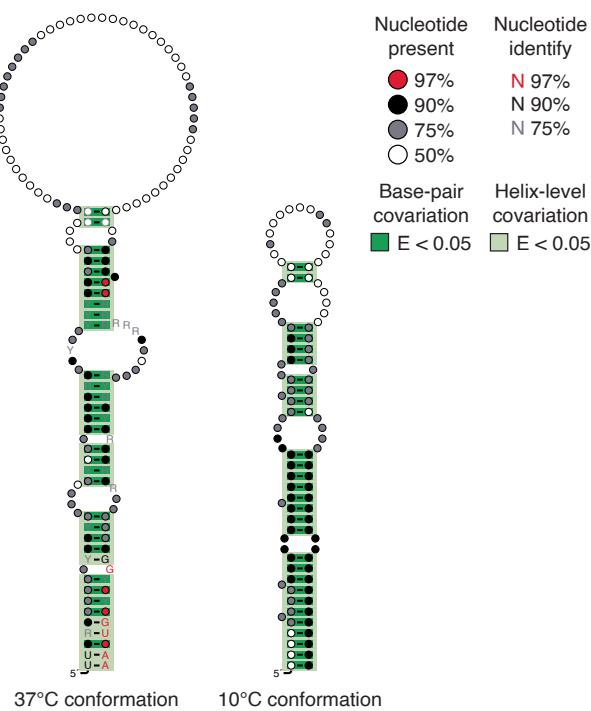
Extended data is available for this paper at https://doi.org/10.1038/s41587-025-02739-0.

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41587-025-02739-0.

Correspondence and requests for materials should be addressed to Danny Incarnato.
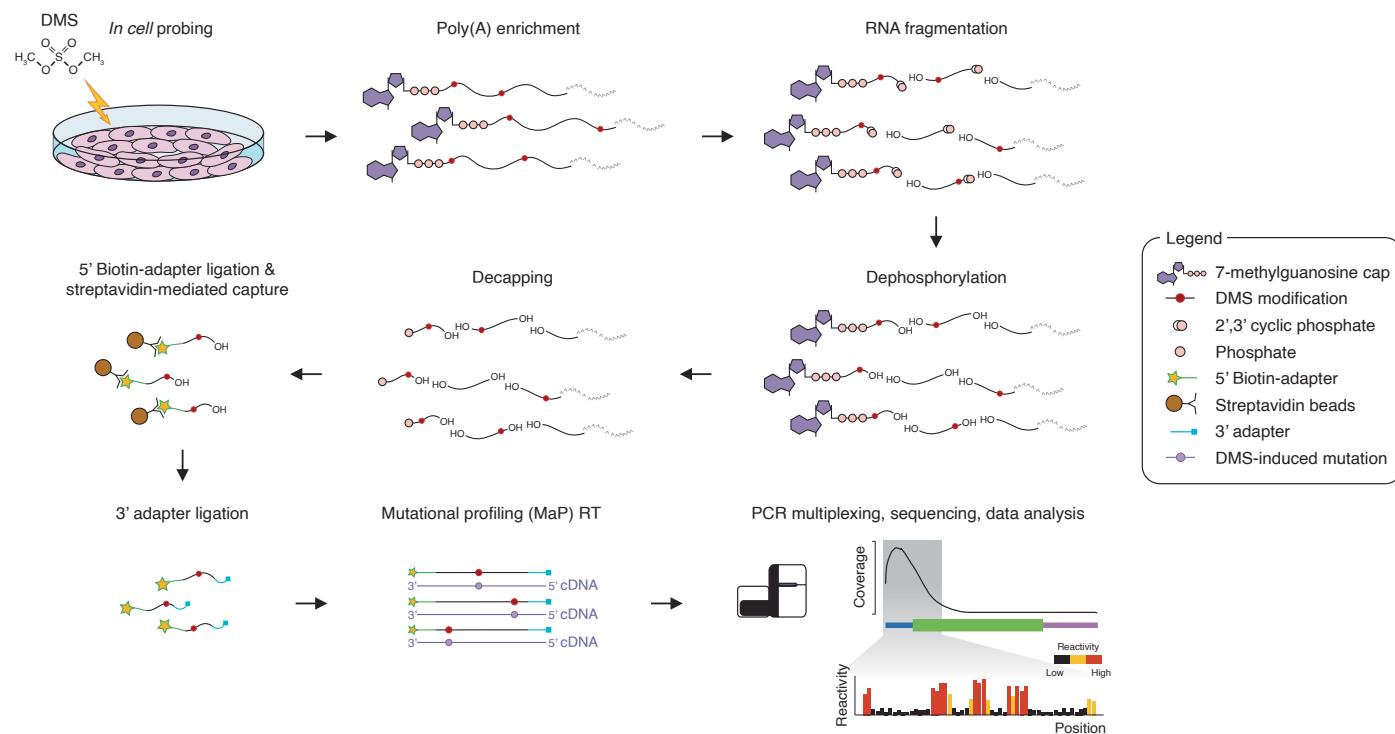
Peer review information *Nature Biotechnology* thanks Hashim Al-Hashimi and Yiliang Ding for their contribution to the peer review of this work.

Reprints and permissions information is available at www.nature.com/reprints.
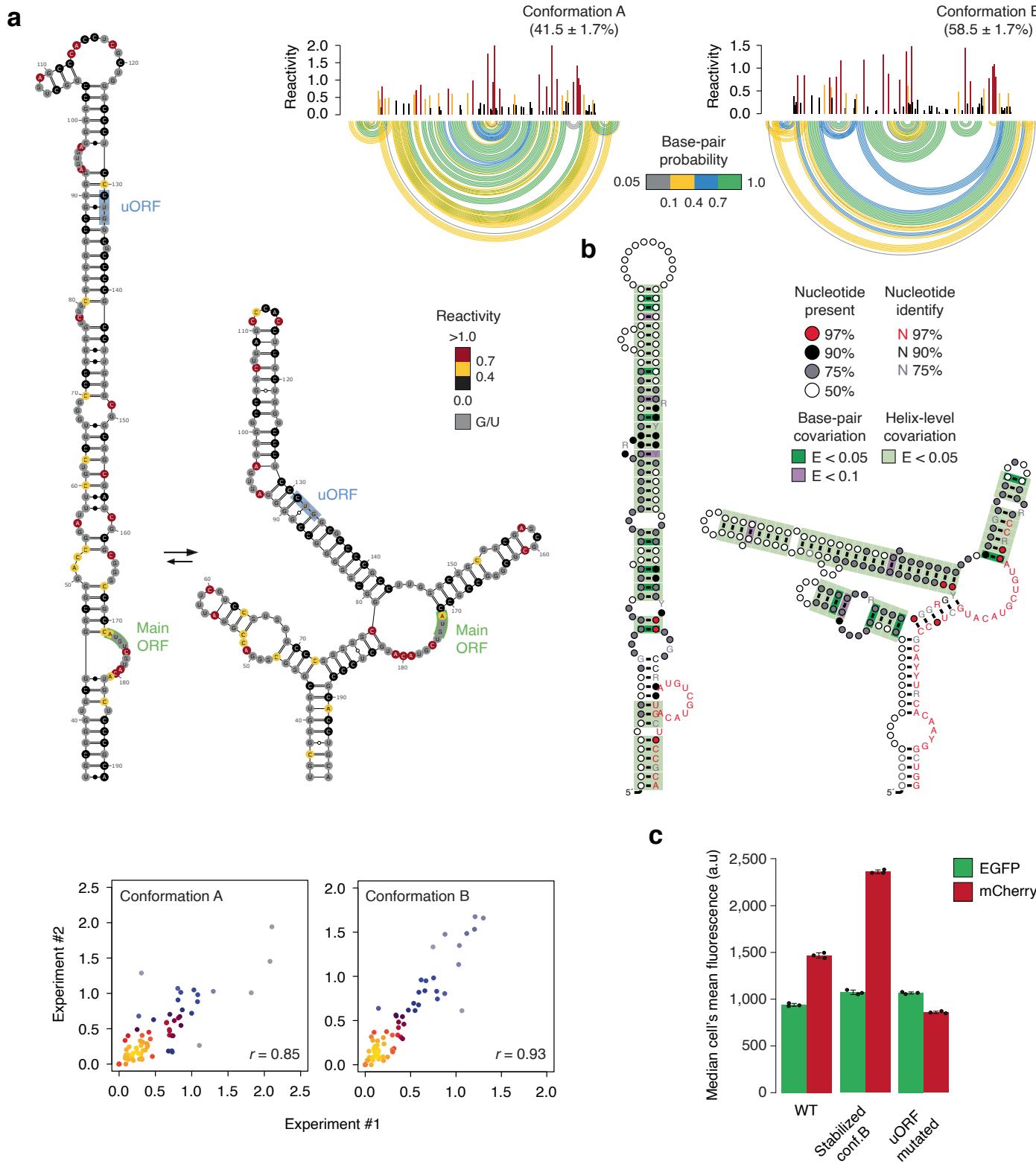
Extended Data Fig. 1 | Structure and conservation of the *cspG* RNA thermometer. (**a**) Secondary structure models of the 5′ UTR of *cspG* at 37 °C and 10 °C, with overlaid *in vitro* DMS reactivities, along with reactivity profiles and base-pairing probabilities for both conformations. Reactivities are averaged across two independent experiments. (**b**) Structure models for the two conformations of the identified *cspG* RNA thermometer, inferred by phylogenetic analysis. Base-pairs showing significant covariation (as determined by R-scape) are boxed in dark green (E-value < 0.05). Helices showing helix-level covariation support (E-value < 0.05) are boxed in light green.

Extended Data Fig. 2 | Overview of the 5′UTR-MaP library preparation strategy. Cells are subjected to in vivo probing, after which poly(A)+ RNA is isolated and fragmented. Fragments are dephosphorylated to remove any endogenous 5′ phosphate and to resolve 2′-3′-cyclic phosphates generated by chemical fragmentation. Capped RNA fragments are decapped, leaving a 5′ phosphate that can be used to ligate a biotinylated adapter. Ligated fragments are captured via streptavidin-coated beads, after which an adapter is ligated to the 3′ end, and library is prepared as per standard MaP protocol.

Extended Data Fig. 3 | Characterization of the *TXNL4A* uORF-regulating RNA structural switch. (**a**) Secondary structure models for the two conformations of the *TXNL4A* 5′ UTR as identified via ensemble deconvolution from targeted DMS-MaPseq analysis, with overlaid in vivo DMS reactivities, along with reactivity profiles and base-pairing probabilities for both conformations. Reactivities are averaged across the two replicate experiments. The scatter plots depict the correlation of base reactivities for the deconvolved conformations across the two replicate experiments. (**b**) Structure models for the two conformations of the

TXNL4A 5′ UTR, inferred by phylogenetic analysis. Base-pairs showing significant covariation (as determined by R-scape) are boxed in dark green (E-value < 0.05) or purple (E-value < 0.1). Helices showing helix-level covariation support are boxed in light green (E-value < 0.05). (**c**) Histogram depicting the median of cell's mean fluorescence in HEK293 cells expressing a dual-frame vector, harboring the 5′ UTR of *TXNL4A*, either wild type, with the uORF start codon mutagenized from CUG to CCG, or mutagenized to stabilize conformation B. Error bars represent the standard deviation of 3 independent biological replicates.

nature portfolio

Corresponding author(s):   Danny Incarnato

Last updated by author(s):   May 19, 2025

Reporting Summary

Nature Portfolio wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Portfolio policies, see our Editorial Policies and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a	Confirmed	
☐	☒	The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☐	☒	A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒	The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☒	☐	A description of all covariates tested
☐	☒	A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☐	☒	A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☐	☒	For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
☒	☐	For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☒	☐	For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☐	☒	Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection	Sequencing data was generated using either DNBSEQ-G400 (MGI) or NextSeq 1000 (Illumina)
Data analysis	Cutadapt v4.4 (https://github.com/marcelm/cutadapt/) PEAR v0.9.11 (https://cme.h-its.org/exelixis/web/software/pear/) RNA Framework v2.8.3 (https://www.rnaframework.com/) Bowtie v1.3.1 (https://bowtie-bio.sourceforge.net/index.shtml) Bowtie v2.3.5.1 (https://bowtie-bio.sourceforge.net/bowtie2/index.shtml) Samtools v1.15.1 (https://www.htslib.org/) DRACO v1.2 (https://github.com/dincarnato/draco) BEDTools v2.31.0 (https://github.com/arq5x/bedtools2) DAVID (https://david.ncifcrf.gov/tools.jsp) ViennaRNA v2.5.1 (https://www.tbi.univie.ac.at/RNA/) Mugsy v1.2.3 (https://mugsy.sourceforge.net/) Infernal v1.1.3 (http://eddylab.org/infernal/) R-scape v2.0.0.q (http://eddylab.org/R-scape/) Custom scripts: filterMM, consensusFold, cm-builder, stockholmPolish (https://github.com/dincarnato/labtools)

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Portfolio guidelines for submitting code & software for further information.

Data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:
- Accession codes, unique identifiers, or web links for publicly available datasets
- A description of any restrictions on data availability
- For clinical datasets or third party data, please ensure that the statement adheres to our policy

> Sequencing data has been deposited to the Gene Expression Omnibus (GEO) database, under the accession GSE247244. Additional processed files (including the browsable set of conserved structurally-heterogeneous regions) are available at https://www.incarnatolab.com/datasets/EcoliEnsembles_Borovska_2024.php.

Research involving human participants, their data, or biological material

Policy information about studies with human participants or human data. See also policy information about [sex, gender (identity/presentation), and sexual orientation](sex, gender) and race, ethnicity and racism.

Reporting on sex and gender	The study did not involve human participants
Reporting on race, ethnicity, or other socially relevant groupings	The study did not involve human participants
Population characteristics	The study did not involve human participants
Recruitment	The study did not involve human participants
Ethics oversight	The study did not involve human participants

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	No sample size calculations were performed. Consistency of replicates was checked using n=2 biologically-independent experiments. All findings from this paper are limited to transcriptome regions showing consistent results across both biological replicates.
Data exclusions	Sequencing reads failing to meet the criteria described in Methods were excluded from the analysis.
Replication	For transcriptome-wide chemical probing experiments, n=2 experiments were performed. This was deemed sufficient given the extremely high reproducibility (Pearson Correlation Coefficient ≥ 0.94). For each western-blot assay, two biological replicates were performed starting from two independent colonies for E. coli, or from two or three independent transfections for HEK293 cells
Randomization	No randomization was necessary in this study because all specimens came from the same source.
Blinding	Blinding was irrelevant to this study because all data were analyzed using the same exact pipeline.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
☐	☒ Antibodies
☐	☒ Eukaryotic cell lines
☒	☐ Palaeontology and archaeology
☒	☐ Animals and other organisms
☒	☐ Clinical data
☒	☐ Dual use research of concern
☒	☐ Plants

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☒	☐ Flow cytometry
☒	☐ MRI-based neuroimaging

Antibodies

Antibodies used	ANTI-FLAG® M2 antibody (cat. F1804, Merck), anti-LacI [9A5] universal antibody (cat. EG1501, Kerafast), EGFP polyclonal antibody (cat. CAB4211, ThermoFisher Scientific), HA tag polyclonal antibody (cat. PA1-985, ThermoFisher Scientific), GAPDH monoclonal antibody (cat. 60004-1-Ig, Proteintech)
Validation	Validation information is available from the respective vendors' websites: ANTI-FLAG® M2 antibody (https://www.sigmaaldrich.com/EN/en/product/sigma/b3111) anti-LacI [9A5] universal antibody (https://www.kerafast.com/item/740/anti-laci-9a5-universal-antibody) EGFP polyclonal antibody (https://www.thermofisher.com/antibody/product/eGFP-Antibody-Polyclonal/CAB4211) HA tag polyclonal antibody (https://www.thermofisher.com/antibody/product/HA-Tag-Antibody-Polyclonal/PA1-985) GAPDH monoclonal antibody (https://www.ptglab.com/products/GAPDH-Antibody-60004-1-Ig.htm)

Eukaryotic cell lines

Policy information about cell lines and Sex and Gender in Research

Cell line source(s)	Cells were gifted by the Cell Culture Facility of the GBB institute (University of Groningen)
Authentication	Cells were not authenticated
Mycoplasma contamination	Cells were tested for Mycoplasma contamination at passage 2 post-revival
Commonly misidentified lines (See ICLAC register)	*Name any commonly misidentified cell lines used in the study and provide a rationale for their use.*

Plants

Seed stocks	*Report on the source of all seed stocks or other plant material used. If applicable, state the seed stock centre and catalogue number. If plant specimens were collected from the field, describe the collection location, date and sampling procedures.*
Novel plant genotypes	*Describe the methods by which all novel plant genotypes were produced. This includes those generated by transgenic approaches, gene editing, chemical/radiation-based mutagenesis and hybridization. For transgenic lines, describe the transformation method, the number of independent lines analyzed and the generation upon which experiments were performed. For gene-edited lines, describe the editor used, the endogenous sequence targeted for editing, the targeting guide RNA sequence (if applicable) and how the editor was applied.*
Authentication	*Describe any authentication procedures for each seed stock used or novel genotype generated. Describe any experiments used to assess the effect of a mutation and, where applicable, how potential secondary effects (e.g. second site T-DNA insertions, mosiacism, off-target gene editing) were examined.*