

University of Nebraska - Lincoln

DigitalCommons@University of Nebraska - Lincoln

---

Department of Chemical and Biomolecular  
Engineering: Faculty Publications

Chemical and Biomolecular Engineering,  
Department of

---

12-18-2023

## Machine Learning and Metabolic Model Guided CRISPRi Reveals a Central Role for Phosphoglycerate Mutase in *Chlamydia trachomatis* Persistence

Niaz Bahar Chowdhury

Nick Pokorzynski


Elizabeth A. Rucks

Scot P. Ouellette

Rey A. Carabeo

*See next page for additional authors*

Follow this and additional works at: <https://digitalcommons.unl.edu/chemengall>

 Part of the [Biochemical and Biomolecular Engineering Commons](#), and the [Biomedical Engineering and Bioengineering Commons](#)

---

This Article is brought to you for free and open access by the Chemical and Biomolecular Engineering, Department of at DigitalCommons@University of Nebraska - Lincoln. It has been accepted for inclusion in Department of Chemical and Biomolecular Engineering: Faculty Publications by an authorized administrator of DigitalCommons@University of Nebraska - Lincoln.

---

**Authors**

Niaz Bahar Chowdhury, Nick Pokorzynski, Elizabeth A. Rucks, Scot P. Ouellette, Rey A. Carabeo, and Rajib Saha

# **Machine Learning and Metabolic Model Guided CRISPRi Reveals a Central Role for Phosphoglycerate Mutase in *Chlamydia trachomatis* Persistence**

Niaz Bahar Chowdhury <sup>1</sup>, Nick Pokorzynski <sup>2, #</sup>, Elizabeth A. Rucks <sup>2</sup>, Scot P. Ouellette <sup>2</sup>, Rey A.  
Carabeo <sup>2</sup>, & Rajib Saha <sup>1, \*</sup>

<sup>1</sup> Chemical and Biomolecular Engineering, University of Nebraska-Lincoln, Lincoln, Nebraska,  
68508, USA.

<sup>2</sup> Department of Pathology, Microbiology, and Immunology, University of Nebraska Medical  
Center, Omaha, Nebraska, 68198, USA.

<sup>#</sup> Present address: Department of Microbial Pathogenesis, Yale University School of Medicine,  
New Haven, Connecticut, 06510, USA.

\* Address correspondence to Rajib Saha, [rsaha2@unl.edu](mailto:rsaha2@unl.edu)

## SUMMARY

Upon nutrient starvation, *Chlamydia trachomatis* serovar L2 (CTL) shifts from its normal growth to a non-replicating form, termed persistence. It is unclear if persistence is an adaptive response or lack of it. To understand that transcriptomics data were collected for nutrient-sufficient and nutrient-starved CTL. Applying machine learning approaches on transcriptomics data revealed a global transcriptomic rewiring of CTL under stress conditions without having any global stress regulator. This indicated that CTL's stress response is due to lack of an adaptive response mechanism. To investigate the impact of this on CTL metabolism, we reconstructed a genome-scale metabolic model of CTL (*iCTL278*) and contextualized it with the collected transcriptomics data. Using the metabolic bottleneck analysis on contextualized *iCTL278*, we observed phosphoglycerate mutase (*pgm*) regulates the entry of CTL to the persistence. Later, *pgm* was found to have the highest thermodynamics driving force and lowest enzymatic cost. Furthermore, CRISPRi-driven knockdown of *pgm* and tryptophan starvation experiments revealed the importance of this gene in inducing persistence. Hence, this work, for the first time, introduced thermodynamics and enzyme-cost as tools to gain deeper understanding on CTL persistence.

## KEYWORDS

*C. trachomatis*, persistence, nutrient starvation, global stress response, metabolic bottleneck.

## INTRODUCTION

*Chlamydia trachomatis* serovar L2 (CTL) is a Gram-negative obligate intracellular human pathogen<sup>1</sup> causing an estimated 1.7 million new genital tract infections annually<sup>2</sup>. This obligate intracellular pathogen replicates in a specialized membrane compartment, named inclusion, and

uses different strategies to survive in the host intracellular environment<sup>3</sup>. During its development cycle, CTL can alternate between the extracellular infectious elementary body (EB) and the intracellular non-infectious reticulate body (RB)<sup>4</sup>. EBs enter epithelial cells and differentiate into RBs inside the inclusion. After several cycles of replication, RBs undergo secondary differentiation to create new EBs, which promote subsequent rounds of host cell infection<sup>5</sup>.

When RBs sense unfavorable growth environments, such as lack of essential nutrients or immune responses, it enters a reversible state of growth-arrest, termed persistence<sup>6</sup>. The RB in the persistent state exhibits a distinct morphological form, referred to as an aberrant reticulate body (ARB). ARBs are metabolically active with an enlarged and aberrant morphology<sup>7</sup>. Several studies identified stimuli that induce CTL persistence. Some of these stimuli are the addition of the cytokine interferon- $\gamma$ <sup>8</sup>, tryptophan starvation<sup>9</sup>, iron starvation<sup>10</sup>, and infection with immune cells<sup>11</sup>. Despite these studies, the fundamental nature of persistence is still puzzling. One hypothesis is, CTL enters persistence as it lacks an adaptive response strategy<sup>12-14</sup>. An extension of this hypothesis is that persistence is effectively a substitute for the lack of a conventional global response regulator, when encountering nutritional stress. A recent study<sup>15</sup> compared iron and tryptophan starvation to the nutrient-replete conditions. While there were significant transcriptomic differences between control and starvation conditions, it is still not clear if those changes are due to CTL's lack of adaptive response on nutrient starvation or not. To gain a holistic understanding, it is pertinent to understand the metabolic landscape corresponding to the transcriptional changes. However, there is a dearth of experimental studies investigating the whole CTL metabolic landscape. Thus, a systems biology approach is required to understand the metabolism of CTL under different stress conditions.

To develop such a systems biology approach, genome-scale metabolic models (GSMs) have been widely used<sup>16,17</sup>. A GSM captures annotated metabolic reactions within a biological system and can predict reaction fluxes using flux balance analysis (FBA)<sup>18</sup>, flux variability analysis (FVA)<sup>19</sup>, and parsimonious FBA (pFBA)<sup>20</sup>. To date, GSMs developed for different pathogens were successful in predicting metabolic adaptations, functionalities, and properties of *Mycobacterium tuberculosis*<sup>21</sup>, *Acinetobacter baumannii*<sup>22</sup>, *Klebsiella pneumoniae*<sup>23</sup>, and *Helicobacter pylori*<sup>24</sup>. GSMs were also successful in predicting different virulence factors such as lipid A modifications of *Pseudomonas aeruginosa*<sup>25</sup>, and substrate utilization pattern of *Staphylococcus aureus*<sup>26</sup>. Moreover, GSMs successfully predicted novel drug targets for *Salmonella typhimurium*<sup>27</sup>, and *Campylobacter jejuni*<sup>28</sup>. Hence, to understand the metabolic landscape of CTL, GSM can be useful. There is a recently published GSM of *C. trachomatis*<sup>29</sup>. However, the biochemical network of the model was incomplete. To address this, we reconstructed a new GSM of CTL, *iCTL278*.

To study the nature of persistence, persistence-specific ‘omics’ data needed to be overlaid with the *iCTL278*. This process is called contextualization of the GSM. Without contextualization, GSM may predict unrealistic reaction fluxes<sup>30</sup>, erroneous cellular phenotype<sup>31</sup>, or inaccurate growth rate patterns<sup>32</sup>. Two different approaches are available for contextualization, switch (e.g., GIMME<sup>33</sup>, iMAT<sup>34</sup>, MADE<sup>35</sup>, and RIPTiDe<sup>36</sup>) or valve (e.g., E-Flux<sup>37</sup> and PROM<sup>38</sup>, EXTREAM<sup>39</sup>) approach. While switch approaches display a binary nature, resulting in the active-or-inactive status of a reaction, valve approaches provide more flexibility by making the reaction flux proportional to the abundance of associated transcripts/proteins.

In this work, to perform systems-level investigations of CTL persistence, transcriptomics data were collected for nutrient-sufficient and nutrient-starved conditions. To gain further insights

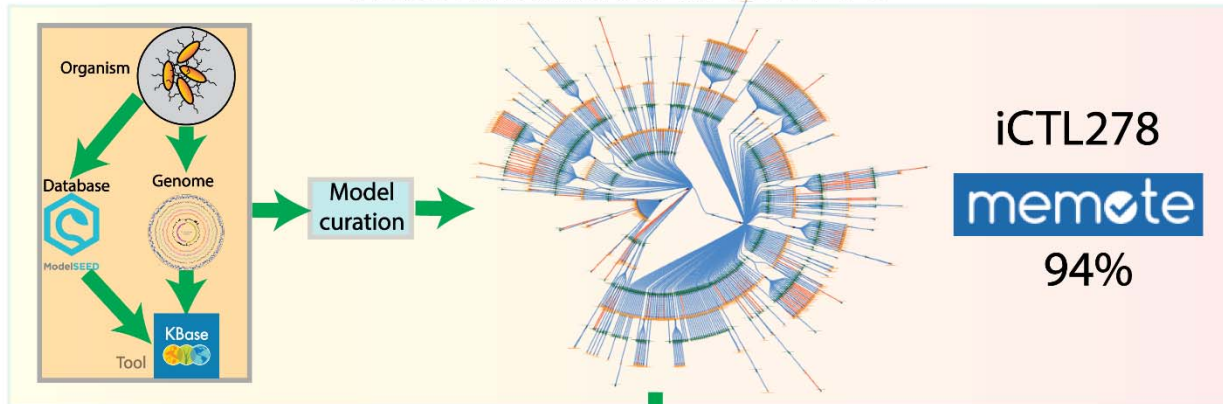
from these transcriptomics data, an unsupervised machine learning approach called K-means clustering algorithm was implemented resulting in the identification of four core components of the CTL transcriptome. This along with a statistical correlation study revealed a global transcriptomic rewiring of CTL under nutritional stress conditions. This, along with a lack of global stress regulator in CTL revealed lack adaptive response as the primary reason of CTL's entry to persistence. To further understand how this global transcriptomic response impacts CTL metabolism, contextualized GSMs were reconstructed using the E-flux algorithm. Our recently developed Metabolic bottleneck analysis (MBA)<sup>39</sup> identified phosphoglycerate mutase (*pgm*) regulated the entry of CTL into persistence, which had the highest thermodynamics driving cost and lowest enzymatic cost. Furthermore, we used CRISPRi to block transcription of *pgm* and subsequently performed starvation experiments. The outcome supported the prediction from MBA, thermodynamics driving force, and enzyme cost analysis. Additional systems-level investigation pin-pointed the cellular fitness of the persistence, which is to prime itself for a rapid growth upon availability of nutrient. Overall, this combined machine learning and metabolic model-guided study, for the first time, decoded CTL persistence through the lens of thermodynamic driving force and enzymatic cost and will work as a blueprint to investigate phenotypical and genotypical changes associated with other microbial infection.

## **RESULTS AND DISCUSSION**

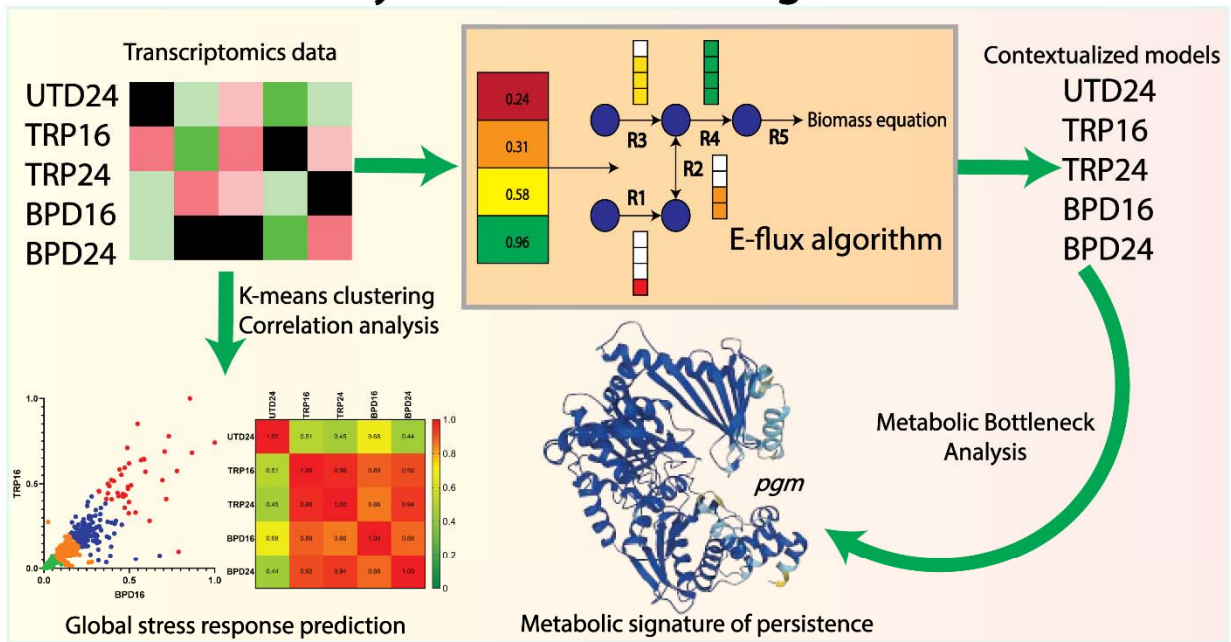
### ***Chlamydia trachomatis* L2 genome-scale metabolic model development and analysis:**

To analyze the metabolism of CTL in different stress conditions, we needed to reconstruct a GSM of CTL. A draft GSM was reconstructed using the NCBI RefSeq genome annotation from the Kbase<sup>40</sup>.

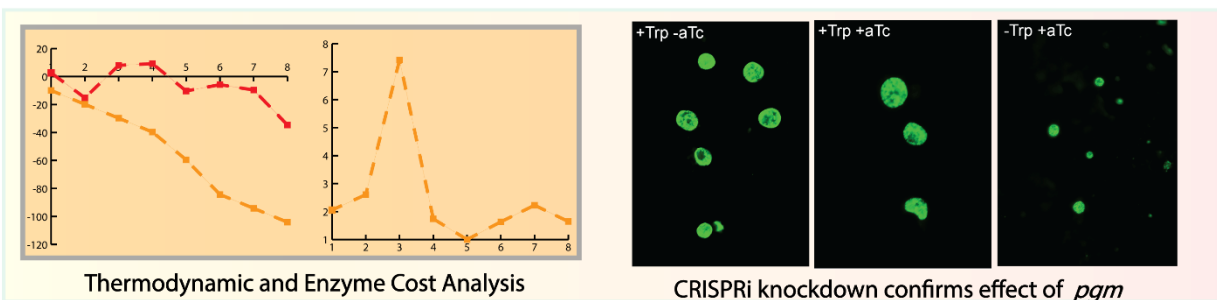
## Reconstruction of iCTL278



## Chlamydial Persistence Regulator Prediction



## *In Silico* Analysis and CRISPRi Validation of *Pgm*



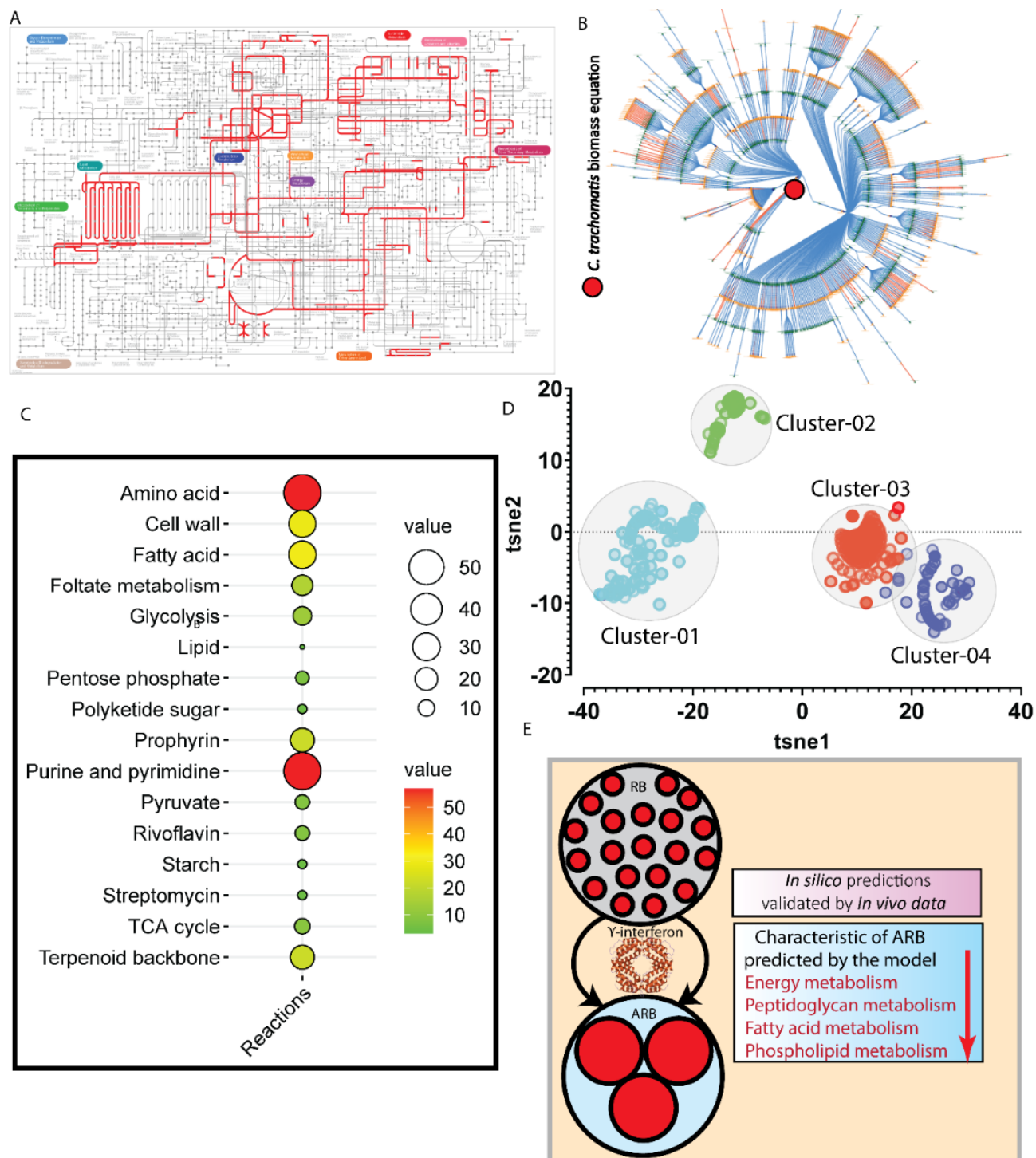


**Fig. 1 Overall process of the model reconstruction, refinement, and subsequent analysis/validation.** Kbase, NCBI RefSeq, and Modelseed database was used to reconstruct the initial model. After standard model curations, a high-quality GSM was ready to use. Later, condition specific transcriptomics data were incorporated with GSM through E-flux algorithm, which resulted in five contextualized models. These models were used to find metabolic bottlenecks and persistence mechanism. K-mean clustering algorithm applied on the transcriptomics data revealed global stress response mechanism. Thermodynamics and enzyme cost analysis, along with *in vitro* experimentation reveal the regulatory role of *pgm* in CTL persistence.

After reconstructing the draft GSM, an extensive literature search was performed regarding the CTL-specific biomass composition. However, with information unavailable, a template Gram-negative biomass equation was used as the objective function of FBA. This draft model did not capture some known metabolic functionality of CTL. For example, CTL has an incomplete TCA cycle and must uptake malate<sup>41</sup>. However, the malate uptake reaction was missing in the model. Similarly, CTL can uptake glucose 6-phosphate from the host cell<sup>41</sup>. To account for these, we performed gap-filling using our previously developed tool OptFill<sup>42</sup> and added transport reactions for both malate and glucose 6-phosphate to the model. CTL partially depends on the host cell for ATP and NAD<sup>+</sup> uptake<sup>41</sup>. We also added transport reactions for CTL to uptake these metabolites when necessary. Overall, the curated model, *iCTL278*, consisted of 729 metabolites, 692 reactions, and 278 genes. We performed the MEMOTE testing<sup>43</sup> for *iCTL278* and it returned an impressive score of 94%. The model is fully mass balanced and stoichiometrically consistent. The full MEMOTE report can be found in the supplementary S1 Text. The MEMOTE score comparison between *iCTL278* and the previously published CTL model is shown in Fig. S1A. In addition to performing MEMOTE, we also verified other known metabolic traits of CTL, such as the *iCTL278* can uptake several nucleotides including GTP, CTP, and UTP to sustain biomass growth<sup>44</sup>. CTL is an auxotroph for

most amino acids and can only biosynthesize alanine, aspartate, and glutamine using dicarboxylates from the TCA cycle<sup>41</sup>. The rest of the amino acids need to be acquired from the host cell. *iCTL278* recapitulated these metabolic traits of amino acids. Overall, *iCTL278* captured all the known metabolic characteristics of CTL. The model reconstruction process is shown in Fig. 1, the metabolic networks of the model are shown in Fig. 2A and Fig. 2B and the number of reactions in different pathways is shown in Fig. 2C.

We also wanted to understand how the interconnectedness of the major pathways of *iCTL278*. Hence, we performed flux sampling of *iCTL278*, followed by the dimensionality reduction and clustering of those flux samples. The analysis yielded four different clusters (Fig. 2 D). Cluster-01 reactions are mostly from glycolysis and pentose phosphate pathways. In cluster-02, most of the reactions are from the ubiquinone and terpenoid pathways. Cluster-03 consists of primarily pyrimidine and fatty acid pathways. Finally, cluster-04 contains reactions mostly from purine and riboflavin pathways. From these clustering, we now have a better understanding how metabolic pathways in CTL are connected to one another in a higher dimensional space. Further details on the clusters can be found in supplementary S1 Table.



**Fig. 2** An overview of the metabolic network captured in the CTL genome-scale metabolic model and result of model validation. A) Visualization of the GSM metabolic network was performed using iPath3<sup>45</sup>. This metabolic network provides an idea regarding the limited metabolic capability of CTL; B) Metabolic network of CTL is centered around its biomass equation, network generate using fluxer<sup>46</sup>; C) A balloon plot representing number of

reactions in different pathways in the *iCTL278*; D) tSNE plot of the flux sampling data obtained from *iCTL278*; E) Predictions for contextualized *iCTL278* matched well with the experimental observations.

### **Model Validation using experimental proteomics data**

Although the MEMOTE testing of *iCTL278* returned an impressive score of 94%, it was still necessary to determine if *iCTL278* could capture the known phenotypes of CTL upon contextualization. In a previous study<sup>37</sup>, a valve-based approach, E-flux algorithm that connects reaction activities linearly with the gene expression levels accurately predicted different cellular phenotypes regarding fatty acid biosynthesis of *M. tuberculosis*, a well known pathogenic bacterium. Therefore, a previously published study on quantitative protein profiling of interferon- $\gamma$ -treated CTL<sup>8</sup> was used to reconstruct two contextualized models using the E-flux algorithm<sup>37</sup>, reticulate body (RB) and aberrant reticulate body (ARB) models. The latter was under stress induced by interferon- $\gamma$ .

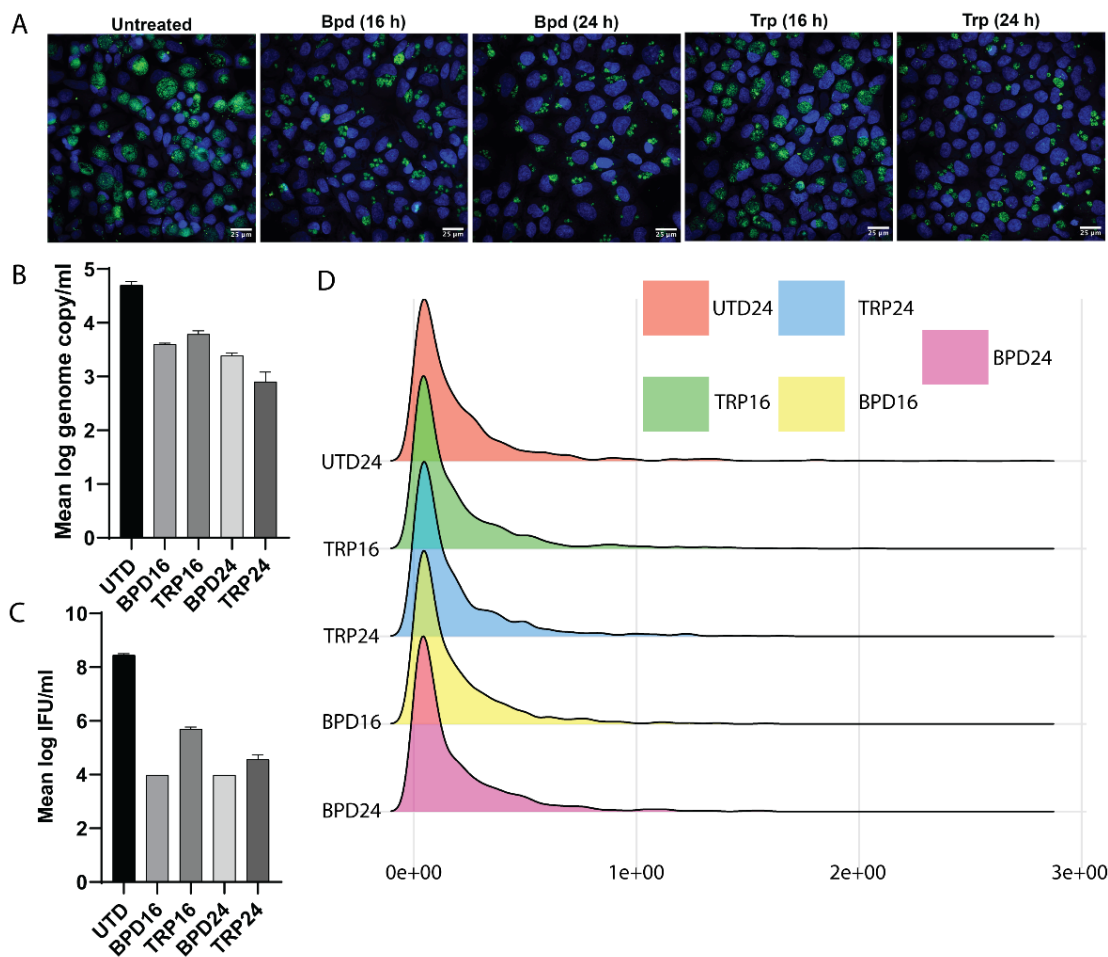
When flux distribution was calculated for both RB and ARB, similar to previous studies, flux through energy metabolism was reduced in the ARB compared to RB<sup>8</sup>. As CTL enters persistence, cell division is arrested, and metabolism is reduced to a minimum level such that only essential cellular functions are maintained. As GSMs are unable to directly calculate the concentration of different metabolites involved in a biochemical network, we used the flux-sum analysis (FSA) approach to predict the metabolic pool size of ATP in both RB and ARB for further verifications, and found a lower ATP pool size in the ARB condition. The details of flux-sum analysis can be found in our previous work<sup>32</sup> and also in the materials and methods section. Thus, the reduced energy metabolism in ARB compared to the RB, a distinguishing feature of the persistence state, was successfully captured by the *iCTL278*.

As a further verification, we explored the peptidoglycan biosynthesis pathway of CTL. Peptidoglycan is essential cell division in CTB but, unusually, is not a component of the cell wall. During stress conditions, CTL arrests its cell division<sup>6</sup>. As cell division is arrested in the ARB, the model predicted reduced peptidoglycan biosynthesis in ARB compared to RB, which is consistent with a previous transcriptomic study<sup>47</sup>. CTL relies on *de novo* fatty acid and phospholipid biosynthesis to produce its membrane phospholipids, which are also crucial for cell division<sup>48</sup>. When fatty acid biosynthesis reactions in both RBs and ARBs were compared, we found reduced activity in the fatty acid biosynthesis in ARBs compared to RBs, which reiterates what has previously been reported<sup>49</sup>. This was expected given the predicted reduced requirement for fatty acid biosynthesis in the absence of bacterial replication.

Overall, *iCTL278* recapitulates known phenotypes of ARBs and RBs upon contextualization. Therefore, *iCTL278* is well suited to further contextualize and analyze the metabolism of CTL under different stress conditions. The summary of the validation result is shown in Fig. 2 E.

### **Effects of nutrient starvation on CTL growth and development**

Our understanding of the molecular underpinnings behind the persistence of CTL is still limited. Despite employing numerous methods to induce persistence, the result consistently shows similar ARB morphology, as well as disruptions in the transcriptome and proteome. Thus, severe alterations to transcriptional and translational profiles may cause aberrant growth.



**Fig. 3** Morphology, genome-copy number, and infectious progeny data for different stage of CTL tryptophan and iron starvation A) CTL exhibits duration-dependent sensitivity to iron limitation and tryptophan (trp) starvation. Infected cells were starved for iron by treatment with the chelator 2,2-bipyridyl (Bpd) starting at either the time of infection or at 8 h post-infection. Inclusions as indicated by staining for the chlamydial major outer membrane protein (MOMP) were allowed to develop for a total of 24 h. Tryptophan starvation was started at the same time points as above, with indicated durations. Note the more significant delay in inclusion development by Chlamydia starved for 24 h. B) Genome copy number data of *euo* gene for UTD24, BPD16, TRP16, BPD24, TRP24. C) Infectious progeny data for UTD24, BPD16, TRP16, BPD24, TRP24 D) Distribution of gene expression values.

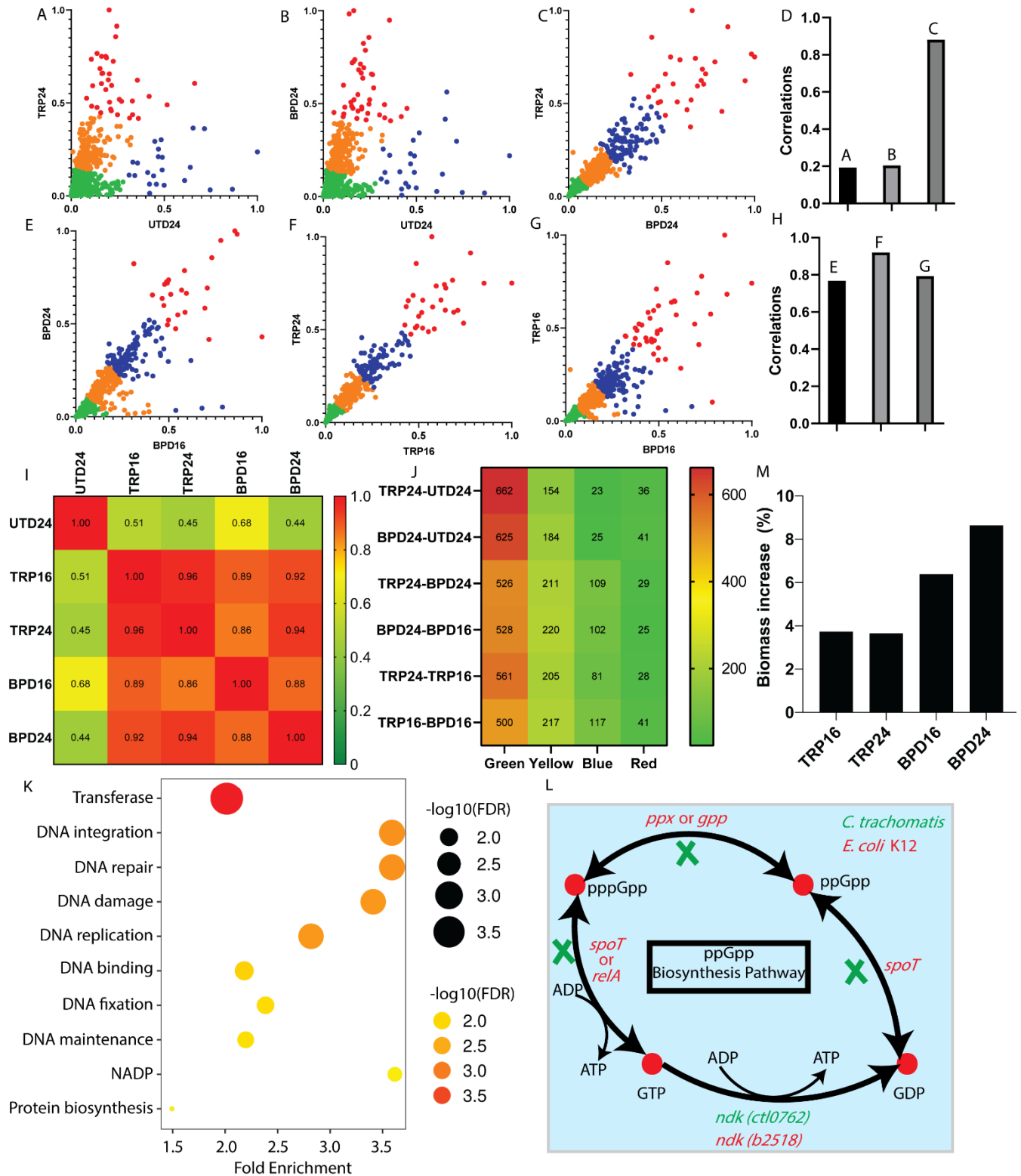
We assayed several distinguishing features of CTL persistence across the different treatment conditions, i.e., untreated (UTD24), 16 hours of iron (BPD16) and tryptophan starvation

(TRP16), and 24 hours of iron (BPD24) or tryptophan starvation (TRP24). One of those was the morphology of CTL, which was monitored by immunofluorescent confocal microscopy. Here we confirmed that, in contrast to the UTD24, all starvation strategies yielded smaller inclusions (Fig. 3 A), which was associated with lower yields of inclusion-forming units, indicating severely delayed development that compromised the regeneration of infectious particles. We observed that BPD treatments resulted in aberrantly enlarged organisms. TRP inclusions showed smaller inclusions, indicated growth defect. We also analyzed genome copy number across all stress conditions. We observed significantly reduced genome equivalents for all stress conditions compared to UTD24. Unsurprisingly, genome copy numbers of chlamydial late-stage transcriptional regulator, *euo*, between BPD and TRP were similar (Fig. 3 B). With regards to generation of infectious progeny, we observed a much more dramatic effect of BPD treatment than TRP starvation, indicating that iron starvation is a much more potent trigger of CTL developmental arrest relative to tryptophan depletion (Fig. 3 C).

To further analyze the stress conditions of CTL, RNA sequencing was performed for each of the stress conditions, including an untreated control<sup>15</sup>. Many genes fall in the low expression level and gene expression distribution also skewed heavily toward the low expression range (Fig. 3 D). Subsequently, a pair-wise comparison of profiles was conducted to determine the level of similarities of transcriptomes across different stress conditions. Using scatter plots, we were able to visualize correlations between pairs of transcriptomics data. From comparisons of TRP24-UTD24 (Fig 4. A), and BPD24-UTD24 (Fig 4. B), we found a weak correlation for each case (Fig 4. D). However, TRP24-BPD24 (Fig 4. C) showed a strong correlation (Fig 4. D). This pointed to CTL may having a common global transcriptomic response to different stress conditions. To reveal more about the pattern of global transcriptomic response, we implemented

an unsupervised machine learning technique called K-mean clustering on TRP24-UTD24, BPD24-UTD24, and TRP24-BPD24. K-mean clustering identified four distinct clusters for each plot. Details of each cluster can be accessed in the supplementary S2 Table.





**Fig. 4** Pairwise K-mean clustering plots for different stress conditions. A) Scatter plot of 24 hour of tryptophan starved (TRP24) and untreated (UTD24) CT; B) Scatter plot of 24 hour of iron starved (BPD24) and untreated

(UTD24) CT; C) Scatter plot of 24 hour of tryptophan starved (TRP24) and 24 hour of iron starved (BPD24); D) Correlation of different scatterplots. These scatterplots and correlation between different stress conditions reveal CTL may behave similarly under different stress conditions; E) Scatter plot of 24 hour of iron starved (BPD24) and 16 hour of iron starved (BPD16) CTL; F) Scatter plot of 24 hour of tryptophan starved (TRP24) and 16 hour of tryptophan starved (TRP16) CT; G) Scatter plot of 16 hour of tryptophan starved (TRP16) and 16 hour of iron starved (BPD16); H) Correlation of different scatterplots. These scatterplots and correlation between different stress conditions reveals, CTL may have a global transcriptomics response; I) Pearson correlation matrix among different conditions. The correlation was calculated using two-tail test with 95% confidence interval; J) Heat map for number of reactions in each cluster for different conditions; K) Gene ontology enrichment chart for common genes in green cluster across all conditions; L) Hypothetical ppGpp pathway in CTL; M) Increased biomass growths for stress condition after introducing ppGpp pathway in CTL.

Genes were clustered based on function, with clusters in green indicating metabolism-related genes; yellow clusters consisting of cell signaling-related genes, and red and blue clusters indicating transcription- and translation-related genes, respectively. From Fig. 4 (A, C), it is evident that most of the genes in green clusters remained similarly expressed. However, genes in the yellow and red cluster were moderately or highly upregulated in the TRP24/BPD24 compared to the UTD24 (Fig 4 A and B). In contrast, genes in the blue clusters were upregulated in the UTD24 compared to the TRP24/BPD24 (Fig 4 A and B).

As green clusters are tightly clustered across all the conditions (Fig 4. A, C), these can be called “core” components, as identified previously<sup>15</sup>. In contrast, blue, yellow, and red clusters were upregulated/downregulated across all conditions (Fig 4. A, C) and can be labeled as stress-specific “accessory” components<sup>15</sup>. Some of the known tryptophan utilization genes such as *trpR*, *trpC*, and *trpS* fell in the yellow clusters and were thus moderately upregulated in TRP24 (Fig 4. A). Several studies predicted similar upregulation of tryptophan utilization genes under tryptophan starvation<sup>14</sup>. *In vivo*, the amount of tryptophan available is often depleted by

the pro-inflammatory cytokine interferon- $\gamma$  through the transcriptional upregulation by the *ido1* gene, encoding the enzyme IDO that catabolizes tryptophan to kynurenine, which cannot be used by CTL. However the microbiome in the relatively hospitable niche of the lower genital tract may provide tryptophan or more likely indole, which CTL can use for tryptophan biosynthesis via the salvage pathway<sup>50</sup>. Indeed, several studies have shown that under tryptophan starvation, CTL can uptake indole supplemented in the media for conversion to tryptophan<sup>50,51</sup>.

Several stress-related genes, such as *ahpC* and *euo* were also upregulated in the TRP24 compared to the UTD24 (Fig 4. A). Between these two genes, *ahpC* was significantly upregulated and fell in the red cluster, while *euo* was moderately upregulated and included in the yellow cluster. *AhpC* is a thio-specific antioxidant peroxidase gene and is involved in the redox homeostasis of CTL. A similar severe upregulation of *ahpC* is also observed in CTL exposed to interferon- $\gamma$  stress<sup>47</sup>. For *Chlamydia pneumoniae*, it was reported that severe upregulation of *ahpC* would protect the bacteria against cytokine-induced reactive nitrogen intermediates, thus allowing *C. pneumoniae* to cause long-term infection<sup>52</sup>. For *euo*, encoding a DNA-binding protein, a moderate upregulation was observed in CTL under interferon- $\gamma$  induced stress. *Euo* was predicted to be a negative regulator of CTL genes involved in RB-to-EB differentiation<sup>53</sup>. *IhfA* is a cell division gene of CTL and was downregulated in the TRP24, and thus, possibly contributing to the lack of cell division observed in TRP24-treated samples (Fig 4. A).

In persistence, glycolysis mostly supports the increased production of starch and carbohydrate, thus CTL mostly depending on the host cell for energy<sup>41</sup>. As a result, reduced expression of TCA cycle genes, such as *sucA*, *sucB*, *sucC*, and *sucD* was expected, which fell in the bottom left portion of the green cluster. A similar overall result was observed for BPD24-UTD24 (Fig 4. B).

In the TRP24-BPD24 (Fig 4. C) scatterplot, the key difference is, that the stress-related accessory genes expressed more in chorus with the core genes, thus establishing a strong relation between two different stress conditions. Overall, the K-mean clustering predicted clusters whose functionality matched closely with the literature. Thus, this clustering analysis will further serve as a “genome-wide library” to identify core and stress-related genes of CTL.

To gather more insights into the presence of a global stress response in CTL, we plotted other stress conditions, such as BPD24-BPD16 (Fig 4. E), TRP24-TRP16 (Fig 4. F), and TRP16-BPD16 (Fig 4. G). Interestingly we noticed a very strong correlation in all the cases (Fig 4. H), supporting the proposed existence of global stress response of CTL. A Pearson correlation heat map (Fig 4. J) of normalized read counts of all genes among different conditions also indicated strong correlation between all the stress conditions. Number of reactions in each cluster for different conditions is shown in Fig. 4 I. Common genes in green, yellow, blue, and red clusters were identified using Venn diagrams (Fig. S1B, Fig. S1C, Fig. S1D, and Fig. S1E respectively). Gene ontology enrichment analysis for the common genes across green clusters is shown in Fig. 4 K. The same analysis for the yellow cluster (Fig. S2A), and the red clusters (Fig. S2B) are provided in the supplementary information. Since the blue cluster has only one common gene (*ctl0256*) we could not perform the gene ontology enrichment analysis.

Notably, the strong correlation between transcriptomes associated with two distinct stresses is not unique to CTL. Analysis of publicly available transcriptomes from differently stressed *Mycobacterium tuberculosis* and *Escherichia coli* revealed a similar correlation. We collected the transcriptional profile of *M. tuberculosis* exposed to *in vitro* lysosomal stress for 24 hours and 48 hours<sup>54</sup>. From the scatterplot (Fig. S3A), we noticed a very high correlation (Fig. S3D). We also collected the transcriptional profile of *M. tuberculosis* exposed in zinc limited medium

and zinc replete medium<sup>55</sup>. Similar to the previous case, the scatterplot (Fig. S3B) indicated a very high correlation (Fig. S3D). However, when we scatterplot (Fig. S3C) *M. tuberculosis* transcription data for lysosomal stress of 24 hours and zinc limited medium, it showed a very low correlation (Fig. S3D). Thus, *M. tuberculosis* response to divergent stresses is customized to each stress. However, the transcriptional response for a single type of stress in *M. tuberculosis* progresses as a function of severity (i.e. duration) of the stress.

Unlike *M. tuberculosis*, *E. coli* showed a global stress response mechanism. Transcriptomics profile of 50 minutes of cold stress against 90 minutes of cold stress (Fig. S4A) for *E. coli* from the literature<sup>56</sup> showed a very strong correlation (Fig. S4D). Similarly, transcriptomics profile of 40 minutes of oxidative stress against 90 minutes of oxidative stress (Fig. S4B) for *E. coli* from the literature<sup>56</sup> also showed a very strong correlation (Fig. S4D). Thus, like *M. tuberculosis*, for temporal progression of similar stress, *E. coli* may have a similar stress response mechanism. Interestingly, transcriptomics profile of 90 minutes of cold stress against 90 minutes of oxidative stress (Fig. S4C) showed a very strong correlation (Fig. S4D). Thus, under different stress responses, the transcriptional response of *E. coli* is very similar to that of CTL.

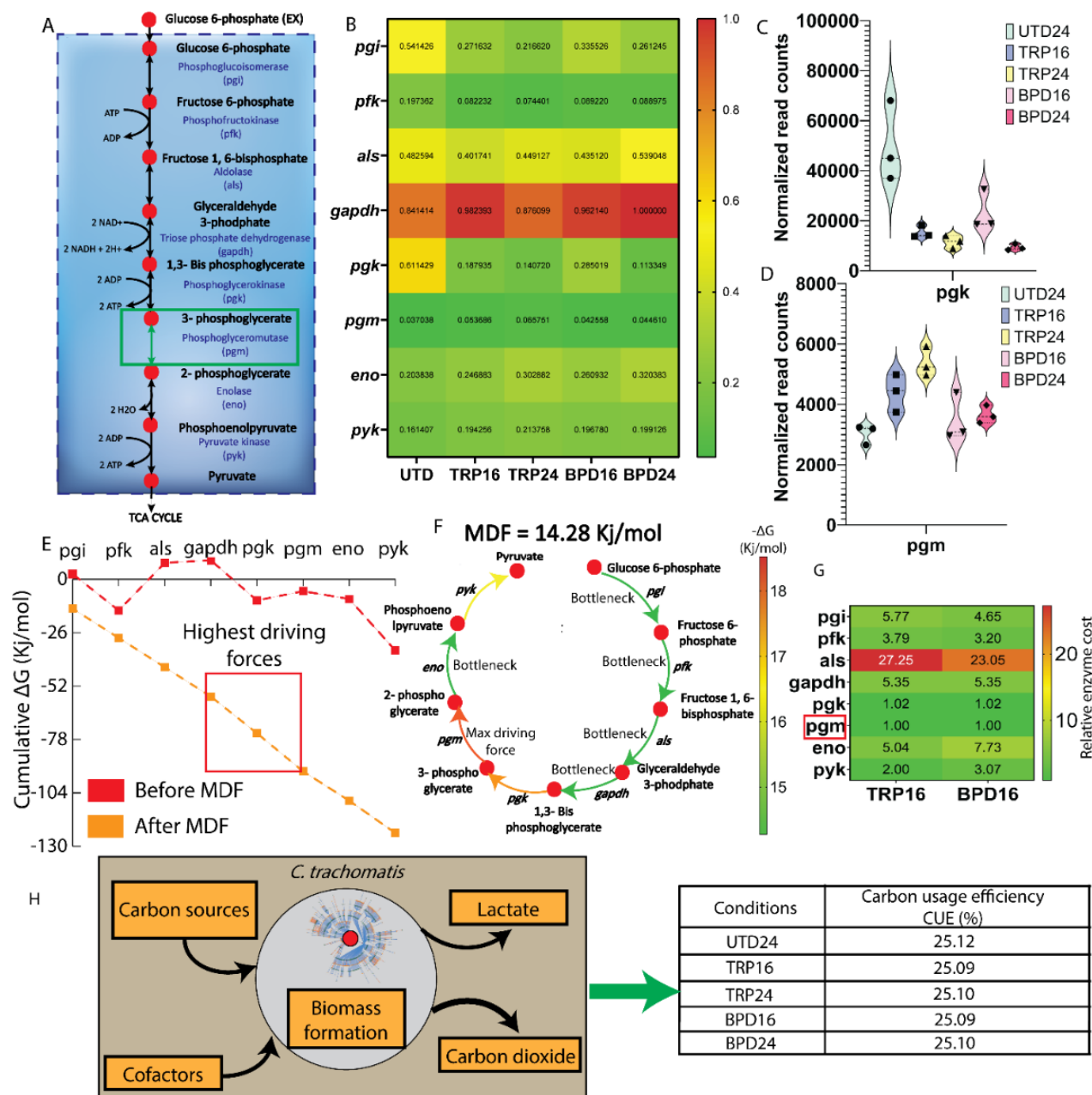
For *E. coli*, ppGpp is a global stress regulator<sup>57</sup>. When an uncharged tRNA binds in the ribosome, the ribosome-associated *relA* protein is activated to synthesize ppGpp<sup>58</sup>, which acts as a global regulator of transcription by modulating transcription complexes at promoters<sup>57</sup>. The stringent response serves to stop the synthesis of stable RNA species, such as rRNA and tRNA, to increase protein degradation pathways to maintain growth rate<sup>58</sup>. Collectively, these responses serve to overcome the starvation. *SpoT* is a cytosolic bifunctional enzyme with ppGpp synthase and hydrolase activity that helps control the levels of ppGpp. CTL does not have homologues of *relA* and *spoT* and does not synthesize ppGpp<sup>59</sup>. The loss of these genes has likely occurred

through reductive evolution as a means for adapting to obligate intracellular environment. Thus, highly correlated gene expression under different stress conditions, the absence of *relA* and *spoT* homologues, the inability to synthesize ppGpp, and the upregulation of stable RNA<sup>12</sup> indicate CTL does not engage in a stringent response during starvation, manifesting as an overlapping transcriptional response during iron and tryptophan starvation.

To further scrutinize the effect of ppGpp for CTL, we reconstructed a model of CTL, capable of producing ppGpp, by adding *spoT*, *relA*, *ppx*, and *gpp* proteins to the *iCTL278*. This was performed by adding three reactions to the *iCTL278* as shown in Fig 4. L. CTL *dksA* does not have the binding pocket for ppGpp. Thus, we also assumed that *dksA* now has a binding site for ppGpp. With the capability to produce ppGpp, it was expected to see an increased biomass growth rate in all tested stress conditions compared to the strain with no capacity to produce ppGpp. Indeed, relative biomass growth rates for all tested stress conditions increased significantly after adding the *spoT* and *relA* proteins to *iCTL278* (Fig 4. L), compared to the original conditions (4% to 8%). This is a clear indication that ppGpp produces a stringent action under stress conditions by maintaining the biomass growth rates. As CTL lacks the capability to produce ppGpp, thus the nature of its response to nutrient stress is not adaptation, but rather induction of abnormal growth.

### **Impact of global stress response on *Chlamydia trachomatis* metabolism**

We next sought to determine if this lack of stringent action translated to novel insights into the metabolic landscape of tryptophan- or iron-starved CTL. To answer that, we contextualized *iCTL278* for UTD24, TRP16, TRP24, BPD16, and BPD24 using the E-flux algorithm.



**Fig. 5** Regulatory reactions in CTL to enter persistence. A) Phosphoglycerate mutase in glycolysis pathway; B) Normalized gene expression heatmap of genes involved in the glycolysis; C) Transcripts level of *pgk* in TRP16-UTD24 and BPD16-UTD24; D) Transcripts level of *p gm* in TRP16-UTD24 and BPD16-UTD24; E) MDF analysis showing the cumulative driving force of glycolysis; F) Driving force plot showing phosphoglycerate mutase (*p gm*) and phosphoglycerate kinase (*pgk*) are two top most driving force reactions; G) Individual driving force of each

reactions; G) Enzyme cost of glycolysis for TRP16 showing phosphoglycerate mutase has the lowest enzyme costs; H) Carbon usage efficiency for UTD24, TRP16, TRP24, BPD16, and BPD24.

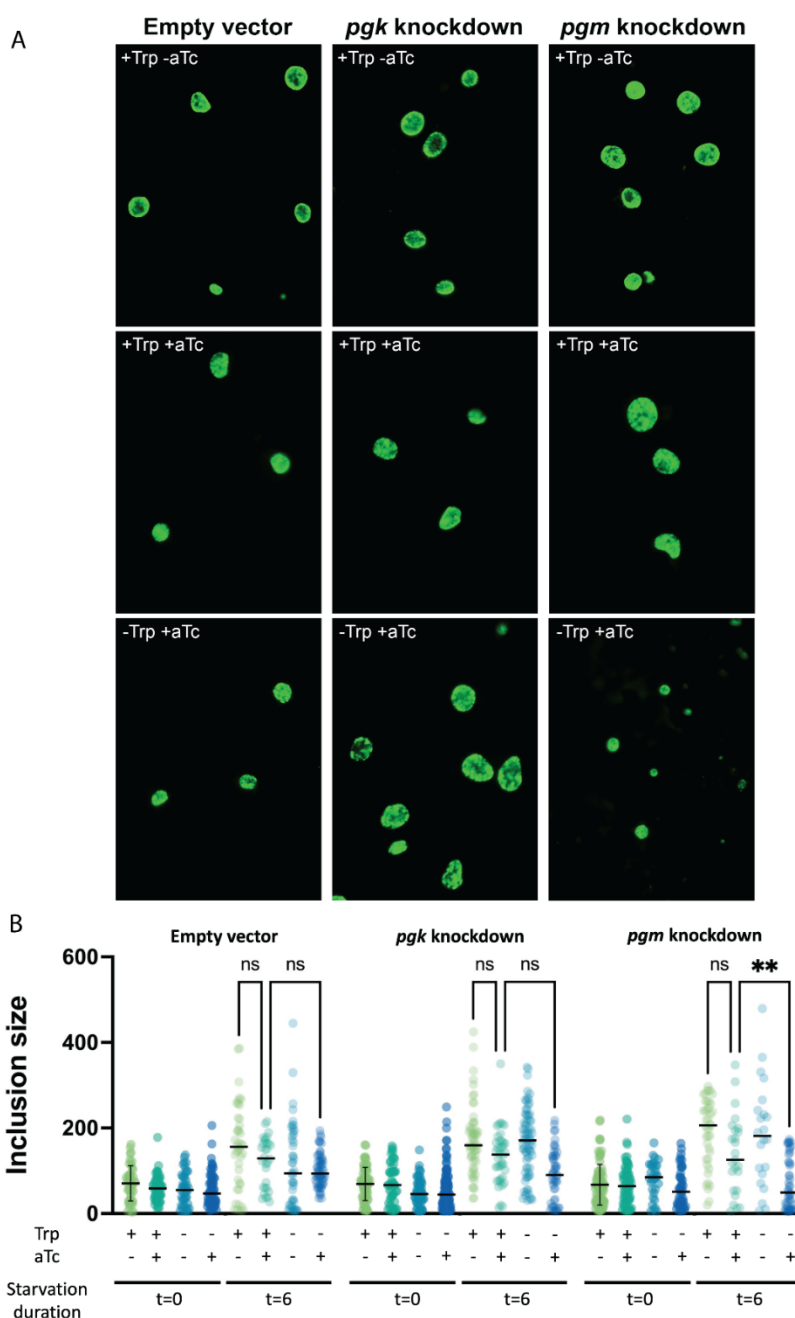
To confirm that the results from contextualized models are merely not the artifact of transcriptomics data, we calculated the correlation matrix between transcriptomics data predicted flux (more information regarding calculating transcriptomics data predicted flux can be found in the E-flux algorithm sub-section of Materials and Method section) and model predicted flux and found very weak correlation between them (Fig. S5 A). We also found that model predicted flux distribution for different stress conditions are highly correlated (Fig. S5 B). This supports the previously found high correlation among different stress conditions from the transcriptomics data analysis (Fig. S4 I).

As CTL enters persistence, the metabolic difference between UTD24 and TRP16/BPD16 can give insight into the impact of CTL global transcriptome rewiring on its metabolism. To find out the metabolic differences between UTD24 and TRP16, we took the TRP16 model and implemented metabolic bottleneck analysis (MBA)<sup>39</sup>. MBA revealed that, for TRP16, allowing phosphoglycerate mutase (*pgm*; *ctl0091*) from the glycolysis pathway to carry a similar reaction flux as UTD24 resulted in the same biomass growth rates for UTD24 and TRP16 conditions. The resulting reaction fluxes of other reactions of UTD24 and TRP16 were also found to be similar. This result indicates that potential regulation of *pgm* is the metabolic trait of rewiring of the CTL global transcriptomics response as it enters persistence (Fig. 5 A, B). The *pgm* and the upstream reaction phosphoglycerate kinase (*pgk*) showed opposite transcript profiles under TRP16 condition, with the former being increased (Fig. 5 C, D). Similar results were obtained for BPD16. However, increased level of transcripts may not necessarily translate into more proteins<sup>6,12</sup>.



To dissect further into the significance of *pgm* transcriptional regulation in persistence, we performed max/min driving force (MDF) analysis<sup>60</sup> on the glycolysis pathway of CTL. MDF analysis maximizes the total driving force of a given pathway within the biologically relevant concentration of different metabolites. Fig. 5 E indicates the driving force of glycolysis before and after the MDF analysis. Among all the reactions in glycolysis, MDF analysis predicted that *pgm* has the highest driving force (Fig. 5 F), while *pgk* has the second-highest driving force (Fig. 5 F). Concentrations of different metabolites, predicted from MDF analysis, are shown in Fig. S5 C. In the context of MDF, shadow price<sup>61</sup> of a reaction indicates the impact of small perturbation of Gibb's free energy of that reaction on the overall pathway thermodynamics. Similarly, shadow price of concentrations for each metabolites indicates the impact of small perturbation of metabolite concentration on the overall pathway thermodynamics. Therefore, we calculated the shadow price of driving forces of each of the reaction (Fig. S5 D) and found the first three steps of the glycolysis (*pgi*, *pfk*, and *als*) having the most impact on the overall driving force with changing reaction fluxes. Next, we calculated the shadow price of concentrations (Fig. S5 E) of each of the metabolites and found H<sup>+</sup>, glyceraldehyde 3-phosphate, and glycerone phosphate having the most impact on the overall driving force with small changes in concentrations. Flux-force efficacy relationship (Fig. S5 F) indicated a high proportion of each reaction in the forward direction and thus the glycolysis pathway was enzymatically highly efficient. In addition, we calculated the enzyme cost<sup>62</sup> of each reaction from their enzyme turnover rate,  $k_{cat}$  (supplementary S3 Table). The turnover rates were calculated using DLKcat for CTL<sup>63</sup>. From the enzyme cost analysis (Fig 5. G), *pgm* has the lowest enzyme costs compared to the other reactions of glycolysis and *pgk* has the second lowest enzyme costs, indicating a low carbon investment to catalyze those reactions compared to other glycolysis reactions. To ensure the

observed thermodynamics driving force and enzyme cost implications are not an artifact of reduced carbon usage efficiency (CUE) of CTL, we calculated CUE for all the conditions<sup>64</sup> and found that CUE remained 25% for both unstressed and stressed conditions (Fig 5. H). Therefore, it is evident that the thermodynamics and enzyme cost of *p<sub>g</sub>m* could potentially dictate its regulatory role in CTL persistence, rather than its CUE.



**Fig. 6** *In vitro* experiments support model prediction that *pgm* plays a major role in pushing *C. trachomatis* to persistence. (A) Tryptophan deprivation for 6 hours resulted in a small reduction in inclusion size. However, with *pgm* knockdown, the decline became more severe in tryptophan-limited and knockdown-inducing settings. (B) Inclusion reduction became more evident in tryptophan-limited and knockdown-inducing settings, with the *pgm* knockdown showing a statistically significant difference in inclusion size, corroborating the modeling studies' findings.

### **Validation of thermodynamics and enzyme cost analysis through CRISPRi-based suppression and starvation experiment**

Using systems biology approaches, we found that *pgm* was regulated at the levels of transcription (from the MBA), thermodynamics driving force, and enzyme cost. However, other glycolysis enzymes, including *pgk*, were regulated only at the level of thermodynamic driving cost and enzyme cost. Therefore, we predicted a *pgm*-mutant would have greater impact on chlamydial growth compared to a *pgk*-mutant, as compared to the wild type.

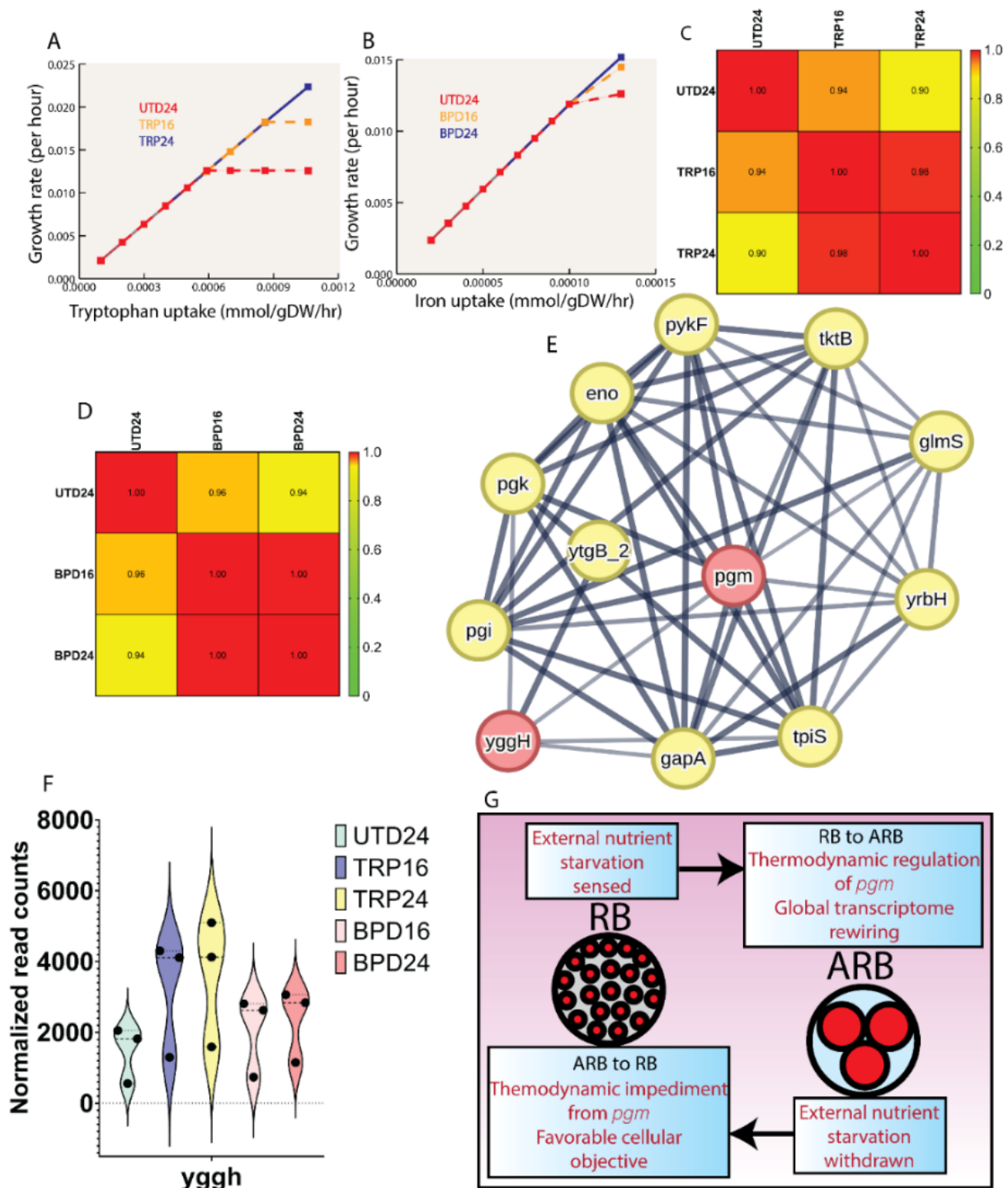
To test these predictions, *C. trachomatis* inducible knock-down transformants targeting *pgk* or *pgm* were created using CRISPRi as previously described<sup>14</sup>. In these transformants, the guide RNA (gRNA) that targets the promoter region of either *pgk* or *pgm* is constitutively expressed, while the expression of an inactive Cas9 endonuclease (dCas9) is induced with anhydrotetracycline (aTc). The RNA-DNA hybrid formed by the gRNA and genomic DNA is recognized by dCas9, where it binds and blocks RNA polymerase processivity. A control *C. trachomatis* was created by transforming *C. trachomatis* with the L9CRia plasmid that lacks gRNA. The knockdown strains and the empty-vector control were validated for dCas9 induction by aTc, reduction in expression of the target genes, chlamydial development, and bacterial

replication (Fig. S6 A, B, and C). In the absence of induction under normal growth conditions, all strains grew equally well (Fig. 6 A).

Computational modeling indicated the importance of *pgm* in chlamydial entry into persistence. To validate this prediction, the strains were grown in HeLa cells for 10 h, followed by a 4-h induction of knockdown by treatment with 5 nM aTc. The experimental groups were further subjected to either mock- or tryptophan-starvation for an additional 6 h, with aTc maintained in the growth media. At the end of the experiment (22 h post-infection), samples were fixed and processed for immunofluorescence staining of the chlamydial inclusions. Images were collected by confocal microscopy, and inclusion size measured by NIH ImageJ particle analysis plug-in. Inclusion size correlates with bacterial growth and replication, in that increase inclusion volume is necessary to accommodate increase in bacterial numbers. As shown in Fig. 6 A, all strains yielded similar inclusion sizes in tryptophan-replete media without induction of knockdown (i.e., +Trp, -aTc). Tryptophan starvation for 6 h led to a slight reduction in inclusion size, as expected. However, reduction became more pronounced in tryptophan-limited and knockdown-inducing conditions (i.e., -Trp, +aTc), with the *pgm* knockdown exhibiting statistically significant difference in inclusion size (Fig. 6 B), thus validating the results obtained from the modeling studies. Collectively, these data implicate *pgm* as a critical determinant of sensitivity to tryptophan starvation-mediated persistence.

### **Cellular objective of global stress response**

As CTL enter the persistence due to the lack of stringent action against the nutrient starvation by regulating *pgm*, it is also critical to investigate the cellular objective of this global transcriptome rewiring of CTL.



**Fig. 7** Reaction halting CTL to undergo secondary differentiation to EB. A) The growth rate vs. tryptophan uptake (UTD24, TRP16, TRP24); B) The growth rate vs. iron uptake (UTD24, BPD16, BPD24); C) Correlation matrix calculated among growth rates of UTD24, TRP16, TRP24 using 95% confidence interval and two tail test; D) Correlation matrix calculated among growth rates of UTD24, BPD16, BPD24 using 95% confidence interval and

two tail test; E) Protein interaction map for the *pgm* enzyme. Minimum required interaction score was 0.50 and top five interactions allowed for first shell; F) Normalized read counts of *yggH* in different conditions; G) Schematic of the mechanism proposed to enter persistence and factors that keep CTL in the persistence mode.

To gain insight into the significance of metabolic rewiring associated with tryptophan starvation, we plotted biomass growth rate against different tryptophan uptake rates. If we made more tryptophan available for uptake in the case of TRP24, then it could grow at a higher growth rate compared to TRP16 and UTD24 (Fig. 7 A). A similar result was obtained for BPD16 and BPD24 (Fig. 7 B). We also calculated the correlation matrix of growth patterns (Fig. 7 C and Fig. 7 D) under tryptophan and iron starvation conditions, and like the previous analysis (Fig. 4 I and Fig. S5 A), stress conditions were better correlated.

However, previous study<sup>65</sup> suggested that, with a shorter starvation timeline, CTL should better restore its growth and start differentiating to the RB upon availability of nutrients. In other words, it should be easier to restore the growth of CTL for 16 hours of nutrient starvation compared to 24 hours of nutrient starvation. Thus, we explored the model further to investigate this phenomenon. We again applied MBA in the TRP16 metabolic model. Surprisingly, we found that if *pgm* reaction flux was relaxed to its value obtained from TRP24 model (Fig. 5 A), the biomass of TRP16 and TRP24 became the same. A similar result was obtained for the BPD16 and BPD24. This analysis indicates that the cellular objective of the CTL global transcriptome rewiring under the persistence state is to reach such a cellular phenotype so that, when the stress is withdrawn, CTL can exit persistence immediately and re-enter the normal development cycle. CTL can do that by relaxing its regulation of *pgm*, which is critical. This final objective also supports the evolutionary selection of CTL, which is to maximize its own fitness function. The protein interaction network suggests that *pgm* has a strong interaction with

*yggH* (Fig. 7 E). *YggH* catalyzes the *S*-adenosyl-l-methionine-dependent formation of *N*<sup>7</sup>-methylguanosine at position 46 (m<sup>7</sup>G46) in tRNA and can impact the activity of tRNA<sup>66</sup>. It was previously reported that inhibition of tRNA synthesis can induce persistence in CTL<sup>67</sup>. The normalized read counts of *yggH* shows significantly higher level of transcript in all the stress conditions compared to the UTD24 (Fig. 7 F). Thus, the regulation of *pgm* may be mediated in part by the *yggH*. The proposed mechanism that led CTL to enter persistence is shown in Fig. 7 G. Quantitative proteomics from the literature<sup>68</sup> indicated that the ARB is primed for a burst in metabolic activity upon nutrient availability, whereas the RB is geared towards nutrient utilization, a rapid increase in cellular mass, and securing the resources for an impending transition into the EB form. In addition to confirming previously published findings, this study indicates that *pgm* may be the possible barrier to ARB-to-RB transition.

In this work, transcriptomics datasets were generated for untreated and nutrient starved CTL. We then used K-mean clustering on those transcriptomics data to identify core and stress-specific components of CTL transcriptome which led to the finding that CTL might have a global transcriptomic rewiring under different stress conditions. This global transcriptomics rewiring along with the absence of ppGpp lead to the lack of stringent action, which is the likely reason of CTL persistence. To further investigate how that global transcriptomic rewiring impacted the metabolism of CTL, we reconstructed its genome-scale model *iCTL278* and later contextualized it for normal and stress conditions with transcriptomics data. Using MBA on these contextualized *iCTL278*, we proposed *pgm* regulates the entry of CTL to persistence. Later, thermodynamics driving force, enzymatic cost, CRISPRi-based suppression, and subsequent starvation experiments supported this finding. Furthermore, we investigated the cellular objective of the persistence and found that priming itself for a metabolic burst upon availability of nutrient

dictates the nature of the persistence. MBA further implicated *pgm* preventing ARB to convert back to the RB. Overall, this study showed how a combined systems and synthetic biology approach can lead to a better understanding of CTL persistence. Future research effort will be geared towards studying the interaction between CTL and different host epithelial cells (e.g., endocervical and endovaginal epithelial cell) and similar to this study, finding the metabolic signatures influencing the CTL infection in these cells.

## **MATERIALS AND METHODS**

### **Cell lines**

Human female cervical epithelial adenocarcinoma HeLa cells (RRID: CVCL\_1276) were cultured at 37°C with 5% atmospheric CO<sub>2</sub> in Dulbecco's Modified Eagle Medium (DMEM; Gibco, Thermo Fisher Scientific) supplemented with 10 µg/mL gentamicin, 2 mM l-glutamine, and 10% (vol/vol) filter sterilized fetal bovine serum (FBS). For all experiments, HeLa cells were cultured between passage numbers 3 and 15. HeLa cells were originally authenticated by ATCC via STR profiling and isoenzyme analysis per ATCC specifications.

### **Bacterial strains**

*Chlamydia trachomatis* serovar L2 was originally obtained from Dr. Ted Hackstadt (Rocky Mountain National Laboratory, NIAID). Chlamydial EBs were isolated from infected HeLa cells at 36 to 40 hpi and purified by density gradient centrifugation essentially as described<sup>69</sup>. For infections, at 80 to 90% confluence, HeLa cells were first washed with Hanks Buffered Saline Solution (HBSS; Gibco, Thermo Fisher Scientific) and ice-cold inoculum prepared in HBSS at the indicated multiplicity of infection was overlaid onto the cell monolayer. To synchronize the infection, inoculated cells were then centrifuged for 15 min at 500xRCF, 4°C in an Eppendorf



5810 R tabletop centrifuge with an A-4-81 rotor. The inoculum was then aspirated and pre-warmed DMEM (or relevant media with treatment supplementation) was added to the cells. Infected cultures were then returned to the tissue culture incubator until the indicated time post-infection.

### **Treatment conditions and induction of knockdown of *pgk* and *pgm* transcription by CRISPRi**

HeLa cells were infected at a multiplicity of infection (moi) of 0.5 (t=0 hour post-infection (hpi)), and were maintained in complete DMEM for 14 h. The infected cells were then exposed to anhydrous tetracycline (aTc) at 5 nM for 4 h to induce knockdown of expression (t=14 hpi). Tryptophan depletion was performed at t=18 hpi by first washing cells with HBSS and then replacing complete DMEM with tryptophan-depleted DMEM-F12 (U.S. Biological Life Sciences). Treated cells were then returned to the tissue culture incubator for the remainder of the experimental time course. At t=22 hpi, samples were collected and fixed with freshly prepared 4% paraformaldehyde and permeabilized for 10 min with 0.1% Triton X-100 in PBS. Samples were immunostained with mouse antibody against the *C. trachomatis* major outer membrane protein (1:1000 dilution). Samples were incubated at 4C overnight with constant rocking. The next day, the primary antibody solution was removed and samples rinsed 3x with 1x PBS, and incubated with 1:1000 dilution of goat anti-mouse IgG conjugated with Alexa-488 for 1 h. Samples were rinsed and visualized by fluorescence microscopy. Images were processed and inclusion size measured using NIH ImageJ.

### **Genome-scale metabolic model reconstruction of CTL**

The genome-scale metabolic network reconstruction was based on available genome information of the model strain *C. trachomatis* L2, according to the ModelSEED databases<sup>70</sup>, BRENDA<sup>71</sup>, and available literature. Additional reactions were added to the model based on the previously published genome-scale metabolic model of CTL<sup>29</sup>, experimental data<sup>8</sup>, and previously published literature on metabolic traits of CTL. Gap filling for pathways in our model were first conducted in Kbase<sup>40</sup> and then by OptFill<sup>42</sup> based on the complete media, as complete media was used in the experimental settings as well to grow CTL *in vitro*. In total, 95 reactions were added to the draft model to fill the gaps. The model was further checked for elemental mass balance and MEMOTE reports confirms that all the reactions in the model are mass balanced. GPR for all the reactions were manually curated from the KEGG database. As a result, 525 out of 692 reactions has GPR relationship. We used GAMS platform along with CPLEX solver for solving all the optimization problems. NEOS server can be used to run GAMS codes without having to buy the license. Details procedure of running GAMS codes in NEOS server can be found in the literature<sup>72</sup>. However, for the convenience of COBRAPy users, SBML version of the *iCTL278* is also provided.

### **Flux sampling and tSNE plot of *iCTL278***

Flux sampling was performed using the ACHR sampling toll of COBRAPy. For tSNE, plot we used a perplexity of 45 and random state of 42. Clusters from the tSNE plot was generated using K-mean clustering algorithm.

### **Parsimonious flux balance analysis**

pFBA<sup>20</sup> is constrained based optimization technique to model GSMs. The pseudo-steady state mass balance in pFBA is represented by a stoichiometric matrix, where the columns represent

metabolites, and the rows represent reactions. For each reaction, upper and lower bounds is imposed based on thermodynamic information. pFBA provides the flux value for each reaction in the model according by solving the following optimization problem:

$$\min \sum_{j \in J} v_j - 0.0001 v_{biomass}$$

*Subject to:*

$$\sum_{j \in J} S_{ij} v_j = 0, \quad \forall i \in I \quad [1]$$

$$a_j \leq v_j \leq b_j \quad [2]$$

In this formulation,  $I$  is the set of metabolites and  $J$  is the set of reactions in the model.  $S_{ij}$  is the stoichiometric matrix with  $i$  indicating metabolites and  $j$  indicating reactions, and  $v_j$  is the flux value of each reaction. The objective function,  $v_{biomass}$ , is the proxy of the growth rate of an individual cell.  $a_j$  and  $b_j$  are the lower and upper bounds of flux values for each reaction. For forward reactions, the highest possible bounds were 0 mmol/gDW/h to 1000 mmol/gDW/h. For the reversible reactions, the highest possible bounds were -1000 mmol/gDW/h to 1000 mmol/gDW/h.

### Flux variability analysis

Result from FBA may include degenerate optimal solutions, this flux variability analysis (FVA)<sup>19</sup> was used to find out the alternate flux distributions. The formulation is the following:

$$\max/\min v_j$$

*Subject to:*

$$\sum_{j \in J} S_{ij} v_j = 0, \quad \forall i \in I \quad [3]$$

$$v_{biomass} = v_{biomass}^{\max} \quad [4]$$

$$a_j \leq v_j \leq b_j, \quad \forall j \in J \quad [5]$$

FVA maximizes and minimizes each of the reaction fluxes subject to the pseudo-steady mass balance, fixing biomass growth rate for a specific condition, and upper and lower bounds on reaction fluxes. In this manuscript, all the FVA analysis resulted in very tight bounds (changes can only be observed after two decimal place).

### **E-Flux algorithm**

E-Flux is an extension of FBA/pFBA that uses transcriptomic data to further constrain the feasible space based on the transcriptomics data<sup>37</sup>. The E-flux algorithm involved solving the following linear optimization problem:

$$\min \sum_{j \in J} |v_j| - 0.0001 v_{biomass}$$

*Subject to:*

$$\sum_{j \in J} S_{ij} v_j = 0, \quad \forall i \in I \quad [6]$$
$$a_j \leq v_j \leq b_j \quad [7]$$

where  $a_j$  and  $b_j$  are the minimum and maximum allowed fluxes through reaction  $j$ , based on the transcriptomics data. The E-Flux method calculates the upper bound,  $b_j$ , for the  $j^{th}$  reaction according to the following function of the gene expression:

$$b_j = (\text{expression level of genes associated with reaction } j) \quad [8]$$

In this manuscript,  $b_j$  is the exact level of each reaction that was calculated through its GPR association (i.e., based on “OR” (addition of gene expressions), and “AND” (minimum of gene expressions) relation). If the reaction catalyzed by the corresponding enzyme was reversible then  $a_j = -b_j$ , otherwise  $a_j = 0$ .

### **Flux sum analysis**

Metabolite pool size of ATP in the validation section was determined based on the flux-sum analysis (FSA) method<sup>32</sup>. The flux-sum is a measure of the amount of flux through the reactions associated with either the production or consumption of the metabolite. The range of the flux-sum can be calculated as follows as follows:

$$\left[ \begin{array}{l} \text{Max/Min } 0.5 \sum_{j=1}^m |S_{ij} v_j| \\ \text{Subject to:} \\ \sum_{j=1}^m S_{ij} v_j = 0, i \in I \quad [9] \\ a_j \leq v_j \leq b_j \quad [10] \\ v_{biomass} = v_{biomass}^{\max} \quad [11] \end{array} \right], \forall i \in I$$

Here, set  $I$  represents the set of metabolites for which flux sum will be calculated. By linearizing the objective function, the resulting formulation became a mixed-integer linear programming problem. Therefore, the basic idea was to determine the range of the flux-sum of a metabolite under a given condition by fixing the maximum biomass growth of that condition (equation 11).

### **K-mean clustering**

K-mean clustering algorithm was used to classify different genes into different clusters. Number of clusters was determined using the Elbow method (Fig. S7). The whole K-mean clustering was implemented in Python, using numpy, pandas, and sklearn modules. Different plots for the K-mean clustering was generated using matplotlib module. Default setting of K-mean clustering, mentioned in the sklearn, was not changed in this study.

### **Metabolic bottleneck analysis**

To determine the metabolic bottleneck in a GSM metabolic bottleneck analysis<sup>39</sup> was used. The formulation is as following.

$$\left[ \begin{array}{l} \text{Max } v_{biomass} \\ \text{Subject to:} \\ \sum_{j=1}^m S_{ij} v_j = 0, i \in I \quad [12] \\ a_j \leq v_j \leq b_j, j/\{j'\} \in J \quad [13] \\ v_{j',min} \leq v_{j'} \leq v_{j',max}, j' \in J \quad [14] \end{array} \right], \forall j' \in J$$

Here  $a_j$  is the lower bound reaction  $v_j$  and  $b_j$  is the upper bound of reaction  $v_j$ . Both  $a_j$  and  $b_j$  were calculated from the transcriptomics data and gene-protein-reaction association.  $v_{j',min}$  is the expanded lower bound of the reaction  $j'$  and  $v_{j',max}$  is the expanded upper bound of the reaction  $j'$ . In this case, we set  $v_{j',min} = -1000 \frac{mmol}{gDW.hr}$  and  $v_{j',max} = 1000 \frac{mmol}{gDW.hr}$ . We solved the optimization problem by maximizing the biomass  $v_{biomass}$  for the new expanded flux space of each reaction  $j'$  in an iterative manner and then recorded the biomass growth rate. From this biomass growth rate collections, we can check for which  $j'$  biomass growth rate increased significantly. Then that  $j'$  can be considered as the metabolic bottleneck of a given metabolic network.

### Max/Min driving force analysis

To find out the thermodynamic driving force and thermodynamic bottleneck of a given pathway, we used max/min driving force analysis (MDF) from the literature<sup>60</sup>. In this method, we maximized the driving force of each reaction in a given pathway within the biologically relevant concentration and found the maximum possible driving force of a pathway. The formulation of MDF analysis is given below:

$$\begin{aligned}
 & \max B \\
 & \text{Subject to:} \\
 -G & \geq B & [15] \\
 G & = (G^0 + RT \cdot S^T \cdot x) & [16] \\
 \ln(C_{\min}) & \leq x \leq \ln(C_{\max}) & [17] \\
 x_{ATP} & \geq 10 x_{ADP} & [18] \\
 x_{NADH} & \geq 0.1 x_{NAD} & [19] \\
 x_{ADP} & = 0.4 & [20] \\
 x_{NAD} & = 6.5 & [21] \\
 x_{3PGA} & = 13.0 & [22]
 \end{aligned}$$

$G^0$  is the standard Gibbs free energy,  $R$  is the gas constant,  $T$  is the temperature, and  $x$  and  $C$  indicates concentration. This analysis is particularly useful when metabolomics information is not fully available for a given pathway. Thus biologically relevant metabolite concentration range can be used to infer information about a pathway, whether it will be thermodynamically feasible or not. Furthermore, this analysis will indicate reaction(s) with the highest and lowest thermodynamic driving forces and will also provide concentration of metabolites to support the maximum pathway driving force. Here we fixed the ATP/ADP and NADH/NAD ratio as mentioned in the original MDF article<sup>60</sup>. Besides concentrations (mM) of ADP, NAD, and 3-PGA were fixed in the MDF optimization problem based on literature evidence<sup>73,74</sup>. Shadow price of different constraints were calculated using the built-in features of GAMS. The temperature in the optimization problem was set to 37<sup>0</sup> C, which is similar to the temperature of the growth culture. Ratio of forward to backward reaction fluxes for each of the glycolysis reaction was calculated using the following equation.

$$\frac{v^+ - v^-}{v^+ + v^-} = \frac{e^{-\frac{\Delta G}{RT}} - 1}{e^{-\frac{\Delta G}{RT}} + 1} \quad [23]$$

Here,  $\Delta G$  was calculated from the MDF analysis. Also,  $v^+$  and  $v^-$  indicate reaction flux in forward and backward directions respectively.

## Enzyme cost calculation

To find out enzyme synthesis costs for different enzymes of a given pathway, we used the following equation from the literature<sup>62</sup> assuming all the enzymes are fully saturated.

$$v = E \cdot k_{cat} \left( 1 - e^{-\frac{\Delta G}{RT}} \right) \quad [24]$$

Here,  $v$  is the reaction flux obtained from the contextualized GSM,  $E$  is the enzyme cost,  $k_{cat}$  is the enzyme turnover rate,  $\Delta G$  is the gibbs free energy obtained from the MDF analysis,  $R$  is the gas constant, and  $T$  is the temperature. Relative enzyme cost was calculated using the following equation.

$$E_{i,relative} = \frac{E_i}{E_{lowest}} \quad [25]$$

Here,  $E_{i,relative}$  is the relative cost of an enzyme in a given pathway,  $E_i$  is the actual cost of an enzyme in a given pathway, and  $E_{lowest}$  is the lowest cost of an enzyme in a given pathway.

## Simulation platform

The General Algebraic Modeling System (GAMS) version 24.7.4 with IBM CPLEX solver was used to run pFBA and FVA, E-Flux, FSA, MBA, and MDF algorithm on the model. Each of the algorithm was scripted in GAMS and then run on a Linux-based high-performance cluster computing system at the University of Nebraska-Lincoln. Furthermore, flux sampling, tSNE plot, and K-mean clustering analysis of transcriptomics data was performed in Python using an Intel(R) Core(TM) i5-8250U CPU @ 1.60GHz HP laptop with 8.00 GB of RAM and 64-bit operation with Windows 11 Home operating system.

## SUPPORTING INFORMATION



S1 Text. MEMOTE report of *iCTL278*.

S1-S7 Figure. Supporting figures.

S1 Table. Details about clusters from flux sampling.

S2 Table. Details about each clusters.

S3 Table. Details of  $k_{cat}$  for each reaction in glycolysis.

## **ACKNOWLEDGMENTS**

RS gratefully acknowledge the funding support from National Science Foundation (NSF) CAREER grant (1943310), National Institute of Health (NIH) R35 MIRA grant (5R35GM143009), and Nebraska Collaboration Initiative grant. RC, SO, and EAR acknowledge AI132406 from the NIH.

## **AUTHOR CONTRIBUTION**

Conceptualization: Rajib Saha, Rey A. Carabeo

Data curation: Niaz Bahar Chowdhury

Formal analysis: Niaz Bahar Chowdhury

Funding acquisition: Rajib Saha, Rey A. Carabeo, Elizabeth A. Rucks, Scot P. Ouellette

Methodology: Niaz Bahar Chowdhury, Nick Pokorzynski

Project administration: Rajib Saha, Rey A. Carabeo, Elizabeth A. Rucks, Scot P. Ouellette

Software: Niaz Bahar Chowdhury

Supervision: Rajib Saha, Rey A. Carabeo

Validation: Niaz Bahar Chowdhury

Visualization: Niaz Bahar Chowdhury

Writing – original draft: Niaz Bahar Chowdhury

Writing – review & editing: Rajib Saha, Rey A. Carabeo

## **DATA AVAILABILITY**

The data that support for the findings of this study can be found in the related cited articles and/or in the supplementary data. All the codes used to generate these results can be accessed in the GitHub repository ([https://github.com/ssbio/c\\_trachomatis](https://github.com/ssbio/c_trachomatis)).

## **REFERENCES**

1. S., W. S., Evelyn, M., Aikaterini, A., Julie, L. & M., L. I. Chlamydia trachomatis: the Persistent Pathogen. *Clin. Vaccine Immunol.* **24**, e00203-17 (2017).
2. Wilson, M. & Wilson, P. J. K. Chlamydia BT - Close Encounters of the Microbial Kind: Everything You Need to Know About Common Infections. in (eds. Wilson, M. & Wilson, P. J. K.) 379–390 (Springer International Publishing, 2021). doi:10.1007/978-3-030-56978-5\_27.
3. Elwell, C., Mirrashidi, K. & Engel, J. Chlamydia cell biology and pathogenesis. *Nat. Rev. Microbiol.* **14**, 385–400 (2016).
4. Bastidas, R. J., Elwell, C. A., Engel, J. N. & Valdivia, R. H. Chlamydial Intracellular Survival Strategies. *Cold Spring Harb. Perspect. Med.* **3**, a010256 (2013).
5. Chavda, V. P., Pandya, A., Kypreos, E., Patravale, V. & Apostolopoulos, V. *Chlamydia*

- trachomatis*: quest for an eye-opening vaccine breakthrough. *Expert Rev. Vaccines* **21**, 771–781 (2022).
6. Riffaud, C. M., Rucks, E. A. & Ouellette, S. P. Persistence of obligate intracellular pathogens: alternative strategies to overcome host-specific stresses. *Front. Cell. Infect. Microbiol.* **13**, (2023).
  7. Mpiga, P. & Ravaoarino, M. Chlamydia trachomatis persistence: An update. *Microbiol. Res.* **161**, 9–19 (2006).
  8. Østergaard, O. *et al.* Quantitative Protein Profiling of Chlamydia trachomatis Growth Forms Reveals Defense Strategies Against Tryptophan Starvation. *Mol. & Cell. Proteomics* **15**, 3540–3550 (2016).
  9. Beatty, W. L., Byrne, G. I. & Morrison, R. P. Morphologic and antigenic characterization of interferon gamma-mediated persistent Chlamydia trachomatis infection in vitro. *Proc. Natl. Acad. Sci.* **90**, 3998–4002 (1993).
  10. Thompson, C. & Carabeo, R. An Optimal Method of Iron Starvation of the Obligate Intracellular Pathogen, Chlamydia Trachomatis. *Front. Microbiol.* **2**, (2011).
  11. Gracey, E., Lin, A., Akram, A., Chiu, B. & Inman, R. D. Intracellular Survival and Persistence of Chlamydia muridarum Is Determined by Macrophage Polarization. *PLoS One* **8**, e69421 (2013).
  12. Ouellette, S. P. *et al.* Global transcriptional upregulation in the absence of increased translation in Chlamydia during IFN $\gamma$ -mediated host cell tryptophan starvation. *Mol. Microbiol.* **62**, 1387–1401 (2006).

13. P., O. S., J., R. K. & A., R. E. Tryptophan Codon-Dependent Transcription in *Chlamydia pneumoniae* during Gamma Interferon-Mediated Tryptophan Limitation. *Infect. Immun.* **84**, 2703–2713 (2016).
14. P., O. S., D., H. N., A., W. N., L., H. A. & S., C. M. Codon-Dependent Transcriptional Changes in Response to Tryptophan Limitation in the Tryptophan Auxotrophic Pathogens *Chlamydia trachomatis* and *Streptococcus pyogenes*. *mSystems* **6**, e01269-21 (2021).
15. D., P. N., R., A. M. & A., C. R. Host Cell Amplification of Nutritional Stress Contributes To Persistence in *Chlamydia trachomatis*. *MBio* **13**, e02719-22 (2022).
16. Sertbas, M. & Ulgen, K. O. Genome-Scale Metabolic Modeling for Unraveling Molecular Mechanisms of High Threat Pathogens . *Frontiers in Cell and Developmental Biology* vol. 8 (2020).
17. Alsiyabi, A., Chowdhury, N. B., Long, D. & Saha, R. Enhancing in silico strain design predictions through next generation metabolic modeling approaches. *Biotechnol. Adv.* 107806 (2021) doi:10.1016/j.biotechadv.2021.107806.
18. Orth, J. D., Thiele, I. & Palsson, B. O. What is flux balance analysis? *Nat. Biotechnol.* **28**, 245–248 (2010).
19. Mahadevan, R. & Schilling, C. H. The effects of alternate optimal solutions in constraint-based genome-scale metabolic models. *Metab. Eng.* **5**, 264–276 (2003).
20. Lewis, N. E. *et al.* Omic data from evolved *E. coli* are consistent with computed optimal growth from genome-scale models. *Mol. Syst. Biol.* **6**, 390 (2010).
21. López-Agudelo, V. A. *et al.* A systematic evaluation of *Mycobacterium tuberculosis*

- Genome-Scale Metabolic Networks. *PLOS Comput. Biol.* **16**, e1007533 (2020).
22. Norsigian, C. J., Kavvas, E., Seif, Y., Palsson, B. O. & Monk, J. M. iCN718, an Updated and Improved Genome-Scale Metabolic Network Reconstruction of *Acinetobacter baumannii* AYE . *Frontiers in Genetics* vol. 9 (2018).
  23. Yu-Chieh, L. *et al.* An Experimentally Validated Genome-Scale Metabolic Reconstruction of *Klebsiella pneumoniae* MGH 78578, iYL1228 . *J. Bacteriol.* **193**, 1710–1717 (2011).
  24. Ines, T., D., V. T., D., P. N. & Ø., P. B. Expanded Metabolic Reconstruction of *Helicobacter pylori* (iIT341 GSM/GPR): an In Silico Genome-Scale Characterization of Single- and Double-Deletion Mutants. *J. Bacteriol.* **187**, 5818–5830 (2005).
  25. Zhu, Y. *et al.* Genome-scale metabolic modeling of responses to polymyxins in *Pseudomonas aeruginosa*. *Gigascience* **7**, giy021 (2018).
  26. Mazharul Islam, M. *et al.* An integrated computational and experimental study to investigate *Staphylococcus aureus* metabolism. *npj Syst. Biol. Appl.* **6**, 3 (2020).
  27. Hartman, H. B. *et al.* Identification of potential drug targets in *Salmonella enterica* sv. Typhimurium using metabolic modelling and experimental validation. *Microbiology* **160**, 1252–1266 (2014).
  28. Metris, A., Reuter, M., Gaskin, D. J. H., Baranyi, J. & van Vliet, A. H. M. In vivo and in silico determination of essential genes of *Campylobacter jejuni*. *BMC Genomics* **12**, 535 (2011).
  29. Yang, M., Rajeeve, K., Rudel, T. & Dandekar, T. Comprehensive Flux Modeling of *Chlamydia trachomatis* Proteome and qRT-PCR Data Indicate Biphasic Metabolic

Differences Between Elementary Bodies and Reticulate Bodies During Infection .  
*Frontiers in Microbiology* vol. 10 (2019).

30. Schellenberger, J., Lewis, N. E. & Palsson, B. Ø. Elimination of Thermodynamically Infeasible Loops in Steady-State Metabolic Models. *Biophys. J.* **100**, 544–553 (2011).
31. Reed, J. L. & Palsson, B. Ø. Genome-Scale In Silico Models of *E. coli* Have Multiple Equivalent Phenotypic States: Assessment of Correlated Reaction Subsets That Comprise Network States. *Genome Res.* **14**, 1797–1805 (2004).
32. Chowdhury, N. B. *et al.* Dissecting the metabolic reprogramming of maize root under nitrogen-deficient stress conditions. *J. Exp. Bot.* **73**, 275–291 (2022).
33. Becker, S. A. & Palsson, B. O. Context-specific metabolic networks are consistent with experiments. *PLoS Comput. Biol.* **4**, (2008).
34. Zur, H., Ruppin, E. & Shlomi, T. iMAT: An integrative metabolic analysis tool. *Bioinformatics* **26**, 3140–3142 (2010).
35. Jensen, P. A. & Papin, J. A. Functional integration of a metabolic network model and expression data without arbitrary thresholding. *Bioinformatics* **27**, 541–547 (2011).
36. Jenior, M. L., Moutinho Jr., T. J., Dougherty, B. V & Papin, J. A. Transcriptome-guided parsimonious flux analysis improves predictions with metabolic networks in complex environments. *PLOS Comput. Biol.* **16**, e1007099 (2020).
37. Colijn, C. *et al.* Interpreting expression data with metabolic flux models: Predicting Mycobacterium tuberculosis mycolic acid production. *PLoS Comput. Biol.* **5**, (2009).
38. Chandrasekaran, S. & Price, N. D. Probabilistic integrative modeling of genome-scale

- metabolic and regulatory networks in *Escherichia coli* and *Mycobacterium tuberculosis*. *Proc. Natl. Acad. Sci. U. S. A.* **107**, 17845–17850 (2010).
39. Chowdhury, N. B. *et al.* A multi-organ maize metabolic model connects temperature stress with energy production and reducing power generation. *iScience* **26**, (2023).
  40. Arkin, A. P. *et al.* KBase: The United States Department of Energy Systems Biology Knowledgebase. *Nat. Biotechnol.* **36**, 566–569 (2018).
  41. Mehlitz, A. *et al.* Metabolic adaptation of *Chlamydia trachomatis* to mammalian host cells. *Mol. Microbiol.* **103**, 1004–1019 (2017).
  42. Schroeder, W. L. & Saha, R. OptFill: A Tool for Infeasible Cycle-Free Gapfilling of Stoichiometric Metabolic Models. *iScience* **23**, 100783 (2020).
  43. Lieven, C. *et al.* MEMOTE for standardized genome-scale metabolic model testing. *Nat. Biotechnol.* **38**, 272–276 (2020).
  44. J., T. *et al.* Two Nucleotide Transport Proteins in *Chlamydia trachomatis*, One for Net Nucleoside Triphosphate Uptake and the Other for Transport of Energy. *J. Bacteriol.* **181**, 1196–1202 (1999).
  45. Darzi, Y., Letunic, I., Bork, P. & Yamada, T. iPath3.0: interactive pathways explorer v3. *Nucleic Acids Res.* **46**, W510–W513 (2018).
  46. Hari, A. & Lobo, D. Fluxer: a web application to compute, analyze and visualize genome-scale metabolic flux networks. *Nucleic Acids Res.* **48**, W427–W435 (2020).
  47. Belland, R. J. *et al.* Genomic transcriptional profiling of the developmental cycle of *Chlamydia trachomatis*. *Proc. Natl. Acad. Sci.* **100**, 8478–8483 (2003).

48. Ouellette, S. P. *et al.* Localized cardiolipin synthesis is required for the assembly of MreB during the polarized cell division of *Chlamydia trachomatis*. *PLOS Pathog.* **18**, e1010836 (2022).
49. Yao, J. *et al.* Type II fatty acid synthesis is essential for the replication of *Chlamydia trachomatis*. *J. Biol. Chem.* **289**, 22365–22376 (2014).
50. Fehlner-Gardiner, C. *et al.* Molecular basis defining human chlamydia trachomatis tissue tropism. *J. Biol. Chem.* **277**, 26893–26903 (2002).
51. Caldwell, H. D. *et al.* Polymorphisms in *Chlamydia trachomatis* tryptophan synthase genes differentiate between genital and ocular isolates. *J. Clin. Invest.* **111**, 1757–1769 (2003).
52. Sanghamitra, M. *et al.* Protein Expression Profiles of *Chlamydia pneumoniae* in Models of Persistence versus Those of Heat Shock Stress Response. *Infect. Immun.* **74**, 3853–3863 (2006).
53. Rosario, C. J. & Tan, M. The early gene product EUO is a transcriptional repressor that selectively regulates promoters of *Chlamydia* late genes. *Mol. Microbiol.* **84**, 1097–1107 (2012).
54. Wenwei, L. *et al.* Transcriptional Profiling of *Mycobacterium tuberculosis* Exposed to In Vitro Lysosomal Stress. *Infect. Immun.* **84**, 2505–2523 (2016).
55. Dow, A. *et al.* Zinc limitation triggers anticipatory adaptations in *Mycobacterium tuberculosis*. *PLOS Pathog.* **17**, e1009570 (2021).
56. Jozefczuk, S. *et al.* Metabolomic and transcriptomic stress response of *Escherichia coli*.



- Mol. Syst. Biol.* **6**, 364 (2010).
57. Magnusson, L. U., Farewell, A. & Nyström, T. ppGpp: a global regulator in *Escherichia coli*. *Trends Microbiol.* **13**, 236–242 (2005).
  58. Chatterji, D. & Kumar Ojha, A. Revisiting the stringent response, ppGpp and starvation signaling. *Curr. Opin. Microbiol.* **4**, 160–165 (2001).
  59. Mittenhuber, G. Comparative genomics and evolution of genes encoding bacterial (p)ppGpp synthetases/hydrolases (the Rel, RelA and SpoT proteins). *J. Mol. Microbiol. Biotechnol.* **3**, 585–600 (2001).
  60. Noor, E. *et al.* Pathway Thermodynamics Highlights Kinetic Obstacles in Central Metabolism. *PLOS Comput. Biol.* **10**, e1003483 (2014).
  61. Schroeder, W. L., Harris, S. D. & Saha, R. Computation-Driven Analysis of Model Polyextremo-tolerant Fungus *Exophiala dermatitidis*: Defensive Pigment Metabolic Costs and Human Applications. *iScience* **23**, (2020).
  62. Noor, E., Flamholz, A., Liebermeister, W., Bar-Even, A. & Milo, R. A note on the kinetics of enzyme action: A decomposition that highlights thermodynamic effects. *FEBS Lett.* **587**, 2772–2777 (2013).
  63. Li, F. *et al.* Deep learning-based kcat prediction enables improved enzyme-constrained model reconstruction. *Nat. Catal.* **5**, 662–672 (2022).
  64. Saifuddin, M., Bhatnagar, J. M., Segrè, D. & Finzi, A. C. Microbial carbon use efficiency predicted from genome-scale metabolic models. *Nat. Commun.* **10**, 3568 (2019).
  65. J., B. A., R., W. M. & A., C. R. Genomewide Transcriptional Responses of Iron-Starved

- Chlamydia trachomatis Reveal Prioritization of Metabolic Precursor Synthesis over Protein Translation. *mSystems* **3**, 10.1128/msystems.00184-17 (2018).
66. S., D. B. L. G. *et al.* The *yggH* Gene of *Escherichia coli* Encodes a tRNA (m7G46) Methyltransferase. *J. Bacteriol.* **185**, 3238–3243 (2003).
67. D., H. N. & P., O. S. Inhibition of tRNA Synthetases Induces Persistence in Chlamydia. *Infect. Immun.* **88**, 10.1128/iai.00943-19 (2020).
68. Saka, H. A. *et al.* Quantitative proteomics reveals metabolic and pathogenic properties of Chlamydia trachomatis developmental forms. *Mol. Microbiol.* **82**, 1185–1203 (2011).
69. Caldwell, H. D., Kromhout, J. & Schachter, J. Purification and partial characterization of the major outer membrane protein of Chlamydia trachomatis. *Infect. Immun.* **31**, 1161–1176 (1981).
70. Seaver, S. M. D. *et al.* The ModelSEED Biochemistry Database for the integration of metabolic annotations and the reconstruction, comparison and analysis of metabolic models for plants, fungi and microbes. *Nucleic Acids Res.* **49**, D575–D588 (2021).
71. Chang, A. *et al.* BRENDA, the ELIXIR core data resource in 2021: new developments and updates. *Nucleic Acids Res.* **49**, D498–D508 (2021).
72. Niaz, B. C., Adil, A. & Rajib, S. Characterizing the Interplay of Rubisco and Nitrogenase Enzymes in Anaerobic-Photoheterotrophically Grown *Rhodospseudomonas palustris* CGA009 through a Genome-Scale Metabolic and Expression Model. *Microbiol. Spectr.* **10**, e01463-22 (2022).
73. Yongjin, Z. *et al.* Determining the Extremes of the Cellular NAD(H) Level by Using an

*Escherichia coli* NAD<sup>+</sup>-Auxotrophic Mutant . *Appl. Environ. Microbiol.* **77**, 6133–6140 (2011).

74. Takaine, M., Imamura, H. & Yoshida, S. High and stable ATP levels prevent aberrant intracellular protein aggregation in yeast. *Elife* **11**, e67659 (2022).