Figures and figure supplements

Essential metabolism for a minimal cell

Marian Breuer et al
Figure 1. Comparison of protein coding genes in the genomes of JCVI-syn3A (NCBI GenBank: https://www.ncbi.nlm.nih.gov/nuccore/CP016816.2 (Glass, 2017)), M. pneumoniae (NCBI GenBank: https://www.ncbi.nlm.nih.gov/nuccore/U00089.2 (Himmelreich et al., 2014)), and E. coli (NCBI GenBank: https://www.ncbi.nlm.nih.gov/nuccore/NC_012967.1 (Jeong et al., 2017)) with 452, 688, and 4637 coding genes, respectively. Each color represents a primary functional class, each contiguous shaded region corresponds to a secondary functional class, within each of the shaded regions the bold lines separate tertiary functional classes, finally each polygonal cell represents a single gene. The functional class hierarchy is presented in Supplementary file 1A. The ratio of metabolic to genetic information processing genes—0.67, 0.79, and 2.23 respectively—is smallest for JCVI-syn3A. The JCVI-syn3A genome contains both the smallest absolute number of genes of unclear function and the smallest percentage, 91 (20%), compared to M. pneumoniae with 311 (45%) and E. coli with 1780 (38%).

DOI: https://doi.org/10.7554/eLife.36842.003
Figure 2. Classification of gene essentiality from transposon insertion data using a Poisson mixture model for a representative region of the JCVI-syn3A genome. Coding regions are colored by their predicted class: red (essential), yellow (quasi-essential), blue (non-essential). Lavender regions denote RNA and light brown regions are pseudogenes. The distributions of transposon insertions in passage 1 and passage 4 are represented by yellow and dark green histograms, respectively (bin size of 50 bp). The overlap of the two histograms is highlighted in blue. When a common gene name is not available, the four-digit locus tag for JCVI-syn1.0 is used instead. Locus number identifiers with the (3A) suffix represent newly identified open reading frames in JCVI-syn3A which are missing from the JCVI-syn1.0 annotation. Asterisks mark genes with unknown functionality.

DOI: https://doi.org/10.7554/eLife.36842.004
Figure 2—figure supplement 1. Classification of gene essentiality from transposon insertion data using a Poisson mixture model for 0–275,000 bp. Coding regions are colored by their predicted class: red (essential), yellow (quasi-essential), blue (non-essential). Lavender regions denote RNA, light brown regions are pseudogenes, and green regions are markers used to construct and implant the genome. The distributions of transposon insertions in passage 1 and passage 4 are represented by yellow and dark green histograms respectively (bin size of 50 bp). The overlap of the two histograms is shown. Figure 2—figure supplement 1 continued on next page.
Figure 2—figure supplement 1 continued

highlighted in blue. When a common gene name is not available, the four-digit locus tag for JCVI-syn1.0 is used instead. Locus number identifiers with
the (3A) suffix represent newly identified open reading frames in JCVI-syn3A which are missing from the JCVI-syn1.0 annotation. Asterisks
mark genes with unknown functionality.

DOI: https://doi.org/10.7554/eLife.36842.005
Figure 2—figure supplement 2. Classification of gene essentiality from transposon insertion data using a Poisson mixture model for 275,000–543,379 bp. DOI: https://doi.org/10.7554/eLife.36842.006
Figure 2—figure supplement 3. Distribution of transposon insertion counts for P₁ (panel a) and P₄ (panel b) compared to the distribution inferred through the Poisson mixture model. To separate genes labeled ‘non-essential’ by the mixture model, but that showed a significant decrease in insertion counts from P₁ to P₄, k-means clustering was used on the ratios of transposon insertion rates in P₁ and P₄ for the genes labeled ‘non-essential’. Panel c shows how the genes were divided into two clusters such that the first cluster (blue) contains quasi-essential genes and the second contains truly non-essential genes.

DOI: https://doi.org/10.7554/eLife.36842.007
Figure 3. Essential, quasi-essential, and non-essential protein coding genes in JCVI-syn3A across four functional classes. (a) Distribution across genome (cell areas all equal). (b) Distribution across proteome (cell areas proportional to protein copy number in an average cell). Among non-essential proteins, the three most abundant ones are ftsZ/0522, the peptidase 0305 and 0538 (unclear function). A detailed breakdown of the JCVI-syn3A genome into these classes is available in Table 1.

DOI: https://doi.org/10.7554/eLife.36842.008
Figure 4. Biomass reaction equation for JCVI-syn3A. This reaction consumes biomass precursors (macromolecules, lipids, capsule, small molecules) (black) and consumes energy in the form of ATP (red) to produce biomass (blue). Values in parentheses are the stoichiometric coefficients in mmol compound per gram cellular dry weight (mmol gDW⁻¹). The macromolecular compositions are highlighted in green (stoichiometric coefficients within the macromolecule, unitless) and the compositions of lipids and small molecule pools are highlighted in gray (mmol gDW⁻¹). ATP expenses within green boxes denote total macromolecular synthesis costs (based on macromolecular fractions in the biomass) and the ATP expense in the main equation denotes the nonquantifiable part of the growth-associated maintenance cost (GAM; see Section ‘GAM/NGAM’).

DOI: https://doi.org/10.7554/eLife.36842.011
Figure 5. Overview of the metabolic reconstruction of JCVI-syn3A, drawn with Escher (King et al., 2015). Orange nodes represent metabolites, labeled by their short names in the model (black); the suffixes ‘_c’ and ‘_e’ denote cytoplasmic and extracellular compartments, respectively. For clarity, H₂O, H⁺, PPᵢ, and Pᵢ are generally omitted as reactants. Blue edges represent (enzymatic or spontaneous) reactions, labeled by reaction name (gray labels) and associated gene loci (gene-protein-reaction (GPR) rules, turquoise, omitting ‘MMSYN1_’ prefix). Blue parenthesized numbers denote reactants (products) which are consumed (produced) in stoichiometric quantities greater than one. In this map and subsequent maps, the following color scheme for highlighted reactions is used—blue: reaction based on new annotation, light green: reaction based on suggested annotation refinement, cyan: specific reaction assumed for generic annotation, light violet: non-enzymatic reaction, orange: reaction not accounted for by gene yet but supported by experimental evidence, and red: reaction included based on gap filling. Small boxes list metabolites that can be taken up (green boxes) or secreted (brown boxes) under physiological conditions.

DOI: https://doi.org/10.7554/eLife.36842.015
Figure 6. Central metabolism in JCVI-syn3A. Map components and labels as in Figure 5. Big arrows denote incoming or outgoing connections to other parts of the metabolic network. For context, the node representing glucose transport has been labeled explicitly and glycolysis has been highlighted in gray.

DOI: https://doi.org/10.7554/eLife.36842.012
Figure 6—figure supplement 1. Steady-state fluxes through central metabolism in JCVI-syn3A. Map components and labels as in Figure 5, with gene loci/gene-protein-reaction rules omitted. Numbers after reaction labels denote steady-state reaction fluxes in mmol gDW⁻¹ h⁻¹; edge color corresponds to the absolute value of the carried flux—gray to blue to purple to red, from low to high flux. For reversible reactions, the reaction progresses from the white to the filled arrowhead.
DOI: https://doi.org/10.7554/eLife.36842.013
Figure 7. Nucleotide metabolism in JCVI-syn3A. Map components and labels as in Figure 5.
DOI: https://doi.org/10.7554/eLife.36842.014
Figure 8. Apparent dead-end of dUMP/deoxyuridine and possible solutions. Internal metabolites are highlighted with cyan boxes, external ones with red boxes. Blue arrows denote reactions incorporated during model reconstruction—no reaction leads away from the dUMP/deoxyuridine pair. Red arrows denote hypothetical reactions that could possibly solve this dead-end. In the model, we have adopted the hypothetical CTP synthase reaction converting dUMP to dCMP (see also Figure 7, CTPSDUMP).

DOI: https://doi.org/10.7554/eLife.36842.016
Figure 9. Cofactor metabolism in JCVI-syn3A. Map components and labels as in Figure 5. DOI: https://doi.org/10.7554/eLife.36842.017
Figure 10. Lipid and capsule metabolism in JCVI-syn3A. Map components and labels as in Figure 5. DOI: https://doi.org/10.7554/eLife.36842.018
Figure 11. Macromolecule metabolism in JCVI-syn3A. Map components and labels as in Figure 5. The detailed (amino acid-specific) stoichiometry of the protein synthesis and degradation reactions can be found in Supplementary file 4. Protein synthesis reactions for the proteins explicitly included in the model (apo-ACP, dUTPase and PdhC) are analogous to the translation reaction shown and are therefore not included in the map.

DOI: https://doi.org/10.7554/eLife.36842.019
Amino acid metabolism in JCVI-syn3A. Map components and labels as in Figure 5. As amino acid metabolism in JCVI-syn3A constitutes sets of analogous reactions (for each amino acid or peptide), we use generic reactions in the upper right part of the map. The ABC importer Opp catalyzes tetrapeptide uptake reactions in the model ([amino acid]4abc in Supplementary file 4); the AA permeases (incl. GltP) catalyze amino acid proton symport reactions ([amino acid]t2[p]r in Supplementary file 4). The peptidases catalyze peptide hydrolysis reactions ([amino acid]4P in Supplementary file 4). The aminoacyl tRNA synthetases (‘aaRS’s’ in the map) catalyze charging of tRNAs ([amino acid]TRS in Supplementary file 4). Synthesis of Gln-tRNA<sub>Thr</sub> requires transamidation of initially mischarged Glu-tRNA<sub>Thr</sub> and the corresponding reactions are shown on the lower left. In the 5'-adenosylmethionine pathway on the lower right, we note that nucleic acid modification reactions (indicated by the edge labeled ‘DNA/RNA modification’) were not included in the model due to lack of sufficient information on kind and abundance of nucleic acid modifications in JCVI-syn3A.

DOI: https://doi.org/10.7554/eLife.36842.020
Figure 13. Ion transport reactions in JCVI-syn3A. Map components and labels as in Figure 5.
DOI: https://doi.org/10.7554/eLife.36842.021
Figure 14. Comparison of growth curves of JCVI-syn1.0 and JCVI-syn3A. JCVI-syn1.0 has a doubling time of 66 min (blue; ‘×’ markers), whereas JCVI-syn3A has a doubling time of 105 min (orange; ‘+’ markers). Doubling times ($t_d$) were calculated as described in Section ‘Materials and methods’, plotting fluorescence staining of cellular DNA vs. time, fitted by exponential regression curves. The regression curves for JCVI-syn1.0 and JCVI-syn3A have $R^2$ values of 0.9986 and 0.9976, respectively.

DOI: https://doi.org/10.7554/eLife.36842.022
Figure 15. Comparison of FBA steady-state fluxes $\nu$ to maximal fluxes $V_{\max}$ obtained from protein abundances and turnover numbers from BRENDA and the literature. Map components and labels as in Figure 5, with reaction highlighting and gene loci/gene-protein-reaction rules omitted. Each edge is colored according to the ratio between $V_{\max}$ and $\nu$: Blue indicates $V_{\max} > \nu$, red indicates $V_{\max} < \nu$ and green indicates that no $V_{\max}$ could be obtained (because of either missing turnover number or missing protein abundance, or because reaction is not enzymatic to begin with).

DOI: https://doi.org/10.7554/eLife.36842.023
**Figure 15—figure supplement 1.** Statistics of FBA steady-state fluxes $\nu$ vs. maximal fluxes $V_{\text{max}}$ comparison (see Figure 15). (A) Summary of $V_{\text{max}}$ vs. $\nu$ comparison over all 253 non-exchange reactions in the model. Red, blue, green: Meaning as in Figure 15. Green-striped: Subset of green set—

Figure 15—figure supplement 1 continued on next page
reactions without $V_{\text{max}}$ that pertain to transport, which usually do not have an EC number associated with them. Gray: Reactions with $v = 0$ in the FBA solution (thus $V_{\text{max}} > v$ always fulfilled). B: Histogram of $V_{\text{max}}/v$ over the blue and red subset in panel A.

DOI: https://doi.org/10.7554/eLife.36842.024
Figure 16. Partitioning of genes classified as essential, quasi-essential, and non-essential by transposon mutagenesis experiments into those which are in silico essential, in silico non-essential, and not modeled ('Non-metabolic'). All genes are included (i.e. also RNA genes and pseudogenes).

DOI: https://doi.org/10.7554/eLife.36842.026
Figure 16—figure supplement 1. In silico double-gene knockouts between genes that are non-essential in single-gene knockouts. Among the individually non-essential genes, a double knockout of the gene pair (0876, 0878) is the only lethal combination (red). This knockout corresponds to simultaneously removing both amino acid permeases, thus preventing cysteine uptake. Simultaneous knockout of the glutamate/aspartate permease gltP/0886 and any Opp gene (oppB/0165 through oppA/0169) is non-lethal in silico, as the model will under these circumstances produce glutamate through the hypothesized dUMP breakdown reaction CTPSDUMP and, to a lesser extent, through the reaction CTPS2 (both catalyzed by pyrG/0129). Glutamate production through pyrG/0129 is not expected to be able to meet cellular demands in vivo. If flux through CTPSDUMP is set to zero in the model, a double knockout of gltP/0886 and Opp becomes lethal in silico.

DOI: https://doi.org/10.7554/eLife.36842.027
Figure 17. Distributions of absolute protein abundances (number of molecules per average cell) in JCVI-syn3A. (a) Breakdown of the JCVI-syn3A proteome into functional classes. The area of each cell is proportional to its relative abundance. (b,c) Histograms of absolute protein abundances. (b) Absolute abundances of model-included metabolic proteins essential or non-essential in silico compared to all protein abundances. ‘Technical non-essential’ proteins are not included (see Section ‘In silico gene knockouts and mapping to in vivo essentiality’). (c) Absolute abundances for proteins classified by in vivo essentiality from transposon mutagenesis experiments. (d,e) Exceedence plots of absolute abundances for proteins classified by in silico or in vivo essentiality. The exceedence at a given protein abundance value x is the fraction of the protein set displaying an abundance higher than x. (d) Model-included proteins (classified by in silico essentiality) compared to all proteins. (e) Proteins classified by in vivo (transposon-based) essentiality.

DOI: https://doi.org/10.7554/eLife.36842.030
Appendix 1—figure 1. Thiamin diphosphate (ThDP) from the Mycoplasma hyorhinis Cypl crystal structure (pdb: 3EKI) overlaid onto the crystal structure of MG289 (pdb: 3MYU). (Structures aligned using STAMP (Russell and Barton, 1992) in VMD (Humphrey et al., 1996; Eargle et al., 2006).) a): Space-filling view, with MG289 in gray and ThDP in color. The pyrophosphate tail of ThDP from the Cypl structure would have an appropriate cavity in MG289 as well. b): Visualization of hydrogen bonds for the same alignment. All possible hydrogen bonds are shown between potential donor and acceptor heavy atoms within 3.5 Å or less of each other. Even in the absence of the residues involved in pyrophosphate binding in Cypl (Sippel et al., 2009), the alignment suggests other side group and backbone interactions could still allow for pyrophosphate binding.

DOI: https://doi.org/10.7554/eLife.36842.048
Appendix 1—figure 2. Sensitivity analysis of model doubling time with respect to model constraints. In each panel, the stated parameter was varied over the indicated range and the model doubling time calculated while keeping all other constraints constant. (A) Maximal glucose uptake. (B) Maximal acetate secretion. (C) ATPase ATP cost. (D) GAM ATP cost. (E) Protein degradation rate. (F) RNA degradation rate. (G) Imposed NADPH consumption. The blue circle marks the value used in the FBA model and resulting doubling time; the orange circle indicates the parameter that would yield the experimental doubling time. If there is no value of the parameter which would yield the experimental doubling time, a horizontal line is plotted.

DOI: https://doi.org/10.7554/eLife.36842.049