Supplementary Discussions

Description of end-enriched RNA-seq (Rend-seq)

End-enriched RNA sequencing (Rend-seq), recently developed in our lab[35], allows for the simultaneous measurement of RNA-levels and single nucleotide mapping of transcripts 5' and 3' ends. In a sparsely fragmented pool of RNA, fragment ends are much more likely (inversely proportional to per base fragmentation probability) to originate from ends initially present in the RNA sample. This end-enrichment provides the basis of Rend-seq's readout. Size selection of short fragments (15-45 nt) prior to cDNA conversion allows retrieval of information for both 5' and 3' ends from a short single read sequencing run. Following sequencing of a quantitatively generated cDNA library and mapping to the genome, the 5' and 3'-mapped read counts of fragments are summed in separate channels (displayed in orange and blue, respectively). The resulting signal from a simple transcript is then a single-nucleotide wide peaks (with around 50× increase in signal relative to the transcript body) in 5' and 3'-mapped channels at respective ends, and largely uniform read coverage across the transcript body (example schematically shown in Extended Data Fig. 6a). Lalanne et al[35] contains detailed experimental methods, mathematical derivations, and description of the computational pipeline used for the analysis of Rend-seq data.

Note on previous computational approaches for intrinsic terminator identification

Given the rich primary sequence features of intrinsic transcription terminators, numerous previous studies have designed algorithms to computationally identify intrinsic terminators from genome sequences, e.g.,[70–73]. While these previous terminator identification strategies attempted to balance false positives and false negatives, our current approach is focused on being stringent at the expense of possible false negatives. Our approach is based on rigorous thermodynamic RNA secondary structure prediction[59] and relies on looking for canonical intrinsic terminator structures (long stretch U residues preceded by a strong hairpin). Importantly, thresholds were selected on a species-by-species basis by generating a null set of RNA secondary structures from folding randomly selected regions in the genome. This provides a species-specific stringent cutoff on hairpin parameters used to call putative intrinsic terminators, mitigating possible difficulties with different genome properties, for example varying GC content. Additionally, an advantage of the current work is that we used transcriptomics data (Rend-seq) in five diverse species to directly benchmark the fraction of false positive of our computational pipeline, demonstrating the false discovery rate at 1% or lower.

Our bioinformatic pipeline searches for canonical terminator features deemed critical for function in model organisms such as *E. coli* and *B. subtilis*, which might miss some terminators in more diverse bacteria given potential mechanistic differences[51]. Indeed, some clades have few identified terminators per genome, using the approach here and with other computational methods[71,74,75] (e.g., Actinobacteria, Mycoplasma, Campylobacterota). It will be an important avenue of future research to validate conclusions derived from proximity of terminator hairpin stems to coding sequences once reliable and experimentally validated *in silico* classification schemes are developed in these species.

On the possibility of translation initiation on nascent mRNAs

Our observations of the delay between mRNA and protein synthesis raise an interesting question of whether ribosomes can even initiate translation on nascent mRNAs in *B. subtilis*. In this model, the delay may be primarily due to slow translation initiation on mRNAs within the nucleoid, instead of due to slow translation elongation. Several lines of evidence suggest that this model is unlikely.

First, we can estimate the translation initiation time for *pycA-lacZα* using data for the truncated construct (Fig. 1e, Extended Data Fig. 2c). We found that it takes 64 s longer to synthesize the full-length construct than the truncated one (1000 aa shorter). These constructs are expected to have similar translation initiation rates because they share the same UTR and 5' CDS. The time difference can therefore be attributed to the time it takes to translate the 1000 aa region that is absent from the truncated version, giving a translation elongation rate of 16 aa/s. Based on this elongation rate, the elongation time for the full-length *pycA-lacZα* would be 1255 aa/(16 aa/s) = 78 s, which is only 1 s shorter than the observed first appearance time of proteins (Fig. 1d). Compared to the first appearance time of full-length mRNA (45 s, Fig. 1d), the estimated translation initiation time of 1 s indicates that ribosome can still engage in translation on nascent mRNA for this gene.

Second, if translation initiated only after the full-length mRNA were transcribed, translation elongation would need to complete in 34 s, i.e. the first appearance time difference between *pycA-lacZα* mRNA and protein (Fig. 1d). This would require a translation elongation rate of 1255 aa/34 s = 37 aa/s for *pycA*, which is much faster than the 16 aa/s measured as described above, and is nearly twice as fast as any translation elongation rate measured to date, e.g., references [25,76], and unlikely to be physiologically possible for the following reasons. First, given the number of ribosomes relative to protein mass per cell, the translation elongation rate would need to be 18 aa/s for *B. subtilis* doubling

every 20 min (estimated using the method in reference [25] with the ribosomal proteins proteome fraction estimated from the ribosome profiling data in reference [35]). Having a much higher elongation rate would imply a wasteful situation such that a large fraction of ribosomes is not participating in protein synthesis. Second, work in *E. coli* has shown that the translation elongation rate is limited by binding of ternary complexes to the ribosome,[25] which is close to being diffusion limited[77]. As the concentrations of translation elongation factors (EF-Tu, EF-G, tRNAs, etc) is very similar between *E. coli* and *B. subtilis* doubling every 20 minutes,[35,41,78] it is unlikely that the translation rate in *B. subtilis* would be much faster than in *E. coli*. All of this is not to rule out the possibility that for some transcripts, full length mRNA is synthesized completely before translation begins (especially for shorter transcripts with low translation efficiency), but rather to demonstrate that translation of nascent mRNAs does still occur.

Transcription rates in mutant backgrounds

To assess the contribution of various non-essential transcription factors and RNAP subunits to the transcription rate, we measured the induction kinetics of our *pycA-lacZα* construct in different knockout backgrounds (Δ*greA*, Δ*nusG*, Δ*rpoZ*, Δ*ykzG*). *greA* encodes the transcription factor GreA, which helps resolve stalled transcription complexes.[79] *nusG* encodes the transcription factor NusG that stimulates RNAP pausing.[22,31,80,81] *rpoZ* encodes the accessory RNAP ω subunit, a universally conserved protein that, although well characterized in *E. coli* for its roles in RNAP assembly and stringent response, has remained largely unstudied in gram-positive bacteria.[82] *ykzG* encodes the accessory RNAP ε subunit, which is specific to gram-positive bacteria, but its role is still poorly understood.[82] The time of first appearance of *pycA-lacZα* mRNA in these mutants was not substantially different from that observed in WT (Extended Data Fig. 3), and not slow enough to promote coupling. Therefore, these factors alone are not sufficient to explain the much faster transcription elongation rate in *B. subtilis*.

**Phylogeny of domain architecture of proteins implicated in coupling**

The mechanistic basis of transcription-translation coupling has been a very active area of research[4–8,10,11,83–88]. Initially proposed to be mediated through binding of the transcription termination factor NusG to ribosomal protein S10[5], additional structural characterizations are now highlighting that multiple different coupling states likely exist[6,10,86]. These states may correspond to different transcripts and/or stages of transcription, and may also illustrate mechanistic differences underlying coupling in different species[11]. While residues involved in the NusG/S10 bridge are widely conserved[5], lineage-specific domains of essential proteins (e.g., C-terminal domain of NusA in *M. pneumonia*[11], sequence insertion βSI2 in β subunit of RNAP in *E. coli*[10]) have recently been implicated in coupling. Multiple sequence alignments of NusA, NusG and RpoB (Extended Data Fig. 4, S10 did not have a diverse domain architecture and is thus not included) from species shown in Fig. 4 of the current work show the relationship between presence/absence of domains in these proteins and proximity of intrinsic terminators from coding sequences. Specifically, the C-terminal domain of NusA (orange box in NusA alignment in Extended Data Fig. 4) is missing in a large number of Gram-positive bacteria[69] (partly present in Mollicutes, which includes Spiroplasmas and Mycoplasmas, red brace in NusA alignment in Extended Data Fig. 4) and Campylobacterota. In addition, as noted in[10,68], sequence insertion βSI2 (green box in RpoB alignment in Extended Data Fig. 4) is largely absent in Gram-positive bacteria. No single structural domain in these proteins perfectly overlaps with the bioinformatic signature of runaway transcription (high fraction of ORF-proximal intrinsic terminators), suggesting a complex landscape of mechanisms for translation-coupled transcription.

Supplemental References

70.   Kingsford, C. L., Ayanbule, K. & Salzberg, S. L. Rapid, accurate, computational discovery of Rho-independent trans cription terminators illuminates their relationship to DNA uptake. *Genome Biol.* **8**, R22 (2007).

71.   De Hoon, M. J. L., Makita, Y., Nakai, K. & Msyasio, S. Prediction of transcriptional terminators in bacillus subtiiis and species. *PLoS Comput. Biol.* **1**, 0212–0221 (2005).

72.   Mitra, A., Kesarwani, A. K., Pal, D. & Nagaraja, V. WebGeSTer DB-A transcription terminator database. *Nucleic Acids Res.* **39**, 1–7 (2011).

73.   Gardner, P. P., Barquist, L., Bateman, A., Nawrocki, E. P. & Weinberg, Z. RNIE: Genome-wide prediction of bacterial intrinsic terminators. *Nucleic Acids Res.* **39**, 5845–5852 (2011).

74.   Washio, T., Sasayama, J. & Tomita, M. Analysis of complete genomes suggests that many prokaryotes do not rely on hairpin formation in transcription termination. *Nucleic Acids Res.* **26**, 5456–5463 (1998).

75.   Mitra, A., Angamuthu, K., Jayashree, H. V. & Nagaraja, V. Occurrence, divergence and evolution of intrinsic terminators across Eubacteria. *Genomics* **94**, 110–116 (2009).

76.   Young, R. & Bremer, H. Polypeptide-chain-elongation rate in Escherichia coli B/r as a function of growth rate. *Biochem. J.* **160**, 185–94 (1976).

77.   Klumpp, S., Scott, M., Pedersen, S. & Hwa, T. Molecular crowding limits translation and cell growth. *Proc. Natl. Acad. Sci. U. S. A.* **110**, 16754–9 (2013).

78.   Pedersen, S., Bloch, P. L., Reeh, S. & Neidhardt, F. C. Patterns of protein synthesis in E. coli: a catalog of the amount of 140 individual proteins at different growth rates. *Cell* **14**, 179–190 (1978).

79.   Kusuya, Y., Kurokawa, K., Ishikawa, S., Ogasawara, N. & Oshima, T. Transcription factor GreA contributes to resolving promoter-proximal pausing of RNA polymerase in Bacillus subtilis cells. *J. Bacteriol.* **193**, 3090–3099 (2011).

80.   Yakhnin, A. V. & Babitzke, P. NusG/Spt5: Are there common functions of this ubiquitous transcription elongation factor? *Current Opinion in Microbiology* **18**, 68–71 (2014).

81.   Yakhnin, A. V., Yakhnin, H. & Babitzke, P. Function of the Bacillus subtilis transcription elongation factor NusG in hairpin-dependent RNA polymerase pausing in the trp leader. *Proc. Natl. Acad. Sci. U. S. A.* **105**, 16131–16136 (2008).

82.   Weiss, A. & Shaw, L. N. Small things considered: The small accessory subunits of RNA polymerase in Gram-positive bacteria. *FEMS Microbiology Reviews* (2015).

doi:10.1093/femsre/fuv005

83.   Demo, G. *et al.* Structure of RNA polymerase bound to ribosomal 30S subunit. *Elife* **6**, 1–17 (2017).

84.   McGary, K. & Nudler, E. RNA polymerase and the ribosome: The close relationship. *Curr. Opin. Microbiol.* **16**, 112–117 (2013).

85.   Artsimovitch, I. Rebuilding the bridge between transcription and translation. *Mol. Microbiol.* **108**, 467–472 (2018).

86.   Webster, M., Takacs, M. & Zhu, C. Structural basis of transcription-translation coupling and collision in bacteria . *bioRxiv* 1–32 (2020).

87.   Washburn, R. S. *et al.* Escherichia coli NusG links the lead ribosome with the transcription elongation complex. *bioRxiv* 871962 (2019). doi:10.1101/871962

88.   Saxena, S. *et al.* Escherichia coli transcription factor NusG binds to 70S ribosomes. *Mol. Microbiol.* **108**, 495–504 (2018).
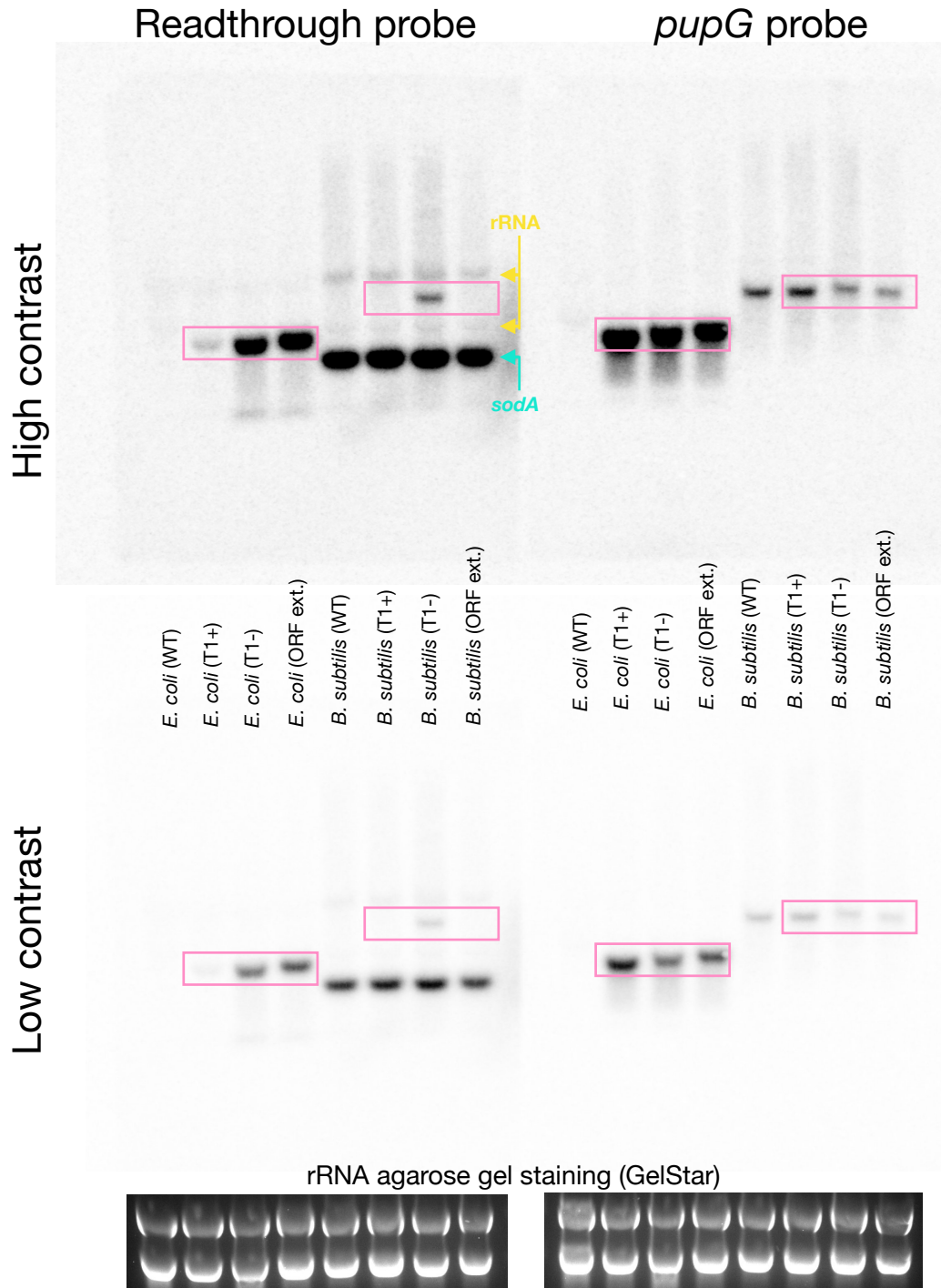
Supplementary Figure 1. Uncropped Northern blot raw data for ORF extension experiment. Left blot is with probe to the long (readthrough) mRNA isoform. Right blot is with the probe upstream of the *pupG* terminator. Strain corresponding to each sample is indicated. Top and bottom show the same blot at two levels of contrast. Pink rectangles indicate cropping regions for Fig. 2b-c. Position of non-specific bands from rRNA (as determined by staining of gel prior to transfer) are indicated in yellow. For B. subtilis samples, terminator T2 was from *sodA*. The band corresponding to the putative *sodA* transcript is indicated in cyan. rRNA bands from staining of agarose RNA gels prior to transfer are shown at bottom.