Molecular Systems Biology 9; Article number 653; doi:10.1038/msb.2013.6
Citation: ***Molecular Systems Biology*** **9:653**
www.molecularsystemsbiology.com

molecular systems biology

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

Judith AH Wodke[1,2,3], Jacek Puchałka[4,11], Maria Lluch-Senar[1,2], Josep Marcos[5,6], Eva Yus[1,2], Miguel Godinho[4,7], Ricardo Gutiérrez-Gallego[5,6], Vitor AP Martins dos Santos[4,8,9], Luis Serrano[1,2,10,*], Edda Klipp[3,*] and 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 and [10] Institució Catalana de Recerca i Estudis Avançats (ICREA), Barcelona, Spain
[11]Present address: Dr. von Hauner Children's Hospital, Department of Pediatrics, Ludwig Maximilian University, Munich, Germany
* Corresponding authors. L Serrano or T Maier, 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; E-mail: luis.serrano@crg.eu or E-mail: tobias.maier@crg.eu or E Klipp, Theoretical Biophysics, Humboldt-Universität zu Berlin, Berlin, Germany. Tel.: +49 30 20939040; Fax: +49 30 20938813; E-mail: edda.klipp@biologie.hu-berlin.de

Received 12.6.12; accepted 20.2.13

***Mycoplasma pneumoniae*, a threatening pathogen with a minimal genome, is a model organism for bacterial systems biology for which substantial experimental information is available. With the goal of understanding the complex interactions underlying its metabolism, we analyzed and characterized the metabolic network of *M. pneumoniae* in great detail, integrating data from different omics analyses under a range of conditions into a constraint-based model backbone. Iterating model predictions, hypothesis generation, experimental testing, and model refinement, we accurately curated the network and quantitatively explored the energy metabolism. In contrast to other bacteria, *M. pneumoniae* uses most of its energy for maintenance tasks instead of growth. We show that in highly linear networks the prediction of flux distributions for different growth times allows analysis of time-dependent changes, albeit using a static model. By performing an *in silico* knock-out study as well as analyzing flux distributions in single and double mutant phenotypes, we demonstrated that the model accurately represents the metabolism of *M. pneumoniae*. The experimentally validated model provides a solid basis for understanding its metabolic regulatory mechanisms.**

Molecular Systems Biology **9:** 653; published online 2 April 2013; doi:10.1038/msb.2013.6
Subject Categories: metabolic and regulatory networks; microbiology & pathogens
Keywords: biomass composition; energy metabolism; *in silico* knock-outs; metabolic modeling; *Mycoplasma pneumonia*

Introduction

Representing cellular networks by mathematical models and gaining novel biological insights from subsequent *in silico* analysis and iterative experimental validation are major hallmarks of systems biology. The integration of experimental data from different sources into suitable mathematical models poses a formidable challenge, but allows to place metabolites and enzymes into their network context and to extract biologically relevant information for the examined system (Kitano, 2002). For this purpose, constraint-based modeling approaches that allow the determination of possible network flux distributions provide a useful framework (Fell and Small, 1986; Savinell and Palsson, 1992a, b; Feist *et al*, 2009; Oberhardt *et al*, 2009).

Flux balance analysis (FBA) is a mathematical method to determine metabolic fluxes within a constraint-based model, that is, fulfilling the steady-state condition. Thereby, the flux distribution is optimized toward so-called objective functions, commonly energy production or growth in form of biomass production, for a given set of available nutrients (Varma and Palsson, 1994a; Kauffman *et al*, 2003; Reed and Palsson, 2003). FBA has been successfully used to, for example, predict the effect of mutations on *Escherichia coli* growth rates (Edwards and Palsson, 2000), to predict active pathways under different growth conditions (Covert *et al*, 2001), to improve biotechnology applications (Puchalka *et al*, 2008), and to understand infection processes (Oberhardt *et al*, 2008). More recently, *in vivo* reaction directionalities were assigned to the *E. coli* metabolic network, considering quantitative metabolite measurements and associated fluxes (Fleming *et al*, 2009). In the past year, several tools for constraint-based modeling, including tools for data integration and extraction of relevant

information from the modelled system, have been developed (Fleming and Thiele, 2011; Schellenberger *et al*, 2011; Thorleifsson and Thiele, 2011). In addition, the first reconstruction of the human metabolism has been published (Duarte *et al*, 2007; Rolfsson *et al*, 2011) being the largest metabolic reconstruction so far.

However, the sheer complexity of living systems and the scarce availability of unbiased, large-scale quantitative data often prevent comprehensive organism-wide studies. Thus, knowledge gain from modeling approaches is limited due to the difficulties in setting up validation experiments for specific model predictions. In recent years, the bacterium *Mycoplasma pneumoniae*, a human pathogen preferentially colonizing the pulmonary epithelium (Waites and Talkington, 2004), has been established as model organism for systems biology, providing the research community with detailed quantitative information on its genome, transcriptome, proteome, and metabolism (Güell *et al*, 2009; Kühner *et al*, 2009; Yus *et al*, 2009; Maier *et al*, 2011). *M. pneumoniae* combines several favorable properties for organism-wide analyses. It evolved by massive genome reduction, resulting in a single, small chromosome of only 816 394 base pairs, encoding 689 proteins (Himmelreich *et al*, 1996; Dandekar *et al*, 2000). Its expressed proteome is of low complexity, spanning only three orders of magnitude in abundance (Maier *et al*, 2011). As a consequence of reduced genome and parasitic life, *M. pneumoniae* lacks many metabolic pathways, forcing it to acquire the necessary cell building blocks, among them amino acids, nucleobases, and fatty acids, from the environment (Yus *et al*, 2009). For ATP generation, *M. pneumoniae* relies on simple organic acid fermentation, due to the absence of TCA cycle and a functional respiratory chain (Pollack *et al*, 1997; Yus *et al*, 2009). The lack of most anabolic processes and rescue pathways known from more complex organisms and, consequently, the expected high linearity of its metabolic network render *M. pneumoniae* an ideal organism to study basic metabolic functions and to dissect energetic expenses. In addition, it can be grown autonomously in a laboratory environment and basic genetic tools, such as transposon mutagenesis to study gene essentiality, are available.

We present here a genome-scale constraint-based model of the *M. pneumoniae* metabolism (*i*JW145—*in silico* model including *145* genes, designed by *JW*). We established the biomass composition of an average *M. pneumoniae* cell based on quantitative experimental data. Performing FBA for growth simulations under a wide variety of conditions, we explored and characterized the metabolic network by an iterative cycle between model predictions and their experimental validation. We corrected the metabolic network annotation and the functional annotation of three metabolic enzymes. In addition, we quantitatively dissected the *M. pneumoniae* energy metabolism. We show that *M. pneumoniae* dedicates most of its energy to cellular homeostasis and not growth, as suggested for other bacteria (Schuetz *et al*, 2012), possibly as a consequence of its small size and its parasitic life style. We analyzed carbon fluxes *in vivo* and predicted metabolic fluxes *in silico*. Finally, we proved the predictive capacity of the model by predicting gene essentiality with high accuracy (95%) and specificity (98%). The analysis of the derived single and double mutant phenotypes provided insight into the adaptive capabilities of the metabolic network.

The presented model and the biological findings herein reported can be used for the design of future experiments and the development of novel engineering tools. Furthermore, this model will provide the basis to design dynamic models for metabolic sub-networks and to relate the gene-regulatory and the protein interaction networks to metabolism in *M. pneumoniae*.

Results

Model generation

A constraint-based metabolic reconstruction is a union of (i) a stoichiometric metabolic model, (ii) a set of constraints for metabolic fluxes. and (iii) a list of genes responsible for the catalysis of reactions included in the model. Based on a curated wiring diagram for the metabolism of the bacterium *M. pneumoniae* (Yus *et al*, 2009), we built a genome-scale constraint-based metabolic reconstruction (*i*JW145_reconstruct—initial reconstruction of the network, Figure 1). The main subsystems of the metabolic network are energy producing pathways, amino acid, nucleotide, lipid, and cofactor metabolism as well as transport reactions. In addition, we considered RNA, DNA, and protein biosynthesis as part of the metabolic network (Supplementary Figure S1; Supplementary Table S1).

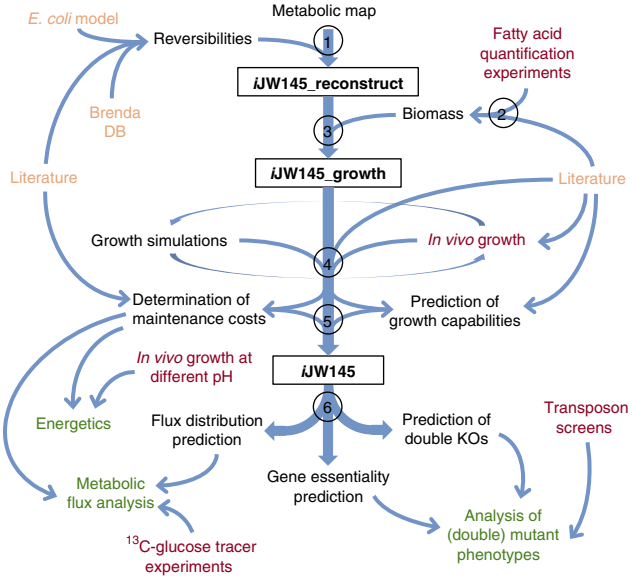

Figure 1 Workflow: 1—Reconstruction of the metabolic network based on the metabolic map from Yus *et al* (2009) and integration of reaction reversibilities ⩾ *i*JW145_reconstruct; 2—experimental analysis of the fatty acid composition in *M. pneumoniae* to allow; 3—Definition of the biomass composition ⩾ *i*JW145_growth; 4—Verification/correction of the network structure based on an iterative cycle between growth simulations and their comparison to *in vivo* findings and literature data; 5—Prediction of growth capabilities for different nutrient conditions (alternative sugar sources) as well as analysis of energy homeostasis *in silico* and subsequent experimental validation (pH stress experiments) ⩾ *i*JW145; 6—Model evaluation and application: (i) prediction of flux distributions *in silico*, (ii) *in silico* knock-out simulations to predict gene essentiality validated with transposon screens, and (iii) prediction of double knock-out phenotypes; experimental data input is shown in red, the different states of the analysis are shown in black, and model outputs are shown in green.

Missing transport reactions were added and reaction reversibilities for the metabolic network were defined (Supplementary Table S2; Supplementary information). We assessed the correct connectivity of the reconstructed network by individually testing the feasibility of the production of all metabolic intermediates *in silico*. Contrary to more complex organisms, the *M. pneumoniae* metabolic network is predominantly composed of linear pathway modules and with the exception of ubiquitous cofactors, such as AMP, ADP, ATP, H^+, H_2O, NAD^+, NADH, P_i, PP_i, only few metabolites interconnect the core metabolic routes (Supplementary Figure S1; Supplementary Table S3). For our study, this low connectivity of the *M. pneumoniae* metabolism facilitates the analysis of inter-pathway crosstalk and enables the prediction of metabolic changes in response to environmental perturbations. Particularly, the absence of most catabolic and also a high number of anabolic routes in *M. pneumoniae* allows the direct relation of external metabolite measurements to intracellular metabolic fluxes and corresponding catabolic activity.

A possible connection of sugar metabolism to aspartic acid via its direct precursor oxaloacetate, which can be synthesized from pyruvate or malate, has been suggested (Manolukas *et al*, 1988). However, in an organism-wide approach to quantify cellular proteins of *M. pneumoniae*, no proof for the presence of the proteins catalyzing the corresponding reactions has been found (Maier *et al*, 2011). To assure the completeness of the reconstructed network, we validated the proposed connection *in vivo* by determining the carbon flux from glycolysis into aspartate. To this end, we monitored the incorporation of ^{13}C-labelled carbon into aspartate. In agreement with previous results (Yus *et al*, 2009), even after 96 h of growth in medium containing $^{13}C_6$-labelled glucose as sole carbon source, no increase in heavy isotope-labelled reporter ions for aspartate was observed (Supplementary Figure S2A). Thus, for *M. pneumoniae* cells grown in rich medium we could discard a link between glycolysis and amino-acid metabolism involving aspartate as connecting metabolite.

To be able to predict metabolic flux distributions and to simulate cell growth with FBA, an accurate, quantitative representation of the biomass composition of an average *M. pneumoniae* cell is required. Therefore, experimental data on DNA content, RNA composition, and protein abundances were considered (Table I; Supplementary information) (Güell *et al*, 2009, 2011; Maier *et al*, 2011). For all three classes of macromolecules, we defined an artificial molecule reflecting their average cellular composition (Supplementary information). To accurately represent the composition of the cell membrane in our model, we experimentally analyzed the fatty acid profile of *M. pneumoniae* and quantified the fatty acid composition of the cytosol, the surrounding growth medium, as well as the cytoplasmic membrane directly (Supplementary Figure S3; Supplementary information). We found predominantly fatty acid chains of 16 and 18 carbons length in both the membrane and the cytoplasm (Supplementary Figure S3A). Based on our results, we designed an artificial molecule representing the average fatty acid composition to describe lipids in the *M. pneumoniae* biomass equation (Supplementary information). To account for the experimentally shown essentiality of vitamins and other cofactors, such as FAD, NAD+/NADH or folate (Yus *et al*, 2009), the end products of the secondary metabolism pathways were included into the biomass qualitatively, that is, in small arbitrary quantities, since a sensitivity analysis showed that a 10-fold change in cofactor amounts does not

Table I Biomass composition

Biomass component	Biomass fraction in % of total cell mass	Biomass fraction in mmol/g of cells	Biomass fraction in molecules/cell	Quantity determined by
DNA	5.00	1624	15 775	Sequence
RNA	6.50	2015	19 573	Sequence
protein	62.00	16 049	155 892	Sequence
ACP	0.003	3	29	LC-MS[a]
Glycolipid	10.00	63 702	618 768	GC/MS (fatty acid chains)
Phosphatidic acid	10.00	148 168	1 439 228	GC/MS (fatty acid chains)
Glycine	0.07	9220	89 558	GC/MS
L-alanine	0.09	9824	95 425	GC/MS
L-arginine	0.07	3913	38 009	GC/MS
L-asparagine	0.001	54	525	GC/MS
L-aspartate	0.12	9318	90 510	GC/MS
L-cysteine	0.001	67	651	GC/MS
L-glutamate	0.27	18 651	181 166	GC/MS
L-glutamine	0.004	294	2856	GC/MS
L-histidine	0.04	2422	23 526	GC/MS
L-isoleucine	0.02	1858	18 048	GC/MS
L-leucine	0.28	21 313	207 024	GC/MS
L-lysine	0.03	1741	16 911	GC/MS
L-methionine	0.02	1370	13 307	GC/MS
L-phenylalanine	0.08	5122	49 752	GC/MS
L-proline	0.08	6837	66 411	GC/MS
L-serine	0.03	3202	31 103	GC/MS
L-threonine	0.03	2489	24 177	GC/MS
L-tryptophan	0.04	1864	18 106	GC/MS
L-tyrosine	0.02	1366	13 269	GC/MS
L-valine	0.03	2793	27 130	GC/MS
Adenosine	0.05	1981	19 242	GC/MS
Cytidine	0.01	503	4886	GC/MS
Guanosine	0.05	1657	16 095	GC/MS
Thymidine	0.03	1242	12 064	GC/MS
Uridine	0.06	2541	24 682	GC/MS
Orthophosphate (Pi)	0.40	41 537	403 474	Colorimetric assay[b]
Thiamin diphosphate	0.00	100	971	Function
NADPH	0.01	100	971	Function
NADP+	0.01	100	971	Function
CoA	0.01	100	971	Function
FAD	0.01	100	971	Function
5fTHF	0.005	100	971	Function
Pyridoxal phosphate	0.002	100	971	Function
S-adenosyl-L-met	0.004	100	971	Function
CDP-CHO	0.0001	100	971	Function
G6P	4.51	174 748	1 814 216	100% biomass—rest

Abbreviations: ACP, acyl carier protein; fTHF, formyl tetrahydrofolate; CDP-CHO, cytidine diphosphate choline; G6P, glycerol-6-phosphate.
The detailed process of biomass definition is described in Supplementary information.
[a]Maier *et al* (2011).
[b]Data from *E. coli* (Amin and Peterkofski, 1995; Neidhardt, 1996).

significantly influence on the growth rate (Supplementary Figure S4; Supplementary Table S4; Supplementary information). Other cofactors, for example organic phosphate, were included based on the literature information (Table I; Supplementary information).

The biomass equation defining the macromolecular composition of an average *M. pneumoniae* cell in a general form reads:

Equation 1: Biomass (eq. 1)

DNA + RNA + proteins + lipids + bases + amino acids + fatty acids + cofactors → Biomass

The assembly of the stoichiometric network (iJW145_reconstruct) and the assignment of reaction reversibilities together with the definition of the biomass composition provided us with a version of our *M. pneumoniae* model that was able to simulate growth: iJW145_growth (Figure 1).

Model assessment and refinement

To validate the network structure of iJW145_growth and to avoid errors in the reconstruction (Reed *et al*, 2006), we determined the network behavior under different nutrition conditions. To this end, we defined several flux constraint sets based on own experimental data (Supplementary Figure S5) and the literature information (Supplementary Table S5; Supplementary information). As a general strategy, we minimized the possible number of flux constraints to not restrict the solution space of the model but to keep a high predictive capacity (Edwards *et al*, 2002; Covert and Palsson, 2003; Price *et al*, 2004). Maximum constraints were set to limit the uptake of all sugar sources and arginine (reflecting the available nutrients for the growth conditions and preventing unlimited ATP synthesis from arginine, Supplementary information), while minimum constraints account for mRNA and protein degradation rates, detoxification, and maintenance energy costs (Supplementary Table S5; Supplementary information).

Initial simulations for different growth conditions verified the network annotation (Supplementary Figure S1; Supplementary information). In addition, the observed *in silico* fluxes confirmed experimental data on mRNA expression of metabolic proteins (Güell *et al*, 2009) and for the design of a defined medium (Yus *et al*, 2009) (Supplementary information). However, we also identified several conflicts between model predictions and available experimental results (Hames *et al*, 2009; Yus *et al*, 2009). These conflicts were resolved by an iterative process, involving protein sequence comparison, additional experiments, literature mining, and repeated simulations using adjusted constraints (Figure 2;

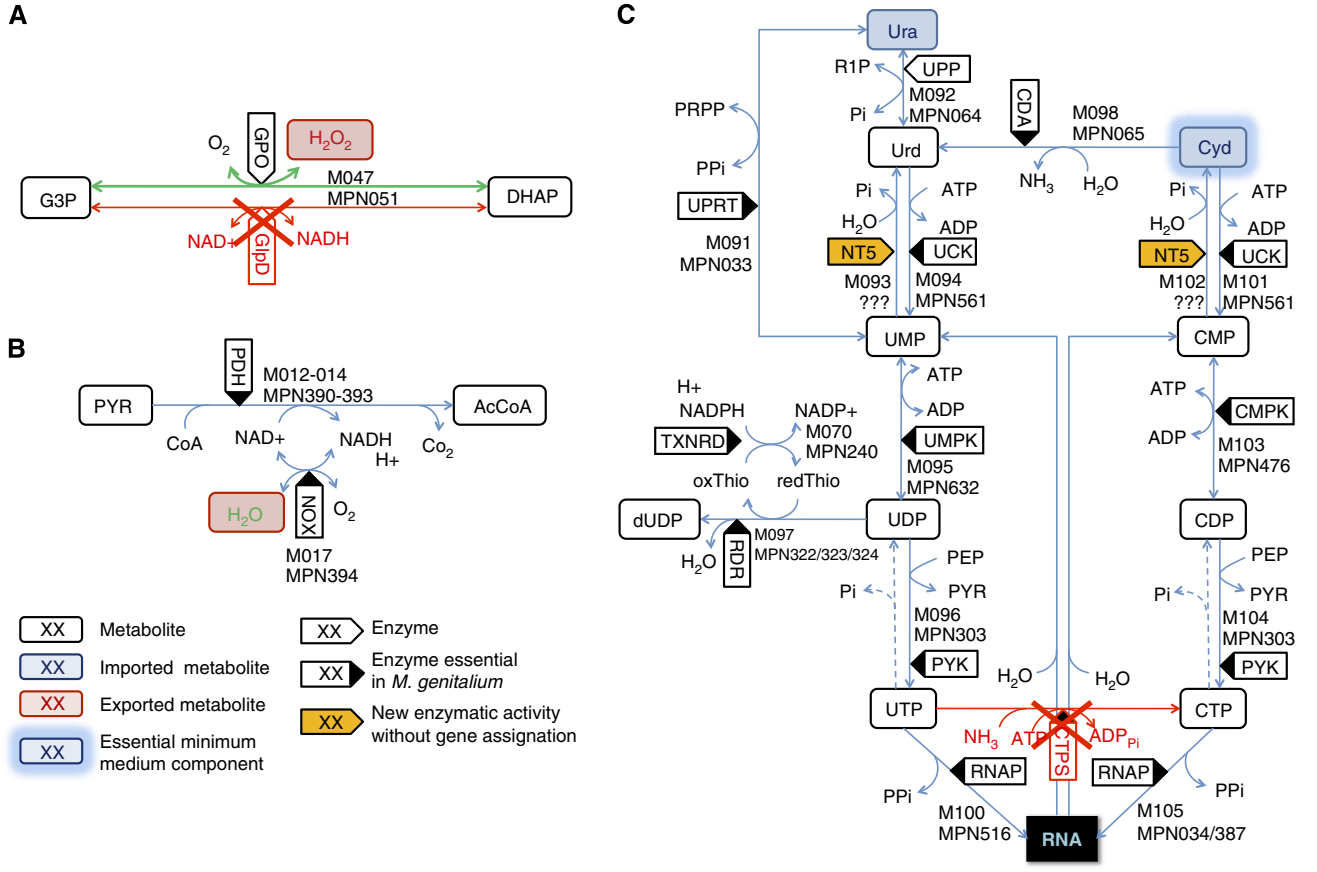

Figure 2 Functional re-annotation of metabolic routes and reactions in *M. pneumoniae* based on experimental data and sequence comparison analysis (Supplementary information). (**A**) Removal of a glycerol 3-phosphate dehydrogenase (GlpD) catalyzed glycerol phosphate oxidase reaction (MPN051) due to a redox balance conflict. (**B**) Functional re-annotation of NADH oxidase (NOX, MPN394) as an O_2 producing enzyme. (**C**) Removal of a CTP synthase (CtpS)-catalyzed reaction in pyrimidine metabolism due to conflicts between *in silico* predictions and *in vivo* growth as well as lack of genetic evidence.

Supplementary information). In summary, the comparison of model results and experimental data referring to cellular redox states and to nucleotide metabolism led to the functional re-annotation of three metabolic enzymes: MPN051—glycerol phosphate oxidase (Figure 2A), MPN394—NADH oxidase (Figure 2B), MPN256—no CTP synthase activity (Figure 2C), and guided the correction of the wiring diagram (correction of one reaction, deletion of two reactions, Supplementary information).

Simulation of biomass production

Batch culture growth experiments on glucose lead to a measurable decrease in medium pH (pH 7.8–5.5 during a 4-day growth course) due to secretion of lactic and acetic acid (Yus *et al*, 2009). A shift from energetically favorable acetic acid production during early growth (4 ATP/glucose) toward predominantly lactic acid production during later growth stages (two ATP/glucose) has been observed (Supplementary Figure S5C–E). Concurrently, cellular lactate dehydrogenase levels (MPN674, LDH) increased five-fold from 203 copies to above 1000 copies per cell during 4 days of growth (Maier *et al*, 2011). To correctly represent this metabolic shift in our model and to account for further carbon incorporated into glycolysis from other sources than glucose, we directly constrained the energetically favored acetic acid production. To this end, we fitted the maximum constraint for acetate production to the experimentally determined ratio of secreted lactic and acetic acid (Supplementary Figure S5E).

In laboratory experiments, *M. pneumoniae* can be grown in rich and defined medium (Chanock *et al*, 1962; Yus *et al*, 2009). In both conditions, it metabolizes glucose as major carbon and energy source. Alternatively, a variety of additional reduced carbon compounds, such as fructose, mannose, ribose, ascorbic acid, glycerol, glycerol 3-phosphate (G3P) (Yus *et al*, 2009), and glycerophosphocholine (G3PC) can be metabolized (Schmidl *et al*, 2011). In addition, arginine can be used to produce ATP but we did not consider it as alternative carbon source since (i) its contribution to energy production is negligible, (ii) *in vivo* it did not permit growth (Yus *et al*, 2009), and (iii) only one enzyme involved in arginine fermentation has been detected (Maier *et al*, 2011). For the other carbon compounds, we adjusted the flux constraints of the model according to rich and defined growth medium compositions as described in Yus *et al* (2009) and applied FBA. We verified *in silico* the experimentally tested metabolic capabilities of *M. pneumoniae* (Table II; Yus *et al*, 2009; Schmidl *et al*, 2011). Under defined medium conditions (Supplementary Table S5) the model predicted growth on glucose, mannose, mannitol, ribose, and ascorbate. However, these strict constraints did not allow growth on fructose, glycerol, G3P, and G3PC (Table II). We identified the inability to provide the pentose phosphate pathway precursor fructose 6-phosphate (F6P) for *de novo* nucleotide synthesis as cause for the observed *in silico* growth limitations (Supplementary Figure S1).

While *in silico* there is no difference in energy yield for *M. pneumoniae* grown on different carbon sources (assuming the same amount of carbon taken up for each sugar source, Supplementary information), *in vivo* the doubling times differed significantly and only glucose and mannose allowed robust growth (Yus *et al*, 2009). This apparent discrepancy can be explained by the abundances of the respective uptake and processing systems. While the glucose-specific uptake protein (MPN207) has high copy numbers (~385/cell), the known transport proteins for other sugars (fructose, ribose, ascorbate, mannitol, mannose, glycerol, and G3P) are 8–20 times less abundant (Maier *et al*, 2011). In fact, to allow relevant growth of *M. pneumoniae* on fructose *in vivo* the cells had to be adapted over several serial passages and showed significant overexpression of the proteins involved in fructose import and metabolism (Yus *et al*, 2009).

Table II Growth simulations for alternative carbon sources: comparison of the metabolic capabilities of *M. pneumoniae* when grown on different carbon sources *in silico* and *in vivo*

	Rich medium		Defined medium
Sugar	*In silico* growth	*In vivo* growth	*In silico* growth
Glucose	+	+[a]	+
Fructose	+	+[a]	0
Mannose	+	+[a]	+
Mannitol	+	−[a]	+
Ribose	+	(+)[a]	+
Ascorbate	+	(+)[a]	+
Glycerol	+	(+)[a]	0
G3P	+	Not tested	0
G3PC	+	+[b]	0

Abbreviations: G3P, Glycerol-3-phosphate; G3PC, glycerophosphocholine.
'+'—growth, '(+)'—at least catabolic activity (growth not examined explicitly), '0'—no growth, but catabolic activity, '−'—no growth and no catabolic activity.
[a]Yus *et al* (2009).
[b]Schmidl *et al* (2011).

Cellular energy balance

Using constraints derived from non-linear fittings to experimentally obtained metabolite quantification data (glucose consumption as well as acetic and lactic acid production for defining the carbon uptake and the acetic acid production, and total protein increase to restrict the growth rate) from batch culture growth (Supplementary Figure S5B–E), we applied FBA to our refined model to determine cellular ATP production at different growth stages (24, 36, 48, and 60 h). We found that during mid exponential growth (36 h after inoculation) ~60 000 ATP molecules per cell and second are synthesized *in silico* (Supplementary Table S6; Supplementary Materials and methods).

Next, we quantitatively assessed the contribution of the available cellular energy to biomass production (eq. 1, Biomass) and other functions specified in the model, for example, protein turnover (Supplementary information). For this purpose, we compared *in silico* and *in vivo* doubling times (Figure 3A) at different growth stages. We found that simulation-based results (between 2.3 and 3.8 h) differed significantly from experimentally measured values during the exponential phase in batch culture growth (between 19.7 and 59.7 h; Figure 3A), as well as from previous reports on microscope studies of single cell division (~8 h; Seybert

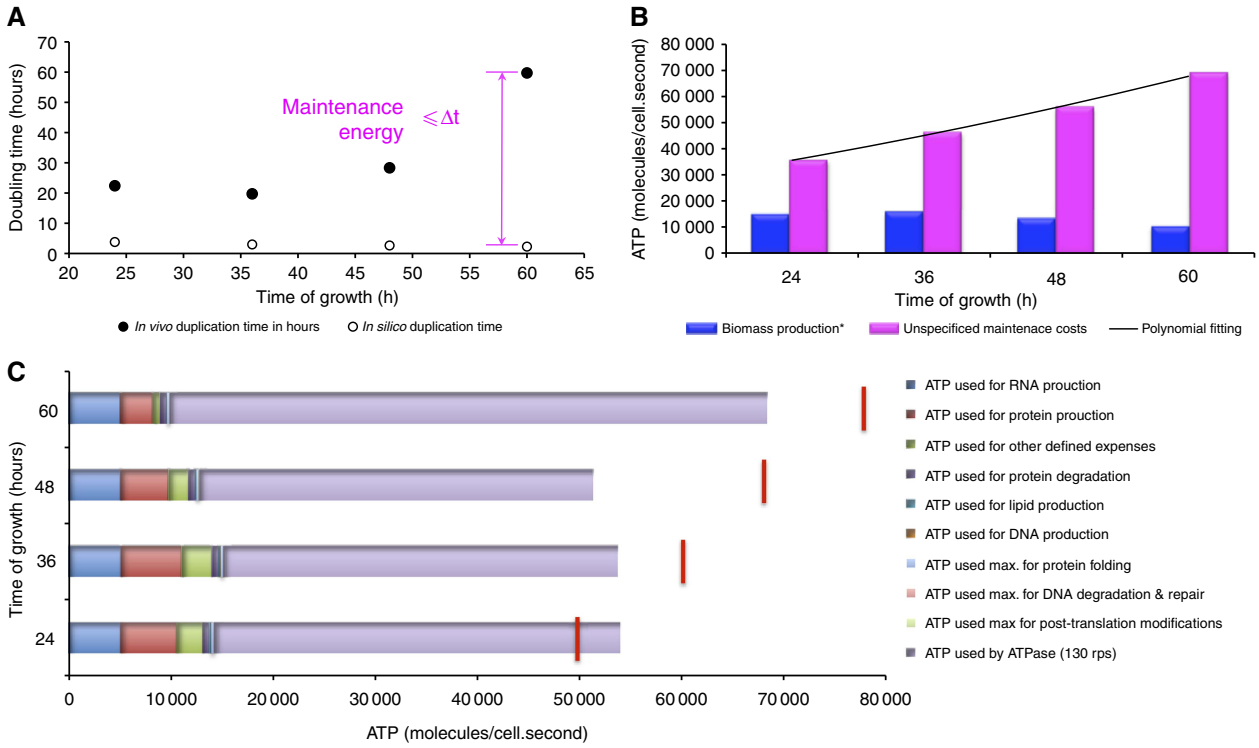

Figure 3 *In silico* analysis of energy usage in *M. pneumoniae*: (**A**) Comparison of *in vivo* (filled circle) and *in silico* (empty circle) doubling times of *M. pneumoniae* at different stages during growth in batch culture. Based on the difference in doubling times Δt, the maintenance energy has been determined; (**B**) Contribution of biomass + defined cellular functions (blue) and yet undefined cellular functions (pink) to the total energy usage. A polynomial fitting to *in silico* maintenance costs (black line) used for back-calculating constraints for different time points (for the variables of the fitting, see Supplementary Table S8); (**C**) comparison of the *in silico* energy consumption for synthesis of the different biomass building blocks and the estimated expenses (upper boundaries) for different maintenance tasks. The red vertical lines indicate the total amount of ATP produced *in silico* at each simulated time of growth.

et al, 2006). These results suggest the existence of yet undefined ATP consuming reactions, contributing considerably to cellular energy homeostasis in batch culture growth.

To quantify the contribution of those additional cellular ATP sinks to the energy metabolism of *M. pneumoniae*, we defined a single unspecific energy consuming reaction. The minimum constraint of this reaction was fitted manually for each simulated time point in order to allow reproduction of *in vivo* doubling times. Strikingly, 71–88% of the total ATP available *in silico* is used for cellular tasks not directly involved in biomass production (Figure 3B; Supplementary Table S6). Consequently, depending on the growth time, only between 12 and 29% of the total available energy is used for cell growth. More precisely, at 36 h of batch culture growth *M. pneumoniae* uses 9.8% of its total energy for protein production and degradation (protein half-life of 23 h; Maier *et al*, 2011) while 8.4% of the total ATP is dedicated to RNA production (mRNA half-life of 1 min; Maier *et al*, 2011) and not even 0.1% to DNA synthesis (Figure 3C; Supplementary Table S6; Supplementary information). Lipid production consumes 0.5% of the available ATP while 5.9% are used for metabolic precursor uptake and the subsequent synthesis of secondary metabolites, such as vitamins, FAD, NAD(H), P_i, folate, and other defined functions such as detoxification (Figure 3C; Supplementary Table S6; Supplementary information).

To characterize the yet undefined energy sinks, we further either classified them as growth-associated maintenance (GAM) and non-growth-associated maintenance (NGAM) (Pirt, 1965; Varma and Palsson, 1994b). To estimate the contribution of GAM to the total energy cost, we calculated upper boundaries for the ATP expenses related to post-translational modifications (0.01%), DNA repair (0.01%), and chaperone-assisted protein folding (0.25%) based on the available literature and experimental data (Figure 3C, Supplementary Table S6; Supplementary information; Drake *et al*, 1998; Naylor and Hartl, 2001; Maier *et al*, 2011; van Noort *et al*, 2012). Additionally, considering protein turnover costs and other defined expenses mainly related to the uptake and processing of cellular building blocks and reaction cofactors (as defined in the model), total expenses on GAM account for a maximum of 7.6% of the total available cellular energy.

Systematic literature screening for further cellular energy sinks identified proton translocation by the cellular ATPase as most significant quantifiable NGAM task (Kobayashi, 1985, Hutkins and Nannen, 1992). In fact, for lactic acid bacteria, the proton ATPase is the major contributor to maintain pH homeostasis (for a review, see Hutkins and Nannen, 1992). In these bacteria, the ATPase is predominantly involved in creating an optimal proton gradient across the cytoplasmic membrane to allow nutrient import and to maintain the intracellular pH in an acidic environment (Kobayashi, 1985; Hutkins and Nannen, 1992; Moreno *et al*, 1998). Considering the experimentally determined ATPase abundance (between 99 and 150 ATPase complexes per cell, Supplementary

information; Maier *et al*, 2011) and assuming constant activity at the maximum catalytic rate reported (130 r.p.s.; Wu *et al*, 2010), we estimated the global ATP hydrolysis rate of the *M. pneumoniae* ATPase to be maximally ~38 500 ATP per cell and second, accounting for up to 57% of the total available ATP at simulated 36 h of growth (Figure 3C; Supplementary Table S6; Supplementary information).

To further analyse the contribution of maintenance tasks to cellular energy homeostasis, we challenged growing *M. pneumoniae* cultures by incubating them for 36 h in medium with pre-adjusted pH, ranging from pH 5.5 to 8.0. After harvesting the cultures, we measured growth medium glucose levels and protein content, reflecting cellular energy production and bacterial growth, respectively (Figure 4A; Supplementary information). Comparing protein production with glucose consumption for those cultures revealed that at physiological (high) pH, generated ATP is utilized for growth to a much larger extent than at low pH (Figure 4B; Supplementary information). These findings support our *in silico* energy analysis, particularly the predicted high cellular maintenance costs at later growth stages in batch culture (Figure 3B). Intriguingly, metabolically inactive cells in medium with low pH values resume activity when shifted back to physiological pH (Figure 4C). This indicates that growth arrest in acidic medium is not predominantly caused by cell death but rather by reversible metabolic stalling as a consequence of an unfavorable ion balance for ATP generation. Thus, in *M. pneumoniae* energy metabolism and biomass

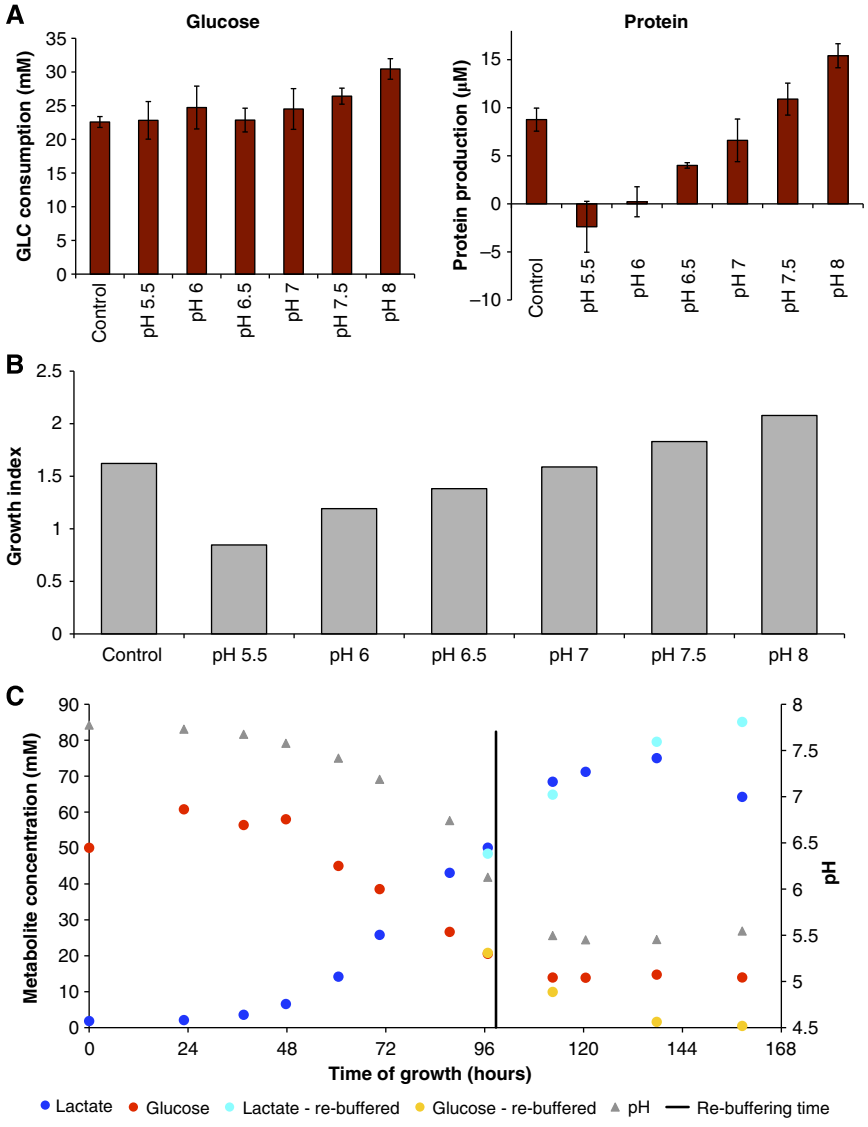

Figure 4 *In vivo* analysis of energy usage. (**A**) Glucose consumption (left) and protein production (right) for *M. pneumoniae* batch cultures grown for 36 h in medium with pre-set pH. *M. pneumoniae* cells were grown for 48 h before the experiment. Error bars represent biological triplicates. Control: Cultures with unchanged growth medium. (**B**) Growth index based on the difference between the ratios of protein (in μM, growth indicator) to glucose (in mM) consumption for *in vivo* growth of *M. pneumoniae* grown in medium at different pH. A negative growth index indicates colony shrinking by cell death. (**C**) Medium pH re-buffering experiment. Glucose (red/yellow, left y axis) and lactic acid concentrations (blue/cyan, left y axis) were determined with enzymatic assays. The pH curve for the not re-buffered samples is shown (grey triangles, right y axis), as well as the time at which the medium pH was adjusted back from pH 6 to 7.7 by NaOH titration (re-buffering time). Source data for this figure is available on the online supplementary information page.

production are decoupled depending on growth medium acidity. Below a critical pH, cells enter into a reversible dormant metabolic state.

Taken together, we used our constraint-based model and experimental results to quantitatively analyze the global energy balance of *M. pneumoniae*. Considering all quantifiable ATP consuming processes we are able to explain about 75–100% of the total growth stage-dependent cellular energy expenses (Supplementary Table S6). Biomass production itself does account for about 11–22%, GAM for about 2–7%, and NGAM for 57–80% of the total ATP generated. This high fraction of energy dedicated to NGAM on first sight contradicts findings in *E. coli* where GAM expenses are far exceeding NGAM costs ~2.5-fold to ~7-fold (Varma and Palsson, 1994a; Feist *et al*, 2007). However, artificially adjusting *E. coli* doubling times to values measured for *M. pneumoniae* (20 h; Figure 3A) resulted in much higher NGAM expenses in *E. coli*, and a GAM/NGAM ratio approaching the value calculated for *M. pneumoniae* (Supplementary Table S7). Furthermore, our results coincide with a recently published whole-cell model for *M. genitalium* (Karr *et al*, 2012) identifying protein and RNA production as the major ATP sinks in biomass production. Movement, which has been shown to be an energy consuming process in *M. mobile* (Jaffe *et al*, 2004), could not be estimated so far due to lacking information about the exact function of the gliding machinery and associated ATP costs.

It is important to note that in *M. pneumoniae in silico* even during the exponential growth phase 78–89% of the available energy is not directly used for the production of cell building blocks but for cellular homeostasis. This finding could also explain the large discrepancy of 44% between energy production and consumption observed in the whole-cell model for *M. genitalium* (Karr *et al*, 2012).

Prediction of flux distributions for varying conditions *in silico* and *in vivo* analysis of central carbon metabolism

Introducing the maintenance energy sink reaction into our model and setting its constraint as described above completed the construction process of our metabolic model. The final model, *i*JW145, contains 306 reactions connecting 216 metabolites and 145 enzymes (Supplementary Figure S1; Supplementary Table S1). Its capability to simulate growth of *M. pneumoniae in silico* allowed us to reproduce experimentally determined doubling times.

Constraint-based models describe cellular metabolism under steady-state assumptions. Integrating experimental data on metabolic parameters and our information about energy homeostasis, we analyzed changes in metabolism during the exponential growth phase *in silico*. For this purpose, in addition to fitting sigmoidal functions to experimentally measured glucose consumption, protein production, as well as to lactate and acetate secretion during a 96- h time course (Supplementary Figure S5), we fitted a polynomial function to the *in silico* maintenance costs (Figure 3B; Supplementary Table S8). The fitted functions allow the calculation of metabolic constraints for any given time of a 4-day batch culture growth experiment (Supplementary information).

Simulating growth under rich medium conditions with respectively determined sets of constraints (Supplementary Table S5; Supplementary information) permitted the prediction of metabolic flux distributions for different times of the exponential growth phase, thus providing information about the flux changes in batch culture growth (Figure 5A; Supplementary Table S9). To this end, we used representative flux distributions instead of considering multiple optima since

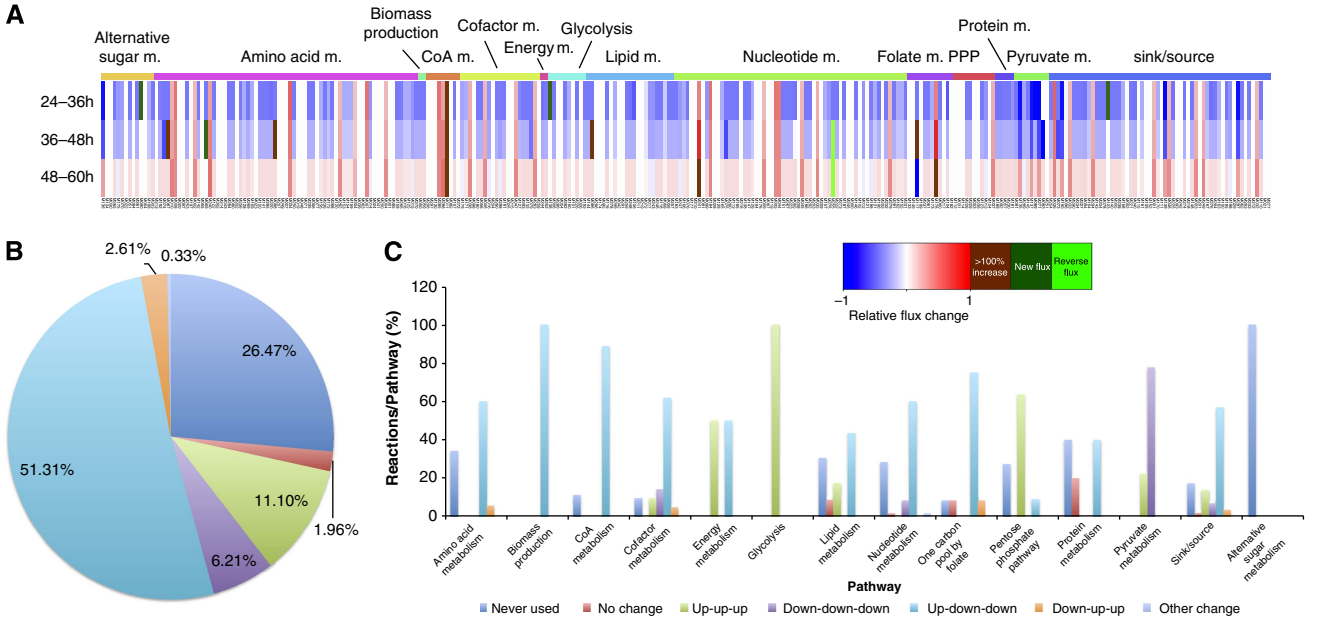

Figure 5 Analysis of *in silico* flux changes along the exponential growth phase: (**A**) *In silico* flux changes from 24 to 36 h, from 36 to 48 h, and from 48 to 60 h. Genes have been clustered according to their pathways (top line of colored boxes with corresponding pathway labels: m. abbreviates metabolism, PPP stands for pentose phosphate pathway). (**B**) Classification of reactions according to their flux changes from 24 to 36 to 48 to 60 h; colors are as indicated in (**C**). (C) Classification of reactions according to their flux changes from 24 to 36 to 48 to 60 h distinguished for the different pathways.

a flux variability analysis (FVA) showed only negligible variability in the majority of the reactions (Supplementary Table S10, see below). We determined those changes between *in silico* flux distributions for 24, 36, 48, and 60 h of batch culture growth and found that the majority of the reactions (51.6%) show the same trend as biomass synthesis, that is, from 24 to 36 h the flux increases and from 36 to 48 h as well as from 48 to 60 h the flux decreases (Figure 5B). Another 2.6% of the fluxes change contrary to biomass synthesis, that is, first decrease (24–36 h) and then increase (36–48–60 h). Those findings support the notion that microorganisms generally optimize for growth (Neidhardt, 1996) and that the reductive genome evolution of *M. pneumoniae* eliminated most cellular functions not associated with growth and survival. For 11.4% of all reactions a constant flux increase is observed while 5.9% of the fluxes constantly decrease over time. The constantly changing reactions either belong to glycolysis, pyruvate metabolism, the energy producing arginine fermentation (amino-acid metabolism) or are associated transport reactions (Figure 5C). This can be explained by the increase in maintenance costs during batch culture growth (Figure 3B) and the subsequent adaptation of the catabolic pathways as well as by the imposed acetate production constraints. Our results suggest the central carbon metabolism to be influenced rather by the external conditions challenging cellular maintenance functions than by the biomass production rate, which further supports our energetics analysis.

Around 1.6% of all reactions have the same flux at all times and two reactions show other changes not being used at all simulated time points. Both of them belong to the nucleotide metabolism. Our FVA revealed that around 14% of all reactions, mostly associated either directly or indirectly (by involving nucleotides as reaction cofactors) to nucleotide metabolism, can be used with differing fluxes to produce optimal solutions for FBA problems (Supplementary Table S10). For example, the two routes to produce deoxy-CDP and deoxy-GDP, respectively, are energetically equal and therefore do not alter growth rate and energy homeostasis, which explains the changes observed during the flux analysis. In general, the FVA revealed that the flux distributions predicted have only very low variability with most of the reactions either allowing no or only very little changes (Supplementary Table S10). A total fraction of 26.1% of the reactions is not used under the simulated rich medium conditions. A higher fraction of unused reactions was observed in amino-acid metabolism, nucleotide metabolism, and pentose phosphate pathway (between 25 and 35%). Hence, these pathways could serve as possible rescue routes for the adaptation to diverse stress conditions. Lipid metabolism also contains 30% non-active reactions, which in addition to being a direct effect of not including cardiolipin in the biomass (its catalyzing reactions are not used) is in agreement with detected protein quantification data of the respective pathway (Maier *et al*, 2011). Due to the simulated conditions, reactions involved in the processing of alternative sugars are not active which is in agreement with the low abundance of involved processing enzymes (Maier *et al*, 2011). In summary, the majority of the reactions of the metabolic network of *M. pneumoniae* are active even under favorable growth conditions reflecting the reduced genome and the simple linear network structure.

Usually, metabolic fluxes are studied in connection with transcriptomics data to reveal regulatory mechanisms, but in *M. pneumoniae* mRNA and protein levels correlate only moderately (Pearson's correlation coefficient of 0.41–0.51; Maier *et al*, 2011). Therefore, we integrated *in silico* flux changes along the exponential growth phase directly with protein abundance changes (Supplementary information). Aligning qualitative changes for protein abundances and fluxes along the exponential growth phase and in addition considering a 25% error rate for protein abundances (a two-fold error was reported by Maier *et al*, 2011), we found that about 86% of all protein abundance changes agree with the *in silico* flux changes (Supplementary information). These findings are in agreement with a recent study analyzing the dynamic adaptation of *Bacillus subtilis* to nutritional shifts (Nicolas *et al*, 2012). The integration of information about post-translational modifications (van Noort *et al*, 2012), however, did not lead to further conclusions about the influence of protein levels on metabolic regulation (Supplementary information).

To further validate the flux distributions predicted by the model, *M. pneumoniae* cells were pulse-fed with heavy isotope-labelled glucose ($^{13}C_6H_{12}O_6$) and the propagation of the labelled carbon atoms through glycolysis was observed by GC–MS (Supplementary information). To this end, we monitored reporter compounds for the pentose phosphate pathway (ribose 5-phosphate (R5P)) and for lipid metabolism (G3P and glucose 1-phosphate (G1P)) (Supplementary Figure S2A). Under rich medium conditions, we found a large excess of heavy isotope reporter ions for all detected intermediates of glycolysis already 15 s (earliest possible time point for reproducible measurements) after supplying ^{13}C-labelled glucose (Supplementary Figure S2B). This finding confirms the high fluxes predicted for glycolysis reactions *in silico*.

The model predicts slow influx from G3P and the pentose phosphate pathway into glycolysis and even slower outflux from glycolysis into lipids via G1P when compared with the overall speed of glycolysis *in silico*. We used R5P, a key intermediate of the pentose phosphate pathway, to check for outflux from glycolysis into the pentose phosphate pathway *in vivo*. Contrasting the high fluxes observed for glycolysis with ^{13}C-carbon saturation after 15 s, we did not observe a significant accumulation of labelled reporter ions for R5P on comparable timescales (Supplementary Figure S2A). However, after about 24 h supply of $^{13}C_6$-labelled glucose, intracellular R5P pools were fully labelled, indicating a slow, albeit steady flux from glycolysis into the pentose phosphate pathway *in vivo*.

Two independent branches connect glycolysis to lipid biosynthesis. The first branch involves G1P, a precursor providing hexose sugars for glycolipid synthesis. G1P gets synthesized from glucose 6-phosphate (G6P) in a reaction catalyzed by the phosphoglucomutase (MPN066, 80 copies per cell; Maier *et al*, 2011). G1P is rapidly detected during flux experiments. However, the conversion of G6P into G1P is reversible and it cannot be excluded that G1P is only used as an extension of the G6P pool. For the second connection, G3P provides the polar head group to which fatty acids are covalently attached by ester bonds during phospho- and glycolipid synthesis (Supplementary Figure S1). When not

imported from the growth medium or obtained from its precursor glycerol, G3P can be generated from dihydroxyacetonephosphate (DHAP) in the corrected GPO catalyzed reaction (compare Figure 2A; MPN051: 62 protein copies per cell; Maier *et al*, 2011). Analysis of the incorporation of labelled reporter ions in G3P revealed a comparatively slow conversion of DHAP into G3P, reaching saturation after 24 h of incubation (Supplementary Figure S2A) suggesting low flux between glycolysis and lipid biosynthesis as compared with flux velocity in glycolysis.

Summing up, our experimental analysis, in consistency with previous suggestions (Yus *et al*, 2009) and predictions from our own model as well as from the whole-cell model for *M. genitalium* (Karr *et al*, 2012), showed that the flux through glycolysis far exceeds those shuffled to adjacent pathways such as the pentose phosphate pathway or lipid biosynthesis (Figure 6). Comparing *in vivo* results to flux distributions predicted *in silico*, we found that the model predicts well the major carbon flux through glycolysis. However, the flux directions between glycolysis and other metabolic pathways (namely between DHAP and G3P or between FBP + GAP and the pentose phosphate pathway) are inverted (Figure 6). This suggests a slight overestimation of the import of ribose and glycerol/G3P, which however should not significantly affect growth rates and overall energy metabolism due to the low fluxes in the branching reactions when compared with the speed of glycolysis (Figure 6).

Prediction of gene essentiality

To test the accuracy of the refined metabolic network on a global scale, we performed an *in silico* knock-out study for genes encoding 131 metabolic proteins (Supplementary information; Supplementary Table S11). We systematically silenced, that is, limited to zero, all reactions catalyzed by the individual gene products and recorded the resulting growth ability for rich medium conditions (Supplementary Table S11; Supplementary information). Seventy-three genes (56% of metabolic enzymes included in the prediction, Supplementary information) were identified as essential, because the

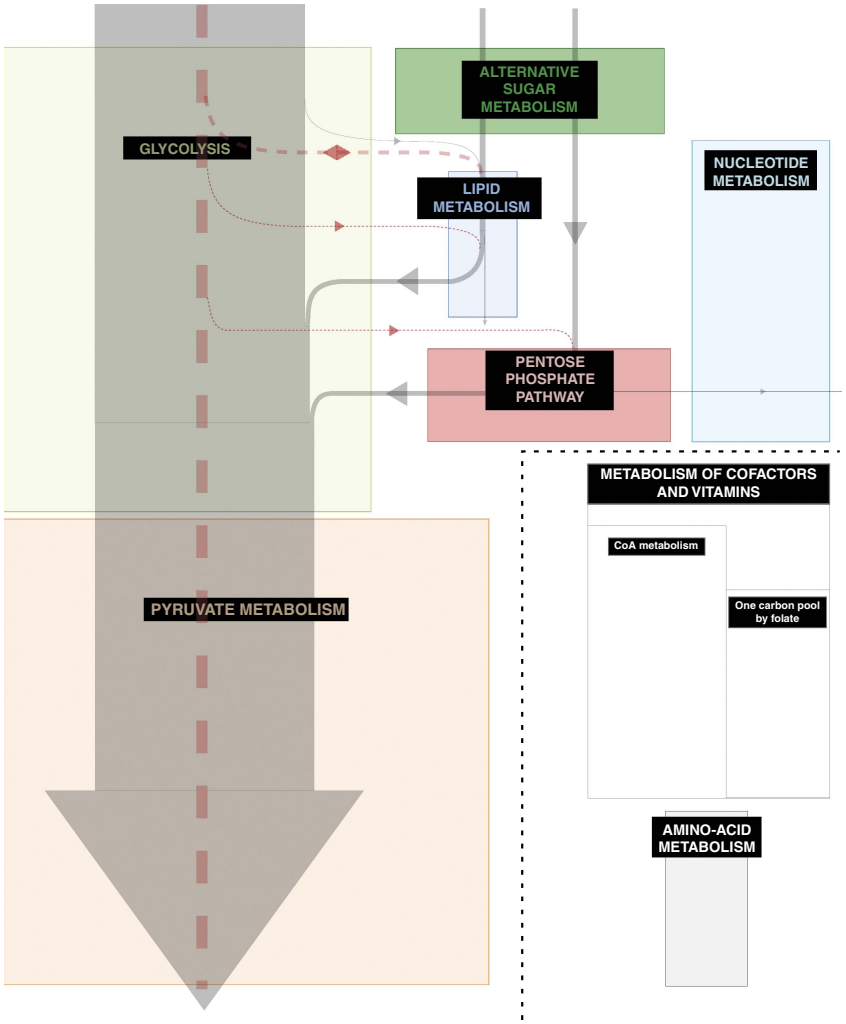

Figure 6 Central carbon metabolism in *M. pneumoniae* is represented by quantitative *in silico* fluxes (grey arrows, arrow widths correspond to *in silico* determined reaction rates) and qualitative *in vivo* fluxes (dashed red arrows, arrow width corresponds to fast or slow saturation of metabolite pools with ^{13}C-labelled carbon, respectively). The size of underlying pathway boxes reflects the average protein copy number determined for each pathway (Maier *et al*, 2011).

respective *in silico* knock-out either led to growth arrest (22%) or rendered the FBA infeasible (34%), that is, at least one minimum constraint could not be satisfied (Figure 7A; Supplementary Table SII). Conversely, 58 artificial gene knock-outs (44%) resulted in objective values larger than zero, either showing no change as compared with the wild type (19%) or with a lower objective value as in the wild type representing a reduced fitness phenotype (25%), thereby predicting these genes to be not essential for growth and survival in *M. pneumoniae* (Figure 7A; Supplementary Table SII).

We evaluated the predictions by comparing them with a genome-wide transposon mutagenesis-based knock-out screen in the closely related bacterium *Mycoplasma genitalium* (Glass *et al*, 2006) (Table III; Supplementary information). This comparison was based on the assignment of functional orthologs in the two mycoplasmas (Supplementary Table S12; Supplementary information). In a first, unbiased analysis of the *in silico* knock-out results using gene essentiality in *M. genitalium* as sole criterion, we achieved 86% accuracy (correct predicted/total predicted) and 98% specificity (true negatives/(true negatives + false positives)) of our *in silico* essentiality prediction (Table III; Supplementary information). In case of a contradiction between prediction and this transposon study, we screened an *M. pneumoniae* transposon library (Halbedel and Stülke, 2007) for respective mutants (Supplementary Figure S7).

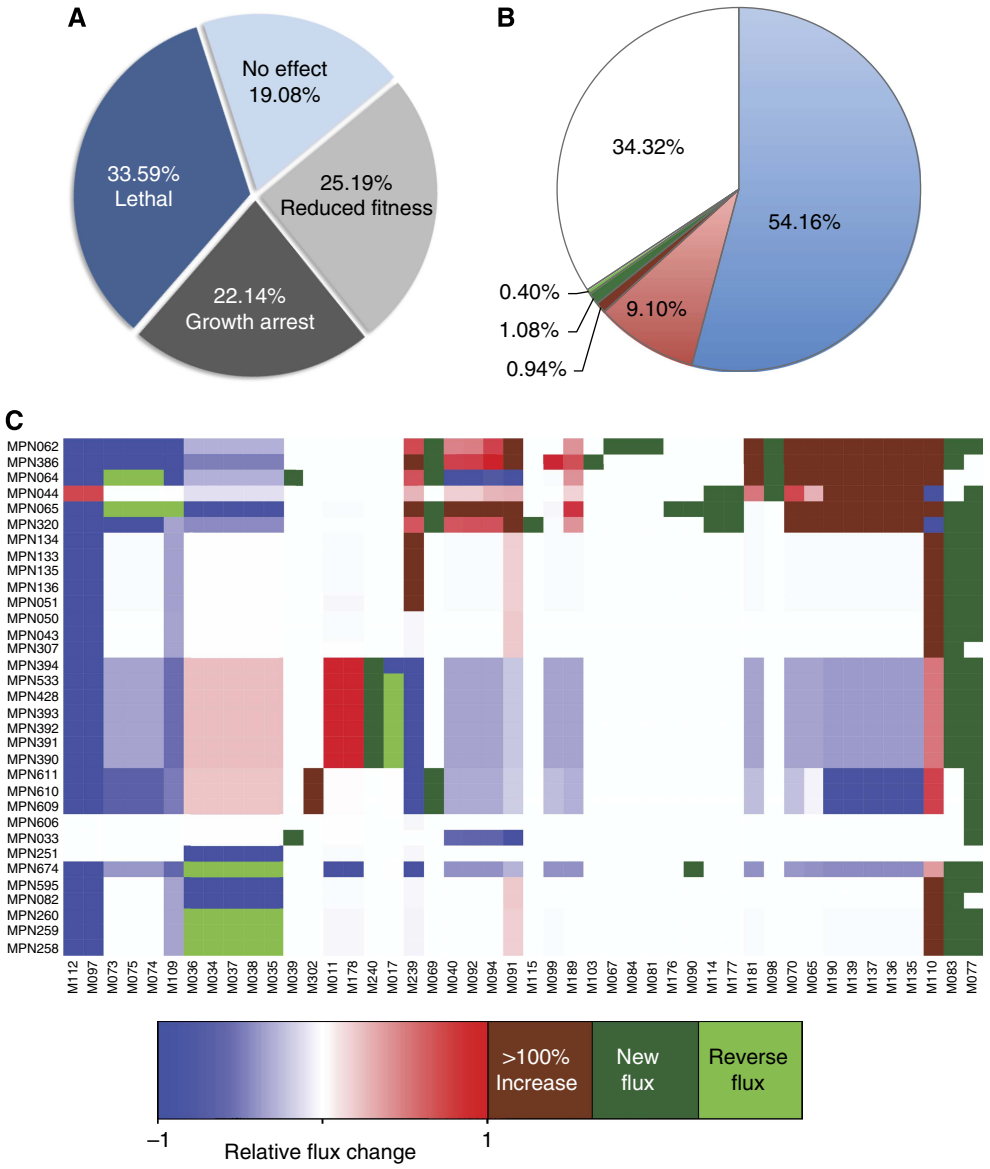

Figure 7 Essentiality prediction. (**A**) KO effects on growth separated into lethal (dark blue), growth inhibiting (light blue), reduced fitness (light grey), and growth not affecting (dark grey) mutant phenotypes; (**B**) relative flux changes for single reactions in reduced fitness knock-out versus. WT simulations (sink/source reactions are excluded): downregulation (blue), upregulation up to 100% (red), upregulation > 100% (brown), inversed fluxes (light green), new fluxes (dark green), no flux change (white); (**C**) relative flux changes of reactions predicted to have inverse, new or highly upregulated flux under at least one KO condition (reaction IDs refer to Supplementary Table S1 and gene names are listed in Supplementary Table S14).

Table III Statistics for the gene essentiality prediction

	Comparison with Glass *et al* (2006)	Comparison with Glass *et al* (2006) and mutants	Taking conditions into account
TP (true positive)	72	72	72
TN (true negative)	41	48	53
FP (false positive)	1	1	1
FN (false negative)	17	10	5
ACC (prediction accuracy)	0.8626	0.9160	0.9542
SPC (prediction specificity)	0.9762	0.9796	0.9815
ACC in %	*86.26*	*91.60*	*95.42*
SPC in %	*97.62*	*97.96*	*98.15*

The gene essentiality prediction for 131 metabolic genes has been evaluated (i) based on a genome-wide transposon study in *M. genitalium* (Glass *et al*, 2006), (ii) also taking into account single transposon screens in *M. pneumoniae*, and (iii) in addition, taking into account simulation conditions and biomass assumptions (for details, see Supplementary information). Italic ACC & SPC in % highlight the numbers referred to throughout the text.
TP—essential in *vivo* and *in silico*, TN—not essential *in vivo* and *in silico*, FP—essential *in silico*, not essential *in vivo*, FN—essential *in vivo*, not essential *in silico*.

Thus, we could confirm five model predictions that were not supported by the *M. genitalium* study alone, raising the prediction accuracy to 92%. Considering additional parameters, namely the literature data, and the simulation conditions, allowed the classification of another eight false-negative hits (Supplementary information). Thus, the model predicts gene essentiality with a final accuracy of 95% and a specificity of 98% (Table III; Supplementary information). We conclude that our refined model of the *M. pneumoniae* metabolism (*i*JW145) possesses high predictive power regarding metabolic phenotypes.

To gain a more detailed quantitative understanding of the resulting phenotypic changes, we assayed the relative *in silico* flux changes of representative flux distributions for the individual reduced fitness knock-out strains. As expected, most individual reaction fluxes (i.e., fluxes of all reactions under all different conditions) are downregulated (54%) or do not change (34%) in response to the different gene deletions *in silico* (Figure 7B). However, we also observed several (highly) upregulated reactions (10.4% of all fluxes) and a few changes in flux direction for reversible reactions (0.4%). Only 88 new fluxes in 32 reduced fitness mutant strains (1.1% of all fluxes) have been observed and they are all found in only 15 different reactions, highlighting—in agreement with our FVA—the limited ability of *M. pneumoniae* to dynamically adapt to perturbation conditions by using alternative metabolic routes for the synthesis of biomass components and ATP. We analyzed highly upregulated reactions as well as those with new or reversed fluxes in at least one *in silico* knock-out strain further (Figure 7C). As expected, lactic acid production (M011) gets upregulated in all strains with gene knock-outs in the acetate branch of pyruvate metabolism. Interestingly, the other 28 examined reactions belong either to the nucleotide metabolism (and reactions involving nucleotides as cofactors) or to the pentose phosphate pathway. This suggests that the non-essential genes in those pathways allow *M. pneumoniae* to adapt to environmental changes and hence lead to a more robust metabolic network.

We used our final model (*i*JW145) to perform an *in silico* genetic interaction screen by predicting metabolic phenotypes for double knock-outs of the 58 non-essential metabolic genes (Supplementary Table S13). Genetic interaction screens are performed to assay network connectivity and to link functionally related genes of different metabolic pathways (Tong *et al*, 2001, 2004; Szappanos *et al*, 2011). Especially, the analysis of synthetic lethal and sick interactions, that is, those double mutants that cause cell death or reduced fitness, allows the identification of gene products that impinge on the same biological process (Hartman *et al*, 2001). The analysis of the synthetic lethal and sick interactions predicted *in silico* shows that genes involved in pyruvate metabolism have a global effect on the metabolic network behavior (Figure 8, for a list of gene names see Supplementary Table S14). Particularly, *mpn674* encoding the lactate dehydrogenase has a strong influence on the phenotype fitness due to the limited acetic acid production. Pyruvate metabolism is central to cellular energy metabolism and mutations in the involved genes limit ATP availability for all cell functions. As expected, the double knock-outs of the lactate dehydrogenase together with a gene involved in acetic acid synthesis are lethal for *M. pneumoniae*. Additionally, the genes coding for proteins involved in sugar uptake and processing can limit the synthesis of ATP and are also found enriched among synthetic lethal interactions (*P*-value 6.4E − 07). Furthermore, pentose phosphate pathway and folate metabolism genes, respectively, are found to be enriched among the synthetic lethal interactions (*P*-values 0.00192 and 0.009981), confirming the existence of rescue routes for parts of these pathways suggested by the single knock-out analysis.

In summary, flux activity analysis at different time points of the exponential growth phase and in *in silico* knock-outs coincide suggesting the same metabolic pathways to be responsible for adaptation to perturbations in *M. pneumoniae*, namely nucleotide metabolism and pentose phosphate pathway. This can be explained by the relatively unchanging environmental conditions *M. pneumoniae* encounters in its natural habitat leading to the elimination of pathways not required for life in the lung. In agreement with a recently published study relating metabolic and gene co-expression networks to predict gene essentiality in *M. pneumoniae* (Güell *et al*, 2012), our metabolic gene essentiality analysis identified the genes associated with the main catabolic pathways (glycolysis + pentose phosphate pathway) among the non-essential genes as most important for the metabolic performance further highlighting the simplicity of the metabolic network of *M. pneumoniae*.

Discussion

We developed a comprehensive metabolic model, *i*JW145, for *M. pneumoniae*. First, we reconstructed the metabolic network based on a manually curated metabolic map (Yus *et al*, 2009) giving rise to *i*JW145_recocnstruct. In the second step, we assigned reaction reversibilities and semi-quantitatively determined the biomass composition of *M. pneumoniae* resulting in *i*JW145_growth. Third, the model was improved in an iterative process of *in silico* growth simulations and their evaluation

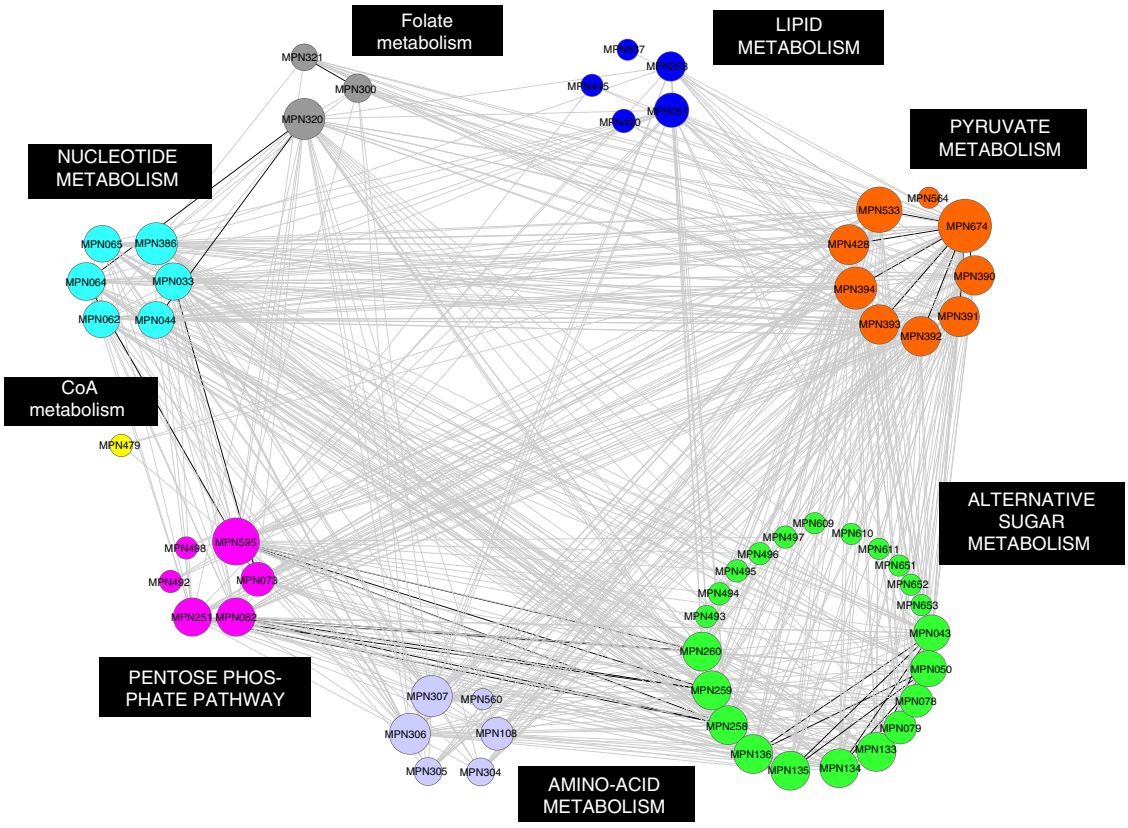

Figure 8 Protein pairs producing sick (grey edges) and synthetic lethal (black edges) interactions in *in silico* double knock-outs. Genes are clustered according to their pathways; the node size reflects the number of interactions.

based on a variety of experimental data and literature information. We were able to correct the structure of the metabolic network allowing the final model, *i*JW145, to reproduce experimental findings along the exponential growth phase (Yus *et al*, 2009).

The number of constraint-based metabolic models of different organisms has been constantly increasing over the past years (for a list of validated models, see Feist *et al*, 2009 and Supplementary Table 2). For mycoplasmas, an automatic reconstruction for *M. genitalium* has been published (Suthers *et al*, 2009). However, since it has been shown that automatic reconstructions are often highly error prone (Reed *et al*, 2006; Henry *et al*, 2010), we integrated different experimental data directly during the manual reconstruction process. This enabled us not only to obtain an accurate metabolic reconstruction but also to revise the wiring diagram and the functional annotation of key enzymes.

After establishing the biomass composition of an average *M. pneumoniae* cell, we quantitatively analyzed its energy metabolism. Comparing our results with that of the recently published whole-cell model for *M. genitalium* showed that both models identified protein and RNA production as well as their maintenance as the major energy sinks in biomass production (Karr *et al*, 2012). Most strikingly, we showed that *M. pneumoniae*—at least under laboratory conditions—dedicates only a small fraction of its ATP directly to biomass production. Alternative quantified ATP sinks include chaperone assisted protein folding, DNA maintenance, and post-translational modifications. The combination of *in silico* calculations taking into account the ATPase protein copy number in *M. pneumoniae* and catalytic rates from other organisms reported in the literature, as well as the *in vivo* analysis of growth in medium at different pH suggest that the ATPase uses about 57–80% (growth stage depending) of the total generated energy to maintain a favorable proton gradient across the membrane and the intracellular pH. On the one hand, this surprising finding can be explained by the small size of *M. pneumoniae*. Membrane leaking and the transport of molecules across the membrane have a higher effect on cytoplasmic homeostasis in small organisms (*M. pneumoniae* has surface-to-volume ratio 2500 times higher than *E. coli*, Supplementary information; Supplementary Table S7). On the other hand, the continuous, growth-associated secretion of acids poses an increasing pH maintenance burden on *M. pneum*oniae grown in batch culture. Therefore, in consistency with the abundance of the core components of the ATPase protein complex (Maier *et al*, 2011) and our experimental results for *M. pneumoniae* growth under pH stress (Figure 4A–B), we propose that at later growth stages most of the available energy is diverted toward cellular maintenance with the ATPase as a major energy sink of *M. pneumoniae* in batch

culture growth, as also suggested for other lactic acid bacteria (Hutkins and Nannen, 1992).

In contrast to the high ATPase costs, the costs for protein folding and maintenance by molecular chaperones are surprisingly low taking into account their high cellular abundance (~10% of the total cellular protein mass; Maier *et al*, 2011), as is the total amount of energy used for GAM (~2–7% of the total generated ATP, ~2.5–10% of NGAM). Despite it has been shown that GAM/NGAM estimations show considerable variance depending on the experimental data used (Varma and Palsson, 1994b; Feist *et al*, 2007; Orth *et al*, 2011), our finding is in agreement with recent results from the whole-cell model in *M. genitalium* (Karr *et al*, 2012). Furthermore, we showed that the doubling time has major impact on the ratio between GAM and NGAM since assuming different doubling times while providing the same amount of nutrients leads to a significant alteration of this ratio (Supplementary Table S7). Summing up, we were able to explain 75–100% of the total ATP consumption during the exponential growth phase. Movement and attachment are the only known major energy consuming processes we could not assess in our analysis. *M. mobile* has been shown to use ATP for gliding (Jaffe *et al*, 2004) but no details on related energy consumption are available. The whole-cell model for *M. genitalium* also showed that other cellular processes which are not considered in our metabolic model, such as chromosome condensation, RNA modification and processing, ribosome assembly, protein translocation, and the replication initiation only contribute to a minor extent to the cellular energy balance (<3%; Karr *et al*, 2012). The missing expenses can alternatively be attributed to (i) measurement errors in absolute cellular protein quantifications (a two-fold error has been reported in Maier *et al*, 2011), (ii) the determined protein and mRNA turnover rates (we used average half-life times Maier *et al*, 2011), or (iii) the estimation of cell doubling times (protein quantities in the beginning of a growth experiment, that is, up to 36 h after inoculation, are near the lower detection limit; Supplementary Figure S5A). When comparing our results with those of the whole-cell model of *M. genitalium* (Karr *et al*, 2012), we find that both models agree in the general predictions on central carbon metabolism. Our findings on cellular energy homeostasis (Figures 3B and 4B) can also explain the discrepancy between produced and consumed energy found for *M. genitalium in silico* (44%; Karr *et al*, 2012).

The metabolism of most bacteria follows a single objective function: maximizing growth (Neidhardt, 1996; Buescher *et al*, 2012). A recently applied multidimensional optimality analysis revealed that metabolic flux states in bacteria additionally evolved to minimize the costs for adjusting to different environmental conditions (Schuetz *et al*, 2012). Our energy calculations and their integration with the model of *M. pneumoniae* represent a novel approach toward a more detailed understanding of cellular metabolism. They allow a quantitative dissection of cellular performance into energy gain, biomass production and other cellular, growth and non-growth-associated energy sinks. A comparison of maintenance energy costs between *M. pneumoniae* and *E. coli* revealed fundamental differences in their energy sink reactions, suggesting individual and characteristic energy expense profiles for different bacteria. We identified four parameters governing the composition of these energy expense profiles: the topology of the metabolic network, environmental conditions, growth rate, and cell size. In the case of *M. pneumoniae*, one can speculate if the large amount of energy diverted toward maintenance tasks and the associated slow growth in the laboratory additionally represent an adaption to its parasitic lifestyle on epithelial cells of the human lung.

Integrating flux distributions predicted *in silico* and an experimental analysis of the central carbon metabolism by ^{13}C-labelled glucose tracer experiments, we showed that glycolysis is directly connected to the pentose phosphate pathway and lipid biosynthesis but not to amino-acid metabolism in *M. pneumoniae in vivo*. Model predictions and experimental results further agree that most of the carbon taken up is shuffled through glycolysis for the production of ATP during organic acid synthesis. However, in the flux directions of the minor inter-pathway fluxes they disagree, indicating a slight overestimation of imported ribose and glycerol/G3P. For an exact *in silico* representation of also the minor fluxes, further experiments addressing nutrient uptake of *M. pneumoniae* from the environment are necessary but technically not feasible yet. Taking into account the comparatively small amounts of carbons shuffled via the routes interconnecting different metabolic pathways, we estimate the influence on the overall growth rate and energy balance to be very small.

One interesting general finding with respect to network dynamics is that under all simulated conditions (rich/defined medium, alternative sugars, knock-outs) oxygen consumption is tightly coupled to acetic acid production. This prediction is in agreement with findings in other organisms, such as *Lactococcus lactis*, in which limited oxygen availability at later growth stages prevents the regulation of the cellular redox imbalance associated with acetic acid production, while the lactate dehydrogenase is released from its supposed oxygen-dependent inhibition (Gottschalk, 1986; Neves *et al*, 2005). We propose that oxygen could also have a regulatory role in pyruvate processing in *M. pneumoniae* explaining the metabolic shift from mainly acetic to mainly lactic acid fermentation observed *in vivo* during a 4-day batch culture growth experiment (Yus *et al*, 2009).

In silico knock-out studies have been used in other organisms to predict gene essentiality (Reed and Palsson, 2003; Feist *et al*, 2007). With our prediction for *M. pneumoniae* we achieve slightly higher accuracy and specificity that has been obtained for *E. coli* so far (Feist *et al*, 2007). *M. pneumoniae* has an exceptionally high percentage of essential metabolic genes (56.6% versus. 19% in *E. coli*; Baba *et al*, 2006; Joyce *et al*, 2006) for which consequently no other gene can buffer the loss of function caused from gene deletion. We found that in *M. pneumoniae* nucleotide metabolism and the pentose phosphate pathway are the only pathways preserving rescue routes for gene deletion events. The most likely explanation for this lack of rescue routes in most pathways is the adaptation to parasitism in a specific relatively unchanging niche accompanied by the reductive genome evolution.

In vivo, synthetic genetic array analysis has been shown to allow the automating of the isolation and analysis of double

mutants (Tong *et al*, 2001, 2004). We used *in silico* prediction of double mutant phenotypes to extract information about the specific effects caused by different gene deletions on the metabolic network, and the adaptive capabilities of *M. pneumoniae*. The results confirm the general findings of the single knock-out simulations that have been evaluated based on independent experimental data. Especially in organisms for which no engineering tools for *in vivo* analysis of double mutant phenotypes are available, the applied *in silico* analysis provides a promising alternative to experimental approaches.

Materials and methods

The model

The used model in sbml format including MIRIAM annotation is provided for download at http://nin.crg.es/serranolab/mycomap/ (user name: mycomap, password: bicha987) and can be accessed in the BioModels database with ID MODEL1301290000.

Modeling procedures

Constraint-based modeling and flux balance analysis

Constraint-based modeling is an approach to analyze a (metabolic) system under steady-state conditions. This means that the concentrations of the network components x_i do not change over time, that is,

$$dx_i/dt = Nv = 0$$

where N is the stoichiometric matrix of size mxn with m being the number of reactants and n the number of reactions and v an n-dimensional vector containing the fluxes through the n reactions of the network. FBA is a mathematical method to determine a set of metabolic fluxes (the vector v) fulfilling the steady-state condition and, at the same time, maximizing an objective function such as growth for a given set of available nutrients using linear programming (Varma and Palsson, 1994a; Kauffman *et al*, 2003). For our model, maximization of biomass production has been chosen as single objective function, since no other objectives have been revealed for *M. pneumoniae* so far.

Metabolic reconstruction

We used the reconstruction and modeling platform ToBiN (*To*olbox for *Bi*ochemical *N*etworks; http://github.com/miguelgodinho/tobin). The initial reaction network was based on the list of reactions found in Yus *et al* (2009). Some changes had to be introduced to keep elements and charges balanced and to cope with reactions that can be represented in stoichiometric models only with considerable impact on the model complexity, such as RNA and DNA elongation reactions (see below). To simulate the exchange of compounds with the environment reactions producing or consuming given compounds have been defined (source and sink reactions). Appropriate reaction reversibilities and minimum–maximum flux constraints were imposed based on the experimental data and literature information (Supplementary Tables S2 and S5; Supplementary information). We used CellDesigner 4.1 (Kitano *et al*, 2005) to visualize the model (Supplementary Figure S1) and the abbreviations used can be found in Supplementary Table S15.

Definition of biomass equation

Based on the general biomass equation (Results), the different components were identified and, if possible, quantified. According to Razin *et al* (1963), mycoplasma cells are composed of 54–62% protein, 12–20% lipids, 3–8% carbohydrates, 8–17% RNA, and 4–7% DNA. Assuming that one *M. pneumoniae* cell contains 10 fg of protein (Yus *et al*, 2009) and assuming proteins to compose 62% of the total cell mass, we were able to calculate the fractions of RNA and DNA. Based on sequence information and the biophysical properties of *M. pneumoniae* (Yus *et al*, 2009; Maier *et al*, 2011), we determined the DNA to account for 5% and the RNA to account for 6.5% of the total cell mass. Further, we assume that 20% lipids and 6.5% carbohydrates and other metabolites make up the missing 26.5% of the cell mass (Supplementary information). The lipid composition of mycoplasmas varies depending on the fatty acids provided with the medium (McElhaney and Tourtellotte, 1969; Pollack *et al*, 1970; Rottem, 1980). As these variations cannot be represented in a static model, an artificial molecule has been defined taking into account the average fatty acid composition determined (Supplementary Figure S3; Supplementary information, dataSheet). The synthesis of each macromolecule (RNA, DNA, and protein) is represented by a dedicated artificial reaction (Supplementary information). To consider mRNA and protein half-lives, minimum constraints have been set on the respective degradation reactions. DNA repair as well as rRNA and tRNA degradation have been accounted for qualitatively by small overhang quantities of the respective components in the biomass equation (Supplementary information). Free bases and free amino acids (together accounting for ~1.5% of the total biomass) were quantified by GC-MS analysis (unpublished results). To account for the essentiality of vitamins and other cofactors, the end products of the secondary metabolism pathways are included into the biomass equation qualitatively (i.e., in small arbitrary quantities), whenever their precursors have been shown to be essential in the defined medium (Yus *et al*, 2009). This qualitative consideration is possible, as they are supposed to be low abundant and we proved in a sensitivity analysis that a 10-fold change of their quantities does not change the general model behavior but only introduces a negligible change in the objective value and thus the *in silico* doubling yield. The missing proportion of biomass is accounted for by including a respective amount of G6P (representative for carbohydrates) into the biomass equation. To overcome missing experimental information and to restrict the complexity of the model, a number of approximations have been made which are fully listed in Supplementary information.

Sensitivity analysis

We varied different macromolecule fractions of the biomass based on the ranges reported by Razin *et al* (1963), namely protein from 54 to 62%, RNA from 6.5 to 10% (a further increase in the RNA fraction was not possible without altering other biomass fractions), and lipids from 12 to 20%. Furthermore, we tested the impact of a change in the amount of cofactors included in the biomass from 0.1- to 10-fold as compared with the value given in the *M. pneumoniae* biomass function. In all cases, the total cell weight was maintained constant by accounting for the applied changes by a respective increase or decrease in G6P (representing carbohydrates). We tested the influence on the growth rate (keeping fixed all normal growth constraints, including the minimum constraint for maintenance costs) as well as on the maintenance energy (in this case, manually fitted the minimal constraint for maintenance energy to reproduce *in vivo* doubling times.

Growth simulations

Growth simulations for the exponential growth phase (24–60 h after inoculation) have been accomplished using biomass production as objective function for the FBA. The resulting objective value ov gives information about the doubling time t_{doub} of an average *M. pneumoniae* cell. The relation of the t_{doub} and ov when assuming a constant cell number (1 g of cells is the default in the used modeling platform ToBiN) can be described by $t_{doub} = 1/ov$. Thus, it is possible to distinguish between growth (ov larger than zero), catabolic activity (growth arrest) (ov equal to zero), and cell death (infeasibility of the FBA). The FBA solution is considered as infeasible if at least one of the minimum requirements specified cannot be satisfied under the given nutrient conditions.

Flux variability analysis

FVA is a method to assess the potential alternative flux distributions supporting a given FBA objective. After performing FBA optimization, the objective flux (e.g., Biomass production) was fixed to the value 0.01% lower than the FBA result (this slight decrease in the objective

flux aimed at avoiding model infeasibilities due to rounding errors) and for each model reaction two FBA runs were performed, one setting the flux maximization and one setting the flux minimization through the reaction as the objective. Reactions for which these two values differ span the space of alternative flux distributions. As these optimizations are not independent of each other, not every flux combination lying within the hypercube spanned by the minimal and maximal flux values supports the initial value of the objective (or even valid flux distribution).

Energy calculations

All constraints and fluxes in ToBiN have the unit mmol $\times$ g(cells)$^{-1}$ $\times$ h^{-1}. To calculate the total ATP produced by *M. pneumoniae in silico*, we summed up the ATP production of the different catalytic pathways in *M. pneumoniae* using the following equation:

flux(prod(ATP_{total})) = flux(prod(LAC)) + 2 $\times$ flux(prod(ACE)) + flux (prod(ornithine))

with prod:= production. To know the production per *M. pneumoniae* cell, we converted the obtained fluxes into molecules $\times$ cell^{-1} $\times$ s^{-1}.

Comparison of qualitative changes in fluxes and protein abundances

First, linear fittings to the *in silico* reaction fluxes obtained at $t = 24$, 36, 48, and 60 h and to protein abundances measured at $t = 24$, 36, 48, and 72 h during batch culture growth experiments *in vivo* (Maier *et al*, 2011, unpublished results). Second, we determined the qualitative overall change of fluxes and protein abundances during the exponential growth phase, considering proteins to change only if the measured abundance difference exceeds 25% of the abundance at $t = 24$ h, thus accounting for the reported experimental error that would otherwise have a high impact especially on the changes of low-abundant proteins (Maier *et al*, 2011). Finally, we aligned protein concentration changes with the change of the sum of fluxes of reactions catalyzed by the respective enzyme (Supplementary Figure S6).

Gene essentiality prediction

For the gene essentiality prediction, the gene–protein relationship has been determined for all reactions for which the catalyzing enzyme is known. In each *in silico* knock-out, all reactions catalyzed by the corresponding gene product have been limited to zero flux. A gene is considered as essential when its knock-out leads to an objective value of zero (no growth but minimum constraints can be matched) or the infeasibility of the FBA (minimum constraints are not fulfilled). The genes encoding proteins catalyzing DNA degradation, protein folding and the ATPase reaction have been excluded from the essentiality prediction since their corresponding functions have not been modeled explicitly. All simulations of this section have been performed using rich medium conditions for 36 h growth time (Supplementary Table S5). Since the amount of energy used for maintenance tasks has been determined by manually fitting the minimum constraint of the respective energy consuming reaction to allow reproduction of the experimentally determined doubling time, those expenses have been neglected for the essentiality prediction. Otherwise, a knock-out leading to significant slower energy production would result in infeasibility of the FBA. Subsequently, the obtained objective values give no information about the absolute doubling times but only about the relative changes in the growth rate between wild type and knock-out simulation.

For the prediction of double mutant phenotypes, we applied the same strategy as for the single *in silico* gene knock-outs, but simultaneously silenced the reactions catalyzed by two different non-essential enzymes at a time. Double knock-outs resulting in reduced fitness, that is, the objective value is smaller than for the two corresponding single knock-outs alone, or in cell death, that is, the objective value equals zero or the FBA is infeasible, were considered for the analysis of synthetic lethal and sick interactions.

For the statistical analysis of accuracy and specificity of the gene essentiality prediction, we evaluated the prediction results based on a genome-wide transposon study in *M. genitalium* (Glass *et al*, 2006), transposon screens in *M. pneumoniae* (this study), and the simulation conditions. Computationally and experimentally essential genes are considered as true positives, true negatives are computationally and experimentally not essential, computationally essential and experimentally non-essential genes are defined as false positives and computationally non-essential and experimentally essential genes accordingly false negative hits.

Sequence comparison

All sequence analyses have been performed using the Basic Local Alignment Search Tool (BLAST) for proteins (pBLAST) (Altschul *et al*, 1997). pBLAST (algorithm pblast) was used as *M. pneumoniae* uses the TGA codon to encode for tryptophan instead of indicating the end of a gene as in most other organisms. Protein sequences of related organisms (ordered for preference: other mycoplasmas, *B. subtilis*, *L. lactis*, *E. coli*) were obtained from KEGG (Kanehisa and Goto, 2000) or the National Center of Biotechnology Information (NCBI) (Tatusova *et al*, 1999) and used to perform pBLAST against the *M. pneumoniae* proteome. Alternatively, *M. pneumoniae* protein sequences were aligned to the nr-DB to detect possible homologies. This has been done so (i) to identify the cofactors used by the GPO (MPN051), (ii) to shed light on the NOX isoform (MPN394), (iii) to confirm that a reaction converting UTP into CTP does not exist in *M. pneumoniae*, and (iv) to search for proteins possibly catalyzing phospholipid production. All pBLAST results are shown in Supplementary Information, pBLAST Results.

Experimental procedures

GC–MS analysis of fatty acids

Fatty acids were targeted specifically by using tailored protocols. Depending on the case, growth medium, total cell content, cell pellet, or cytoplasm was analyzed as described in each protocol.

Fatty acid analysis was conducted on all three cellular preparations according to Ghanem *et al* (1991). Briefly, to a sample preparation (lyophilized in the case of cytoplasm and growth medium, humid in the case of cell pellet or total cell content) 1 ml of NaOH solution (3.75 M in a 1:1 (v/v) MeOH:H_2O mixture) was added. The suspension was incubated for 5 min at 100°C, vortexed for 10 s and incubated for another 25 min at 100°C. Samples were cooled to room temperature and added with 1 ml MeOH:HCl solution (0.46:0.54 (v/v) of MeOH and 6 M HCl) followed by 1 ml MeOH:H_2SO_4 solution (0.46:0.54 (v/v) of MeOH and 50% H_2SO_4). This mixture was vortexed for 10 s, incubated at 80°C for 10 ± 1 min, and cooled on ice immediately. Subsequently, 1.25 ml of a hexane:ether (1:1 v/v) mixture was added and mixed end-over-end for 10 min. The water layer was removed after centrifugation for 2 min at 3000 r.p.m. Then, to the organic layer 3 ml of a NaOH:NaCl mixture (0.3 M NaOH in 4.75 M NaCl) was added and the solution mixed end-over-end for 5 min. The aqueous phase was frozen and the organic phase transferred to a new tube. The content evaporated under nitrogen stream at 30°C, and reconstituted in 100 µl hexane/tert-butyl methyl ether (1:1 v/v). GC-MS was carried out on a 6890N gas chromatograph coupled with a 5973 MSD (Agilent Technologies, Palo Alto, CA, USA). The fatty acids methyl esters were separated on a Phenomenex Zebron ZB-5 crosslinked 5% phenyl polymethyl siloxane column (15 m $\times$ 0.25 mm i.d., 0.25 µm film thickness). Helium was used as the carrier gas at a constant pressure of 5 p.s.i. A 1-µl aliquot of the extract was injected into the system operated in split mode (split ratio 70:1). The GC temperature is ramped as follows: initial 150°C, increased to 240°C at 5°C/min, held for 3 min, thereafter increased to 310°C at 30°C/min, and held for 2 min. The injector and transfer line are kept at 280°C, the MS source at 230°C, and the quadrupole at 150°C. The mass range scanned is from 50 to 650 Da. Conventional fatty acids were identified by comparison with the analysis of 1 µl of a reference mixture (Supelco 37 Component FAME Mix) under the same conditions. Other fatty acid methyl esters were identified by monitoring the typical loss of 43 mass units (CH_3-CO^-) from the parent ion.

GC–MS analysis of glycolytic intermediates and reporter compounds for adjacent pathways

For the analysis of glycolysis intermediates (cytosolic extract), samples added with the internal standard (5 µl of methyltestosterone 10 ng/µl) were lyophilized and subsequently dried in a vacuum oven (500 mbar, 50°C) in the presence of diphosphorus pentoxide for at least 4 h. Then, aldehyde and ketone functional groups were converted into methyl oximes with 75 µl of methoxyamine hydrochloride in pyridine (2 g/L) at 40°C for 90 min with intermediate mixing. Subsequently, hydroxyl groups were converted into trimethylsilyl groups with 100 µl MSTFA (N-Methyl-N-trifluoroacetamide) at 40°C for 50 min. Samples were transferred to glass inserts, spun for 5 min at 5000 r.p.m., and the supernatant transferred to a new vial for GC–MS analysis (Chan *et al*, 2011). In all, 4 µl of glycolytic products was analyzed (split ratio 1:10) using a HP-Ultra1 crosslinked methyl-silicone column, 16.5 m × 0.2 mm i.d., film thickness 0.11 µm (J&W Scientific, Folsom, CA, USA) in an Agilent 6890N gas chromatograph coupled to an Agilent 5973 mass selective detector. Helium was used as carrier gas at a constant pressure of 5 p.s.i. The GC temperature is ramped as follows: initial 70°C, held for 1 min, increased to 280°C at 6 °C/min, and held for 1 min at 280°C. For optimal sensitivity, the acquisition, performed in SIM mode, was split into four time segments with three characteristic ions per compound (plus the corresponding ^{13}C isotopes using 15 ms dwell times) in each. Four different time segments ranged from 6 to 16 min (13 ions: *m/z* 103.1, 117.1, 211.1, 232.2, **236.2**, 299.2, 315.2, 369.2, **371.3**, 384.3, **386.3**, 445.4, **448.4** for phospho-groups, PEP, GAP, DHAP, and G3P), 16 to 19.5 min (15 ions: *m/z* 103.1, **104.1**, 147.1, 160.1, **162.1**, 191.1, 205.1, **207.1**, 217.1, **220.1**, 307.2, **310.2**, 319.2, **323.2**, 409.3, fructose, I.S., Glucose, and G1P), 19.5 to 28 min (11 ions: *m/z* 103.1, **104.1**, 299.2, 315.2, 357.2, **359.2**, 387.2, 459.3, **462.3**, 471.3, **475.3**: F6P, G6P, and R5P), and 28 to 32 min (9 ions: *m/z* 103.1, **104.1**, 299.2, 315.2, 357.2, **359.2**, 387.2, 459.3, **462.3**: FBP) (bolded are the ions corresponding to the ^{13}C-labelled compounds). Identification of the sample constituents was based on the theoretical fragmentation pattern expected for each compound. Quantification was performed with respect to the corresponding calibration curve standardized against the internal standard (methyltestosterone-MO-TMS). Samples and the calibration curves were analyzed at least in triplicate.

Transposon screens

The 64 pools of an ordered collection of *M. pneumoniae* transposon mutants generated by 'haystack mutagenesis' (Halbedel *et al*, 2006) were assorted into 10 groups. Then, genomic DNA extractions were performed using Illustrabacteria genomic KIT (GE). The disruptive insertions in genes *mpn133*, *mpn321*, *mpn392*, *mpn533*, and *mpn595* were detected by PCR (Supplementary Figure S7). The fragments corresponding to junctions between the genes and the mini-transposon were amplified using the primer 3JpMT85 and the primers 5MPN133, 5MPN321, 5MPN392, 5MPN533, and 5MPN595, respectively (Supplementary Table S16). The position of the transposon insertion in the different genes was determined by DNA sequencing.

pH Re-buffering experiment

To test for the influence of the medium pH on growth performance, *M. pneumoniae* cells were grown in batch culture in 75 cm^2 culture flasks. Cells were grown in pre-culture for 96 h in glucose containing medium, harvested by scraping and diluted into fresh growth medium. Medium pH was adjusted back to pH 7.7 after 4 days of growth by titration with sterile 1 M NaOH. Samples from growth medium supplemented with 1% glucose (55.5 mM) were taken at indicated time points (Figure 4C). Glucose and lactic acid concentrations were determined with enzymatic assays.

M. pneumoniae growth in medium at different pH

M. pneumoniae cells were grown for 48 h in 75 cm^2 culture flasks with Hayflick medium supplemented with 55 mM glucose. The medium of each growth flask was exchanged with Hayflick medium with a pre-set pH (between pH 5.5 and 8) and grown for additional 36 h. Triplicates have been used for each pH. Metabolite and protein measurements were done as in Yus *et al* (2009).

Supplementary information

Supplementary information is available at the *Molecular Systems Biology* website (www.nature.com/msb).

Acknowledgements

We thank Jörg Stülke from the University Göttingen for allowing us to screen their transposon library. We thank past and present colleagues for helpful discussions; in particular, Javier Delgado for help on scripting; Konstantinos Michalodimitrakis for experimental supervision; and Hinnerk Eilers for information on protein activity. We thank Toni Hermoso for the conversion of Supplementary Figure S1 into a clickable metabolic map. This work was supported by the European Research council (ERC) advanced grant, the Fundacion Marcelino Botin and the Spanish Ministry of Research and Innovation to the ICREA researcher LS. JW is recipient of a laCaixa-CRG fellowship.

Author contributions: JW developed and organized the project, designed the model, analyzed the results, prepared figures, and wrote the manuscript. JP designed the model and analyzed the results. ML-S performed the mutant screens and prepared Supplementary Figure S7. JM conducted the mass spectrometry experiments. EY performed the pH stress experiments to validate the ATPase hypothesis. MG provided data extraction and analysis tools. RGG analyzed the mass spectrometry results. VMdS initiated the project and contributed to its development. LS and EK contributed to the project design and development and discussed results. TM participated in the project development, performed experiments, contributed to the figures, and wrote the manuscript.

Conflict of interest

The authors declare that they have no conflict of interest.

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

Baba T, Ara T, Hasegawa M, Takai Y, Okumura Y, Baba M, Datsenko KA, Tomita M, Wanner BL, Mori H (2006) Construction of Escherichia coli K-12 in-frame, single-gene knockout mutants: the Keio collection. *Mol Syst Biol* **2:** 0008

Buescher JM, Liebermeister W, Jules M, Uhr M, Muntel J, Botella E, Hessling B, Kleijn RJ, Le Chat L, Lecointe F, Mäder U, Nicolas P, Piersma S, Rügheimer F, Becher D, Bessieres P, Bidnenko E, Denham EL, Dervyn E, Devine KM *et al* (2012) Global network reorganization during dynamic adaptations of Bacillus subtilis metabolism. *Science* **335:** 1099–1103

Chan ECY, Pasikanti KK, Nicholson JK (2011) Global urinary metabolic profiling procedures using gas chromatography-mass spectrometry. *Nat Protoc* **6:** 1483–1499

Chanock RM, Hayflick L, Barile MF (1962) Growth on artificial medium of an agent associated with atypical pneumonia and its identification as a PPLO. *Proc Natl Acad Sci USA* **48:** 41–49

Covert MW, Palsson BØ (2003) Constraints-based models: regulation of gene expression reduces the steady-state solution space. *J Theor Biol* **221:** 309–325

Covert MW, Schilling CH, Palsson B (2001) Regulation of gene expression in flux balance models of metabolism. *J Theor Biol* **213:** 73–88

Dandekar T, Huynen M, Regula JT, Ueberle B, Zimmermann CU, Andrade MA, Doerks T, Sánchez-Pulido L, Snel B, Suyama M, Yuan YP, Herrmann R, Bork P (2000) Re-annotating the Mycoplasma pneumoniae genome sequence: adding value, function and reading frames. *Nucleic Acids Res* **28:** 3278–3288

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

Draper NR, Smith H (1998) *Applied Regression Analysis.* New York: Wiley

Duarte NC, Becker SA, Jamshidi N, Thiele I, Mo ML, Vo TD, Srivas R, Palsson BØ (2007) Global reconstruction of the human metabolic network based on genomic and bibliomic data. *Proc Natl Acad Sci USA* **104:** 1777–1782

Edwards JS, Covert M, Palsson B (2002) Metabolic modelling of microbes: the flux-balance approach. *Environ Microbiol* **4:** 133–140

Edwards JS, Palsson BØ (2000) Metabolic flux balance analysis and the in silico analysis of Escherichia coli K-12 gene deletions. *BMC Bioinformatics* **1:** 1

Feist AM, Henry CS, Reed JL, Krummenacker M, Joyce AR, Karp PD, Broadbelt LJ, Hatzimanikatis V, Palsson BØ (2007) A genome-scale metabolic reconstruction for Escherichia coli K-12 MG1655 that accounts for 1260 ORFs and thermodynamic information. *Mol Syst Biol* **3:** 121

Feist AM, Herrgård MJ, Thiele I, Reed JL, Palsson BØ (2009) Reconstruction of biochemical networks in microorganisms. *Nat Rev Microbiol* **7:** 129–143

Fell DA, Small JR (1986) Fat synthesis in adipose tissue. An examination of stoichiometric constraints. *Biochem J* **238:** 781–786

Finch L, Mitchell A (1992) Micoplasmas. In *Molecular Biology and Pathogenesis. Source of Nucleotides*. Maniloff J, McElhany RN, Finch LR, Baseman JB (eds), Washington, DC, American Society for Microbiology

Fleming RMT, Thiele I (2011) von Bertalanffy 1.0: a COBRA toolbox extension to thermodynamically constrain metabolic models. *Bioinformatics* **27:** 142–143

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

Ghanem FM, Ridpath AC, Moore WE, Moore LV (1991) Identification of Clostridium botulinum, Clostridium argentinense, and related organisms by cellular fatty acid analysis. *J Clin Microbiol* **29:** 1114–1124

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 USA* **103:** 425–430

Gottschalk G (1986) *Bacterial Metabolism* Springer

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

Güell M, Yus E, Lluch-Senar M, Serrano L (2011) Bacterial transcriptomics: what is beyond the RNA horiz-ome? *Nat Rev Microbiol* **9:** 658–669

Güell O, Sagués F, Serrano MÁ (2012) Predicting effects of structural stress in a genome-reduced model bacterial metabolism. *Sci Rep* **2:** 621

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

Halbedel S, Stülke J (2007) Tools for the genetic analysis of Mycoplasma. *Int J Med Microbiol* **297:** 37–44

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

Hartman 4th JL, Garvik B, Hartwell L (2001) Principles for the buffering of genetic variation. *Science* **291:** 1001–1004

Henry CS, DeJongh M, Best AA, Frybarger PM, Linsay B, Stevens RL (2010) High-throughput generation, optimization and analysis of genome-scale metabolic models. *Nat Biotechnol* **28:** 977–982

Himmelreich R, Hilbert H, Plagens H, Pirkl E, Li BC, Herrmann R (1996) Complete sequence analysis of the genome of the bacterium Mycoplasma pneumoniae. *Nucleic Acids Res* **24:** 4420–4449

Hutkins RW, Nannen NL (1992) pH Homeostasis in Lactic Acid Bacteria. *J Dairy Science* **76:** 2354–2365

Jaffe JD, Miyata M, Berg HC (2004) Energetics of gliding motility in Mycoplasma mobile. *J Bacteriol* **186:** 4254–4261

Joyce AR, Reed JL, White A, Edwards R, Osterman A, Baba T, Mori H, Lesely SA, Palsson BØ, Agarwalla S (2006) Experimental and computational assessment of conditionally essential genes in Escherichia coli. *J Bacteriol* **188:** 8259–8271

Kanehisa M, Goto S (2000) KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Res* **28:** 27–30

Karr JR, Sanghvi JC, Macklin DN, Gutschow MV, Jacobs JM, Bolival Jr B, Assad-Garcia N, Glass JI, Covert MW (2012) A whole-cell computational model predicts phenotype from genotype. *Cell* **150:** 389–401

Kauffman KJ, Prakash P, Edwards JS (2003) Advances in flux balance analysis. *Curr Opin Biotechnol* **14:** 491–496

Kitano H (2002) Computational systems biology. *Nature* **420:** 206–210

Kitano H, Funahashi A, Matsuoka Y, Oda K (2005) Using process diagrams for the graphical representation of biological networks. *Nat Biotechnol* **23:** 961–966

Klipp E, Herwig R, Kowald A, Wierling C, Lehrach H (2005) *Systems Biology in Practice.* Berlin: Wiley-VCH

Kobayashi H (1985) A proton-translocating ATPase regulates pH of the bacterial cytoplasm. *J Biol Chem* **260:** 72–76

Kühner S, van Noort V, Betts MJ, Leo-Macias A, Batisse C, Rode M, Yamada T, Maier T, Bader S, Beltran-Alvarez P, Castaño-Diez D, Chen W-H, Devos D, Güell M, Norambuena T, Racke I, Rybin V, Schmidt A, Yus E, Aebersold R *et al* (2009) Proteome organization in a genome-reduced bacterium. *Science* **326:** 1235–1240

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

Manolukas JT, Barile MF, Chandler DK, Pollack JD, Reactions A (1988) Presence of anaplerotic reactions and transamination, and the absence of the tricarboxylic acid cycle in mollicutes. *J Gen Microbiol* **134:** 791–800

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

Moreno SN, Zhong L, Lu HG, Souza WD, Benchimol M (1998) Vacuolar-type H+-ATPase regulates cytoplasmic pH in Toxoplasma gondii tachyzoites. *Biochem J* **330:** 853–860

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

Neidhardt FC (1996) *Escherichia coli and Salmonella: Cellular and Molecular Biology.* Washington, DC, ASM Press

Neves AR, Pool WA, Kok J, Kuipers OP, Santos H (2005) Overview on sugar metabolism and its control in Lactococcus lactis - the input from *in vivo* NMR. *FEMS Microbiol Rev* **29:** 531–554

Nicolas P, Mader U, Dervyn E, Rochat T, Leduc a, Pigeonneau N, Bidnenko E, Marchadier E, Hoebeke M, Aymerich S, Becher D, Bisicchia P, Botella E, Delumeau O, Doherty G, Denham EL, Fogg MJ, Fromion V, Goelzer a, Hansen a *et al* (2012) Condition-dependent transcriptome reveals high-level regulatory architecture in Bacillus subtilis. *Science* **335:** 1103–1106

Oberhardt MA, Palsson BØ, Papin JA (2009) Applications of genome-scale metabolic reconstructions. *Mol Syst Biol* **5:** 320

Oberhardt MA, Puchalka J, Fryer KE, dos Santos VAPM, Papin JA (2008) Genome-scale metabolic network analysis of the opportunistic pathogen Pseudomonas aeruginosa PAO1. *J Bacteriol* **190:** 2790–2803

Orth JD, Conrad TM, Na J, Lerman JA, Nam H, Feist AM, Palsson BØ (2011) A comprehensive genome-scale reconstruction of Escherichia coli metabolism. *Mol Syst Biol* **7:** 535

Pirt SJ (1965) The maintenance energy of bacteria in growing cultures. *Proc R Soc Lond B Biol Sci* **163:** 224–231

Pollack JD, Somerson NL, Senterfit LB (1970) Isolation, characterization, and immunogenicity of Mycoplasma pneumoniae membranes. *Infect Immun* **2:** 326–339

Pollack JD, Williams MV, McElhaney RN (1997) The comparative metabolism of the mollicutes (Mycoplasmas): the utility for taxonomic classification and the relationship of putative gene annotation and phylogeny to enzymatic function in the smallest free-living cells. *Crit Rev Microbiol* **23:** 269–354

Price ND, Reed JL, Palsson BØ (2004) Genome-scale models of microbial cells: evaluating the consequences of constraints. *Nat Rev Microbiol* **2:** 886–897

Puchalka J, Oberhardt MA, Godinho M, Bielecka A, Regenhardt D, Timmis KN, Papin JA, Martins dos Santos VAP (2008) Genome-scale reconstruction and analysis of the Pseudomonas putida KT2440 metabolic network facilitates applications in biotechnology. *PLoS Comput Biol* **4:** e1000210

Razin S, Argaman M, Avigan J (1963) Chemical composition of mycoplasma cells and membranes. *J Gen Microbiol* **33:** 477–487

Reed JL, Famili I, Thiele I, Palsson BØ (2006) Towards multidimensional genome annotation. *Nat Rev Genet* **7:** 130–141

Reed JL, Palsson BØ (2003) Thirteen years of building constraint-based in silico models of Escherichia coli. *J Bacteriol* **185:** 2692–2699

Rolfsson O, Palsson BØ, Thiele I (2011) The human metabolic reconstruction Recon 1 directs hypotheses of novel human metabolic functions. *BMC Syst Biol* **5:** 155

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

Sakamoto M, Uchimera T, Komagata K (1996) Comparison of H_2O-forming NADH oxidase from Leuconostoc mesenteroides subsp. mesenteroides NRIC 1541 T and H_2O_2-forming NADH oxidase from Sporolactobacillus inulinus NRIC 1133 T. *J Ferment Bioeng* **82:** 531–537

Savinell JM, Palsson BØ (1992a) Optimal selection of metabolic fluxes for *in vivo* measurement. I. Development of mathematical methods. *J Theor Biol* **155:** 201–214

Savinell JM, Palsson BØ (1992b) Optimal selection of metabolic fluxes for *in vivo* measurement. II. Application to Escherichia coli and hybridoma cell metabolism. *J Theor Biol* **155:** 215–242

Schellenberger J, Que R, Fleming RMT, Thiele I, Orth JD, Feist AM, Zielinski DC, Bordbar A, Lewis NE, Rahmanian S, Kang J, Hyduke DR, Palsson BØ (2011) Quantitative prediction of cellular metabolism with constraint-based models: the COBRA Toolbox v2.0. *Nat Protoc* **6:** 1290–1307

Schmidl SR, Otto A, Lluch-Senar M, Piñol J, Busse J, Becher D, Stülke J (2011) A trigger enzyme in Mycoplasma pneumoniae: impact of the glycerophosphodiesterase GlpQ on virulence and gene expression. *PLoS Pathog* **7:** e1002263

Schuetz R, Zamboni N, Zampieri M, Heinemann M, Sauer U (2012) Multidimensional optimality of microbial metabolism. *Science* **336:** 601–604

Seybert A, Herrmann R, Frangakis AS (2006) Structural analysis of Mycoplasma pneumoniae by cryo-electron tomography. *J Struct Biol* **156:** 342–354

Suthers PF, Dasika MS, Kumar VS, Denisov G, Glass JI, Maranas CD (2009) A genome-scale metabolic reconstruction of Mycoplasma genitalium, iPS189. *PLoS Comput Biol* **5:** e1000285

Szappanos B, Kovács K, Szamecz B, Honti F, Costanzo M, Baryshnikova A, Gelius-Dietrich G, Lercher MJ, Jelasity M, Myers CL, Andrews BJ, Boone C, Oliver SG, Pál C, Papp B (2011) An integrated approach to characterize genetic interaction networks in yeast metabolism. *Nat Genet* **43:** 656–662

Tatusova TA, Karsch-Mizrachi I, Ostell JA (1999) Complete genomes in WWW Entrez: data representation and analysis. *Bioinformatics* **15:** 536–543

Thorleifsson SG, Thiele I (2011) rBioNet: A COBRA toolbox extension for reconstructing high-quality biochemical networks. *Bioinformatics* **27:** 2009–2010

Tong AH, Evangelista M, Parsons AB, Xu H, Bader GD, Pagé N, Robinson M, Raghibizadeh S, Hogue CW, Bussey H, Andrews B, Tyers M, Boone C (2001) Systematic genetic analysis with ordered arrays of yeast deletion mutants. *Science* **294:** 2364–2368

Tong AHY, Lesage G, Bader GD, Ding H, Xu H, Xin X, Young J, Berriz GF, Brost RL, Chang M, Chen Y, Cheng X, Chua G, Friesen H, Goldberg DS, Haynes J, Humphries C, He G, Hussein S, Ke L *et al* (2004) Global mapping of the yeast genetic interaction network. *Science* **303:** 808–813

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

Varma A, Palsson BØ (1994a) Metabolic flux balancing: basic concepts, scientific and practical use. *Nat Bio/Technol* 12

Varma A, Palsson BØ (1994b) Stoichiometric flux balance models quantitatively predict growth and metabolic by-product secretion in wild-type Escherichia coli W3110. *Appl Environ Microbiol* **60:** 3724–3731

Waites KB, Talkington DF (2004) Mycoplasma pneumoniae and its role as a human pathogen. *Clin Microbiol Rev* **17:** 697–728table of contents available at http://dx.doi.org/10.1128/CMR.17.4.697-728.2004

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