Dissecting the energy metabolism in Mycoplasma pneumoniae through genome-scale metabolic modeling

Supplementary Information

Judith A H Wodke[1,2,3], Jacek Puchałka[4,a], Maria Lluch-Senar[1,2], Josep Marcos[5,6], Eva Yus[1,2], Miguel Godinho[4,7], Ricardo Gutiérrez-Gallego[5,6], Vitor A P Martins dos Santos[4,8,9], Luis Serrano[1,2,10], Edda Klipp[3] & Tobias Maier[1,2]

1. EMBL/CRG Systems Biology Research Unit, Centre for Genomic Regulation (CRG), Barcelona, Spain
2. Universitat Pompeu Fabra, Barcelona, Spain
3. Theoretical Biophysics, Humboldt-Universität zu Berlin, Berlin, Germany
4. Synthetic and Systems Biology Group, Helmholtz Center for Infection Research (HZI), Braunschweig, Germany
5. Department of Experimental and Health Sciences, Pompeu Fabra University, Barcelona, Spain
6. Bio-analysis Group, Neuroscience Research Program, IMIM-Parc Salut Mar, Barcelona, Spain
7. Lifewizz Lda, Porto, Portugal
8. Systems and Synthetic Biology, Wageningen University, Wageningen, The Netherlands
9. LifeGlimmer GMBH, Berlin, Germany
10. Institució Catalana de Recerca i Estudis Avançats (ICREA), Barcelona, Spain

Correspondence to: Luis Serrano1,2,10 EMBL/CRG Systems Biology Research Unit Centre for Genomic Regulation (CRG), Dr Aiguader 88, Barcelona 08003, Spain. Tel.:+34 933 160 101; Fax:+34 933 160 099; Email: luis.serrano@crg.eu

Correspondence to: Tobias Maier1,2 EMBL/CRG Systems Biology Research Unit Centre for Genomic Regulation (CRG), Dr Aiguader 88, Barcelona 08003, Spain. Tel.:+34 933 160 101; Fax:+34 933 160 099; Email: tobias.maier@crg.eu

Correspondence to: Edda Klipp3 Theoretical Biophysics, Humboldt-Universität zu Berlin, Berlin, Germany. Tel.:+49 30 20939040; Fax:+49 30 20938813; Email: edda.klipp@biologie.hu-berlin.de

Table of Contents

Supplementary Information

Assignment of Reaction Reversibilities. For the definition of reaction reversibilities, we integrated consensus information obtained from the BRENDA enzyme information database (Scheer *et al*, 2011) with published reaction directions assigned to the metabolic network of *E. coli* (Fleming *et al*, 2009) and the metabolic map of *M. pneumoniae* (Yus *et al*, 2009) thus covering 65.1% of all model reactions (table S2). In case of contradicting information and no proven need for reversibility, reactions were set irreversible (accounts for 8.5% of all reactions, among them all tRNA-biosynthesis reactions as well as six lipid metabolism reactions). Note that the lipid metabolism is the least studied part of the metabolic network of *M. pneumoniae* and that a final assignment of reversibilities would need an extensive biochemical characterization of the enzymes involved. The reversibility of the NADH oxidase reaction (M017) has been changed along with the functional re-annotation of the NADH oxidase (MPN394, see below). Transport reactions were included in the network as required for producing all network components and account for 12.7% of all reactions. Model-specific reactions, such as protein and biomass synthesis reactions or exchange reactions, i.e. reactions to simulate the metabolite exchange with the environment, account for the missing 21.8% of the reactions and were included in the model as required to reproduce experimental findings.

Assessment of iJW145_reconstruct. By individually maximizing the production of each component of the metabolic network of *M. pneumoniae*, we tested the initial reconstruction for completeness and correct connectivity. Despite some transport reactions and the source/sink reactions (model specific reactions to exchange components with the environment) no further reactions had to be added to the network, i.e. it was possible to synthesize all metabolic intermediates if respective precursors could be taken up *in silico*.

***In Vivo* Assessment of Fatty Acid Profiles in *M. pneumoniae*.** The exact lipid composition of *M. pneumoniae* is not known and has been shown to vary depending on the fatty acids provided with the medium (McElhaney & Tourtellotte, 1969; Pollack *et al*, 1970; Rottem, 1980). To at least estimate the lipid composition, we

determined the composition of fatty acids incorporated into lipids of the cytoplasmic membrane and quantified the fatty acid spectrum present in both the *M. pneumoniae* cytosol and the surrounding growth medium (figure S3A). 80% of all fatty acids measured from the growth medium contain C_{18} chains. In contrast, for both the cytosolic and the cell membrane profiles, fatty acids with C_{16} chain length are the most prominent species (41% and 49% of all fatty acids, respectively) (figure S3A). In general, cytosolic and membrane fatty acid profiles are similar, suggesting regulated import of fatty acids but unregulated usage of available membrane building blocks. This finding is in agreement with results showing that the membrane composition of Mycoplasma highly depends on fatty acid availability (Rottem, 1980; Pollack *et al*, 1970a; McElhaney & Tourtellotte, 1969).

A comparison of fatty acid profiles before and after batch culture growth showed no significant depletion in the growth medium, hence the availability of fatty acids from the medium is not growth limiting (figure S3B). We detected minor rearrangements of the cytoplasmic fatty acid profile in samples taken at regular intervals during batch culture growth (figure S3A). At early growth stages longer chain fatty acids are more commonly found in the *M. pneumoniae* cytosol than towards the end of the growth course (17% and 3.5%, respectively, figure S3A).

Interestingly, intracellular and membrane incorporated fatty acids are specifically enriched in saturated fatty acids. While 59% of the growth medium fatty acids are singly or multiply unsaturated, the membrane and the cytosol only contain 29% and 30% unsaturated fatty acids, respectively. We reason that the cellular preference for saturated fatty acids aids membrane stability and integrity, since unsaturated (*cis*) fatty acids incorporated into lipids of the cell membrane disrupt its structure. Cell membrane stability is of particular importance for *M. pneumoniae*, as they lack a protective cell wall.

Definition of the average *M. pneumoniae* Biomass Composition. The biomass contains all macromolecules necessary to build up an entire cell in their respective quantities and a biomass function is the basis for *in silico* growth simulations. In case of *M. pneumoniae* the biomass composition differs significantly from that of other bacteria due to the reduced metabolism and the lack of a cell wall. We determined the major building blocks of an *M. pneumoniae* cell quantitatively based on sequence

information (DNA, RNA, proteome) and on GC-MS data (proteome (Maier *et al*, 2011), bases, amino acids (unpublished data), fatty acids (figure S3). Cofactors, as for example FAD, pyridoxal phosphate, or coenzyme A, were included in the biomass equation in a qualitative manner based on the essentiality of their precursors in defined medium experiments (Yus *et al*, 2009) or, in case of orthophosphate, based on findings in *E. coli* (Amin & Peterkofsky, 1995; Neidhardt, 1996) (table 1). We accomplished a sensitivity analysis to test the impact of the fraction size of protein, RNA, lipids, and cofactors (figure S4) confirming that assumptions for lipids and cofactors do not significantly impinge on the growth rate in *M. pneumoniae in silico*. The total list of biomass components and their respective considered quantities are found in the main text, table 1.

Due to technical limitations in static modeling approaches and missing experimental evidence, the following assumptions and technical tricks have been made:

1. DNA and all RNA types have to be synthesized once *de novo* which is accomplished by building artificial DNA and RNA molecules of 100 bases length each, displaying the natural GC content of the respective molecules. mRNA turnover is considered by setting a minimal constraint on the degradation reaction (0.0028 $mmol*g^{-1}*h^{-1}$) reflecting the minimum average mRNA half-life of about 1 minute (Maier *et al*, 2011). The sensitivity analysis for the biomass composition showed a major impact of the total amount of RNA on the growth rate, which was mainly attributed to the short half-life used. However, due to a lack of more exact experimental data, this choice prevents an underestimation of RNA costs. The costs for DNA degradation and repair are included in the maintenance costs and estimated based on literature data (compare main text).

2. Protein production is modeled via building of artificial protein molecules of 345 amino acids length (average protein length according to Yus *et al*, 2009) reflecting the amino acid composition found in the quantified proteome (Maier *et al*, 2011). An exception is the acyl-carrier-protein (ACP) that is modeled exactly, i.e. sequence-dependent, and also included in the biomass in the quantity it has been detected on the second day of growth (Maier *et al*, 2011)) as it plays an important role in CoA metabolism. *M. pneumoniae* is known to take

up peptides using the Opp transporter and amino acids using ABC systems to import and proton symport to export them (Yus *et al*, 2009). Since nothing is known about the specificity and activity of the proteases located on the surface of the *M. pneumoniae* cells and the up-take of peptides of varying amino acid composition is more complicated to simulate, we decided to consider only ABC transporters as the source of amino acids.

3. Based on the results of the fatty acid analysis and literature data (McElhaney & Tourtellotte, 1969; Pollack *et al*, 1970; Rottem, 1980) we made the following assumptions for lipids in our model:
   a. Lipids provide 20% of the total cell mass (assumption based on (Razin *et al*, 1963).
   b. The ”average” lipid has two fatty acid chains, one composed of 16 and the other of 18 carbon atoms. Fatty acids with carbon chains of 16 and 18 carbons length, respectively, have been found to be most abundant in *M. pneumoniae* contributing about 95% of the total fatty acids encountered (figure S3). In addition, the defined medium designed for *M. pneumoniae* (Yus *et al*, 2009) only contains fatty acids with 16 and 18 carbon atoms length. Cholesterol, despite having been shown to be essential for *M. pneumoniae* growth (Rottem *et al*, 1971; Johnson & Somerson, 1980) and found to be abundant in the cytosol (figure S3), has not been included explicitly, since no information about the specific up-take mechanism could be found.
   c. Phosphatidic acid (PAC) and glycolipids provide half of the total lipid mass each.
   d. Cardiolipin (CL) was not included in the biomass as the phospholipid branch (conversion of PAC into CL) is not essential in *M. genitalium* (Glass *et al*, 2006) and the functionality of this pathway has not been shown for *M. pneumoniae*. None of the intermediary metabolites nor CL have been detected (unpublished data), as neither have the two proteins assigned to this pathway (Maier *et al*, 2011).
   e. For glycolipids we assumed the attachment of three galactose and three glucose molecules to each diacylglycerol molecule to account for the diversity of glycolipids.

f. The total amount of lipids needed to duplicate one cell was calculated based on the molecular masses of the ”average” glycolipid and PAC, respectively.

4. Free amino acids and bases were determined by GC/MS-experiments (unpublished results) and included in respective amounts in the biomass function, accounting in total for about 1.5% of the cell mass.

5. Orthophosphate has been included based on the concentration measured in *E. coli* (Amin & Peterkofsky, 1995; Neidhardt, 1996).

6. Glucose 6-phosphate (G6P) has been quantitatively included in the biomass as model compound in order to account for all free organic metabolites (carbohydrates). To this end, we calculated the mass sum of all other defined biomass components, subtracted it from the total cell mass, and assigned the missing fraction to G6P (table 1).

7. 5- formyltetrahydrofolate (5fTHF) was included in the biomass equation for its supposed inhibitory function on serine hydroxymethyltransferase (SHMT, mpn576) (known from plants (Goyer *et al*, 2005; Kruger *et al*, 1999)).

8. The biomass composition reflects the composition of *M. pneumoniae* cells during the exponential growth phase, not considering significant alterations in their fractions at the beginning (lag phase) and the end (stationary phase) of a four days batch culture growth experiment.

9. Because ToBiN allows only integer values in reactions and the fraction of different compounds in biomass differs by several orders of magnitude, we rescaled the biomass mass units by a factor of 1.000.000 in order to avoid numerical problems of the solver.

Sensitivity Analysis. To further study the impact of the biomass composition on growth and maintenance costs, we accomplished a sensitivity analysis for the qualitatively included cofactors and for the major biomass fractions, namely for proteins, RNA, and lipids (table S4). We found that even a 10fold change of cellular reaction cofactors does not significantly influence doubling time (at no time more

than 1.4 min, *i.e.* < 0.1% change in doubling time) nor maintenance energy (~3 ATP*cell^{-1}*sec^{-1}, figure S4). For the major biomass fractions, namely proteins, RNA, and lipids, we varied their proportion of the total cellular mass according to the ranges reported in literature (Razin *et al*, 1963) in order to make sure that we do not underestimate their production costs (figure S4). We found that RNA has the largest effect on growth rate and energy metabolism, altering the doubling time by at least 10 hours (figure S4). This on first sight surprisingly high impact of the RNA fraction size (varied from 6.5 to 10% of the total cell mass) on the doubling time can be explained by the slow growth of *M. pneumoniae.* Taking into account that the maintenance energy was fitted to match the *in vivo* doubling times and the rapid mRNA turnover, the influence of the RNA fraction increases along with an increasing doubling time, even preventing growth at 60 hours after inoculation if other maintenance costs are not overestimated in the model (figure S4). Decreasing the protein fraction from 62 to 54% results in a few hours faster doubling times while the lipid fraction size is found to be not growth limiting (figure S4), which is further supported by the analysis of the fatty acid spectrum in the growth medium before and after four days of batch culture growth (figure S2C). Decreasing the lipid fraction from 20 to 12% does not have a significant effect on doubling time and energy metabolism (figure S4).

Definition of Flux Constraints. To allow the correct reproduction of experimental findings, several constraints have to be imposed on the metabolic reconstruction (table S5). Those constraints can define either nutrient availability or specific cellular functions the model has to account for, amongst them the spontaneous generation of (toxic) metabolic compounds or the maintenance and turnover of cellular macromolecules. To simulate batch culture growth of *M. pneumoniae*, we applied nonlinear fitting of sigmoidal curves to metabolite (glucose, lactic acid, acetic acid) and protein quantification data (figure S5). Based on those functions we calculated the ratio of lactic acid to acidic acid and, subsequently, constraints for glucose up-take and acetic acid production. For the glucose constraint we converted the number of molecules taken up per cell and second *in vivo* as described by the mathematical fitting (figure S5B) into mmol per gram of cells and hour: X glucose molecules * 3600 * #ofCellsPerGram * 1000 / Avogadro constant. For the acetic acid production we observed the total ATP produced from all carbon sources and adjusted the

maximum constraint for acetic acid production to the ratio of fitted lactic and fitted acidic acid (figure S5E).

Glycerol and glycerol 3-phosphate (G3P) were limited to 2.5% of the respective glucose constraint based on the defined medium composition from (Yus *et al*, 2009). Ribose was limited to 2.5% of the respective glucose constraint, too, in order to account for ribosylated bases in rich medium. All other sugar sources were set to 0 since it is known that bacteria generally use up one sugar source before switching to another (Monod, 1966) and proteins involved in up-take and processing of alternative sugars have been shown to be low abundant or not detectable (Maier *et al*, 2011). Arginine availability was arbitrarily constrained to 0.25 $mmol*g^{-1}*h^{-1}$ in order to prevent unlimited energy production from arginine and reflect the low copy numbers of involved proteins (Maier *et al*, 2011). mRNA and protein turnover have been accounted for by minimum constraints on the respective degradation reactions based on experimentally determined mRNA and protein half-lives (Maier *et al*, 2011). To represent known detoxification events, minimum constraints were set on dihydroxyacetone production (spontaneous) and 5-formyl tetrahydrofolate (regulatory function). Based on the facts that the conversion of DHAP into G3P is very slow (figure 2A) and that glycerol is essential for growth in defined medium (Yus *et al*, 2009) we constrained the conversion of DHAP into G3P arbitrarily to 0.25 $mmol*g^{-1}*h^{-1}$ to allow reproduction of the experimental data by the model. The ATPase reaction and two reactions describing the activity of the chaperones DnaK/DnaJ/GrpE and GroEL/GroES have been included in the model for completeness but have not been constrained, since there was no information available about their exact contributions to energy consumption. Instead, a minimum constraint for maintenance energy costs has been fitted manually, thus allowing the reproduction of *in vivo* doubling times. Nevertheless, the model can be easily adapted if new experimental data on energy consuming processes in *M. pneumoniae* becomes available. The constraint sets for different times of batch culture growth in rich medium and for defined medium can be found in table S5.

For the simulation of growth on alternative sugars, glucose has been limited to 0 and the respective alternative sugar source was limited to provide the same amount of carbon as when using the constraint for glucose for 36 hours of batch culture growth.

Finally, we used the total protein measured for *M. pneumoniae* cell cultures *in vivo* in order to constrain the objective value thus allowing the reproduction of the experimentally determined doubling times.

Network Refinement.

We simulated *M. pneumoniae* growth under defined medium conditions and analyzed the resulting flux distributions. We found that in general the reaction activity did not contradict the expression profiles for metabolic genes, i.e. highly active reactions are associated with highly expressed genes and reactions with low or no activity are mainly catalyzed by proteins translated from sparsely or not expressed genes (Güell *et al*, 2009). Furthermore, many findings from growth curve experiments in defined medium (Yus *et al*, 2009) could be confirmed *in silico*. Nevertheless, also a few contradictions could be indentified when comparing growth and metabolite production *in vivo* and *in silico*. Those contradictions have been resolved in an iterative cycle between model simulations, their comparison to experimental findings, and the subsequent adjustment of the model.

Transport Reactions: In order to allow the *in silico M. pneumoniae* to grow, several transport reactions had to be added to the metabolic network published by Yus *et al*, 2009. Since it has been shown that ABC transporters are the most prominent transport systems in different mycoplasma species (Nicolas *et al*, 2007), we preferentially use ABC transporters for the import of sugars, amino acids, or other metabolites. Only when evidence for alternative transport systems exist those mechanisms, for example proton symport for the export of lactic and acidic acid, have been included in the model (table S1). ABC transport systems have been used for:

- orthophosphate (Pi) that is imported via the ptsA/ptsC-like ABC transport system composed ofMPN411 (ptsA) and MPN412 (ptsC),
- the import of all amino acids,
- phosphatidylcholine import.

Redox Balance: As a consequence of extensive evolutionary genome reduction, *M. pneumoniae* lacks the TCA cycle and a functional respiratory chain (Yus *et al*, 2009; Himmelreich *et al*, 1996; Dandekar *et al*, 2000). For ATP generation, it relies on glycolysis and subsequent low-energy gain fermentation reactions, producing lactic

and acetic acid. Pyruvate dehydrogenation, ultimately leading to acetic acid production, causes intracellular accumulation of reducing equivalents (NADH). To account for this redox imbalance, the model initially predicted a massive interconversion of dihydroxyacetone phosphate (DHAP) and glycerol 3-phosphate (G3P) involving a coupled glycerol 3-phosphate dehydrogenase (GlpD) and a glycerol phosphate oxidase (GPO) (MPN051, figure 2A). In a cycle DHAP was converted into G3P thereby compensating the redox imbalance produced during acetic acid production and G3P was converted back into DHAP using oxygen (O_2) as a cofactor and producing H_2O_2. This was possible since in the network published by Yus *et al*, 2009, two reversible reactions converting DHAP into G3P and vice versa existed. However, in ^{13}C-glucose tracer experiments we observed only comparably low metabolite conversion rates involving this reaction (figure S2A), thereby dismissing DHAP and G3P interconversion as major redox balance maintaining reaction in *M. pneumoniae*. In addition, sequence analysis (SI) and experimental results (Hames *et al*, 2009) unambiguously identified MPN051 as a GPO with no GlpD similarity. Consequently, the GlpD reaction (R050 in Yus *et al*, 2009) has been removed from the network (figure 2A).

Subsequently, the model predicted the NADH oxidase (NOX, MPN394) to maintain the cellular redox balance, thereby producing major quantities of hydrogen peroxide (H_2O_2). This hypothesis is not supported experimentally, since it has been shown that the GPO is the major source for H_2O_2 production *in vivo* (Hames *et al*, 2009). Still, the high abundance of the NOX (1763 copies per cell (Maier *et al*, 2011)), suggests high activity. Literature research revealed that NOX exists in H_2O_2 and H_2O producing isoforms (Sakamoto *et al*, 1996). pBLAST analysis allowed the re-annotation of the *M. pneumoniae* NOX as the H_2O producing isoform, based on both global sequence similarity and, more specifically on conservation of a characteristic FAD binding motif (figure S8, SI). Integrating this information into the model, i.e. changing reaction M017 accordingly (figure 2B), confirmed the central role for the NOX in maintaining the cellular redox balance *in silico*.

Base Requirements: Experimental results identified the nucleoside cytidine as essential and sufficient to synthesize all pyrimidine nucleotides (CTP, UTP, TTP) (Yus *et al*, 2009). In contradiction to these findings but in agreement with literature data on other mycoplasma species (Pachkov *et al*, 2007; Finch & Mitchell, 1992), our

simulations initially suggested that the nucleobase uracil could additionally serve as precursor for the synthesis of all pyrimidine nucleotides. To determine what allows the *in silico M. pneumoniae* to grow on uracil being the only pyrimidine base provided, we silenced the individual reactions of the pyrimidine metabolism. We identified the direct conversion of UTP into CTP through the CTP synthase (EC number 6.3.4.2, R098 in (Yus *et al*, 2009), catalyzed by MPN256) to be responsible for the production of all other pyrimidine bases from uracil (figure 2C). Through a sequence alignment with pBLAST versus the non-redundant database (nr-DB) no similarity could be detected between MPN256 and any CTP synthase from other organisms. To ensure that this reaction is not just catalyzed by another *M. pneumoniae* enzyme we additionally performed pBLAST of *M. gallisepticum*, *M. hyopneumoniae*, *M. mobile*, *M. mycoides*, *M. penetrans*, *M pulmonis*, *E. coli*, *B. subtilis* and *L. lactis* CTP synthases (pyrG gene) versus the *M. pneumoniae* proteome. No significant similarity of the CTP synthases to any *M. pneumoniae* protein could be detected (Sequence Alginments). We therefore removed the corresponding reaction from the metabolic network (figure 2C), leading to congruent results from model predictions and experimental data (Yus *et al*, 2009). This deletion contradicts the assumptions of (Pachkov *et al*, 2007) who included the CTP synthase reaction in the metabolic network of *M. pneumoniae* as a putative activity based on experimental findings in other mycoplasmas (Finch & Mitchell, 1992). As our experimental results could not prove uracil-sufficiency (Yus *et al*, 2009), we claim that *M. pneumoniae* differs from other mycoplasmas in pyrimidine metabolism in missing a CTP synthase. It would be important to check back the genome annotations of other mycoplasmas in order to identify errors that have been made when assigning gene function based on sequence similarity.

Essentiality of Glycerol: *M. pneumoniae* requires glycerol for growth in defined medium (Yus *et al*, 2009). Yet, *in silico* growth simulations could not confirm this requirement. Glucose could be used for lipid production via conversion of DHAP into G3P even after removing one of the two reactions doing so. Therefore, alterations of the metabolic network were identified that would render glycerol essential for growth. The only way to achieve this was setting the remaining reaction interconverting G3P to DHAP irreversible (G3P − > DHAP), since this results in glucose being insufficient for lipid production. Thus, glycerol, G3P and fructose would be the only metabolites

from which phosphatidic acid, the precursor of all lipids, can be synthesized. Sequence analysis with pBLAST showed that *M. pneumoniae* has no classical glycerol 3-phosphate dehydrogenase but a glycerol 3-phosphate oxidase instead (Sequence Alignments). This oxidase is attached to the membrane and converts G3P to DHAP by using O2 and producing H_2O_2 (Hames *et al*, 2009) which is exported from the cell and probably plays a major role in *M. pneumoniae* pathogenicity (Cohen & Somerson, 1967; Hames, 2008). However, in ^{13}C-tracer experiments it has been shown that glucose is converted into G3P although this conversion compared to the glycolysis flux is insignificantly slow (figure S2A). One could speculate that due to the export of toxic H_2O_2 the substrate concentration for the backward reaction is not sufficient to produce G3P from DHAP in sufficient amounts for lipid synthesis.

Furthermore, *in silico* glycerol could be substituted by fructose under defined medium conditions, which was not possible *in vivo* (Yus *et al*, 2009). This possible substitution is achieved through the reaction that converts fructose-1-phosphate (F1P) into glyceraldehyde (GA), the precursor of glycerol, and DHAP catalyzed by fructose bisphosphate aldolase (aldolase, MPN025). But, in *E. coli* and *Helix pomatia* the aldolase has an affinity towards F1P of only 3.5% and 5%, respectively, of its affinity towards fructose bisphosphate (FBP) (Kochman *et al*, 1982; Szwergold *et al*, 1995). This could lead to a shift of the equilibrium constant from F1P to FBP and to an extremely limited incorporation of carbon into the lipid metabolism, thus explaining why glycerol in the defined medium cannot be substituted by fructose although the pathway exists in *M. pneumoniae*. In addition, it has been shown that glycerol plays a regulatory role in glucose up-take via the phosphorylation of the phosphocarrier protein HPr (MPN053) (Halbedel *et al*, 2006). Since a constraint-based model cannot account for regulatory functions directly, we constrained the conversion of DHAP into G3P arbitrarily to 0.25 mmol*g^{-1}*h^{-1}.

Lipid Metabolism: We tried to shed light on the phospholipid branch, where only the two upstream enzymes of the supposed four catalyzing enzymes are known and both could be disrupted in *M. genitalium* (Glass *et al*, 2006). For the corresponding genes in *M. pneumoniae*, namely *mpn637* and *mpn253*, mRNA expression near the detection limit and changes in the expression levels along a four days batch culture growth experiment have been published (Güell *et al*, 2009). However, the respective proteins have not been detected (Maier *et al*, 2011). The two downstream enzymes of

the phospholipid branch (catalyzing reactions M054 & M055, R057 & R058 in Yus *et al*, 2009) could not be identified by sequence comparison of *E. coli* phosphatidylglycerol phosphatases (pgpA and pgpB) and cardiolipin synthase with the *M. pneumoniae* proteome using pBLAST. These enzymes could be neither identified in any other mycoplasma species nor in *B. subtilis* or *L. lactis*. It remains questionable if *M. pneumoniae* is able to synthesize phospholipids or just still has some genes encoding for proteins once involved in phospholipid synthesis. Accordingly, phosphatidic acid has been included in the biomass instead of cardiolipin.

Sequence Alignments. We used NCBI pBLAST (Altschul *et al*, 1997) to determine sequence similarity for different *M. pneumoniae* proteins to those of other organisms. We used the nr-DB when searching for similarity to *M. pneumoniae* proteins and the *M. pneumoniae* proteome when using protein sequences from other organisms to check for similarities in *M. pneumoniae*. In case of significant hits (e-value < 1e-25) the alignment of the best hit is shown, in case of no significant hits the 5 first entries from the hit list are shown.

1. Glycerol 3-phosphate dehydrogenase/oxidase (MPN051):

Query= glycerol 3-phosphate dehydrogenase (GlpD) {Bacillus subtilis}
Length=555

Sequences producing significant alignments:	Score (Bits)	E Value
ref|NP_109739.1| glycerol-3-phospate dehydrogenase [Mycoplasm...	53.5	4e-10

ALIGNMENTS
>ref|NP_109739.1| glycerol-3-phospate dehydrogenase [Mycoplasma pneumoniae M129]
sp|P75063.1|Y051_MYCPN RecName: Full=Uncharacterized protein MG039 homolog
gb|AAB95751.1| glycerol-3-phospate dehydrogenase [Mycoplasma pneumoniae M129]
Length=384

Score = 53.5 bits (127), Expect = 4e-10, Method: Compositional matrix adjust.
Identities = 86/359 (24%), Positives = 143/359 (40%), Gaps = 72/359 (20%)

```
Query  21  KTYDLFIIGGGITGAGTALDAASRGMKVALSEMQDFAAG-TSSRSTKLVHGGLRYLKQFE  79
           +T D+ I+GGG+ G  TA + +   +KV L E   + A  TS  ++ ++H G+
Sbjct   2  ETRDVLIVGGGVIGCATAYELSQYKLKVTLVEKHHYLAQETSHANSGVIHTGI-------  54

Query  80  VKMVAEVGKERAIVYENGPHVTTPEWMLLPFHKGGTFGSFTTSIGL---RVYDFLAGVKK  136
                       +  PH  T ++ +L   K     ++   +G    ++   +    +
```

```
Sbjct  55  ---------------DPNPHKLTAKYNIL--GKKLWLNTYFKRLGFPRQKIRTLIVAFNE  97

Query 137  SERRS-------------------MLSAKETLQKEPLVKKDGLKG----GGYYVEYRTDD  173
            ER                     MLS +ETL+ EP V  + + G    G + ++
Sbjct  98  MEREQLEVLKQRGIANQINLEDIQMLSKEETLKLEPYVNPEIVAGLKIEGSWAIDPVLAS  157

Query 174  ARLTIEVMKEAVKF--GAEPVNYSKVKELLYEKGKAVGVLIEDVLTKKEYKVYAKKIVNA  231
              L +   +  V+     E  N SK  +  Y       ++  +  T   +KV  KKI++A
Sbjct 158  KCLALAAQQNKVQICTNTEVTNISKQVDGTY-------LVWTNNETTPSFKV--KKIIDA  208

Query 232  TGPWVDQLREKDHSKNGKHLQHTKGIHLVFDQSVFPLKQAVYF-DTPDGRMVFAIPR-EG  289
            G + D L     + + +         +V +Q    L   V+   T  G+ V   P  +G
Sbjct 209  AGHYADYLAHLAKADDFEQTTRRGQYVVVTNQGELHLNSMVFMVPTIHGKGVIVSPMLDG  268

Query 290  KTYVGTT--DTVYKEALEHPRMTTEDRDYVIKSI-NYMFPELNITANDIESSWAGLRPL  345
               VG T  D V KEA    R  T+D   ++  I  +M P LNI  N+   S+AG RP+
Sbjct 269  NFLVGPTALDGVDKEAT---RYITKDAPCMLTKIGKHMVPSLNI--NNALISFAGSRPI  322
```

2. NADH oxidase (NOX, MPN394):

a) H_2O-forming NOX of *S. mutants* vs. *M. pneumoniae* proteome

Query= NADH oxidase (H2O-forming) (NaoX) {Streptococcus mutans}

Length=457

Sequences producing significant alignments:	Score (Bits)	E Value
lcl|58781 MPN394 NADH oxidase (nox) {Mycoplasma pneumoniae M129}	286	1e-81

ALIGNMENTS

>lcl|58781 MPN394 NADH oxidase (nox) {Mycoplasma pneumoniae M129}

Length=479

Score = 286 bits (731), Expect = 1e-81, Method: Compositional matrix adjust.

Identities = 171/472 (36%), Positives = 281/472 (59%), Gaps = 20/472 (4%)

```
Query   1  MSKIVIVGANHAGTAAINTILDNYGSENEVVVFDQNSNISFLGCGMALWIGKQISGPQGL  60
           M K++++G NHAGT+ I T+L     + +V  +D+N+NISFLGCG+AL +   +   + L
Sbjct   1  MKKVIVIGVNHAGTSFIRTLLSK-SKDFQVNAYDRNTNISFLGCGIALAVSGVVKNTEDL  59

Query  61  FYADKESLEAKGAKIYMESPVTAIDYDAKRVTA--LVNGQEHVESYEKLILATGSTPILP  118
           FY+  E L+A GA ++M   V  +D D K+V    L  G+E V+ Y++L++A+G+ PI
Sbjct  60  FYSTPEELKAMGANVFMAHDVVGLDLDKKQVIVKDLATGKETVDHYDQLVVASGAWPICM  119

Query 119  PIKGAAIK-----EGSRDFEATLKNLQFVKLYQNAEDVINKLQ-DKSQNLNRIAVVGAGY  172
            ++            +  +   +KNL   KLYQ+A  +I+  + DKS  +  +A+VG+GY
Sbjct 120  NVENEVTHTQLQFNHTDKYCGNIKNLISCKLYQHALTLIDSFRHDKS--IKSVAIVGSGY  177

Query 173  IGVELAEAFKRLGKEVILIDVVDTCLAGYYDQDLSEMMRQNLEDHGIELAFGETVKA--I  230
           IG+ELAEA  + GK+V +ID++D      +D++ +  + + ++  GI L  G  VK   +
Sbjct 178  IGLELAEAAWQCGKQVTVIDMLDKPAGNNFDEEFTNELEKAMKKAGINLMMGSAVKGFIV  237

Query 231  EGDGKVER-IVTDKASHDVDMVILAVGFRPNTAL--GNAKLKTFRNGAFLVDKK-QETSI  286
           + D  V + + TDK   D D+VI ++GFRPNT     + + +  RNG+  V++  Q  +
Sbjct 238  DADKNVVKGVETDKGRVDADLVIQSIGFRPNTQFVPKDRQFEFNRNGSIKVNEYLQALNH  297

Query 287  PDVYAIGDCATVYDNAINDTNYIALASNALRSGIVAGHNAAGHKLESL-GVQGSNGISIF  345
            +VY IG  A +YD A      I LA+NA++SG+VA  + G K   L  + G+N + +F
Sbjct 298  ENVYVIGGAAAIYDAASEQYENIDLATNAVKSGLVAAMHMIGSKAVKLESIVGTNALHVF  357

Query 346  GLNMVSTGLTQEKAKRFGYNPEVTAFTDFQKASFIEHDNYPVTLKIVYDKDSRLVLGAQM  405
           GLN+ +TGLT+++AK  G++  V+   D  +  F+   +  V  K++YDK +  +LGAQ+
```

```
Sbjct 358  GLNLAATGLTEKRAKMNGFDVGVSIVDDNDRPEFMGTFD-KVRFKLIYDKKTLRLLGAQL  416

Query 406  AS-KEDMSMGIHMFSLAIQEKVTIERLALLDYFFLPHFNQPYNYMTKAALKA  456
            S   + S  I   +LA+Q+K+ I  L L+D +FLPH+N+P+N++  A L+A
Sbjct 417  LSWNTNHSEIIFYIALAVQKKMLISELGLVDVYFLPHYNKPFNFVLAAVLQA  468
```

b) H_2O_2-forming NOX of *S. mutants* vs. *M. pneumoniae* proteome

Query= NADH oxidase (H2O2-forming) {Streptococcus mutans}

Length=510

Sequences producing significant alignments:	Score (Bits)	E Value
lcl|2291 MPN394 NADH oxidase (nox) {Mycoplasma pneumoniae M129}	23.1	0.022

ALIGNMENTS

>lcl|2291 MPN394 NADH oxidase (nox) {Mycoplasma pneumoniae M129}

Length=479

Score = 23.1 bits (48), Expect = 0.022, Method: Compositional matrix adjust.

Identities = 31/127 (24%), Positives = 53/127 (41%), Gaps = 16/127 (12%)

```
Query 348  KKVAVIGGGNSGLEAAIDLAGLASHVYILEFLPELKADKILQDRAEALDN------ITIL  401
            K VA++G G  GLE A         V +++ L +   +   ++    L+       I ++
Sbjct 168  KSVAIVGSGYIGLELAEAAWQCGKQVTVIDMLDKPAGNNFDEEFTNELEKAMKKAGINLM  227

Query 402  TNVATKEII---GNDHVEGLRYSDRTTNEEYLLDLEGVFVQIGLVPSTDWL-KDSGLALN  457
               A K  I     + V+G+  +D+       +D + V   IG  P+T ++ KD     N
Sbjct 228  MGSAVKGFIVDADKNVVKGVE-TDKGR-----VDADLVIQSIGFRPNTQFVPKDRQFEFN  281

Query 458  EKGEIIV  464
              G I V
Sbjct 282  RNGSIKV  288
```

3. CTP synthase: As an example for the negative results in the attempt of identifying a CTP synthase in *M. pneumoniae* the alignment of *M. gallisepticum* pyrG protein sequence vs. the *M. pneumoniae* proteome is shown.

Query= CTP synthase (pyrG) {Mycoplasma gallisepticum}

Length=540

Sequences producing significant alignments:	Score (Bits)	E Value
lcl|57601 MPN001 DNA polymerase III beta subunit (dnaN) {Myco...	7.3	0.84
lcl|57603 MPN003 DNA gyrase subunit B (gyrB) {Mycoplasma pneu...	18.5	0.30
lcl|57604 MPN004 DNA gyrase subunit A (gyrA) {Mycoplasma pneu...	17.3	0.66
lcl|57605 MPN005 seryl-tRNA synthetase (serS) {Mycoplasma pne...	16.2	1.7

lcl|57606 MPN006 thymidylate kinase {Mycoplasma pneumoniae M129} 15.0 3.4

Energy Metabolism: To analyze the energy metabolism of *M. pneumoniae* we first determined the amount of energy needed to adjust the *in silico* doubling time to *in vivo* measurements by manually fitting the constraint on the ATP-consuming maintenance energy reaction reflecting cellular homeostasis costs (M304). We then analyzed *in silico* usage of ATP for different functions, such as protein, DNA, RNA, and lipid production, protein turnover, and cofactor synthesis/detoxification functions for different simulated time points *t* of batch culture growth *t* = 24, 36, 48, 60 hours (table S6). To further specify the cellular homeostasis costs, we applied literature data to calculate upper boundaries for the costs for DNA maintenance, chaperone-catalyzed protein folding, post-translational modification of proteins, and ATPase function. In the main text the numbers for *t* = 36 hours are given, all other results are shown in table S6.

DNA maintenance costs: DNA-based microbes have mutation rates < 0.005 per genome per duplication (Drake *et al*, 1998). In order to calculate the upper boundary for DNA maintenance costs we assumed a mutation rate of 0.05 for *M. pneumoniae* and find that DNA maintenance can only account for maximally about 0.01% of the total energy produced.

Chaperone-catalyzed protein folding costs: We integrated the abundances determined for the two *M. pneumoniae* chaperones (GroEL, a 14mer and DnaK) (Maier *et al*, 2011) with catalytic rates described for *E. coli* (Naylor & Hartl, 2001). Assuming constant activity of all chaperones we calculated the upper boundary for the ATP consumption for protein folding.

Post-translational modification costs: 93 phosphorylation sites and 720 acetylation sites on 72 and 221 proteins, respectively, have been detected in *M. pneumoniae in vivo* (van Noort *et al*, 2012). Integrating this information with the abundances of the respective proteins (Maier *et al*, 2011), we calculated the upper boundary for post-translational modifications. On one hand, we assumed that all phosphorylation and acetylation sites on all found proteins are modified once. This accounts for 0.01% of the total ATP expenses. Adding another 0.02% of the total ATP consumed at each time point to account for cyclic phosphorylation-dephosphorylation events (Shacter *et*

al, 1984), we find that post-translational modifications account for 0.03% of the total energy. For lack of information about de-acetylation we did not consider them further.

ATPase costs: We determined the amount of ATPases per *M. pneumoniae* cell based on the abundance of the β-subunit (MPN598) of which three copies are contained in the ATPase core (Maier *et al*, 2011). In thermophylic *Bacillus* the ATPase has been shown to hydrolyze 3 ATP per rotation (Watanabe *et al*, 2008) and a maximum catalytic rate of 130 rps has been reported (Wu *et al*, 2010). We calculated the amount of ATP maximally used by the ATPase at each time point of batch culture growth assuming constant full occupancy at maximum catalytic rate (table S6).

Curve Fittings.

Metabolite assay fittings: We used KaleidaGraph 4.0 to fit sigmoidal curves to the *in vivo* data for glucose consumption, acetic acid production, lactic acid production and protein production (figure S5). The general equation for sigmoidal curves reads:

$$f(x) = a + (b - a) / (1 + c^{(d - x)}) \quad \text{(eq. S1)}$$

For glucose consumption this equation has been adapted to represent the inverted curve progression:

$$f(x) = a + (b - a) / (1 + c^{(d - (180 - x))}) \quad \text{(eq. S2)}$$

To describe the progression of maintenance costs (figure 3B) we fitted a polynomial function to the manually determined maintenance costs at different times of growth in order to calculate the respective constraints as exact as possible with the available data (despite that a linear fit also describes the data very well):

$$f(x) = ax^3 + bx^2 + cx + d \quad \text{(eq. S3)}$$

The variable values for each fitted curve can be found in table S8. The resulting curves allow to calculate constraints for $0 \leq x \leq 180$ with x being the growth time of a four days batch culture in hours. Still, we recommend to not use the presented model with biomass production as single objective function for $x \leq 24$ and $x \geq 60$ since this is the determined exponential growth phase (figure S5A).

Assignment of *In Vivo* Protein Abundance Changes and *In Silico* Flux Changes. When analyzing the flux changes during exponential growth phase we find that most of the qualitative flux changes (~51.3%) correlate with biomass production, i.e. from 24 hours to 36 hours the flux increases and from 36 hours to 48 hours to 60 hours the flux decreases. These flux changes can be explained due to the growth behavior of *M. pneumoniae* observed in batch culture growth, which is characterized by a lag phase in the beginning, followed by an exponential growth phase and a growth halt at later growth stages (figure S5A, Yus *et al*, 2009). The *in vivo* doubling times suggest that *M. pneumoniae* at 24 hours after inoculation, i.e. at the beginning of the exponential growth phase still does not grow at optimal rate under laboratory conditions.

In general, the description and analysis of a cellular subsystem in isolation holds the risk to miss important regulatory influences from the other subsystems of the cell. To obtain information about the relation between protein abundances and metabolic activity we integrated experimentally determined protein quantification data (Maier *et al*, 2011, unpublished results) with *in silico* flux predictions for different time points of batch culture growth. For 120 reactions catalyzed by a single enzyme for which abundance data was available we calculated the flux differences between 24 and 36 hours, between 36 and 48 hours, and between 48 and 60 hours (figure 5A) of *in silico* growth. Using protein abundances for 79 metabolic enzymes that on their own catalyze a metabolic reaction, we applied linear fitting to the experimentally determined abundances. A protein has been considered to change only if the overall change of the respective linear fit exceeds 25% of the initial protein abundance. We compared the qualitative trends of flux changes and protein abundance changes and found that they match in 86% of the cases (figure S6A). Analyzing these hits according to metabolic pathways revealed that except in amino acid metabolism, nucleotide metabolism, folate metabolism and pyruvate metabolism none of the protein abundance changes directly contradicts the predicted changes in metabolic fluxes (figure S6B). Interestingly, we found that in pyruvate metabolism one subunit of the pyruvate dehydrogenase increases significantly in abundance while all acetic acid producing reactions are predicted to be down-regulated *in silico.* Glycolysis and pyruvate metabolism enzymes belong to the most abundant proteins in *M. penumoniae* (Maier *et al*, 2011), thus suggesting not only high importance but also a smaller experimental error compared to enzymes that have low concentrations. In

case of sugar processing the complete match mainly results from the *in silico* silencing of alternative sugar sources. Interestingly, all proteins involved in uptake and processing of alternative sugars have been either not detected or shown to be very low abundant (see above, Simulation of biomass production), thus confirming that *M. pneumoniae* uses glucose as preferred sugar source and has to adapt protein expression to use other sugar sources. Unfortunately, the integration of information about post-translational modification sites identified (van Noort *et al*, 2012) did not allow drawing further conclusions about metabolic regulation since we could not find any significant enrichment of modified proteins among the proteins showing abundance changes contradicting the predicted flux changes. Summing up, we suggest that a detailed analysis of the protein interaction network and the occurrence of post-translational modifications will be indispensible to connect protein interaction and metabolic network.

Gene Essentiality Prediction. We predicted gene essentiality by individually knocking out each metabolic gene *in silico*. Out of totally 145 genes coding for proteins assigned to metabolism, we included 131 in our essentiality prediction. We did not consider genes whose encoded proteins are involved in DNA degradation, chaperone-catalyzed protein folding and the ATPase complex since these functions have been included in the model but are not modeled explicitly, i.e. the corresponding reactions have not been constrained yet due to missing data on their exact contribution to energy consumption. To 114 metabolic genes we could assign a functional ortholog in *M. genitalium* (table S12). The other 17 genes are assumed to be not essential for growth due to their absence in *M. genitalium* and the high similarity of the metabolic networks of both organisms (Yus *et al*, 2009).

Assuming further that genes coding for proteins that form part of a metabolic complex with at least one non-essential component are metabolically not essential if they do not catalyze other metabolic reactions independently, we initially obtained an overall prediction accuracy of 86.26% and a specificity of 97.62% when comparing our results to a genome-wide transposon study in *M. genitalium* (Glass *et al*, 2006) as sole criterion. Screening a *M. pneumoniae* transposon library allowed the confirmation of five predicted non-essential genes for which no mutant had been identified in *M. genitalium* (figure S7). Thus, the prediction accuracy could be raised to 91.6% and the specificity to 97.96%. Finally, taking into account the simulation conditions (rich

medium, growth on glucose), the model predicts metabolic gene essentiality with an overall accuracy of 95.42% and specificity of 98.15%.

Comparative Calculations for *M. pneumoniae* and *E. coli*. In order to obtain an estimate for the energy that is used for pH maintenance in *M. pneumoniae*, we accomplished a comparison of the biophysical properties and the energy consuming process in *M. pneumoniae* and *E. coli*. The numbers for biophysical properties of *M. pneumoniae* are taken from (Yus *et al*, 2009) and the numbers for *E. coli* have been obtained from the bionumbers database (Milo *et al*, 2010). The list of biophysical properties used and the calculation results can be seen in table S7.

Supplementary Figure Legends

Figure S1: Implemented reaction network of *M. pneumoniae* metabolism: Full names of network components can be found in table S15.

Figure S2: Central carbon metabolism analysis *in vivo*. Glycolytic intermediates and reporter compounds for adjacent pathways have been analyzed for incorporation of heavy isotope labelled carbon. A: Turnover of glycolytic intermediates. B: Labelling of key metabolites in pathways branching off glycolysis. Light grey bars: $^{13}C/^{12}C$ ratios, black lines: % of labelled metabolite.

Figure S3: Fatty acid analysis *in vivo*: A - Distribution of the most prominent fatty acids in the growth medium, the *M. pneumoniae* cytoplasm, and the cell membrane. B - Fatty acid composition in the cytoplasm at different days of batch culture growth. C - Comparison of the fatty acid composition in the growth medium before (BEFORE) and after (AFTER) a batch culture growth experiment: Fatty acids do not deplete from the medium during batch culture growth. The numbers (1,2,3) after the colon in the legends if subfigure A and B refer o the number of unsaturated bonds of the respective carbon chain.

Figure S4: Sensitivity analysis. The amount of A - protein (62% - dashed line, 54% - solid line), B - RNA (6.5% - dashed line, 10% - solid line), C - lipids (20% - dashed line, 12% - solid line), and D - cofactors (0.1*regular biomass value - dashed line, 10*regular biomass value - solid line) has been varied and the influence on the doubling time (left y-axis, red) or the maintenance energy (when reproducing *in vivo* doubling times with altered biomass fractions; right y-axis, blue) examined.

Figure S5: Comparison of *in vivo* measurements (dots with error bars) and the nonlinearly fitted curves that have been used to determine constraints for different time points of growth: A - protein, only the exponential phase (24h - 60h) for which *i*JW145 is defined has white background ; B – glucose (the average glucose has been calculated based on the (consumed glucose + (produced lactic acid + produced acidic acid)/2)/2 in order to minimize the experimental error); C - lactic acid; D - acetic acid; E - lac:ace ratio (based on fittings). Error bars indicate the experimental error between

technical replicates, except for glucose where they indicate the standard deviation for the average glucose.

Figure S6: Comparison of changes in fluxes and protein abundances during the exponential growth phase. A – Classification of *M. pneumoniae* enzymes; blue – metabolic enzymes without available protein data or only catalyzing metabolic reactions in complexes, red – metabolic enzymes for which the concentration changes have been assigned to flux changes, the dark red shadow indicates those enzymes that show concentration changes that qualitatively match the flux changes of their respective reactions (hit enzymes), grey – non-metabolic enzymes. B – Hit enzymes in percentage of the total number of enzymes involved in their respective pathways.

Figure S7: Transposon Screens in *M. pneumoniae*. A – Schematic representation of the disrupted genes (big coloured arrows) with genome positions, insertion site (black triangle) & the length of the PCR fragment (indicated below the respective gene); B – Western Blot for the detected insertions.

Figure S8: Conserved FAD-binding fingerprint of the H_2O-producing *M. pneumoniae* NOX.

Supplementary Tables

Table S1: Reaction List: For all reactions included in the model the used reaction ID, the corresponding reaction ID of the metabolic map (Yus *et al*, 2009), the reaction equation ([c] - cytoplasmic, [e] - external), the gene number of the catalyzing protein(s), the enzyme name (for the complete names see table S12), the EC number, the assigned pathway, and the reaction reversibility are shown.

Table S2: Reaction Reversibilities: Integration of information on reaction reversibilities reported for *E. coli* (Fleming2009 - Fleming *et al*, 2009), obtained from the BRENDA enzyme DB, and published for the metabolic map of *M. pneumoniae* (Yus2009 - Yus *et al*, 2009) in order to determine reaction reversibilities for *i*JW145 (for details see SI).

Table S3: Branching Metabolites: For each metabolite the number of reactions it is involved in has been determined. Sink/source reactions have not been included in the analysis.

Table S4: Sensitivity Analysis: The impact of the biomass composition on the doubling time (t_{doub} in hours) with all constraints fixed as in the regular *M. pneumoniae* growth simulations is shown. Furthermore, the maintenance energy constraint (ATP in $mmol*g^{-1}*h^{-1}$) has been fitted manually to allow reproduction of *in vivo* doubling times, both the minimal constraint for the maintenance energy reaction and the resulting doubling time are shown.

Table S5: Constraints: Final constraints in $mmol*g^{-1}*h^{-1}$ as used for simulating growth of M. pneumoniae at different time points of the exponential growth phase under rich medium conditions and to simulate growth on defined medium. Glucose and acetate have been constrained as for the 36-hour time point under rich medium conditions (for details see SI).

Table S6: Energy Expenses: The ATP produced and used for different cellular functions at different time points of the exponential growth phase in batch culture growth, as well as the source for the respective amount are shown. In silico results have been extracted from the respective predicted flux distributions by summing up the ATPs used and produced in the respective pathways, for example total ATP = net

ATP produced in glycolysis + net ATP produced in arginine fermentation. Upper boundary calculations are described in detail in the main text and the SI.

Table S7: Comparative Data for *M. pneumoniae* and *E. coli*: Upper part: biophysical properties of *M. pneumoniae* and *E. coli* obtained from Yus *et al*, 2009, and the bionumbers database, respectively; lower part: comparative calculations based on the biophysical properties for *M. pneumoniae* and *E. coli*.

Table S8: Variables for Fitting Mathematical Equations to *In Vivo* Metabolite Assays: Introducing the variables into the functions described in SI, Curve fittings, provides the fitting functions used to determine metabolic constraints for different time points of the exponential growth phase.

Table S9: Predicted Fluxes for the Exponential Growth Phase: The fluxes for all model reactions in mmol per gram of cells as predicted for different time points of the exponential growth phase of batch culture growth are shown. Negative fluxes indicate that the flux direction of the respective reaction is opposed to the reaction annotation in table S1.

Table S10: Flux Variability Analysis (FVA): Reaction fluxes in $mmol*g^{-1}*h^{-1}$ showing variability > 0.00001 are highlighted in orange, reactions that can vary between no flux and flux in yellow.

Table S11: Qualitative *In Silico* Knock-Out Results: The objective values (ov) of the FBA when simulating *in silico* knock-outs of the listed genes as described in the main text and the SI are shown. ov > 0 - growth, ov = 0 - no growth but catabolic activity, no ov = '-' - FBA is infeasible, i.e. at least one minimum requirement defined in the model cannot be matched.

Table S12: Functional Orthologs of *M. pneumoniae* and *M. genitalium*: The assignment of functional orthologs has been used for the evaluation of the gene essentiality prediction (see main text and SI).

Table S13: Synthetic Lethal and Sick Interactions in *In Silico* Double Knock-Outs: Pairs of genes with their respective pathway affiliations and the type of interaction when simulating double knock-outs *in silico* are shown (compare figure 8).

Table S14: Alignment of gene IDs and gene names.

Table S15: Abbreviations: All abbreviations of metabolites and enzymes used along the text, in figures, and tables.

Table S16: Transposon Primer: Primers used for the identification of transposon insertion sites in *M. pneumoniae* genes predicted to be not essential by the model but not found to be disrupted in *M. genitalium* (Glass *et al*, 2006), their sequence, the ORF they have been used for, the PCR fragment, the transposon insertion site, and the percentage of the protein truncated by the insertion are listed.

dataSheet: Experimental raw data for figures 4, S2, and S3.

References

Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, Miller W & Lipman DJ (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res* **25:** 3389-3402

Amin N & Peterkofsky A (1995) A dual mechanism for regulating cAMP levels in Escherichia coli. *J Biol Chem* **270:** 11803-11805

Cohen G & Somerson NL (1967) Mycoplasma pneumoniae: hydrogen peroxide secretion and its possible role in virulence. *Ann N Y Acad Sci* **143:** 85-87

Drake JW, Charlesworth B, Charlesworth D & Crow JF (1998) Rates of spontaneous mutation. *Genetics* **148:** 1667-1686

Finch L & Mitchell A (1992) Micoplasmas. Molecular Biology and Pathogenesis. - Source of nucleotides Maniloff J McElhany RN Finch LR & Baseman JB (eds) American Society for Microbiology

Fleming RMT, Thiele I & Nasheuer HP (2009) Quantitative assignment of reaction directionality in constraint-based models of metabolism: application to Escherichia coli. *Biophys Chem* **145:** 47-56

Glass JI, Assad-Garcia N, Alperovich N, Yooseph S, Lewis MR, Maruf M, Hutchison CA, Smith HO & Venter JC (2006) Essential genes of a minimal bacterium. *Proc Natl Acad Sci U S A* **103:** 425-430

Goyer A, Collakova E, Díaz de la Garza R, Quinlivan EP, Williamson J, Gregory 3rd JF, Shachar-Hill Y & Hanson AD (2005) 5-Formyltetrahydrofolate is an inhibitory but well tolerated metabolite in Arabidopsis leaves. *J Biol Chem* **280:** 26137-26142

Güell M, van Noort V, Yus E, Chen W-H, Leigh-Bell J, Michalodimitrakis K, Yamada T, Arumugam M, Doerks T, Kühner S, Rode M, Suyama M, Schmidt S,

Gavin A-C, Bork P & Serrano L (2009) Transcriptome complexity in a genome-reduced bacterium. *Science* **326:** 1268-1271

Halbedel S, Busse J, Schmidl SR & Stülke J (2006) Regulatory protein phosphorylation in Mycoplasma pneumoniae. A PP2C-type phosphatase serves to dephosphorylate HPr(Ser-P). *J Biol Chem* **281:** 26253-26259

Hames C (2008) Glycerolmetabolismus und Pathogenitaet von Mycoplasma pneumoniae.

Hames C, Halbedel S, Hoppert M, Frey J & Stülke J (2009) Glycerol metabolism is important for cytotoxicity of Mycoplasma pneumoniae. *J Bacteriol* **191:** 747-753

Johnson JK & Somerson NL (1980) Cholesterol as a limiting factor in the growth of Mycoplasma pneumoniae. *Appl Environ Microbiol* **40:** 391-399

Kochman M, Hargrave PA & Buczyłko J (1982) Fructose-bisphosphate aldolase from Helix pomatia. *Methods Enzymol* **90 Pt E:** 259-263

Kruger NJ, Hill SA & Ratcliffe RG (1999) Regulation of primary metabolic pathways in plants Springer

Maier T, Schmidt A, Güell M, Kühner S, Gavin A-C, Aebersold R & Serrano L (2011) Quantification of mRNA and protein and integration with protein turnover in a bacterium. *Mol Syst Biol* **7:** 511

McElhaney RN & Tourtellotte ME (1969) Mycoplasma membrane lipids: variations in fatty acid composition. *Science* **164:** 433-434

Milo R, Jorgensen P, Moran U, Weber G & Springer M (2010) BioNumbers--the database of key numbers in molecular and cell biology. *Nucleic Acids Res* **38:** D750--D753

Monod J (1966) From enzymatic adaptation to allosteric transitions. *Science* **154:** 475-483

Naylor DJ & Hartl FU (2001) Contribution of molecular chaperones to protein folding in the cytoplasm of prokaryotic and eukaryotic cells. *Biochem Soc Symp***:** 45-68

Neidhardt FC (1996) Escherichia coli and Salmonella: Cellular and Molecular Biology ASM press

Nicolas MF, Barcellos FG, Hess PN & Hungria M (2007) ABC transporters in Mycoplasma hyopneumoniae and Mycoplasma synovia: Insights into evolution and pathogenicity. *Gen. Mol. Biol.* **30(1):** 202-211

van Noort V, Seebacher J, Bader S, Mohammed S, Vonkova I, Betts MJ, Kühner S, Kumar R, Maier T, O'Flaherty M, Rybin V, Schmeisky A, Yus E, Stülke J, Serrano L, Russell RB, Heck AJR, Bork P & Gavin A-C (2012) Cross-talk

between phosphorylation and lysine acetylation in a genome-reduced bacterium. *Mol Syst Biol* **8:** 571

Pachkov M, Dandekar T, Korbel J, Bork P & Schuster S (2007) Use of pathway analysis and genome context methods for functional genomics of Mycoplasma pneumoniae nucleotide metabolism. *Gene* **396:** 215-225

Pollack JD, Somerson NL & Senterfit LB (1970) Isolation, Characterization, and Immunogenicity of Mycoplasma pneumoniae Membranes. *Infect Immun* **2:** 326-339

Razin S, Argaman M & Avigan J (1963) Chemical Composition Of Mycoplasma Cells And Membranes. *J Gen Microbiol* **33:** 477-487

Rottem S (1980) Membrane lipids of mycoplasmas. *Biochim Biophys Acta* **604:** 65-90

Rottem S, Pfendt EA & Hayflick L (1971) Sterol requirements of T-strain mycoplasmas. *J Bacteriol* **105:** 323-330

Sakamoto M, Uchimera T & Komagata K (1996) Comparison of H2O-Forming NADH Oxidase from Leuconostoc mesenteroides subsp. mesenteroides NRIC 1541T and H2O2-Forming NADH Oxidase from Sporolactobacillus inulinus NRIC 1133T. *Journal of Fermentation and Bioengineering* **82:** 531-537

Scheer M, Grote A, Chang A, Schomburg I, Munaretto C, Rother M, Söhngen C, Stelzer M, Thiele J & Schomburg D (2011) BRENDA, the enzyme information system in 2011. *Nucleic Acids Res* **39:** D670--D676

Shacter E, Chock PB & Stadtman ER (1984) Energy consumption in a cyclic phosphorylation/dephosphorylation cascade. *J Biol Chem* **259:** 12260-12264

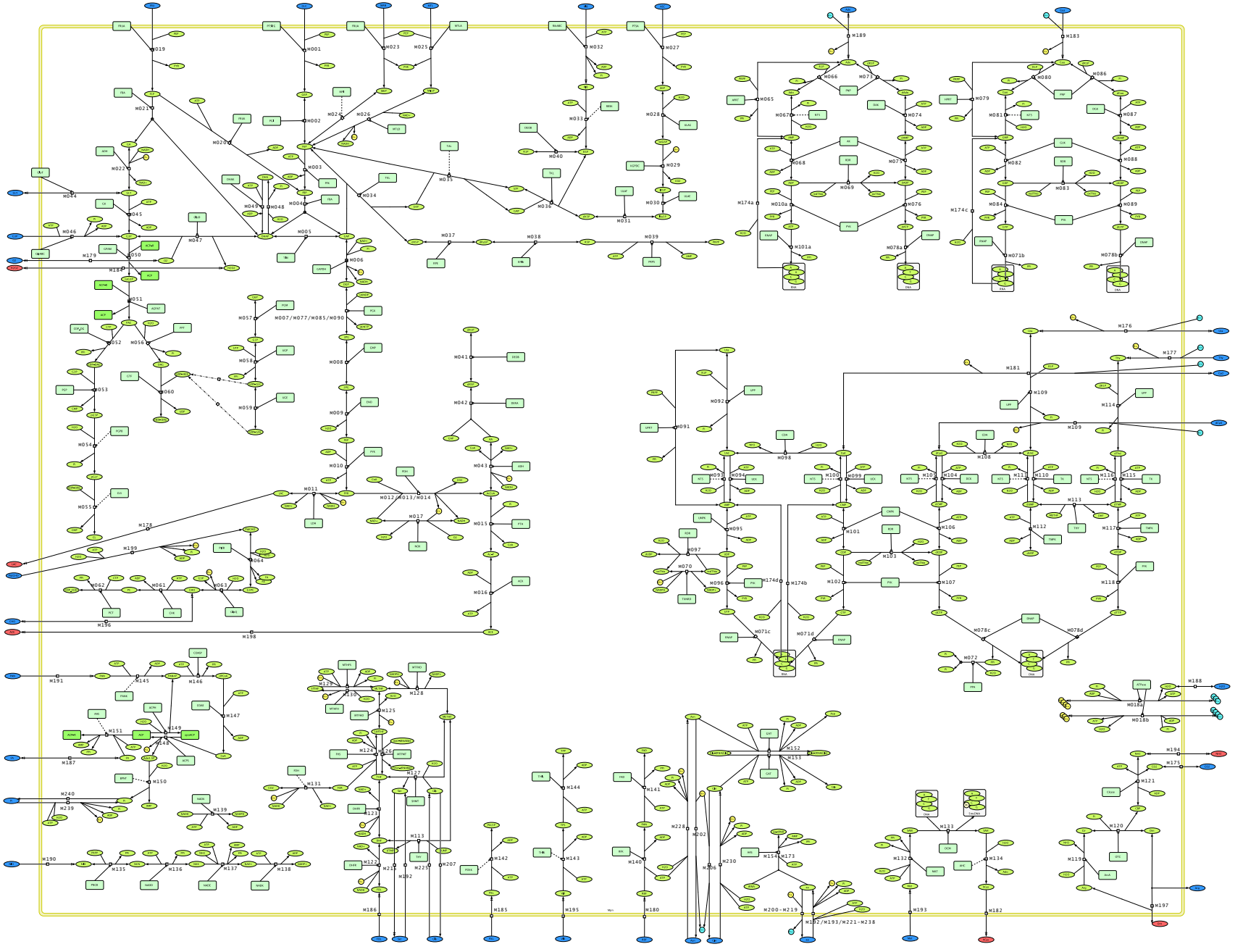
Szwergold BS, Ugurbil K & Brown TR (1995) Properties of fructose-1,6-bisphosphate aldolase from Escherichia coli: an NMR analysis. *Arch Biochem Biophys* **317:** 244-252

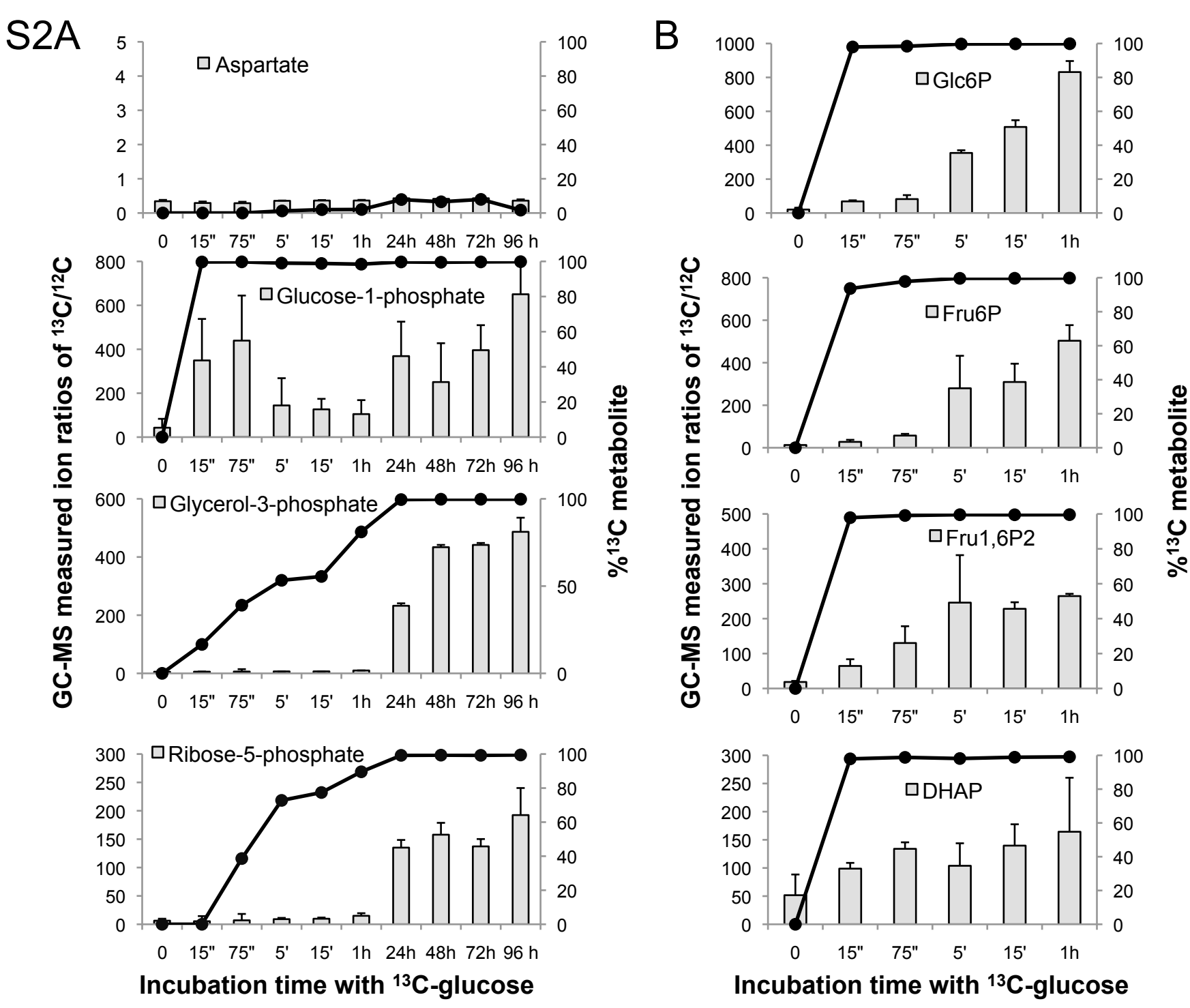
Watanabe R, Iino R, Shimabukuro K, Yoshida M & Noji H (2008) Temperature-sensitive reaction intermediate of F1-ATPase. *EMBO Rep* **9:** 84-90

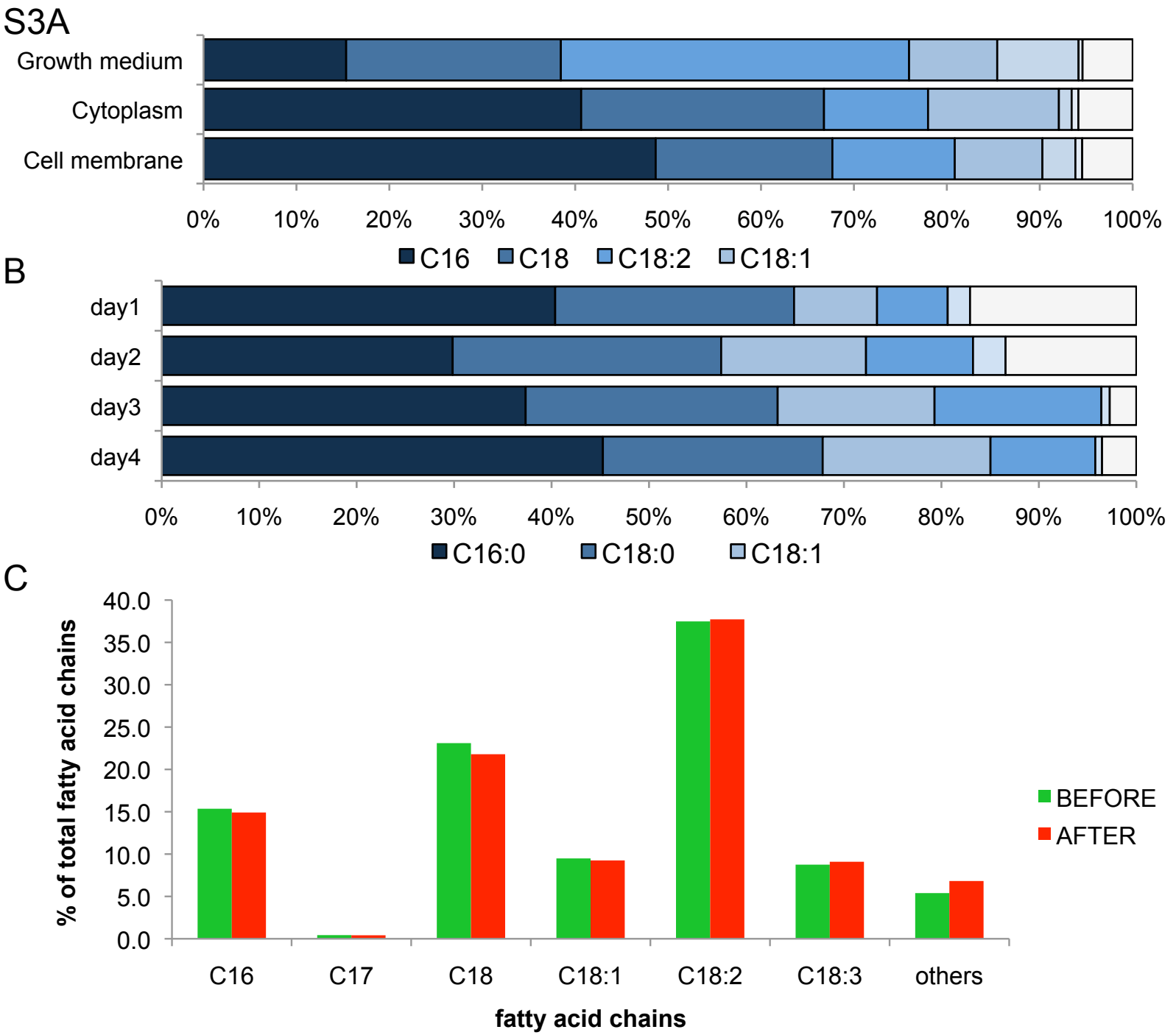
Wu W-X, Zhan Y, Zhao T-J, Han Y-R & Chen Y-F (2010) Stochastic Four-State Mechanochemical Model of F1-ATPase. *Commun. Theor. Phys.* **54:** 630-634

Yus E, Maier T, Michalodimitrakis K, van Noort V, Yamada T, Chen W-H, Wodke JAH, Güell M, Martínez S, Bourgeois R, Kühner S, Raineri E, Letunic I, Kalinina OV, Rode M, Herrmann R, Gutiérrez-Gallego R, Russell RB, Gavin A-C, Bork P, *et al* (2009) Impact of genome reduction on bacterial metabolism and its regulation. *Science* **326:** 1263-1268

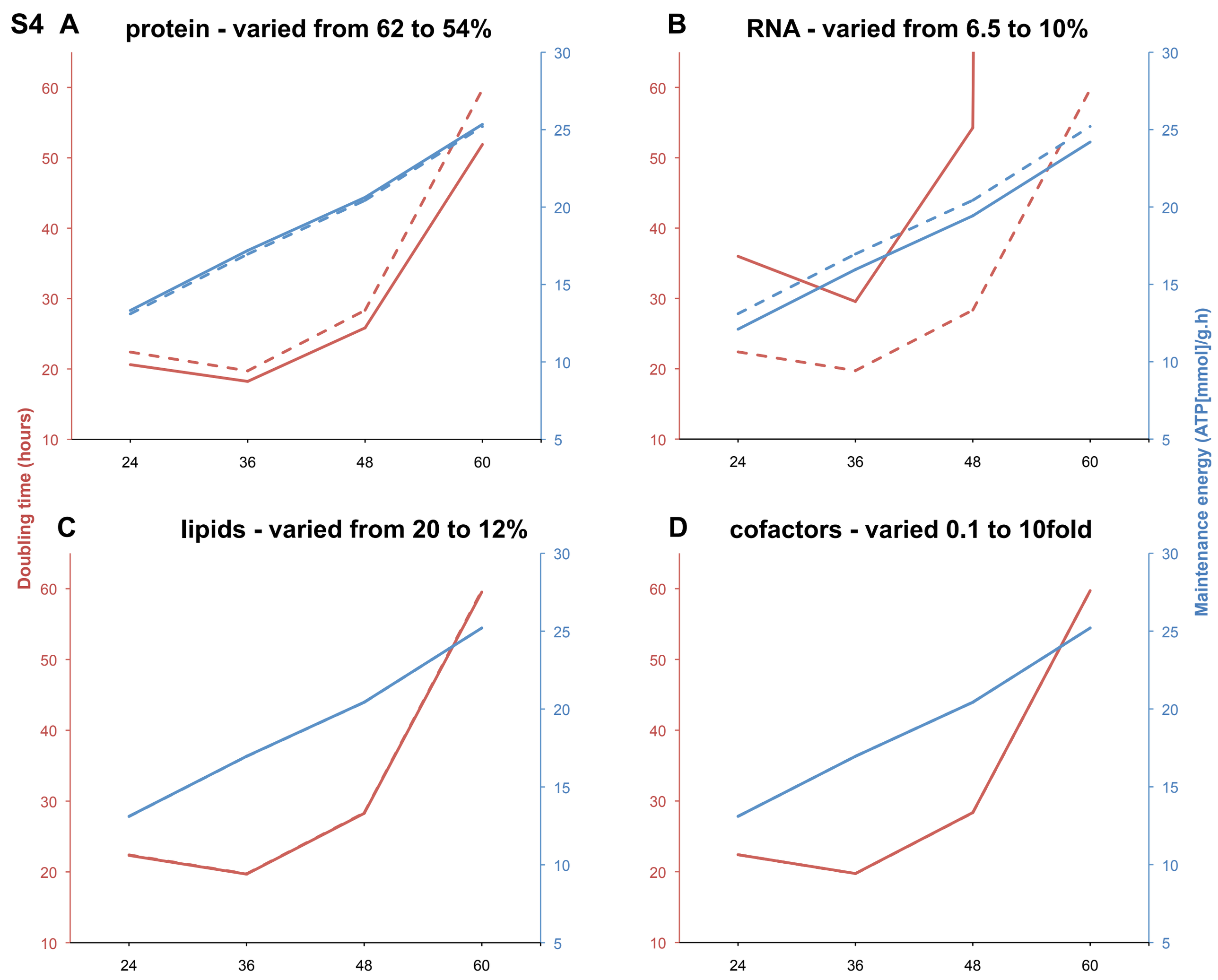
S1

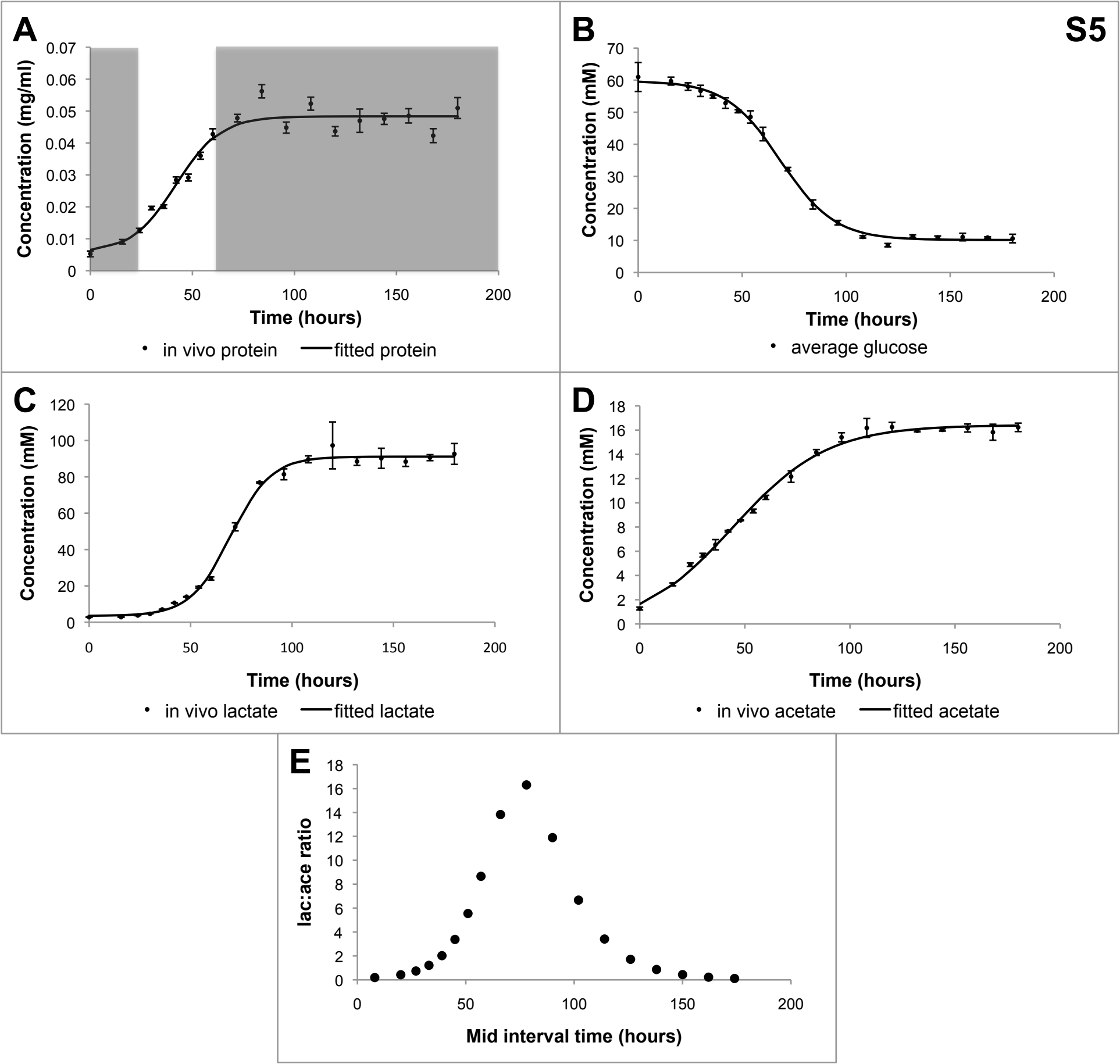



S2A
B
Aspartate
Glucose-1-phosphate
Glycerol-3-phosphate
Ribose-5-phosphate
Glc6P
Fru6P
Fru1,6P2
DHAP
GC-MS measured ion ratios of 13C/12C
%13C metabolite
Incubation time with 13C-glucose
0
15"
75"
5'
15'
1h
24h
48h
72h
96 h

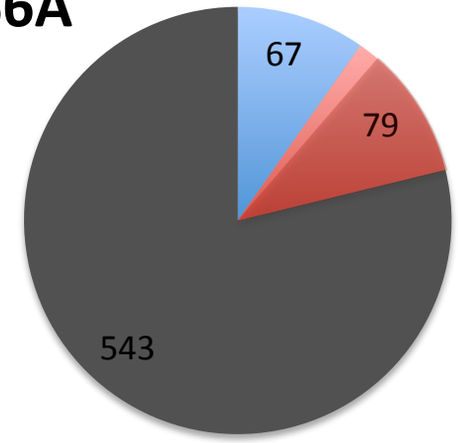

S3A
Growth medium
Cytoplasm
Cell membrane
0%
10%
20%
30%
40%
50%
60%
70%
80%
90%
100%
C16
C18
C18:2
C18:1
B
day1
day2
day3
day4
0%
10%
20%
30%
40%
50%
60%
70%
80%
90%
100%
C16:0
C18:0
C18:1
C
% of total fatty acid chains
40.0
35.0
30.0
25.0
20.0
15.0
10.0
5.0
0.0
C16
C17
C18
C18:1
C18:2
C18:3
others
fatty acid chains
BEFORE
AFTER

S4


A
protein - varied from 62 to 54%
B
RNA - varied from 6.5 to 10%
C
lipids - varied from 20 to 12%
D
cofactors - varied 0.1 to 10fold
Doubling time (hours)
energy (ATP[mmol]/g.h)
Maintenance
10
20
30
40
50
60
5
10
15
20
25
30
24
36
48
60
Simulated time (hours)
Tdoub-BMnorm
maintenance-BMnorm
Tdoub-BMchanged
maintenance-BMchanged

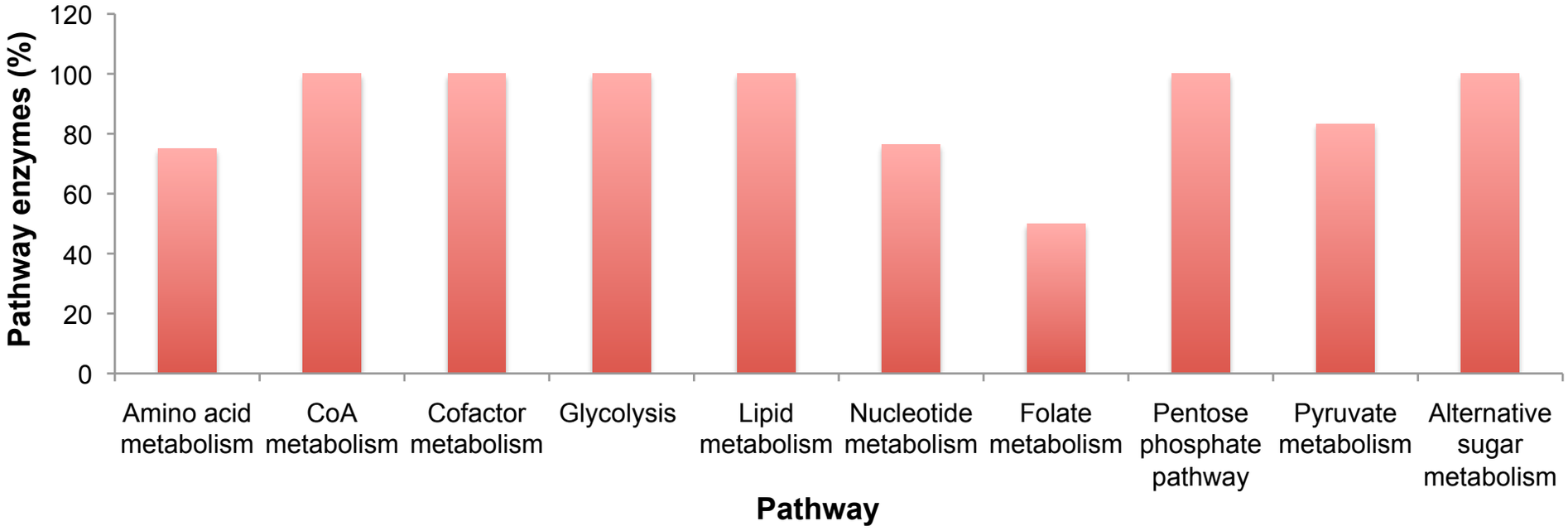

A
Concentration (mg/ml)
0.07
0.06
0.05
0.04
0.03
0.02
0.01
0
0
50
100
150
200
Time (hours)
in vivo protein
fitted protein
B
S5
Concentration (mM)
70
60
50
40
30
20
10
0
0
50
100
150
200
Time (hours)
average glucose
C
Concentration (mM)
120
100
80
60
40
20
0
0
50
100
150
200
Time (hours)
in vivo lactate
fitted lactate
D
Concentration (mM)
18
16
14
12
10
8
6
4
2
0
0
50
100
150
200
Time (hours)
in vivo acetate
fitted acetate
E
lac:ace ratio
18
16
14
12
10
8
6
4
2
0
0
50
100
150
200
Mid interval time (hours)

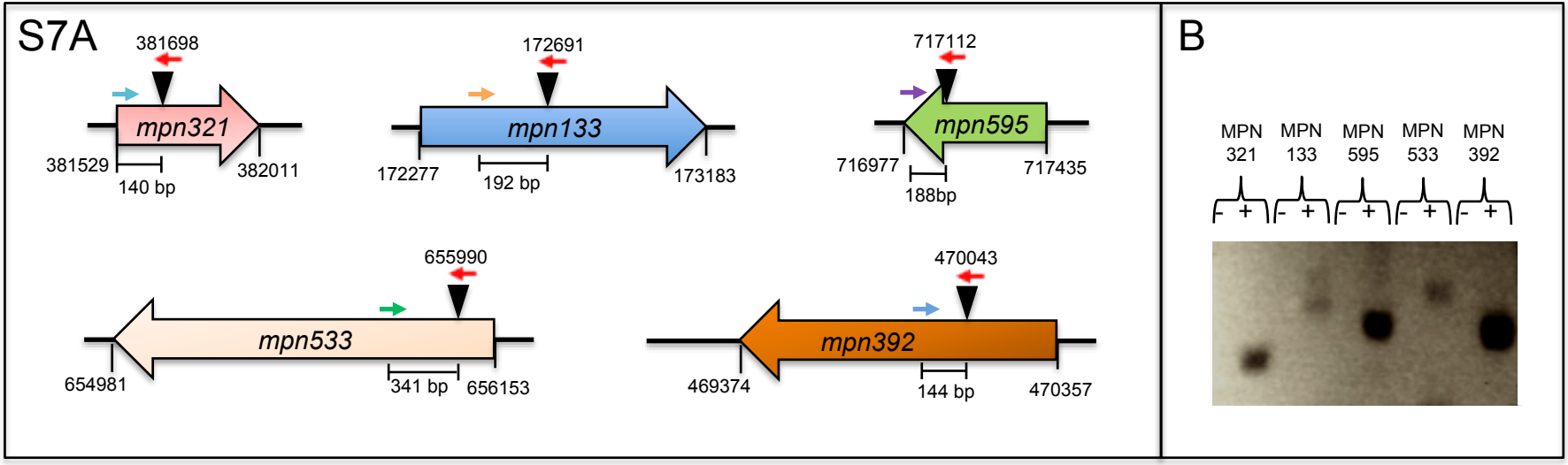
S6A



not analyzed metabolic enzymes
analyzed metabolic enzymes
non-metabolic genes

B




S7A
381698
mpn321
381529
140 bp
382011
172691
mpn133
172277
192 bp
173183
717112
mpn595
716977
188bp
717435
655990
mpn533
654981
341 bp
656153
470043
mpn392
469374
144 bp
470357
B
MPN 321
MPN 133
MPN 595
MPN 533
MPN 392
- + - + - + - + - +

S8