POLITECNICO DI TORINO
Repository ISTITUZIONALE

Systems-level analysis identifies protein-RNA interactions contributing to autotrophy of Clostridium autoethanogenum

Original
Systems-level analysis identifies protein-RNA interactions contributing to autotrophy of Clostridium autoethanogenum / Re, A., Politano, G.M.M., Reinmets, K., Benso, A., Girbal, L., Cocaign-Bousquet, M., Valgepea, K.. - (2025). [10.64898/2025.12.18.695208]

Availability:
This version is available at: 11583/3008424 since: 2026-03-09T13:11:17Z

Publisher:

Published
DOI:10.64898/2025.12.18.695208

Terms of use:

This article is made available under terms and conditions as specified in the corresponding bibliographic description in the repository

Publisher copyright

(Article begins on next page)

11 November 2026

bioRxiv preprint doi: https://doi.org/10.64898/2025.12.18.695208; this version posted December 21, 2025. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Systems-level analysis identifies protein-RNA interactions contributing to autotrophy of *Clostridium autoethanogenum*

Angela Re[1,*], Gianfranco Michele Maria Politano[2], Kristina Reinmets[3], Alfredo Benso[2], Laurence Girbal[4], Muriel Cocaign-Bousquet[4], Kaspar Valgepea[3,*]

[1]Department of Applied Science and Technology, Politecnico di Torino, Torino, Italy

[2]Department of Control and Computer Engineering, Politecnico di Torino, Torino, Italy

[3]Institute of Bioengineering, University of Tartu, Tartu, Estonia

[4] TBI, Université de Toulouse, CNRS, INRAE, INSA, Toulouse, France

* Corresponding author

Abstract

Burning fossil fuels drives climate change with detrimental consequences on humankind while waste accumulation from increasing consumption threatens biosustainability globally. Gas fermentation enables to recycle carbon oxides (CO and CO_2) from industrial waste gases and gasified waste into value-added products using gas-fermenting microbes, namely acetogens. However, our limited understanding of gene function and metabolic regulation is hindering rational engineering of acetogen cell factories. In this work, we identified, for the first time in acetogens, genome-wide protein-RNA interactions contributing to autotrophy in the model-acetogen *Clostridium autoethanogenum* by combining steady-state chemostat cultivation, functional genomics, and computational methods. We first detected limited and uncoupled transcriptional and translational regulation between autotrophy and heterotrophy. Rigorous mapping of genome-wide transcriptional architecture revealed both differential usage and signal strength of transcriptional start and termination sites between genes and growth substrates. We then used computational tools to reconstruct protein-RNA interactions for differentially regulated genes predicting 14 trans-acting regulatory RNA-binding proteins (RBPs) involved in post-transcriptional regulation and contributing to autotrophy. Most RBPs, two of which are translationally regulated, perform RNA modifications and regulate mRNA stability while others target translation-related genes. Our work provides valuable knowledge for metabolic engineering of acetogens and potentially contributes towards understanding primordial life on Earth.

bioRxiv preprint doi: https://doi.org/10.64898/2025.12.18.695208; this version posted December 21, 2025. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Introduction

Our continued reliance on fossil fuels drives greenhouse gas emissions that are accelerating climate change with detrimental consequences on the environment and humankind. In addition, increasing accumulation of waste from increasing global population and prosperity is threatening biosustainability through degradation of ecosystems. Gas fermentation has emerged as an attractive technology to tackle these challenges by recycling waste carbon from industrial off-gases and gasified waste into value-added products using gas-fermenting microbes[1,2]. Acetogens are the preferred biocatalysts for capture and conversion of carbon oxides (CO and CO_2) into fuels and chemicals[3,4] as they can use gas as their sole carbon and energy source[5] by employing the most-efficient CO_2-fixation pathway known to date[6,7], the Wood-Ljungdahl pathway (WLP)[5,8]. However, our limited understanding of gene function and metabolic regulation in acetogens is hindering rational engineering of acetogen cell factories[3,4,9-11].

Most studies aimed at elucidating the relationship between genotypes and phenotypes in acetogens have focused on transcription[3]. Results from a limited set of experimental conditions suggest that metabolic fluxes in acetogens are primarily regulated post-translationally[12-14]. However, comprehensive analysis of acetogen multi-omics datasets[15-18] highlights that post-transcriptional regulation of gene expression has a significant effect on the flow of genetic information from DNA to protein[19], in a similar way to other microbes[20]. Key players in post-transcriptional regulation in microbes are RNA-binding proteins (RBPs)[20], encompassing direct[21] or indirect[22] regulation of transcription[23-25], RNA turnover[26], and translation[23,27-29]. Additionally, RBPs can play key roles in regulating microbial metabolism through distinct RBP-RNA interactions [30,31]. However, no RBPs have been described in acetogens so far, and thus we lack knowledge about RBP-RNA interactions. Hence, fundamental understanding of acetogen metabolism and potentially also engineering of cell factories could be advanced through uncovering the unexplored space of RBPs.

The aim of our work was to identify, for the first time in acetogens, genome-wide RPB-RNA interactions that contribute to autotrophy. We used the model-acetogen *Clostridium autoethanogenum* as its metabolism has been explored through numerous omics datasets[3,32-37], it has been engineered to produce various non-native chemicals[3,38,39], and it is used in commercial-scale gas fermentation for ethanol production[2]. We firstly aimed to experimentally determine RBPs using the RNA interactome capture (RIC) technique[40,41] that was recently adapted for the microbe *Escherichia coli*[4], but after numerous attempts failed to engineer a *C. autoethanogenum* strain expressing a heterologous poly(A) polymerase that is needed for genome-wide RNA polyadenylation and pull-down in RIC. We thus employed an alternative comprehensive approach to predict RBPs and RBP-RNA interactions by combining functional genomics and computational methods (**Fig. 1**).

bioRxiv preprint doi: https://doi.org/10.64898/2025.12.18.695208; this version posted December 21, 2025. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

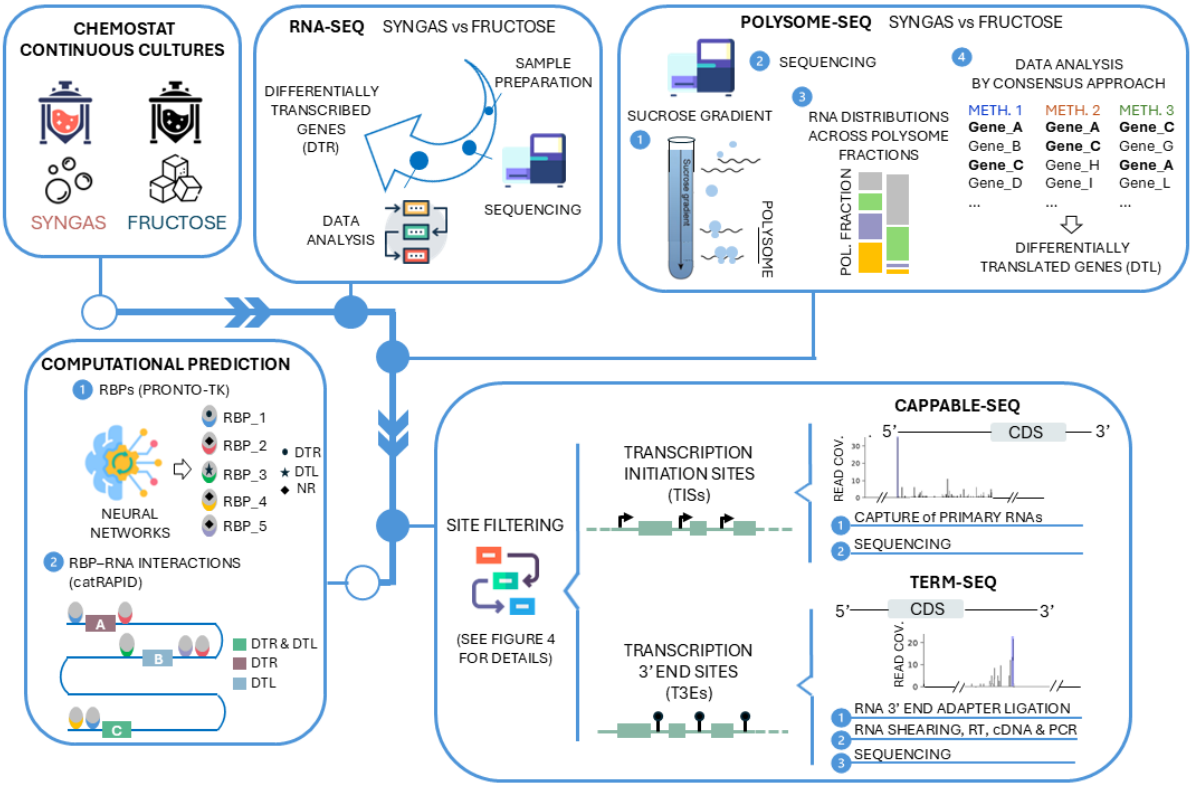

Fig. 1. Integrative functional genomics and computational approach used in this work to predict RNA-binding proteins (RBPs) and protein-RNA interactions contributing to autotrophy of *C. autoethanogenum*. CDS, coding DNA sequence; DTR, differentially transcribed; DTL, differentially translated; NR, not differentially regulated.

Our approach first quantified both genome-wide transcriptional and translational regulation to identify transcripts differentially transcribed and/or translated during autotrophy vs heterotrophy. Transcription levels were quantified using RNA-seq while polysome-seq instead of Ribo-seq was used to evaluate translational status of transcripts since only the former offers resolution at the level of each mRNA copy[43]. Next, we quantified genome-wide transcriptional architecture by accurate mapping of 5'-untranslated regions (UTR) and 3'UTRs at single nucleotide resolution to determine the RNA sequences where RBPs most likely bind[44,45]. 5'UTRs were determined through identification of transcription initiation sites (TISs) by Cappable-seq[46], which simplifies data generation and analysis, and improves reproducibility and robustness[46,47] over the widely used alternative method dRNA-seq[48]. 3'UTRs were identified by mapping transcription 3' ends (T3E) using Term-seq[49] that has an advantage of ensuring that the sequences adjacent to adapters in the sequencing reads represent RNA 3' termini originally present in the samples. We then used PRONTO-TK[50], a neural network-based toolkit, to identify potential RBPs in *C. autoethanogenum* as it uses ProtT5-XL protein embeddings and deep learning to predict GO term-based functions with high precision and computational efficiency. Finally, high-resolution transcript boundaries allowed us to computationally reconstruct RBP-RNA interactions contributing to autotrophy in *C. autoethanogenum* using the inferential tool *cat*RAPID[51], which uniquely combines high-throughput prediction of protein-RNA interactions with multi-parametric rating in an organism-agnostic framework. Importantly, all these analyses were performed from the same autotrophic and heterotrophic cultures that were grown in bioreactor chemostat cultures, thus yielding steady-state data that enable quantification of strictly defined and condition-dependent phenotypes[52,53].

While our work detected limited and uncoupled transcriptional and translational regulation between autotrophy and heterotrophy, *C. autoethanogenum* showed both differential usage and signal strength of

bioRxiv preprint doi: https://doi.org/10.64898/2025.12.18.695208; this version posted December 21, 2025. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

transcriptional start and termination sites between genes and growth substrates. Reconstruction of protein-RNA interactions identified 14 trans-acting regulatory RBPs involved in post-transcriptional regulation of differentially regulated genes for autotrophy. Our work provides valuable knowledge and potential targets for metabolic engineering of acetogens and potentially contributes towards understanding primordial life as the WLP could be first biochemical pathway on Earth[54].

Results

Autotrophic and heterotrophic chemostat cultures of *C. autoethanogenum*

To map genome-wide RPB-RNA interactions involved in enabling autotrophy in acetogens, we grew the *C. autoethanogenum* strain LAbrini[35] on syngas (CO, H_2, CO_2) and fructose in chemostat continuous cultures (**Fig. 1**). Importantly, this allowed us to grow cells at the same specific growth rate (μ; ~0.04 h^{-1} or ~1 day^{-1}), as μ equals dilution rate at steady-state, and thus facilitated unequivocal quantification of transcriptional and translational patterns of autotrophy and heterotrophy without the confounding effects from different μ during exponential autotrophic and heterotrophic batch growth. For instance, from the 2,030 differentially expressed transcripts between *C. autoethanogenum* syngas and fructose batch cultures[55], ~24% were also differentially expressed with faster growth in autotrophic chemostat cultures[33].

Notably, we found that both specific acetate and ethanol production rates (mmol per gram of dry cell weight per day) were ~2.3-fold higher and the specific 2,3-butanediol production rate was 9-fold higher for syngas vs fructose steady-state cultures (**Supplementary Fig. 1a**). The former can be explained by a doubled allocation of carbon into biomass (**Supplementary Fig. 1b**) that is favoured by the more energy-rich nutrient fructose while the latter shows the higher need to regenerate reducing equivalents with by-products during autotrophy when less biomass can be produced. Interestingly, both of our chemostat cultures with the LAbrini strain showed only ~1.5-fold higher acetate production than ethanol (**Supplementary Fig. 1a**) while nearly no ethanol production was detected in fructose-grown batch cultures of *C. autoethanogenum* strain JA1-1[55].

Limited transcriptional and translational control involved in realising autotrophy

We first quantified transcriptional and translational regulation involved in autotrophy (**Fig. 1**) using RNA-seq (**Supplementary Table 1**) and polysome-seq (**Supplementary Table 2**). We note that polysome profiling has been performed only for a few microbes[43,56-59] but not for Clostridia or acetogens. While our RNA-seq analysis quantified transcript levels for 3,557 genes, polysome-seq analysis determined the translation status for 3,875 genes on average across bio-replicates of syngas and fructose cultures. The reproducibility of bio-replicates across syngas and fructose cultures was estimated with Spearman correlation coefficients of 0.99 and 0.71 for RNA-seq and polysome-seq data (**Supplementary Fig. 2a and 3**), respectively. Bio-replicates for RNA-seq data also clustered well showing clearly distinguished transcriptome profiles between syngas and fructose cultures (**Supplementary Fig. 2b and 2c**). Analysis of polysome-seq data (see Methods for details) led to the removal of an outlier bio-replicate culture for both syngas and fructose (**Supplementary Fig. 4**). Analysis of RNA-seq data yielded 1,784 genes with statistically differential expression (q-value < 0.05) between syngas and fructose growth (**Fig. 2a**; **Supplementary Table 3**). We determined the top 5% ranking ones according to q-value (lower value ranks higher) as differentially transcribed genes (DTRs) to ensure consistency with determination of differentially translated genes (DTLs) from polysome-seq data (see next paragraph). Thus, 176 genes were considered as DTR with 109 and 67 being down- or up-regulated, respectively, on syngas vs fructose (**Fig. 2a**; **Supplementary Table 3)**.

Since polysome-seq quantifies transcripts bound to different numbers of ribosomes at single mRNA copy level, it allows to capture distinct traits compared to Ribo-seq that are indicative of translational control. To

bioRxiv preprint doi: https://doi.org/10.64898/2025.12.18.695208; this version posted December 21, 2025. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

take advantage of this for determining DTLs, we established three complementary metrics for translational regulation (**Fig. 2b**; see Methods for details). First, we quantified the difference in translation status per gene between syngas and fructose (TL-STATUS), where the translation status is calculated as the ratio between the sum of the mRNA proportions quantified in the ribosome-bound and ribosome-free fractions. Positive value for TL-STATUS indicates up-regulation and negative down-regulation on syngas vs fructose (**Supplementary Table 4**). The translation status defined here is equivalent to the term 'ribosome occupancy' traditionally used for polysome-seq data[43]. Our data show that >90% of mRNA copies for each gene are ribosome-bound (**Supplementary Table 2**), similar to *E. coli*[58] but higher than in *Lactococcus lactis*[59] and *Saccharomyces cerevisiae*[43,57]. Second, we quantified the difference between the polysome profiles in the two growth conditions for each gene using the variance-weighted Aitchison's distance (TL-PROFILE) (**Supplementary Table 4**). Third, we quantified for each gene the number of polysome fractions for which the difference in mRNA proportions between syngas and fructose was statistically significant by bootstrapping (POLY-FRACTIONS) (**Supplementary Table 4**).

To evaluate these three metrics as indicators of gene-level translational regulation involved in realising autotrophy, we compared the genes within the top 5% ranks of each metric. From the total of 441 genes among the top 5% ranks, ~2% (8) and ~14% (61) were shared between three or two metrics, respectively (**Fig. 2c**). Interestingly, each metric identified on average ~29% unique genes demonstrating the relevance of using multiple metrics to evaluate translational regulation. Consistent with this was the enrichment of unique KEGG Orthology functional categories[60] or Pfam entries[61] within the top 5% gene sets of each metric (**Fig. 2d**). Therefore, to capture the potentially unique information on translational regulation carried by the different metrics while also considering the general overlap between the metrics, we decided to determine DTLs from polysome-seq data as the top 5% genes from ranking the average rank for the three metrics (CONSENSUS). Accordingly, 186 genes were considered as DTLs, all of which were up-regulated on syngas vs fructose (**Supplementary Table 4**). Thus, DTLs covered 42% of all the genes among the top 5% ranks of the three metrics (**Fig. 2c**). Expectedly, DTLs showed strongest enrichment for genes involved in translation (**Fig. 2e**), indicating reasonable definition of DTLs[20].

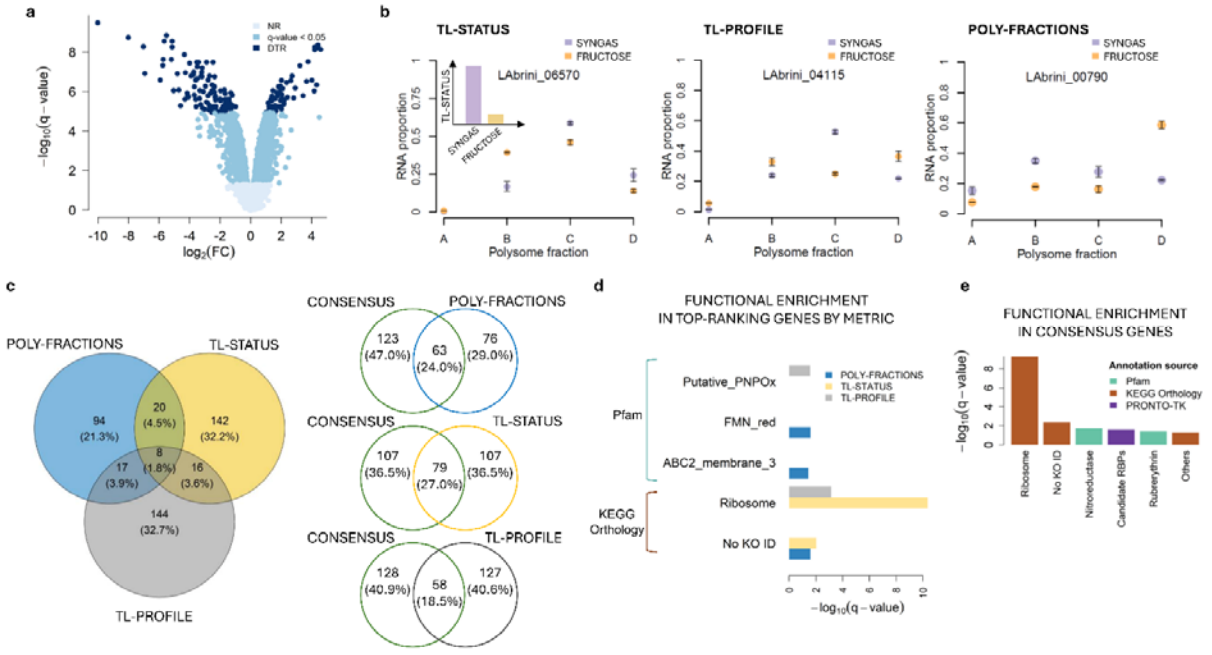

Fig. 2. Limited transcriptional and translational control of autotrophy in *C. autoethanogenum*. a Volcano plot showing genes determined as differentially transcribed (DTR; #176) as a subset of 1,784 genes with statistically different expression (q-value < 0.05) between syngas and fructose chemostats. NR, not differentially regulated; FC,

bioRxiv preprint doi: https://doi.org/10.64898/2025.12.18.695208; this version posted December 21, 2025. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

fold-change **b** Representative polysome profiles of top-ranking genes for each of the three metrics display the mRNA proportions of a gene across the ribosome-free fraction (A) and the polysome-bound fractions B (monosome), C (low ribosome density), and D (high ribosome density). For the TL-STATUS metric, inset shows the calculated translation status for syngas and fructose chemostats. Data are average ± standard deviation between biological duplicate chemostat cultures. **c** The three-set Venn diagram showing the overlaps between the differentially translated genes (DTL) identified by the POLY-FRACTIONS, TL-STATUS, and TL-PROFILE metrics. The two-set Venn diagrams show the overlap between the DTL identified by the CONSENSUS approach and each of the three metrics. **d** Functional classes enriched in DTLs according to each metric by hypergeometric test (FDR = 5%). Pfam, Pfam entries[61] from eggNOG-mapper 2.1.12[62]; KEGG Orthology, KEGG Orthology (KO) functional categories[60] from Valgepea *et al.* (Level 3 in Table S3[13]); NS, not significant. **e** Functional classes enriched in DTLs according to CONSENSUS approach by hypergeometric test (FDR = 5%). PRONTO-TK, computationally predicted RNA-binding proteins by PRONTO-TK[50].

Transcriptional and translational control for autotrophy are uncoupled

Comparing DTRs with DTLs revealed that ~47% of genes (166) were differentially regulated exclusively at transcript levels and ~50% (176) exclusively at the level of translation (**Fig. 3a**). While no DTLs were down-regulated on syngas vs fructose, the concordantly up-regulated six genes included the Fd-NAD$^+$ oxidoreductase Rnf complex subunit RsxC (LABRINI_16130) and the NADH-quinone oxidoreductase subunit NuoF (07815) (**Supplementary Tables 3 and 4**). This very low ~2% overlap clearly shows that transcriptional and translational regulation involved in realising autotrophy in *C. autoethanogenum* are uncoupled. Consistently, no enriched functional categories are shared between DTRs and DTLs (**Fig. 3b**). Furthermore, the enrichment data show that two key metabolic pathways for autotrophy – acetate & ethanol production and energy conservation – are only regulated at transcript levels while ribosomal proteins are only regulated at translational level (**Fig. 3b**). The former includes various alcohol and aldehyde dehydrogenases (including Adh4 (09125), the most abundant Adh enzyme in *C. autoethanogenum*)[13], the bifunctional AdhE1 (18700), and the aldehyde Fd oxidoreductase AOR2 (00495). Transcriptional regulation of these genes was also seen with faster autotrophic growth of *C. autoethanogenum*[33]. Additionally, strictly transcriptional regulation was used to up-regulate components of the Rnf complex (16135–16155), that generates proton motive force to drive the ATPase in *C. autoethanogenum*[63,64] (see also **Fig. 6)**. Transcripts of the Rnf complex are also up-regulated in *C. autoethanogenum* syngas vs fructose batch cultures[55]. Interestingly, among the translationally regulated ribosomal proteins, S7 (09680) is known to act as an autogenous translation repressor[65] while L35 (06570) can act in an extra-ribosomal context to inhibit ornithine decarboxylases[66], which support acid-resistance and maintenance of optimal growth in Clostridia[67]. Additionally, L7Ae (09690) participates in tRNA processing, RNA modification, and translational regulation in archaea[68]. Only four genes showed discordant changes between transcriptional and translational regulation, with 07435 and 00130 putatively involved in iron metabolism[69,70], suggesting potential buffering effects (**Supplementary Tables 3 and 4**).

bioRxiv preprint doi: https://doi.org/10.64898/2025.12.18.695208; this version posted December 21, 2025. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

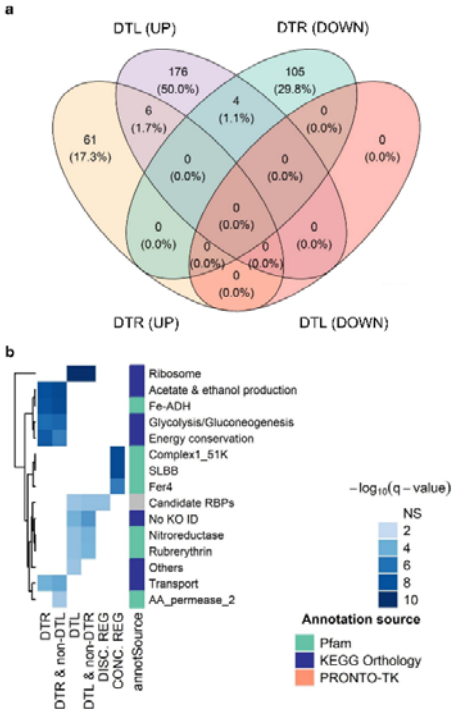

Fig. 3. Uncoupled transcriptional and translational control for autotrophy of *C. autoethanogenum*. **a** Venn diagram showing overlaps between up- or down-regulated differentially transcribed genes (DTR) and differentially translated genes (DTL) between syngas and fructose chemostats. **b** Functional classes enriched in DTRs, DTLs, in genes exclusively regulated at transcriptional (DTR & non-DTL) or translational (DTL & non-DTR) level, and in genes which were both DTRs and DTLs showing concordant or discordant changes by hypergeometric test (FDR = 5%). Pfam[61], Pfam entries from eggNOG-mapper 2.1.12[62]; KEGG Orthology, KEGG Orthology (KO) functional categories[60] from Valgepea *et al.* (Level 3 in Table S3[13]); NS, not significant. PRONTO-TK, computationally predicted RNA-binding proteins by PRONTO-TK[50].

Mapping genome-wide transcriptional architecture

Experimental mapping of transcriptional architecture is essential to accurately identify key RNA sequences – 5'UTR and 3'UTR – through which RBPs typically facilitate post-transcriptional regulation[44,45]. We thus applied Cappable-seq[46] to determine TISs and Term-seq[49] to profile T3Es at genome-scale (**Fig. 1**). Sequencing metrics and read statistics with mappings are in **Supplementary Table 5**. To accurately map TISs and T3Es as primary or internal at the level of transcriptional units[71], we organised *C. autoethanogenum* protein-coding genes into monocistronic operons with a single gene and polycistronic operons operationally defined as sets of protein-coding genes consecutively encoded on the same strand and separated by ≤60 nt. We identified 1,721 single genes while 2221 genes were organised into 781 polycistronic operons (**Supplementary Table 6**).

bioRxiv preprint doi: https://doi.org/10.64898/2025.12.18.695208; this version posted December 21, 2025. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

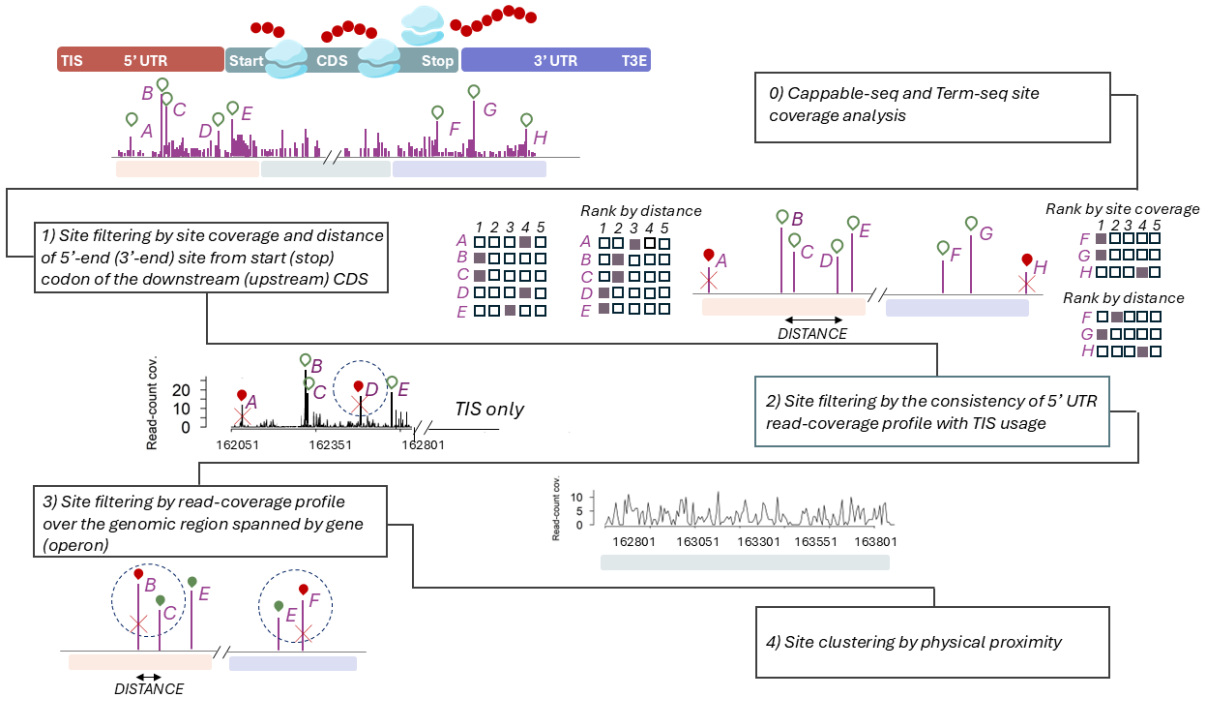

Fig. 4. Heuristic procedure used in this work for genome-wide identification of transcription initiation (TIS) and transcription 3' end (T3E) sites. The procedure consists of multiple filtering steps to ensure reliable identification of TISs and T3Es from preliminary sites extracted from Cappable-seq and Term-seq analysis (see Methods for details). Shortly: Filter 1) selects preliminary sites based on two criteria: site coverage in read counts and distance from the start (stop) codon of the downstream (upstream) coding DNA sequence (CDS). Sites are partitioned into five quantiles according to each criterion and quantiles for each TIS or T3E are combined. A TIS (T3E) passes to next step if the composite quantile accounting for both criteria is sufficiently high; Filter 2) is exclusively applied to 5'-untranslated regions (UTRs) and checks if it contains a stretch of sequence with a length of ≥50% of the total UTR length with a median read coverage ≥5 and ≥50% higher than for 100 nt upstream of the TIS; Filter 3) assesses the genomic region spanned by the gene (operon) for uniformity in read coverage; Filter 4) selects the TIS (T3E) closest to the start (stop) codon of the downstream (upstream) CDS for genes with multiple TISs or T3Es that are closer than 5 nt from each other.

We next employed a rigorous and heuristic multi-step procedure to determine TISs and T3Es for each operon (**Fig. 4**; see Methods for details). Shortly, we first extracted candidate sites upstream and downstream of an operon by removing sites with close-to-background read coverage and then used a series of established criteria – distance from CDS (filter #1 on **Fig. 4**), site coverage (#1), consistency of read coverage for TISs (#2), and uniformity of read coverage across the CDS (#3) – for filtering candidate sites (and clustering if needed; #4) to assign a single primary or multiple (primary and alternative) TISs and T3Es.

Analysis of Cappable-seq data identified 1,179 unique TISs corresponding to 1,588 genes (743 monocistronic and 300 polycistronic operons) that accounts for ~40% of protein-coding genes in *C. autoethanogenum* (**Supplementary Table 7**). From these, 1,003 were identified from syngas and 991 from fructose with 815 from both cultures. In total, around 88% TISs are primary and ~97% map to 5'UTRs (**Fig. 5a**) while ~90% of genes were assigned with a single TIS (**Fig. 5b**). Our results show a median 5'UTR length of 54 nt with 75% being <100 nt (**Fig. 5c; Supplementary Fig. 5a**) similar to values seen for *C. autoethanogenum* batch cultures using dRNA-seq[72]. We took advantage of the experimental TIS data for promoter motif identification using MEME[73] for primary TISs. This supported the quality of our TIS data as

bioRxiv preprint doi: https://doi.org/10.64898/2025.12.18.695208; this version posted December 21, 2025. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

we detected the well-described Pribnow box (TATAAT)[74] motif 10 nucleotides upstream of TISs for ~74% TISs and, at a lesser extent, the −35 motif (TTGTCA) for ~8% TISs (**Fig. 5d**). Notably, the TG dinucleotide in the Pribnow box[75,76] and the AT-rich region downstream of the −35 motif[77] stimulate transcription.

Analysis of Term-Seq data identified 751 unique TE3s, 620 from syngas and 621 from fructose with 490 from both cultures, for a total of 765 genes (**Supplementary Table 7**). Overall, ~70% T3Es are primary (**Fig. 5a**), ~71% of genes were assigned with a single T3E (~23% with two) (**Fig. 5b**), and the median 3' UTR length was 122 nt with 75% being <241 nt (**Fig. 5c; Supplementary Fig. 5b**). Interestingly, ~21% of T3Es mapped to CDSs (**Fig. 5a**), indicating potentially premature termination or processing[78], that was dependent on the growth substrate for 46 monocistronic and 15 polycistronic operons functional in transcription, chemotaxis, transport, amino acid metabolism, and glycolysis/gluconeogenesis (**Supplementary Table 8**). We note that, to the best of our knowledge, determination of T3Es has been performed only once before for acetogens[79].

We show the localization of TISs and T3Es for genes in key pathways for autotrophy on **Fig. 6**. More broadly, **Supplementary Table 9** contains coordinates and annotations of CDSs, TISs, and T3Es to allow intuitive exploration of *C. autoethanogenum* transcript boundaries by any visualisation tool, for instance with the open-source Integrative Genomics Viewer (IGV)[80].

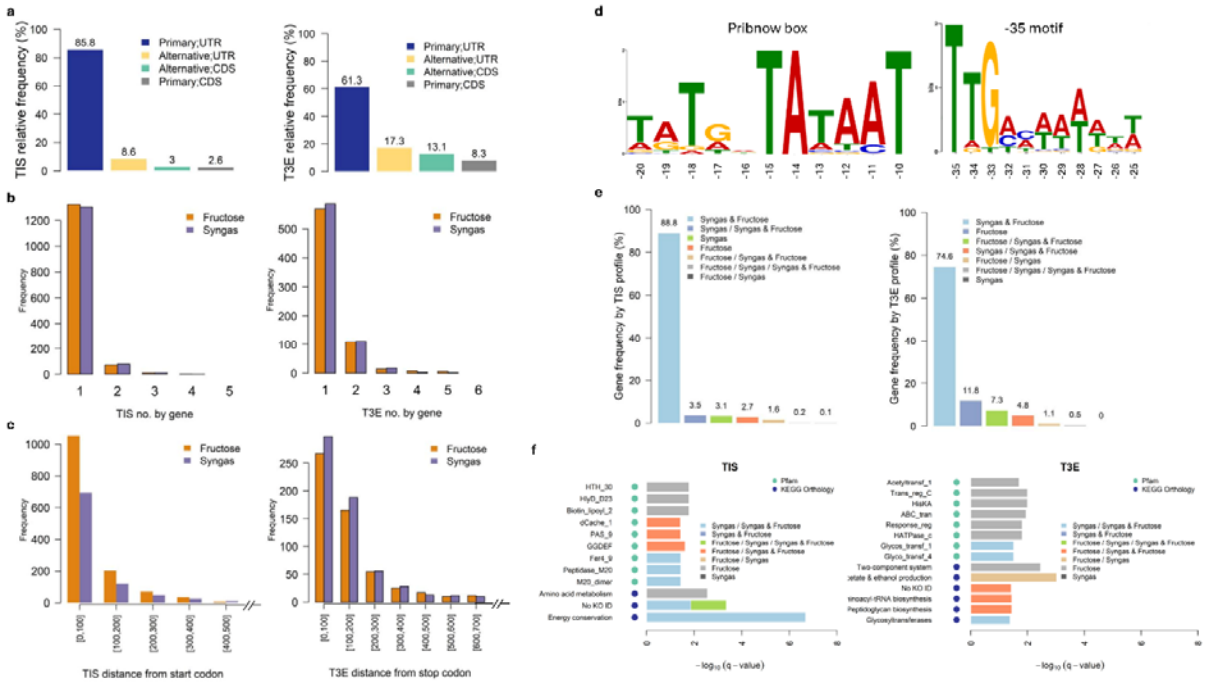

Fig. 5. Characteristics of transcriptional architecture in *C. autoethanogenum*. a Classification of transcription initiation (TIS) and transcription 3' end (T3E) sites according to strength (primary or alternative) and localization (UTR or coding DNA sequence, CDS). Primary site shows highest site coverage and shortest distance from CDS start (stop) codon while remaining sites passing the filtering procedure shown on Fig. 4 are identified as alternative sites. **b** Frequency of TISs and T3Es per gene for syngas and fructose chemostats. **c** Distribution of 5' and 3'UTR lengths for syngas and fructose chemostats (5'UTR < 500 nt; 3'UTR < 700 nt; see **Supplementary Fig. 5** for full distribution). **d** Promoter motifs identified in 5'UTRs by MEME[73]. **e** Frequency of gene-specific TIS or T3E profiles for syngas and fructose chemostats. Bars are aggregate of merged data from biological triplicate chemostat cultures. Syngas & Fructose, site(s) shared between two conditions; Syngas / Syngas & Fructose, site(s) unique to syngas and site(s) shared between two conditions; Syngas, site(s) unique to syngas; Fructose, site(s) unique to fructose; Fructose / Syngas & Fructose, site(s) unique to fructose and site(s) shared between two conditions; Fructose / Syngas / Syngas & Fructose, site(s) unique to syngas, site(s) unique to syngas, and site(s) shared between two conditions; Fructose / Syngas, site(s) unique to syngas

bioRxiv preprint doi: https://doi.org/10.64898/2025.12.18.695208; this version posted December 21, 2025. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

and site(s) unique to syngas. **f** Functional classes enriched in genes defined by TIS and T3E profile by hypergeometric test (FDR=5%). Pfam, Pfam[61] entries from eggNOG-mapper 2.1.12[62]; KEGG Orthology, KEGG Orthology (KO) functional categories[60] from Valgepea *et al.* (Level 3 in Table S3[13]).

Differential usage of TISs and T3Es for autotrophy and heterotrophy

To explore the extent of differential usage of TISs and T3Es between autotrophy and heterotrophy, we next classified genes according to the growth substrate specificity of their TISs and T3Es. The large majority of genes (~89%) feature a single TIS used for both syngas and fructose (**Fig. 5e**; light blue) while only a ~6% minority of genes use TISs exclusive to syngas (~3%; green) or fructose (~3%; orange), which includes a transcription elongation factor (07115) (**Supplementary Table 10**). The remaining set of genes show a composite profile of TIS usage with both TISs exclusive and common for either substrate. Similarly to TIS data, most genes (~75%) feature a single T3E for both syngas and fructose (**Fig. 5e;** light blue) (**Supplementary Table 11**). However, RNA processing at 3'UTRs appears to be more dependent on growth substrates as ~12% of genes show TISs exclusive to fructose (light purple) with another ~7% which show exclusive TISs for fructose along with shared TISs (green). Interestingly, a glyceraldehyde-3-phosphate dehydrogenase (08720) with an accompanying transcriptional regulator (08725) showed the most versatile T3E usage with both T3Es exclusive and common for either substrate (gray). Notably, we did not detect a single gene showing only T3Es exclusive for syngas.

Importantly, we detected enrichment of genes involved in energy conservation in the group of genes featuring syngas-specific TISs in addition to TISs shared for both substrates (**Fig. 5f**). This includes genes encoding the Rnf complex, a key activity for autotrophy in acetogens[63,64] (**Fig. 6**). Furthermore, genes with syngas-only TISs include several genes with Pfam domains involved in two-component signal transduction systems (TCS) as they encode transcriptional regulators (domain 'Trans_reg_C'), histidine kinases (HK; 'HisKA'), and response regulator (RR; 'Response_reg') proteins. Autophosphorylation of a HK by signal sensing drives the phosphorylation and activation of a RR that functions as a molecular switch to control diverse adaptive responses[81,82]. Notably, gene with ID 04855 harbours the ligand-binding domains GAF and PAS, which can bind carbon monoxide[83], and the HK-A domain, which is the phospho-acceptor domain of HKs. The two genes upstream – 04850 and 04855 – both contain the REC domain, which receives the signal from a HK. An additional TCS was enriched here with 16335 containing a REC domain and 16340 containing the HK-A domain. These syngas-only TISs could potentially be controlled by the σ^{54}-dependent Fis family transcriptional regulator (02225) since it also has a syngas-only TIS and a σ^{54} interaction domain occurs repeatedly in TCS genes[84,85]. These are potentially relevant observations to understand transcriptional regulation of autotrophy in acetogens as we previously isolated *C. autoethanogenum* strains with improved phenotypes from autotrophic adaptive laboratory evolution carrying mutations in HKs and RRs[35], which might be part of yet-to-be-described regulatory networks[86].

Only two genes displayed both a syngas-specific and a fructose-specific TIS: the radical SAM protein (11570) and a transcription elongation factor (07115), the latter being consistent with the identification of transcriptional regulators (02225, 12490, 13845, 02980, 18605, 18630) which are associated with substrate-specific TISs. Notably, genes featuring exclusively fructose-specific T3Es were enriched with TCSs, some of which include genes encoding HAMP-domain containing HKs (12620-12625, 03770-03775), which are putatively involved in the coordination of cell wall remodeling during cell division in Gram-positive bacteria[87,88]. For T3Es, the functional category 'acetate & ethanol production' was enriched in genes displaying TE3s specific to either syngas or fructose (**Fig. 5f**).

bioRxiv preprint doi: https://doi.org/10.64898/2025.12.18.695208; this version posted December 21, 2025. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Transcription-initiation signals for the Rnf complex are enhanced for autotrophy

Our analysis detected TISs and T3Es for genes in key metabolic pathways and activities for autotrophic growth of *C. autoethanogenum*: acetate and ethanol production, energy conservation, the WLP, and hydrogenases (**Fig. 6**). Comparison of the strength of transcription-initiation and termination signals between the two substrates could reveal regulation critical for autotrophy.

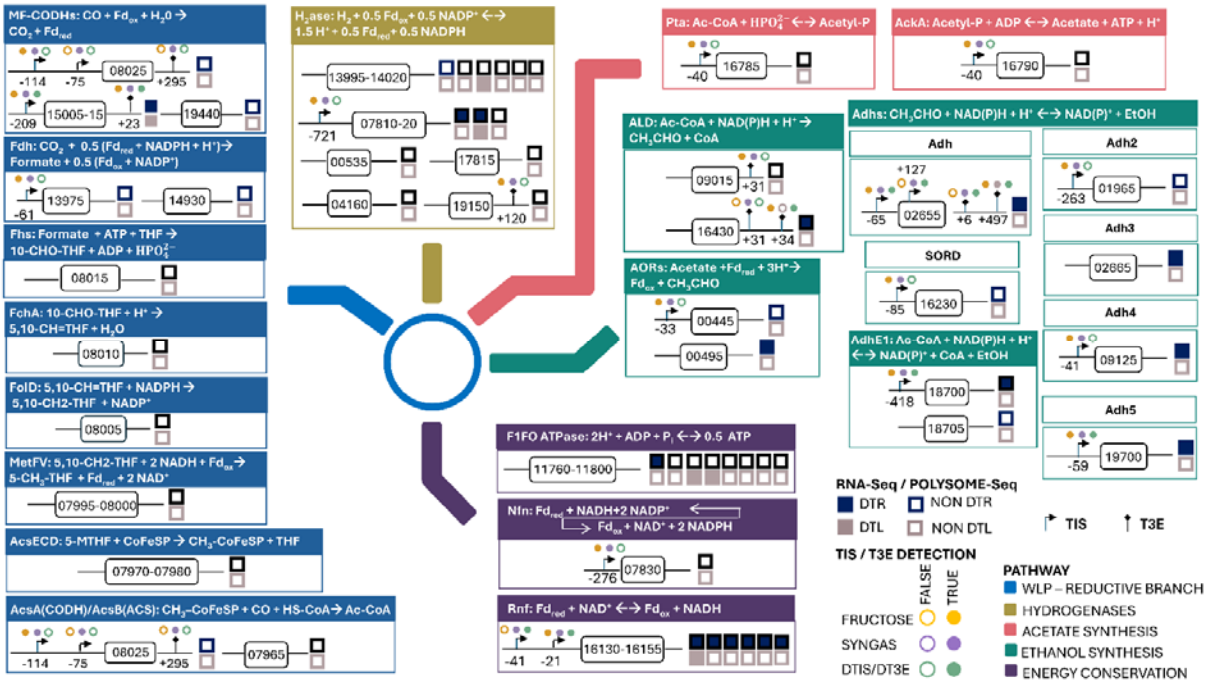

Fig. 6. Transcriptional architecture and transcriptional/translational regulation in key metabolic pathways for autotrophy of *C. autoethanogenum*. Boxed numbers are LAbrini gene IDs. Transcription initiation (TIS) or transcription 3' end (T3E) site distance shown from start codon of downstream coding DNA sequence (CDS) or stop codon of upstream CDS, respectively. DTIS, differentially expressed TIS; DT3E, differentially expressed T3E; MF-CODH, monofunctional carbon monoxide dehydrogenase; Fd, ferredoxin; 10-CHO-THF, 10-formyl tetrahydrofolate; 5,10-CH=THF, 10-methenyl-THF; 5,10-CH2-THF, 10-methylene-THF; 5-CH_3-TFH, 5-methyl-THF; Ac-CoA, acetyl-CoA; EtOH, ethanol. See **Supplementary Table 1** for gene descriptions.

Among the TISs identified in both syngas and fructose, we detected significant differences ($q < 0.05$) in the expression levels of 155 TISs linked to 144 genes between the two substrates (**Supplementary Table 7**). Strikingly, transcription-initiation signals for components of the Rnf complex, which plays a key role in energy conservation for autotrophy[64,89,90], were strongly enhanced on syngas vs fructose (**Fig. 6; Supplementary Fig. 6a**). Thus, the stronger initiation signals most likely lead to the enhanced expression of the Rnf complex CDS transcripts (**Fig. 3B; Supplementary Table 3**). Interestingly, genes with weaker initiation signals on syngas were enriched with CooS1 carbon monoxide dehydrogenase subunits (15005–15015), genes involved in ethanol production (02655, 16430, 18700, and 19700), genes belonging to glycolysis/gluconeogenesis, and numerous transcription factors (**Fig. 6; Supplementary Fig. 6a**). Importantly, overall, the change in transcription-initiation signal strength is consistent with differential expression of respective transcripts, as shown by the enrichment of genes with differential TIS signals

bioRxiv preprint doi: https://doi.org/10.64898/2025.12.18.695208; this version posted December 21, 2025. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

among the top or bottom ranks of transcript expression changes between the two substrates (**Supplementary Fig. 7**).

Within the T3Es detected in both syngas and fructose, we measured significant differences in the expression levels of 439 T3Es linked to 341 genes between the two substrates (**Supplementary Table 12**). Genes with enhanced termination signals for autotrophy included a gene cluster for cobalamin biosynthesis (05425–05445) while genes with repressed termination signals were enriched with ABC transporters, flagellar proteins, and genes involved in cell wall metabolism (**Supplementary Fig. 6b**). The relevance of these changes in termination signal strengths for autotrophy remain unclear.

RNA-binding proteins (RBPs) involved in transcriptional and translational regulation for autotrophy

Accurate mapping of transcriptional architecture defines UTR RNA sequences where RBPs could bind to execute post-transcriptional regulation. We could identify 5'UTRs for ~66% of DTRs and ~61% of DTLs, and 3'UTRs for ~28% of DTRs and ~10% of DTLs. To identify potential RBP binding regions in UTRs and thus RBP-RNA interactions, we first predicted potential RBPs for *C. autoethanogenum* using PRONTO-TK[50], a neural network-based protein function prediction tool (**Fig. 1**; **Supplementary Table 13**). PRONTO-TK uses ProtT5-XL protein language model embeddings to encode sequence-derived biochemical and structural features. We trained a binary classifier using the RNA binding GO term and its descendants, based on experimentally validated protein annotations from the UniProt knowledgebase[91]. We then used the computational tool catRAPID[51] to search for candidate RBP binding sites in the UTRs of DTRs and DTLs (**Fig. 1; Supplementary Table 14**), while excluding structural ribosomal proteins from the RBP list. Overall, 5'UTRs of ~11% of DTRs (20 genes) and ~15% of DTLs (27 genes) were predicted to contain RBP binding sites while the values for 3'UTRs were 12% for DTRs (21) and ~5% for DTLs (9) (**Fig. 7; Supplementary Table 14**). Most RBP target genes were regulated exclusively transcriptionally (34 out of which 13 up-regulated and 21 down-regulated on syngas vs fructose) or translationally (32 up-regulated) (**Fig. 7**). The remaining three RBP targets were a ferritin-encoding gene (00130) showing different regulation between transcription and translation, and a heat-shock protein (10470) and a ketol-acid reductoisomerase (00585) showing concordant regulation. Interestingly, 00585 is among the top 10 most abundant proteins in *C. autoethanogenum*[13]. Most RBPs recognised cis-regulatory elements in both 5' and 3'UTRs of target genes while individual genes were rarely bound by RBPs in both of their UTRs. Within the latter, a potentially relevant example for autotrophy is the alcohol dehydrogenase 02655. While genes involved in coenzyme synthesis for redox reactions, quorum sensing, and base excision repair were mediated by RBPs exclusively via 3'UTRs, translation-related genes (e.g. amino acid metabolism, tRNA processing) and transporters showed binding events in both UTRs.

The above-mentioned UTRs showed binding sites for 14 RBPs while none were found to bind uniquely either to 5'UTRs or 3'UTRs (**Fig. 7**). Furthermore, none of the 69 target genes were linked to regulation by a single RBP. Notably, three RBPs were among the translationally up-regulated DTLs on syngas: the polynucleotide phosphorylase PNPase (17015) involved in mRNA turnover[92] and RNA-mediated regulation[93]; a protein annotated as a single-stranded DNA-binding protein (10520); and the 4-thiouridine synthetase ThiI (04005), which stabilises tRNA folding by realizing post-transcriptional modifications of uracil at position 8[94] and is involved in the biosynthesis of sulfur-containing biomolecules such as thiamin[95].The remaining RBPs included proteins well-known for their roles in RNA metabolism: the S-adenosyl-L-methionine dependent RNA methyltransferase (MTase) (12395)[96]; the ribonucleases (RNases) Y (17055), III (16815), HII (16885), and M5 (11465) involved in RNA decay[97-99]; the transcription attenuation protein NusA (16970)[100,101]; the CCA tRNA nucleotidyltransferase (16090), which catalyzes the addition of the cytosine-cytosine-adenine (CCA) tail to the tRNA species that is indispensable for amino acid charging and tRNA functionality[102]; and two DEAD/DEAH box helicases (07100 and 15020) involved in RNA decay[103].

bioRxiv preprint doi: https://doi.org/10.64898/2025.12.18.695208; this version posted December 21, 2025. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

None of the RBPs were predicted to bind to genes involved in the following key activities for autotrophy of acetogens: C_1-fixation using the WLP, energy conservation through the Rnf complex, redox homeostasis through the Nfn transhydrogenase, or the various hydrogenases. At the same time, multiple RBPs showed UTR interactions with three alcohol dehydrogenases (08985, 08625, and 02655) (**Fig. 6**) that can be critical for maintenance of redox and energy homeostasis for *C. autoethanogenum* autotrophy through regulation of carbon flux between acetate and ethanol[14,104], the two major metabolic by-products (**Supplementary Fig. 1**). Expectedly, several RBPs targeted translation-related processes, including ribosomal structural proteins, tRNA synthetases, genes involved in amino acid metabolism and in biosynthesis of tRNA hypermodified nucleosides, namely queuosine and N^6-threonylcarbamoyladenosine, which are essential for modulating protein translation[105] (**Fig. 7**). Relevant to autotrophic biomass formation through gluconeogenesis was the binding of several RBPs to the 3'UTR of the type I glyceraldehyde-3-phosphate dehydrogenase (08720).

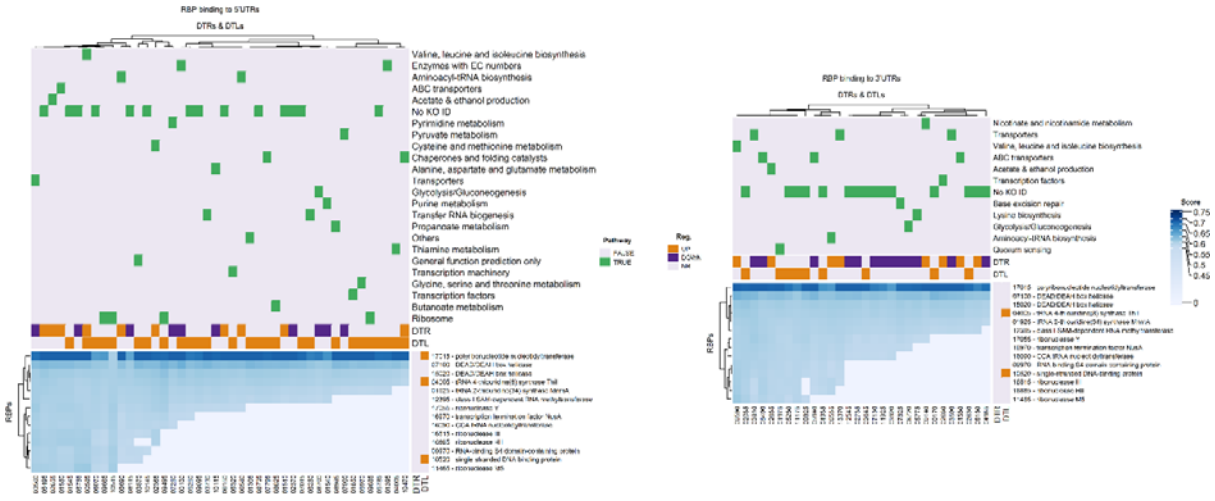

Fig. 7. RNA-binding proteins (RBPs) contributing to autotrophy of *C. autoethanogenum*. Predicted RBP interactions with 5' and 3'UTRs of differentially transcribed genes (DTRs) and differentially translated genes (DTLs) with RBPs in rows and target genes in columns. Blue heatmap shows catRAPID scores for RBP-RNA interactions. Row dendrogram on the left side and column dendrogram on the top of the heatmaps are from complete-linkage hierarchical clustering of the Euclidean distance matrix based on RBP-RNA interaction scores computed by catRAPID. KEGG Orthology functional categories[60] from Valgepea *et al.* (Level 3 in Table S3[13]) for target genes are shown on top. See **Supplementary Table 1** for target gene descriptions.

Discussion

While post-transcriptional regulation of gene expression plays an important role in acetogens[60], none of the key players in this regulatory layer – RBPs – had been described before. We thus also lacked knowledge about RBP-RNA interactions in acetogens that could facilitate global adjustments in metabolism, for instance towards autotrophy. To address these gaps with this work, we employed a comprehensive approach by combining chemostat cultures, functional genomics, and computational methods to identify genome-wide RBPs and RBP-RNA interactions contributing to autotrophy in the model-acetogen *C. autoethanogenum*.

We focused our approach on genes displaying transcriptional and/or translational regulation by their transcripts being differentially transcribed and/or translated during autotrophy vs heterotrophy. Widespread translational regulation could imply its dominant role for regulating protein levels for realising

bioRxiv preprint doi: https://doi.org/10.64898/2025.12.18.695208; this version posted December 21, 2025. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

autotrophy. We determined a low number of DTLs (186), which is consistent with the good correlation (R ~ 0.69 for 1,797 genes) between transcript levels quantified in syngas chemostats here and protein levels in LAbrini syngas chemostats in previous works[34,86] (**Supplementary Fig. 8**), and with the low number of proteins differentially expressed between CO, CO+H_2, and CO+CO_2+H_2 grown *C. autoethanogenum* chemostats[106]. Indeed, metabolic flux adjustments between the latter cultures are dominantly accompanied by changing enzyme catalytic rates rather than concentrations[13]. Such post-translational regulation of fluxes is also seen in self-oscillating *C. autoethanogenum* autotrophic cultures[12,14] and this is presumably the most efficient regulatory mode of fluxes for energy-limited metabolism since close to half of ATP for biomass proliferation is used for protein synthesis[107,108] . Furthermore, post-translational regulation would also allow cells to more rapidly react to gaseous substrates becoming available in natural environments by keeping a higher baseline expression of key proteins. At the same time, nearly half of transcripts showed statistically different expression (q-value < 0.05) here (**Fig. 2a**; **Supplementary Table 3**), consistent with data from *C. autoethanogenum* syngas and fructose batch cultures[55]. Thus, transcriptional regulation seems to be required to ensure the baseline proteome expression for both autotrophy and heterotrophy. Notably, transcriptional control of Rnf complex components and alcohol and aldehyde dehydrogenases detected in our work (**Fig. 6**) suggests attractive strain engineering targets in key metabolic activities of *C. autoethanogenum*.

Overall, ~19% of differentially regulated genes showed interactions with RBPs in our dataset (**Fig. 7**). The fact that we did not identify any RBPs binding only to genes with key activities for autotrophy (e.g. WLP, Rnf, Nfn, hydrogenases) implies that acetogens, at least *C. autoethanogenum*, do not seem to use an RBP-level mechanism for specifically regulating expression of genes exclusively needed for autotrophy. Although multiple RBPs targeted alcohol dehydrogenases, potentially critical for maintenance of redox and energy homeostasis[14,104], elucidation of key RBPs here, however, would be highly challenging due to the large number of candidate RBPs, their binding to other target genes with potentially essential functions, and the difficulties with creating such a number of *C. autoethanogenum* gene deletion mutants. Importantly, our study captured several RBP-RNA interactions well-known in bacteria that are needed for optimal growth in general. These interactions are therefore also involved in mediating differential gene transcription and translation profiles for realising autotrophy in *C. autoethanogenum* and will be thus discussed in more detail next.

Several differentially regulated genes in our dataset showed binding by the RBP NusA (**Fig. 7**; 16970), a factor for transcription attenuation where bacteria direct the RNA polymerase to either terminate transcription or transcribe downstream genes in an operon as a response to metabolic signals[100,101]. In contrast to translation-controlled termination of transcription in *E. coli*, the Gram-positive model microbe *Bacillus subtilis* (*C. autoethanogenum* is Gram-positive) lacks kinetic coupling between the transcription elongation factor and the pioneering ribosome[109]. Therefore, operon-specific RBPs[110-113], or trans-acting factors, such as NusA[78,114] are expected to support transcription attenuation. Our study shows that NusA is involved in realising autotrophy in *C. autoethanogenum*, as it is predicted to bind to genes involved in metabolic processes that are sensitive to transcriptional attenuation such as biosynthesis of thiamine, amino acids, tRNA synthetases, and ribosome biogenesis[115]. Transcription termination is often triggered by the arrangement of one of the two mutually exclusive stem structures in the nascent transcript, where the second stem is followed by a poly-U sequence forming a terminator and the second strand of the first stem base-pairs with the first strand of the second stem[116]. Indeed, we detected the presence of this characteristic arrangement in the 5'UTR of three genes targeted by NusA in our study using PASIFIC[117]: 06495, 01550, and 02365 encoding a tyrosine-tRNA ligase, a transmembrane protein, and a homoserine O-succinyltransferase, respectively. A recent study suggests that NusA-mediated stabilization of transcriptional

bioRxiv preprint doi: https://doi.org/10.64898/2025.12.18.695208; this version posted December 21, 2025. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

pausing and attenuation could be involved in the establishment of transcription-translation coupling through the interaction between NusA and translation initiation factor IF2[118]. This mechanism is unlikely in *C. autoethanogenum* as *B. subtilis* lacks coupling between transcription and translation[109]. Thus, further studies are required to elucidate the role of NusA in *C. autoethanogenum*.

Our results further show binding of an array of RNases to differentially regulated genes (**Fig. 7**). RNases are essential mediators in RNA maturation, quality control, and degradation where changes in tRNA and rRNA maturation impact translational regulation while changes in mRNA stability affect transcript levels. mRNA decay in Gram-positive bacteria relies on the cooperation between endoribonucleases, which cleave mRNAs and provide access to 5′ and 3′ extremities, and 5′→3′ and 3′→5′ exoribonucleases, which degrade the cleaved RNA fragments[26,119]. Our data for *C. autoethanogenum* is consistent with this model as we identified binding of both endoribonucleases (RNase Y, 17055; RNase III, 16815; and RNase M5, 11465) and exoribonucleases (PNPase, 17015 and RNase HII, 16885).

The most target genes were detected for RNase Y, the major decay-initiating endonuclease in *B. subtilis* and many Firmicutes[99,120]. Supporting proteins are proposed to be important for RNase Y target selection in the Gram-positive *Staphylococcus aureus*[121] since RNase Y appears to be strongly influenced by features distal to the position of cleavage, possibly facilitated by RNA secondary structure and RNA-protein interactions[122]. Our study suggests that this could be the case for *C. autoethanogenum* as well since RNase Y target genes were bound also by RNA helicases and RNA methyltransferases (**Fig. 7; Supplementary Table 14**). Interestingly, while RNase M5 in *B. subtilis* is thought to be almost exclusively active in 5S rRNA processing[123], we detected its binding to an alcohol dehydrogenase, multiple transporters, S6 and S10 ribosomal proteins, and genes in folate and thiamine metabolism (**Fig. 7; Supplementary Table 14**). Notably, ~88% of RNase M5 targets were also targeted by RNase III. While RNase III was initially assigned a secondary role in *B. subtilis*[124], it has now been shown to be important in *B. subtilis* for post-transcriptional catalytic and binding activities[98] and to be involved in translational regulation in *S. aureus*[125]. Thus, it is likely that all the three described endoribonucleases are relevant for autotrophy of *C. autoethanogenum*. For degradation of the cleaved RNA, *C. autoethanogenum* seems to primarily rely on the exoribonuclease PNPase as it targeted all the differentially regulated genes that the three endoribonucleases targeted (**Fig. 7**). What is more, PNPase was translationally up-regulated on syngas (**Fig. 7**). Use of the phosphorolytic enzyme PNPase would be beneficial for an energy-limited metabolism as the PNPase degradation products nucleoside diphosphates conserve energy[26]. Interestingly, PNPase can also play a role in complex cell adaptive responses through its recruitment in different multiprotein complexes[92]. RNA degradation could also involve the poorly characterised exoribonuclease RNase HII as it shared an unexpectedly high fraction of targets in common with RNase Y (89%), RNase M5 (67%), and RNase III (90%).

In terms of post-transcriptional RNA modifications for differentially regulated genes, the RBP pseudouridine synthase (09970) could be relevant (**Fig. 7**) as pseudouridine is a ubiquitous bacterial RNA modification, primarily on tRNA and rRNA, and secondarily on mRNA[126]. Whereas the modification plays a role in mammalian transcript stabilization[127] and modulation of transcript translatability[128,129], its potential effects on mRNA function in bacteria are largely unexplored[130]. The fact that 09970 targeted both DTRs and DTLs in our dataset suggests that pseudouridine can affect both transcription and translation in bacteria, similarly to mammals. Its scope seems wide for autotrophy as 09970 targets included an alcohol dehydrogenase, transporters, the threonine-tRNA ligase, and the 30S ribosomal protein S10, which is part of the processive rRNA transcription and antitermination complex.

We believe our work to be important for two reasons. Firstly, we provide the first systems-level steady-state datasets for an acetogen comparing heterotrophic and autotrophic growth. Importantly, this combined

bioRxiv preprint doi: https://doi.org/10.64898/2025.12.18.695208; this version posted December 21, 2025. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

description of transcriptome expression, polysome profiles, and transcriptional architecture advances fundamental understanding of gene function, gene expression regulation, and metabolism of the ancient metabolism of acetogens. Secondly, we report, for the first time for acetogens, genome-wide protein-RNA interactions that contribute to autotrophy in a model acetogen. This is important as it provides novel targets for metabolic engineering of acetogen cell factories from the unexplored space of RBPs. Future work will need to experimentally quantify the effects of genetic perturbations of candidate RBPs on acetogen autotrophy, including potentially synergistic or epistatic links between RBPs.

Methods

Bacterial strain and cultivation conditions

The *C. autoethanogenum* strain LAbrini[35] deposited in the German Collection of Microorganisms and Cell Cultures (DSM 115981) was used in this work.

C. autoethanogenum was grown in biological triplicate chemostat continuous cultures either autotrophically on syngas (50% CO, 20% H_2, 20% CO_2, and 10% Ar; AS Linde Gas) or heterotrophically on fructose (15 g/L) in a chemically defined medium (without yeast extract) as described before[104] with the following difference for fructose cultures: CO_2 sparged at 50 mL/min for 20h post-inoculation followed by sparging N_2 at 50 mL/min for the rest of the fermentation. Shortly, all cultures were grown under strictly anaerobic conditions at 37°C and at pH 5 in 1.4 L Multifors bioreactors (Infors AG) at a working volume of ~750 mL connected to a Hiden HPR-20-QIC mass spectrometer (Hiden Analytical) for online off-gas analysis. We used a syngas flow rate of 50 mL/min and agitation of 675 RPM for autotrophic and 15 g/L of fructose for heterotrophic cultures (agitation of 200 RPM) to achieve steady-state biomass concentrations of 1.37 ± 0.003 and 1.03 ± 0.008 gram of dry cell weight per litre for syngas and fructose cultures, respectively, for chemostats at dilution rates (D) 1.03 ± 0.05 and 0.997 ± 0.003 day^{-1} for syngas and fructose cultures, respectively. Steady-state results and samples were collected after culture optical density, gas uptake, and production rates had been stable for at least three working volumes.

Cappable-seq

C. autoethanogenum chemostat cultures were sampled for Cappable-seq by pelleting either 5 (syngas cultures) or 6.5 mL (fructose cultures) of culture broth (to match biomass amounts) by centrifugation (5,000 × *g* for 3 min at 4°C), discarding the supernatant, and flash-freezing in liquid N_2 followed by storage at −80°C until analysis. Total RNA from frozen samples was extracted by using the RNeasy Mini Kit (Qiagen) according to the manufacturer's instructions with in-column DNase treatment with the RNase-Free DNase Set (Qiagen). Total RNA samples were then converted to Cappable-seq libraries as described before[46]. Following enrichment and decapping, RNA samples were converted to Illumina-ready libraries using the SMARTer smRNA-seq kit (TakaraBio). Library qualities were assessed by the Agilent 2100 Bioanalyzer and finally libraries were sequenced using the Illumina Novaseq 6000 sequencer with 1 x 100 bp cycles (single-end).

Term-seq

The same cell pellets as for Cappable-seq were used for Term-seq (see above for sampling details). Ribosomal RNA was depleted from total RNA using the Illumina RIBO-Zero plus rRNA depletion kit, following manufacturer instructions. Ribosomal RNA depleted-RNA 3' ends were polyadenylated, followed

bioRxiv preprint doi: https://doi.org/10.64898/2025.12.18.695208; this version posted December 21, 2025. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

by RNA fragmentation, end-repair, and library preparation using the SMARTer smRNA-seq kit (TakaraBio). Library qualities were assessed by the Agilent 2100 Bioanalyzer and finally libraries were sequenced using the Illumina Novaseq 6000 sequencer with 2 x 100 bp cycles (paired-end).

Data analysis for genome-wide identification of TIS and T3E sites

Firstly, Term-seq reads (R1 or R2) were all trimmed of sequence elements added during library construction. Of each R1/R2 pair, trimmed-selected reads include either the R1 read, whenever the R1 contained the 3' adapter (i.e., when the read reaches the end of the mRNA fragment), or the inverse-complement of the R2 read, which always includes the end of the mRNA fragment. Whenever possible, R1 was preferred over R2 because R1 generally has higher sequence quality. Trimmed Cappable-seq reads (by the position of their 5'-end) and trimmed-selected Term-seq reads (by 3'-terminal position) were mapped to the genome of *C. autoethanogenum* strain LAbrini (GenBank CP110420.1) using STAR[131]. Read quality control was carried out by Fastq and sequencing metrics and read statistics with mappings are in **Supplementary Tables 34–36**.

Secondly, *C. autoethanogenum* coding DNA sequences (CDS) were organized into monocistronic operons with a single gene and polycistronic operons operationally defined as sets of protein-coding genes consecutively encoded on the same strand and separated by ≤60 nt. For the analysis of Cappable-seq read coverage relative to operons, we implemented the following heuristic procedure:

1. For each operon, we considered genome regions comprising the location of the gene operon and its 5' flanking (upstream) sequence up to the upstream gene operon on the same strand. Based on the genome annotation, we characterized each position within this region as hypothetical '5'-UTR', coding region ('CDS'), or intergenic region ('IG') for polycistronic operons.
2. From the hypothetical '5'UTR' region, we identified the position i with maximum Cappable-seq 5'-end coverage $C(Fructose)_i^M$ summing coverages of the three fructose bio-replicates with maximum coverage $C(Syngas)_i^M$ summing over the three syngas bio-replicates.
3. We identified all positions j within '5'UTR', 'CDS', or 'IG' regions with coverages $C(Fructose)_j$ or $C(Syngas)_j$ such that $C(Fructose)_j \geq T \cdot C(Fructose)_i^M$ and $C(Syngas)_j \geq T \cdot C(Syngas)_i^M$, where $T = 0.1$ is a fractional threshold of the respective maximum coverages identified in the 5'UTR of the operon. We considered all sites identified in the '5'UTR' from fructose or syngas bio-replicates as potential TIS candidates, and all sites identified in 'CDS' and 'IG' regions as representing background coverages.
4. We characterized each site identified within the '5'UTR' region by its ranking relative to the coverages of those selected from 'CDS'+'IG' regions for each operon ('*p*-value'). Thus, if 100 sites were identified as background in 'CDS'+'IG' regions, a '5'UTR' site that had coverage higher than any of those coverages was assigned a *p*-value = 0.0. If, e.g., 10 sites from the 'CDS'+'IG' region had higher coverage, it was assigned a *p*-value = 0.1 (10/100).
5. Among all sites of the '5'UTR' region, those with *p*-value ≤ 0.05 and with average coverage across syngas or fructose bio-replicates > 5 reads were selected.
6. To identify potential TIS candidates within 'CDS' or 'IG' regions for an operon, we selected all sites with coverages ≥50% of the maximum signal in the '5'UTR' region and with average coverage across syngas or fructose bio-replicates > 20 reads.
7. All sites identified in adjacent sequence positions were collapsed into one site of 'size' equal to the number of positions collapsed.

bioRxiv preprint doi: https://doi.org/10.64898/2025.12.18.695208; this version posted December 21, 2025. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

An analogous approach was used to analyze Term-seq read coverage relative to operons, but here we considered for each operon, a genome region comprising the location of the operon and its 3' flanking (downstream) region up to the downstream operon on the same strand. We characterized each position within this region as hypothetical '3'UTR' region, coding region ('CDS') or intergenic region ('IG') for polycistronic operons and then followed the same procedure as Cappable-seq analysis, considering the '3'UTR' region in place of the '5'UTR' region.

Lastly, we used the following multi-step filtering procedure to determine TISs and T3Es for each operon from the candidate sites identified above (see also **Fig. 4**.).

1. *Site filtering by site coverage and distance from start (for TIS) or stop (for T3E) codon from CDS*. Here, candidate TISs and T3Es were filtered based on their distance from the start (for TIS) or stop (for T3E) codon of the downstream (for TIS) or upstream (for T3E) CDS $Dist(site, CDS\ start/stop\ codon)$ and based on their coverage in read counts $SiteCov$. We arranged the sites in descending order based on $SiteCov$ values and into ascending order based on $Dist(site, CDS\ start/stop\ codon)$ and then separately partitioned into five quantiles according to their position and signal intensity. As a result, a hypothetical site i received the quantile $Q(Dist)_i$ corresponding to its distance relative to the start (stop) codon of the downstream (upstream) coding sequence and the quantile corresponding to $\log_2$ site coverage normalized by the average coverage of all sites from all fructose samples $Q(SiteCov)_i^{fructose}$ or from all syngas samples $Q(SiteCov)_i^{syngas}$. Lastly, the $Q(SiteCov)_i^{growth}$ (growth is either fructose or syngas) and the $Q(Dist)_i$ quantiles of the site i were summed up to define the site's composite quantile CQ_i. For each gene (operon), we selected the sites displaying the lowest composite quantile that corresponds to sites which tend to display higher coverage and shorter distances from the respective start or stop codon. If the composite quantile of the site was lower than the central composite quantile, the candidate site passed to the next filtering step. We note that this procedure allows us to retain multiple and different sites per gene (operon) per substrate.
2. *Site filtering by the consistency of 5'UTR read-coverage profile with TIS usage*. Here, candidate 5'UTRs are assessed for consistency with TIS usage as 5'UTRs are expected to display non-uniform read coverage across the TISs. Thus, TISs that contained a stretch of sequence with a length of ≥50% of the total UTR length with a median read coverage ≥5 and ≥50% higher than that for 100 nt upstream of TIS passed to the next filtering steps.
3. *Site filtering by read-coverage profile over the genomic region spanned by gene (operon)*. Here, we assume read coverage over the gene (operon) to be uniform. Thus, a site passed to the next filtering steps if it is linked to a CDS with a median value of the median read-coverage values computed over 200 nt-long sliding windows being <10% different from the median read-coverage over the full CDS (except if an internal TIS (T3E) is present).
4. *Site clustering by physical proximity*. Here, we selected a representative site if multiple close sites were linked to a gene (operon). Thus, from sites closer than 5 nt from each other, the TIS (T3E) closest to the start (stop) codon of the downstream (upstream) CDS was selected as the primary site. All other sites were considered as alternative sites.

Differential expression analysis of TIS and T3E signals

We quantified differences in TIS and T3E signals (i.e. site coverage) between syngas and fructose using EdgeR with the quasi-likelihood F-test procedure[132]. To determine sites with statistically significant

bioRxiv preprint doi: https://doi.org/10.64898/2025.12.18.695208; this version posted December 21, 2025. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

differences (q-value < 0.05), we applied the Benjamini–Hochberg procedure to the F-test results, controlling the false discovery rate (FDR) correction at 5%[133].

Polysome-seq

Sample preparation and sequencing

C. autoethanogenum chemostat cultures were sampled, samples processed, and polysome profiling was performed as previously described[59]. Briefly, translation was arrested by adding chloramphenicol to bioreactors to a final concentration of 0.1 mg/mL. Next, 80 (syngas cultures) or 100 mL (fructose cultures) of culture broth (to match biomass amounts) in four replicate samples were harvested, washed twice, and resuspended in lysis buffer (20 mM Tris HCl pH 8, 140 mM KCl, 40 mM $MgCl_2$, 0.5 mM DTT, 100 µg/mL chloramphenicol, 1 mg/mL heparin, 20 mM EGTA, 1% Triton X-100). After mechanical cell disruption with glass beads, supernatant was flash-frozen in liquid N_2 followed by storage at −80°C until analysis. Then, mRNA-ribosome complexes were size-separated on a sucrose gradient (10 to 50% w/v) in polysome gradient buffer (same composition as lysis buffer except for heparin at a final concentration of 0.5 mg/mL) into 24 subfractions. After protein denaturation and nucleic acid precipitation performed as previously described[59], total RNA was extracted in subfractions using RNeasy midi kits (Qiagen) combined with on-column DNase I treatment. RNA amount was quantified using a Nanodrop spectrophotometer (Thermo Fisher Scientific). External ERCC RNA spikes (Invitrogen) were added to 1 µg of each total RNA sample according to manufacturer instructions. The levels of 16S and 23S rRNAs in subfractions were calculated using the Agilent 2100 Bioanalyzer 2100 and used to pool the subfractions into four fractions labelled A to D. Fraction A (from 1 to 4.5 min) consisted of subfractions containing DNA, free RNAs, and free small and large ribosomal subunits. The first peak (Fraction B from 4.5 to 6 min) was attributed to the monosome, Fraction C (from 6 to 8 min) to polysomes with low ribosome density and Fraction D (from 8 to 12 min) to polysomes with high ribosome density. The average number of ribosomes in Fractions B, C, and D was extrapolated knowing the elution time of the monosome[59] and were found to be similar for syngas and fructose cultures (1.1 ± 0.1 ribosomes in Fraction B, 2.9 ± 0.2 ribosomes in Fraction C, and 7.7 ± 0.5 ribosomes in Fraction D). Total RNA in Fractions A, B, C, and D was ribo-depleted using the RIBO-Zero kit (Illumina). After fragmentation, the first strand cDNA was synthesized using random hexamer primers. During the second strand cDNA synthesis, dUTPs were replaced with dTTPs in the reaction buffer. The directional library was ready after end repair, A-tailing, adapter ligation, size selection, USER enzyme digestion, amplification, and purification. Finally, libraries were sequenced using the Illumina Novaseq X Plus Series platform with 2 x 150 bp cycles (paired-end).

Quantification of transcript abundances

Firstly, reads with >10% of undetermined bases (N) and reads with >50% of bases with Qscore ≤ 5 were excluded. Secondly, the ERCC RNA spikes were used to determine the linearity range between read counts and RNA concentration and thus features among the 4,043 LAbrini genes and 92 ERCC RNAs with <10 read counts were excluded for reliable quantification (259 and 264 features for fructose and syngas, respectively). Reads trimmed of Illumina adapters were then mapped to the genome of *C. autoethanogenum* strain LAbrini (GenBank CP110420.1) and ERCC RNA spikes using Bowtie2[134] and read counting was performed using the featureCounts function within Rsubread[135]. To account for variations in sample preparation and analysis from ribodepletion to sequencing, two normalization steps were used. Firstly, we applied the RUV method to remove unwanted variation[136] using the ERCC RNA spikes introduced in each fraction before the ribodepletion step: we obtained $N_{i,j,k\ NormRUVg}$, with the count of the transcript i, in the fraction j, and bio-replicate k after normalizing with the RUV method, with $i \in \{1,\ldots,3876\}$, $j \in$

bioRxiv preprint doi: https://doi.org/10.64898/2025.12.18.695208; this version posted December 21, 2025. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

{A,B,C,D} and k ∈ {1,2,3} on fructose and with i ∈ {1,…,3871}, j ∈ {A,B,C,D} and k ∈ {1,2,3} on syngas. Secondly, we accounted for the variability in the initial amount of total RNA between fractions as a constant 0.5 µg of total RNA was taken in each fraction for ribodepletion: we obtained the final normalized count of the transcript i, in the fraction j, and bio-replicate k as $N_{i,j,k} = N_{i,j,k\ NormRUVg}$ *x total RNA quantity*$_{j,k}$ */ 0.5*. Finally, for each transcript i, we calculated the proportion of RNA copies in each fraction j for each bio-replicate k:

$$RNA\ proportion\ in\ fraction_{i,j,k}\ (\%) = N_{i,j,k}\ x\ 100\ /\ \sum_{j=A}^{D} N_{i,j,k}$$

Consistency analysis of bio-replicates

Due to the lower-than-expected correlation between fractional mRNA counts of some bio-replicates (**Supplementary Fig. 3**), we assessed data consistency between bio-replicates within both chemostat datasets by three complementary methods suited for compositional data, such as polysome-seq: correlation analysis, compositional distance analysis, and principal component analysis (PCA). Firstly, we computed the Spearman's rank correlation coefficient (ρ) between mRNA counts per polysomal gradient fractions from total counts between bio-replicates. A bio-replicate was determined as an outlier if it showed the highest number of poorly correlated polysomal fractions; a fraction was considered poorly correlated if its data correlated poorly ($\rho < 0.7$) with the other bio-replicates which in turn correlated between each other with $\rho > 0.7$. Secondly, we calculated the average of all mRNA counts for each fraction and bio-replicate, then centered-ln-ratio transformed the averages, and lastly computed the Aitchison's distance between each pair of bio-replicates. A bio-replicate was determined as an outlier if it showed the maximal average Aitchison's distance[137] from the other bio-replicates that was greater than the median distance observed across syngas and fructose. Thirdly, we collated all mRNA counts for each fraction and bio-replicate as a concatenated dataset, then centered-ln-ratio transformed gene compositional data and lastly performed PCA to compute the average distances between the centroids of each pair of bio-replicates. A bio-replicate was determined as an outlier if its centroid showed the maximal average distance from the centroids of the other bio-replicates that was greater than the median distance observed across syngas and fructose. Finally, we assembled the information gathered by the three approaches, and we determined as an outlier the bio-replicate which was found to be outlier according to the highest number of approaches.

Determination of DTLs

We used three complementary metrics to quantify translational regulation and finally determined DTLs using a consensus approach (CONSENSUS) over the three metrics.

First, for the TL-STATUS metric, we firstly calculated for each gene its translation status in both syngas and fructose as the ratio between the sum of the mRNA proportions quantified in the ribosome-bound and ribosome-free fractions. Secondly, subtracting the fructose value from syngas, we quantified the absolute difference of the translation status per gene between syngas and fructose (TL-STATUS). Genes were then ranked by decreasing order of the calculated TL-STATUS values for evaluation with the CONSENSUS approach (see below).

Second, as the TL-PROFILE metric, we calculated for each gene the difference in its polysome profiles between syngas and fructose as the variance-weighted Aitchison's distance of the centered-ln-ratio transformed compositional data by weighing the difference between the individual fractions by pooled variance across syngas and fructose bio-replicates. Genes were then ranked by decreasing order of the calculated TL-PROFILE values for evaluation with the CONSENSUS approach.

bioRxiv preprint doi: https://doi.org/10.64898/2025.12.18.695208; this version posted December 21, 2025. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Third, as the POLY-FRACTIONS metric, we estimated 95% confidence intervals (CI) for each polysome fraction and gene by bootstrapping from a resampling distribution and then identified fractions with differential mRNA proportions if their computed CI did not cross the null value. This metric quantified both the number of statistically different fractions per gene and the maximal difference across fractions between conditions. Genes were then ranked by decreasing order of both features for evaluation with the CONSENSUS approach.

Lastly, to account for the different nature of translational regulation captured by each metric, we used a CONSENSUS approach that ranked genes based on the average of the ranks obtained across the three metrics. We determined DTLs as genes within the top 5% of the CONSENSUS ranking.

RNA-seq

Sample preparation and sequencing

The same cell pellets as for polysome-seq were used for RNA-seq (see above for sampling details). Total RNA from frozen samples was extracted using the RNeasy Midi Kit (Qiagen) combined with on-column DNase I treatment. RNA amount was quantified using the Nanodrop spectrophotometer and RNA integrity validated using the Agilent Bioanalyzer 2100. External ERCC RNA spikes were added to 1 µg of each total RNA sample according to manufacturer instructions. Total RNA was ribo-depleted using the RIBO-Zero kit and size distribution evaluated with the Agilent Bioanalyzer 2100. After fragmentation, the first strand cDNA was synthesized using random hexamer primers. During the second strand cDNA synthesis, dUTPs were replaced with dTTPs in the reaction buffer. The directional library was ready after end repair, A-tailing, adapter ligation, size selection, USER enzyme digestion, amplification, and purification. Finally, libraries were sequenced using the Illumina Novaseq X Plus Series platform with 2 x 150 bp cycles (paired-end).

Quantification of transcript abundances and differential expression analysis

The ERCC RNA spikes were used to determine the linearity range between read counts and RNA concentration and thus transcripts with <10 read counts were excluded for reliable quantification. Additionally, reads with >10% of undetermined bases (N) and reads with >50% of bases with Qscore ≤ 5 were excluded. Reads trimmed of Illumina adapters were then mapped to the genome of *C. autoethanogenum* strain LAbrini (GenBank CP110420.1) and ERCC RNA spikes using Bowtie2[134] and read counting was performed using the featureCounts function within Rsubread[135]. Raw library sizes were normalized across the samples using the *normLibSizes* function within edgeR[138]. Differential expression analysis between syngas and fructose cultures was performed within edgeR using the quasi-likelihood F-test as per the *glmQLFTest* function[139]. To determine genes with statistically different expression (q-value < 0.05), we applied the Benjamini–Hochberg procedure to the F-test results controlling the false discovery rate (FDR) at 5%[133].

Identification of RNA-binding proteins

Potential RNA-binding proteins in *C. autoethanogenum* were identified using PRONTO-TK (10.1093/nargab/lqae112), a neural network-based framework for protein function prediction. Proteins predicted with posterior probabilities above the confidence threshold of 0.75 were retained as candidate RBPs for downstream analyses of RBP–RNA interactions.

bioRxiv preprint doi: https://doi.org/10.64898/2025.12.18.695208; this version posted December 21, 2025. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Prediction of protein-RNA interactions

We employed catRAPID omics v2[51] to predict protein–RNA interactions between the candidate RNA-binding proteins and the genes which turned out to be differently regulated in transcription and translation between growth conditions. The rating system of catRAPID predictions integrates multiple scores to provide final scores ranging in the interval $[0,1]$. Protein-RNA interactions predicted by catRAPID were retained if the corresponding Z-scores pertaining to the interaction propensity and the final scores were ranked in the top 25% positions.

Data availability

All raw sequencing datasets have been deposited at the European Nucleotide Archive under project accession code PRJEB102336. Source data is provided in this paper.

References

1. Fackler, N., et al. Stepping on the Gas to a Circular Economy: Accelerating Development of Carbon-Negative Chemical Production from Gas Fermentation. *Annu. Rev. Chem. Biomol. Engineer.* **12**, 439–470 (2021).
2. Köpke, M., & Simpson, S. D. Pollution to products: recycling of 'above ground' carbon by gas fermentation. *Curr. Opin. Biotechnol.* **65**, 180–189 (2020).
3. Pavan, M., et al. Advances in systems metabolic engineering of autotrophic carbon oxide-fixing biocatalysts towards a circular economy. *Metab. Eng.* **71**, 117–141 (2022).
4. Poehlein, A., et al. Advanced aspects of acetogens. *Bioresour. Technol.* **427**, 131913 (2025).
5. Wood H. G. Life with CO or CO_2 and H_2 as a source of carbon and energy. *Faseb J.* **5**, 156–163 (1991).
6. Drake, H.L., Küsel, K., Matthies, C. Acetogenic Prokaryotes. in: The Prokaryotes (eds. Dworkin, M., Falkow, S., Rosenberg, E., Schleifer, K.H., & Stackebrandt, E.) 354–420 (Springer, New York, 2006).
7. Cotton, C. A., Edlich-Muth, C., & Bar-Even, A. Reinforcing carbon fixation: CO_2 reduction replacing and supporting carboxylation. *Curr. Opin. Biotech.* **49**, 49–56 (2018).
8. Ragsdale, S. W., & Pierce, E. Acetogenesis and the Wood-Ljungdahl pathway of CO(2) fixation. *Biochim. Biophys. Acta* **1784**, 1873–1898 (2008).
9. Heffernan, J. K., Mahamkali, V., Valgepea, K., Marcellin, E., & Nielsen, L. K. Analytical tools for unravelling the metabolism of gas-fermenting Clostridia. *Curr. Opin. Biotechnol.* **75**, 102700 (2022).
10. Bae, J., Park, C., Jung, H., Jin, S., & Cho, B. K. Harnessing acetogenic bacteria for one-carbon valorization toward sustainable chemical production. *RSC Chem. Biol.* **5**, 812–832 (2024).
11. Jin, S., et al. Synthetic Biology on Acetogenic Bacteria for Highly Efficient Conversion of C_1 Gases to Biochemicals. *Int. J. Mol. Sci.* **21**, 7639 (2020).
12. Greene, J., Daniell, J., Köpke, M., Broadbelt, L., Tyo, K. E. J. Kinetic ensemble model of gas fermenting Clostridium autoethanogenum for improved ethanol production. *Biochem. Eng. J.* **148**, 46-56 (2019).
13. Valgepea, K., et al. Absolute Proteome Quantification in the Gas-Fermenting Acetogen *Clostridium autoethanogenum*. *mSystems* **7**, e0002622 (2022).
14. Mahamkali, V., et al. Redox controls metabolic robustness in the gas-fermenting acetogen *Clostridium autoethanogenum*. *Proc. Natl. Acad. Sci.* U. S. A. **117**, 13168–13175 (2020).
15. Song, Y., et al. Genome-scale analysis of syngas fermenting acetogenic bacteria reveals the translational regulation for its autotrophic growth. *BMC Genomics* **19**, 837 (2018).

bioRxiv preprint doi: https://doi.org/10.64898/2025.12.18.695208; this version posted December 21, 2025. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

16. Song, Y., et al. Transcriptome and translatome of CO_2 fixing acetogens under heterotrophic and autotrophic conditions. *Sci. Data* **8**, 51 (2021).
17. Al-Bassam, et al. Optimization of carbon and energy utilization through differential translational efficiency. *Nat. Commun.* **9**, 4474 (2018).
18. Shin, J., et al. Genome-Scale Analysis of *Acetobacterium woodii* Identifies Translational Regulation of Acetogenesis. *mSystems* **6**, e0069621 (2021).
19. Re A. Translational efficiency in gas-fermenting bacteria: Adding a new layer of regulation to gene expression in acetogens. *iScience* **26**, 108383 (2023).
20. Ng Kwan Lim, E., Sasseville, C., Carrier, M. C., & Massé, E. Keeping Up with RNA-Based Regulation in Bacteria: New Roles for RNA Binding Proteins. *Trends Genet*. **37**, 86–97 (2021).
21. Roberts J. W. (1969). Termination factor for RNA synthesis. *Nature* **224**, 1168–1174 (1969).
22. Figueroa-Bossi, N., Schwartz, A., Guillemardet, B., D'Heygère, F., Bossi, L., & Boudvillain, M. RNA remodeling by bacterial global regulator CsrA promotes Rho-dependent transcription termination. *Genes Dev.* **28**, 1239–1251 (2014).
23. Holmqvist, E., & Vogel, J. (2018). RNA-binding proteins in bacteria. *Nat. Rev. Microbiol*. **16**, 601–615 (2018).
24. Mondal, S., Yakhnin, A. V., Sebastian, A., Albert, I., & Babitzke, P. NusA-dependent transcription termination prevents misregulation of global gene expression. *Nat. Microbiol* **1**, 15007 (2016).
25. Kumarevel, T., Mizuno, H., & Kumar, P. K. Structural basis of HutP-mediated anti-termination and roles of the Mg2+ ion and L-histidine ligand. *Nature* **434**, 183–191 (2005).
26. Bechhofer, D. H., & Deutscher, M. P. Bacterial ribonucleases and their roles in RNA metabolism. *Crit. Rev. Biochem. Mol. Biol*. **54**, 242–300 (2019).
27. Ellis, M. J., Trussler, R. S., & Haniford, D. B. Hfq binds directly to the ribosome-binding site of IS10 transposase mRNA to inhibit translation. *Mol. Microbiol.* **96**, 633–650 (2015).
28. Rüttiger, A. S., et al. The global RNA-binding protein RbpB is a regulator of polysaccharide utilization in *Bacteroides thetaiotaomicron*. *Nature Commun*. **16**, 208 (2025).
29. Ignatov, D., et al. RNA-binding protein YebC enhances translation of proline-rich amino acid stretches in bacteria. *Nature Commun.* **16**, 6262 (2025).
30. McQuail, J., et al. Condensation of the RNA chaperone Hfq is coupled to inhibition of glucose uptake and contributes to the stabilization of regulatory RNAs in nitrogen-starved Escherichia coli. *Nucleic Acids Res.* **53**, gkaf1006 (2025).
31. Papenfort, K., & Melamed, S. Small RNAs, Large Networks: Posttranscriptional Regulons in Gram-Negative Bacteria. *Annu Rev Microbiol*. **77**, 23–43 (2023).
32. Piatek, P., et al. Agr Quorum Sensing influences the Wood-Ljungdahl pathway in *Clostridium autoethanogenum*. *Sci Rep* 12, **411** (2022).
33. de Lima, L. A., et al. Faster Growth Enhances Low Carbon Fuel and Chemical Production Through Gas Fermentation. *Front. Bioeng. Biotechnol.* **10**, 879578 (2022).
34. Nwaokorie, U. J., et al. (2023). Deletion of genes linked to the C_1-fixing gene cluster affects growth, by-products, and proteome of Clostridium autoethanogenum. Front. *Bioeng. Biotechnol.* **11**, 1167892 (2023).
35. Ingelman, H., et al. Autotrophic adaptive laboratory evolution of the acetogen *Clostridium autoethanogenum* delivers the gas-fermenting strain LAbrini with superior growth, products, and robustness. *New Biotech*. **83**, 1–15 (2024).
36. Heffernan, J., et al. Adaptive laboratory evolution of *Clostridium autoethanogenum* to metabolize CO_2 and H_2 enhances growth rates in chemostat and unravels proteome and metabolome alterations. *Microb. Biotechnol*. **17**, e14452 (2024).

bioRxiv preprint doi: https://doi.org/10.64898/2025.12.18.695208; this version posted December 21, 2025. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

37. Davin, M. E., et al. Clostridium autoethanogenum alters cofactor synthesis, redox metabolism, and lysine-acetylation in response to elevated H_2:CO feedstock ratios for enhancing carbon capture efficiency. *Biotechnol. Biofuels Bioprod.* **17**, 119 (2024).
38. Liew, F. E., et al. Carbon-negative production of acetone and isopropanol by gas fermentation at industrial pilot scale. *Nat. Biotechnol.* **40**, 335–344 (2022).
39. Dykstra, J. C., et al. Metabolic engineering of *Clostridium autoethanogenum* for ethyl acetate production from CO. *Microb. Cell. Fact.* **21**, 243 (2022).
40. Hentze, M. W., Castello, A., Schwarzl, T., & Preiss, T. A brave new world of RNA-binding proteins. *Nat. Rev. Mol. Cell Biol.* **19**, 327–341 (2018).
41. Van Nostrand, E. L., et al. A large-scale binding and functional map of human RNA-binding proteins. *Nature* **583**, 711–719 (2020).
42. Stenum, T. S., et al. RNA interactome capture in *Escherichia coli* globally identifies RNA-binding proteins. *Nucleic Acids Res.* **51**, 4572–4587 (2023).
43. Arava, Y., Wang, Y., Storey, J. D., Liu, C. L., Brown, P. O., & Herschlag, D. Genome-wide analysis of mRNA translation profiles in *Saccharomyces cerevisiae*. *Proc. Natl. Acad. Sci.* U. S. A. **100**, 3889–3894.
44. Liu, Y. J., Wang, X., Sun, Y., & Feng, Y. Bacterial 5' UTR: A treasure-trove for post-transcriptional regulation. *Biotechnol. Adv.* **78**, 108478 (2025).
45. Menendez-Gil, P., & Toledo-Arana, A. Bacterial 3'UTRs: A Useful Resource in Post-transcriptional Regulation. *Front. Mol. Biosci.* **7**, 617633 (2021).
46. Ettwiller, L., Buswell, J., Yigit, E., & Schildkraut, I. A novel enrichment strategy reveals unprecedented number of novel transcription start sites at single base resolution in a model prokaryote and the gut microbiome. *BMC Genomics* **17**, 199 (2016).
47. Boutard, M., et al. Global repositioning of transcription start sites in a plant-fermenting bacterium. *Nature Commun.* **7**, 13783 (2016).
48. Sharma, C. M., et al. The primary transcriptome of the major human pathogen *Helicobacter pylori*. *Nature* **464**, 250–255 (2010).
49. Dar, D., et al. Term-seq reveals abundant ribo-regulation of antibiotics resistance in bacteria. *Science* **352**, aad9822 (2016).
50. Politano, G., Benso, A., Rehman, H. U., & Re, A. PRONTO-TK: a user-friendly PROtein Neural neTwOrk tool-kit for accessible protein function prediction. *NAR Genom. Bioinform.* **6**, lqae112 (2024).
51. Armaos, A., Colantoni, A., Proietti, G., Rupert, J., & Tartaglia, G. G. catRAPID omics v2.0: going deeper and wider in the prediction of protein-RNA interactions. *Nucleic Acids Res.* **49**, W72–W79 (2021).
52. Adamberg, K., Valgepea, K., & Vilu, R. Advanced continuous cultivation methods for systems microbiology. *Microbiology* **161**, 1707–1719 (2015).
53. Hoskisson, P. A., & Hobbs, G. Continuous culture--making a comeback? *Microbiology* **151**, 3153–3159 (2005).
54. Weiss, M. C., et al. The physiology and habitat of the last universal common ancestor. *Nat. Microbiol.* **1**, 16116 (2016).
55. Marcellin, E., et al. Low carbon fuels and commodity chemicals from waste gases – systematic approach to understand energy metabolism in a model acetogen. *Green Chem.* **18**, 3020-3028 (2016).
56. Jeong, Y., et al. The dynamic transcriptional and translational landscape of the model antibiotic producer Streptomyces coelicolor A3(2). *Nat. Commun.* **7**, 11605 (2016).

bioRxiv preprint doi: https://doi.org/10.64898/2025.12.18.695208; this version posted December 21, 2025. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

57. Lackner, D. H., et al. A network of multiple regulatory layers shapes gene expression in fission yeast. *Mol. Cell* **26**, 145–155 (2007).
58. Nguyen, H. L., Duviau, M. P., Laguerre, S., Nouaille, S., Cocaign-Bousquet, M., & Girbal, L. Synergistic Regulation of Transcription and Translation in *Escherichia coli* Revealed by Codirectional Increases in mRNA Concentration and Translation Efficiency. Microbiol. Spectr. 10, e0204121 (2022).
59. Picard, F., Milhem, H., Loubière, P., Laurent, B., Cocaign-Bousquet, M., & Girbal, L. Bacterial translational regulations: high diversity between all mRNAs and major role in gene expression. *BMC Genomics* **13**, 528 (2012).
60. Kanehisa, M., Sato, Y., & Morishima, K. BlastKOALA and GhostKOALA: KEGG Tools for Functional Characterization of Genome and Metagenome Sequences. *J. Mol. Biol*. **428**, 726–731 (2016).
61. Mistry, J., et al. Pfam: The protein families database in 2021. *Nucleic Acids Res.* **49**, D412–D419 (2021).
62. Cantalapiedra, C. P., Hernández-Plaza, A., Letunic, I., Bork, P., & Huerta-Cepas, J. eggNOG-mapper v2: Functional Annotation, Orthology Assignments, and Domain Prediction at the Metagenomic Scale. *Mol. Biol. Evol*. **38**, 5825–5829 (2021).
63. Hess, V., Gallegos, R., Jones, J. A., Barquera, B., Malamy, M. H., & Müller, V. Occurrence of ferredoxin:NAD(+) oxidoreductase activity and its ion specificity in several Gram-positive and Gram-negative bacteria. *PeerJ* **4**, e1515 (2016).
64. Tremblay, P. L., Zhang, T., Dar, S. A., Leang, C., & Lovley, D. R. The Rnf complex of *Clostridium ljungdahlii* is a proton-translocating ferredoxin:NAD+ oxidoreductase essential for autotrophic growth. *mBio* **4**, e00406–e412 (2012).
65. Meyer, M. M. rRNA Mimicry in RNA Regulation of Gene Expression. *Microbiol. Spectr.* **6**, 10.1128/microbiolspec.rwr-0006-2017 (2018).
66. Kashiwagi, K., & Igarashi, K. Nonspecific inhibition of *Escherichia coli* ornithine decarboxylase by various ribosomal proteins: detection of a new ribosomal protein possessing strong antizyme activity. *Biochim. Biophys. Acta*. **911**, 180–190 (1987).
67. Tan, Q., Gou, L., Fan, T. P., & Cai, Y. Enzymatic properties of ornithine decarboxylase from *Clostridium aceticum* DSM1496. *Biotechnol. Appl. Biochem.* **71**, 525–535 (2024).
68. Daume, M., Uhl, M., Backofen, R., & Randau, L. RIP-Seq Suggests Translational Regulation by L7Ae in Archaea. *mBio* **8**, e00730-17 (2017).
69. Berges, M., et al. Iron Regulation in *Clostridioides difficile*. *Front. Microbiol.* **9**, 3183 (2018).
70. Cernat, R. C., & Scott, K. P. Evaluation of novel assays to assess the influence of different iron sources on the growth of *Clostridium difficile*. *Anaerobe* **18**, 298–304 (2012).
71. Mejía-Almonte, C., et al. Redefining fundamental concepts of transcription initiation in bacteria. *Nat. Rev. Genet*. **21**, 699–714 (2020).
72. de Souza Pinto Lemgruber, R., et al. A TetR-Family Protein (CAETHG_0459) Activates Transcription From a New Promoter Motif Associated With Essential Genes for Autotrophic Growth in Acetogens. *Front. Microbiol.* **10**, 2549 (2019).
73. Bailey, T. L., et al. MEME SUITE: tools for motif discovery and searching. *Nucleic Acids Res.* **37**, W202–W208 (2009).
74. Mandal, P. K., Collie, G. W., Srivastava, S. C., Kauffmann, B., & Huc, I. Structure elucidation of the Pribnow box consensus promoter sequence by racemic DNA crystallography. *Nucleic Acids Res.* **44**, 5936–5943 (2016).
75. Graves, M. C., & Rabinowitz, J. C. *In vivo* and *in vitro* transcription of the *Clostridium pasteurianum* ferredoxin gene. Evidence for "extended" promoter elements in gram-positive organisms. *J. Biol. Chem.* **261**, 11409–11415 (1986).

bioRxiv preprint doi: https://doi.org/10.64898/2025.12.18.695208; this version posted December 21, 2025. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

76. Helmann J. D. Compilation and analysis of Bacillus subtilis sigma A-dependent promoter sequences: evidence for extended contact between RNA polymerase and upstream promoter DNA. *Nucleic Acids Res.* **23**, 2351–2360 (1995).
77. Ross, W., et al. A third recognition element in bacterial promoters: DNA binding by the alpha subunit of RNA polymerase. *Science* **262**, 1407–1413 (1993).
78. Turnbough C. L., Jr Regulation of Bacterial Gene Expression by Transcription Attenuation. *Microbiol. Mol. Biol. Rev*. **83**, e00019-19 (2019).
79. Song, Y., et al. Development of highly characterized genetic bioparts for efficient gene expression in CO_2-fixing *Eubacterium limosum*. *Metab. Eng.* **72**, 215–226 (2022).
80. Robinson, J. T., et al. Integrative genomics viewer. *Nat. Biotechnol.* **29**, 24–26 (2011).
81. Galperin M. Y. (2006). Structural classification of bacterial response regulators: diversity of output domains and domain combinations. *J. Bacteriol*. **188**, 4169–4182 (2006).
82. West, A. H., & Stock, A. M. Histidine kinases and response regulator proteins in two-component signaling systems. *Trends Biochem. Sci*. **26**, 369–376 (2001).
83. Matilla, M. A., Velando, F., Martín-Mora, D., Monteagudo-Cascales, E., & Krell, T. A catalogue of signal molecules that interact with sensor kinases, chemoreceptors and transcriptional regulators. *Fems Microbiol. Rev.* **46**, fuab043 (2022).
84. Shingler V. Signal sensing by sigma 54-dependent regulators: derepression as a control mechanism. *Mol. Microbiol.* **19**, 409–416 (1996).
85. Francke, C., et al. Comparative analyses imply that the enigmatic Sigma factor 54 is a central controller of the bacterial exterior. *BMC Genomics* **12**, 385 (2011).
86. Ingelman, H., Shaikh, K. M., & Valgepea, K. Reverse-Engineered Gas-Fermenting Acetogen Strains Recover Enhanced Phenotypes From Autotrophic Adaptive Laboratory Evolution. *Microb. Biotechnol*. **18**, e70208 (2025).
87. Szurmant, H., Bu, L., Brooks, C. L., 3rd, & Hoch, J. A. An essential sensor histidine kinase controlled by transmembrane helix interactions with its auxiliary proteins. *Proc. Natl. Acad. Sci*. U. S. A. **105**, 5891–5896 (2008).
88. Fukushima, T., Furihata, I., Emmins, R., Daniel, R. A., Hoch, J. A., & Szurmant, H. A role for the essential YycG sensor histidine kinase in sensing cell division. *Mol. Microbiol.* **79**, 503–522 (2011).
89. Woods, C., et al. Required Gene Set for Autotrophic Growth of *Clostridium autoethanogenum*. *Appl. Environ. Microbiol*. **88**, e0247921 (2022).
90. Westphal, L., Wiechmann, A., Baker, J., Minton, N. P., & Müller, V. The Rnf Complex Is an Energy-Coupled Transhydrogenase Essential To Reversibly Link Cellular NADH and Ferredoxin Pools in the Acetogen *Acetobacterium woodii*. *J. Bacteriol.* **200**, e00357-18 (2018).
91. UniProt Consortium. UniProt: the Universal Protein Knowledgebase in 2025. *Nucleic Acids Res.* **53**, D609–D617 (2025).
92. Briani, F., Carzaniga, T., & Dehò, G. Regulation and functions of bacterial PNPase. *Wiley Interdiscip. Rev.-RNA* **7**, 241–258 (2016).
93. Dendooven, T., et al. A cooperative PNPase-Hfq-RNA carrier complex facilitates bacterial riboregulation. *Mol. Cell* **81**, 2901–2913.e5 (2021).
94. Neumann, P., Lakomek, K., Naumann, P. T., Erwin, W. M., Lauhon, C. T., & Ficner, R. Crystal structure of a 4-thiouridine synthetase-RNA complex reveals specificity of tRNA U8 modification. *Nucleic Acids Res.* **42**, 6673–6685 (2014).
95. Mueller, E. G., Palenchar, P. M., & Buck, C. J. The role of the cysteine residues of ThiI in the generation of 4-thiouridine in tRNA. *J. Biol. Chem*. **276**, 33588–33595 (2001).

bioRxiv preprint doi: https://doi.org/10.64898/2025.12.18.695208; this version posted December 21, 2025. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

96. Schubert, H. L., Blumenthal, R. M., & Cheng, X. Many paths to methyltransfer: a chronicle of convergence. *Trends Biochem. Sci.* **28**, 329–335 (2003).
97. Shahbabian, K., Jamalli, A., Zig, L., & Putzer, H. RNase Y, a novel endoribonuclease, initiates riboswitch turnover in *Bacillus subtilis*. *Embo J.* **28**, 3523–3533 (2009).
98. DiChiara, J. M., Liu, B., Figaro, S., Condon, C., & Bechhofer, D. H. Mapping of internal monophosphate 5' ends of *Bacillus subtilis* messenger RNAs and ribosomal RNAs in wild-type and ribonuclease-mutant strains. *Nucleic Acids Res.* **44**, 3373–3389 (2016).
99. Durand, S., Gilet, L., Bessières, P., Nicolas, P., & Condon, C. Three essential ribonucleases-RNase Y, J1, and III-control the abundance of a majority of Bacillus subtilis mRNAs. *PLoS Genet.* **8**, e1002520 (2012).
100. Mandell, Z. F., Vishwakarma, R. K., Yakhnin, H., Murakami, K. S., Kashlev, M., & Babitzke, P. Comprehensive transcription terminator atlas for Bacillus subtilis. *Nat. Microbiol.* **7**, 1918–1931 (2022).
101. Vogel, U., & Jensen, K. F. NusA is required for ribosomal antitermination and for modulation of the transcription elongation rate of both antiterminated RNA and mRNA. *J. Biol. Chem.* **272**, 12265–12271 (1997).
102. Błaszczyk, E., et al. Depletion of tRNA CCA-adding enzyme in *Mycobacterium tuberculosis* leads to polyadenylation of transcripts and precursor tRNAs. *Sci. Rep.* **13**, 20717 (2023).
103. Parthun, M., Long, M. E., & Hemann, E. A. Established and Emerging Roles of DEAD/H-Box Helicases in Regulating Infection and Immunity. *Immunol. Rev.* **329**, e13426 (2025).
104. Valgepea, K., et al. Maintenance of ATP Homeostasis Triggers Metabolic Shifts in Gas-Fermenting Acetogens. *Cell Syst*. 4, 505–515.e5 (2017).
105. de Crécy-Lagard, V., & Jaroch, M. Functions of Bacterial tRNA Modifications: From Ubiquity to Diversity. *Trends Microbiol.* **29**, 41–53 (2021).
106. Valgepea, K., et al. H_2 drives metabolic rearrangements in gas-fermenting *Clostridium autoethanogenum*. *Biotechnol. Biofuels* **11**, 55 (2018).
107. Cox R. A. Quantitative relationships for specific growth rates and macromolecular compositions of *Mycobacterium tuberculosis*, *Streptomyces coelicolor* A3(2) and *Escherichia coli* B/r: an integrative theoretical approach. *Microbiology* **150**, 1413–1426 (2004).
108. Stouthamer A. H. A theoretical study on the amount of ATP required for synthesis of microbial cell material. *Antonie van Leeuwenhoek* **39**, 545–565 (1973).
109. Johnson, G. E., Lalanne, J. B., Peters, M. L., & Li, G. W. Functionally uncoupled transcription-translation in *Bacillus subtilis*. *Nature* **585**, 124–128 (2020)
110. Babitzke, P., & Yanofsky, C. Reconstitution of *Bacillus subtilis* trp attenuation in vitro with TRAP, the trp RNA-binding attenuation protein. *Proc. Natl. Acad. Sci*. U. S. A. **90**, 133–137 (1993).
111. Oda, M., Kobayashi, N., Ito, A., Kurusu, Y., & Taira, K. cis-acting regulatory sequences for antitermination in the transcript of the Bacillus subtilis hut operon and histidine-dependent binding of HutP to the transcript containing the regulatory sequences. *Mol. Microbiol.* **35**, 1244–1254 (2000).
112. Aymerich, S., & Steinmetz, M. Specificity determinants and structural features in the RNA target of the bacterial antiterminator proteins of the BglG/SacY family. *Proc. Natl. Acad. Sci*. U. S. A. **89**, 10410–10414 (1992).
113. Turner, R. J., Lu, Y., & Switzer, R. L. Regulation of the *Bacillus subtilis* pyrimidine biosynthetic (pyr) gene cluster by an autogenous transcriptional attenuation mechanism. *J. Bacteriol*. **176**, 3708–3722 (1994).

bioRxiv preprint doi: https://doi.org/10.64898/2025.12.18.695208; this version posted December 21, 2025. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

114. Julia Dierksheide, K., Battaglia, R. A., & Li, G. W. How do bacteria tune transcription termination efficiency? *Curr. Opin. Microbiol.* **82**, 102557 (2024).

115. Merino, E., & Yanofsky, C. Transcription attenuation: a highly conserved regulatory strategy used by bacteria. *Trends Genet.* **21**, 260–264 (2005).

116. Peters, J. M., Vangeloff, A. D., & Landick, R. Bacterial transcription terminators: the RNA 3'-end chronicles. *J. Mol. Biol.* **412**, 793–813 (2011).

117. Millman, A., Dar, D., Shamir, M., & Sorek, R. Computational prediction of regulatory, premature transcription termination in bacteria. *Nucleic Acids Res.* **45**, 886–893 (2017).

118. Metelev, M., & Johansson, M. A complex between IF2 and NusA suggests early coupling of transcription-translation. *Nature Commun.* **16**, 6906 (2025).

119. Spanka, D. T., Reuscher, C. M., & Klug, G. Impact of PNPase on the transcriptome of *Rhodobacter sphaeroides* and its cooperation with RNase III and RNase E. *BMC Genomics* **22**, 106 (2021).

120. Braun, F., Durand, S., & Condon, C. Initiating ribosomes and a 5'/3'-UTR interaction control ribonuclease action to tightly couple *B. subtilis* hbs mRNA stability with translation. *Nucleic Acids Res.* **45**, 11386–11400 (2017).

121. Khemici, V., Prados, J., Linder, P., & Redder, P. Decay-Initiating Endoribonucleolytic Cleavage by RNase Y Is Kept under Tight Control via Sequence Preference and Sub-cellular Localisation. *PLoS Genet.* **11**, e1005577 (2015).

122. Taggart, J. C., et al. A high-resolution view of RNA endonuclease cleavage in *Bacillus subtilis*. *Nucleic Acids Res.* **53**, gkaf030 (2025).

123. Condon, C., Rourera, J., Brechemier-Baey, D., & Putzer, H. Ribonuclease M5 has few, if any, mRNA substrates in *Bacillus subtilis*. *J. Bacteriol.* **184**, 2845–2849 (2002).

124. DiMari, J. F., & Bechhofer, D. H. Initiation of mRNA decay in *Bacillus subtilis*. *Mol. Microbiol.* **7**, 705–717 (1993).

125. Lioliou, E., et al. Global regulatory functions of the *Staphylococcus aureus* endoribonuclease III in gene expression. *PLoS Genet.* **8**, e1002782 (2012).

126. Schaening-Burgos, C., LeBlanc, H., Fagre, C., Li, G. W., & Gilbert, W. V. RluA is the major mRNA pseudouridine synthase in *Escherichia coli*. *PLoS Genet.* **20**, e1011100 (2024).

127. Dai, Q., et al. Quantitative sequencing using BID-seq uncovers abundant pseudouridines in mammalian mRNA at base resolution. *Nat. Biotechnol.* **41**, 344–354 (2023).

128. Eyler, D. E., et al. Pseudouridinylation of mRNA coding sequences alters translation. *Proc. Natl. Acad. Sci.* U. S. A. **116**, 23068–23074 (2019).

129. Svitkin, Y. V., Cheng, Y. M., Chakraborty, T., Presnyak, V., John, M., & Sonenberg, N. N1-methyl-pseudouridine in mRNA enhances translation through eIF2α-dependent and independent mechanisms by increasing ribosome density. *Nucleic Acids Res.* **45**, 6023–6036 (2017).

130. Gilbert W. V. Recent developments, opportunities, and challenges in the study of mRNA pseudouridylation. *RNA* **30**, 530–536 (2024).

131. Dobin, A., et al. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21 (2013).

132. Lun, A. T., Chen, Y., & Smyth, G. K. It's DE-licious: A Recipe for Differential Expression Analyses of RNA-seq Experiments Using Quasi-Likelihood Methods in edgeR. *Methods Mol. Biol.* **1418**, 391–416 (2016).

133. Benjamini, Y., & Hochberg, Y. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *J. R. Stat. Soc. Ser. B-Stat. Methodol.* **57**, 289-300 (1995).

134. Langmead, B., & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9**, 357–359 (2012).

bioRxiv preprint doi: https://doi.org/10.64898/2025.12.18.695208; this version posted December 21, 2025. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

135. Liao, Y., Smyth, G. K., & Shi, W. The Subread aligner: fast, accurate and scalable read mapping by seed-and-vote. *Nucleic Acids Res.* **41**, e108 (2013).

136. Risso, D., Ngai, J., Speed, T. P., & Dudoit, S. Normalization of RNA-seq data using factor analysis of control genes or samples. *Nat. Biotechnol*. **32**, 896–902 (2014).

137. Aitchison, J. *The Statistical Analysis of Compositional Data*. (Chapman and Hall, London, United Kingdom, 1986).

138. Chen, Y., Chen, L., Lun, A. T. L., Baldoni, P. L., & Smyth, G. K. edgeR v4: powerful differential analysis of sequencing data with expanded functionality and improved support for small counts and larger datasets. *Nucleic Acids Res.* **53**, gkaf018 (2025).

139. Lund, S. P., Nettleton, D., McCarthy, D. J., & Smyth, G. K. Detecting differential expression in RNA-sequence data using quasi-likelihood with shrunken dispersion estimates. *Stat Appl Genet Mol Biol*.**11**, /j/sagmb.2012.11.issue-5/1544-6115.1826/1544-6115.1826.xml (2012).

bioRxiv preprint doi: https://doi.org/10.64898/2025.12.18.695208; this version posted December 21, 2025. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Acknowledgements

This work was funded by the National Recovery and Resilience Plan (NRRP), Mission 4 Component 2 Investment 1.3 - Call for tender No. 1561 of 11.10.2022 of Ministero dell'Universita` e della Ricerca (MUR), funded by the European Union – NextGenerationEU. The project is identified by Project code PE0000021, Concession Decree No. 1561 of 11.10.2022 adopted by Ministero dell'Universita` e della Ricerca (MUR), CUP – E13C22001890001, according to attachment E of Decree No. 1561/2022, Project title ''Network 4 Energy Sustainable Transition – NEST (Spoke 3)''. Furthermore, this work was funded by the European Union's Horizon 2020 research and innovation program under grant agreement N810755.

Author contributions

Conceptualization: A.R. and K.V.; Methodology: A.R., K.R., A.B., and K.V.; Formal analysis: A.R., G.M.M.P., A.B., L.G., M.C-B., and K.V.; Investigation: A.R., G.M.M.P., K.R., L.G., M.C-B., and K.V.; Software: A.R., G.M.M.P., A.B.; Resources: A.R. and K.V.; Writing - Original Draft: A.R. and K.V.; Writing - Review & Editing: A.R., K.R., A.B., L.G., M.C-B., and K.V.; Visualization: A.R and K.V.; Supervision: A.R., M.C-B. and K.V.; Project Administration: A.R and K.V.; Funding acquisition: A.R and K.V..

Competing interests

The authors declare no competing interests.

Materials & Correspondence

Correspondence and requests for materials should be addressed to Angela Re and Kaspar Valgepea.