

# Exploring the adaptability and robustness of the central carbon metabolism of *Mycoplasma pneumoniae*

Niels A. Zondervan<sup>1</sup>, Eva Yus<sup>2</sup>, Daniel C. Sévin<sup>3¶</sup>, Sira Martinez<sup>2</sup>, Carolina Gallo<sup>2</sup>, Peter J.

Schaap<sup>1</sup>, Maria Lluch-Senar<sup>2</sup>, Luis Serrano<sup>2,4,5</sup>, Vitor A. P. Martins dos Santos<sup>1,6,7</sup> and Maria Suarez-Diez<sup>1</sup>

<sup>1</sup> Laboratory of Systems and Synthetic biology, Wageningen University & Research, the Netherlands

<sup>2</sup>Centre for Genomic Regulation (CRG), The Barcelona Institute for Science and Technology, Doctor Aiguader 88, Barcelona, 08003, Spain.

<sup>3</sup> Institute of Molecular Systems Biology, ETH Zürich, 8093 Zürich, Switzerland

<sup>4</sup>Universitat Pompeu Fabra (UPF), Barcelona, Spain

<sup>5</sup>ICREA, Pg. Lluís Companys 23, 08010 Barcelona, Spain

<sup>6</sup>LifeGlimmer GmbH, Markelstrasse 38, 12163 Berlin, Germany

<sup>7</sup> Laboratory of Bioprocess Engineering, Wageningen University & Research, the Netherlands

¶ Current address: Cellzome, GlaxoSmithKline, Meyerhofstrasse 1, Heidelberg 69117, Germany

## Abstract

In this study we explored the adaptability and robustness of glycolysis and pyruvate metabolism of *Mycoplasma pneumoniae* (MPN). We used a dual approach, we analysed metabolomics data collected for a large number of OE and KO mutants and perturbation samples. Furthermore, we trained a dynamic model of central carbon metabolism and tested

the model's capacity to predict these mutants and perturbation samples as well as identify key controlling factors in central carbon metabolism. Our analysis of metabolite data as well as our model analysis indicate *MPN* metabolism is inherently robust against perturbations due to its network structure. Two key control hubs of central carbon metabolism were identified.

## Abbreviation list

AcCoA = Acetyl coenzyme A

ACE = Acetate

ACK = Acetate kinase

ADP = Adenosine diphosphate

ATP = Adenosine triphosphate

ATPase = Adenylpyrophosphatase

CoA = Coenzyme A dehydrogenase

DGP = Diacylglycerol phosphate

ENO = Enolase

F6P = Fructose 6-phosphate

FBA = Fructose-bisphosphate aldolase

FBP = Fructose-1,6-bisphosphatase

G6P = Glucose 6-phosphate

GAP = Glyceraldehyde-3-phosphate

GAPDH = Glyceraldehyde 3-phosphate

GLC\_Ext = Extracellular glucose

GMP = Glycerate phosphomutase

Kcat = Enzyme catalytic rate

Keq = Equilibrium constant

Km = Michaelis-Menten constant

KO = Knock out

LAC = Lactate

LDH = Lactate dehydrogenase

MPN = Mycoplasma pneumonia

NAD = Nicotinamide adenine dinucleotide

NADH = Reduced nicotinamide adenine dinucleotide

OE = Over-expression

PDH = Pyruvate dehydrogenase

PEP = Phosphoenolpyruvate

PFK = Phosphofructokinase

PGI = Phosphoglucose isomerase

PGK = Phosphoglycerate kinase

Pi\_Int = Orthophosphate

PTA = Phosphotransacetylase

PTS\_Glc = Phosphotransferase system

PYK = Pyruvate kinase

PYR = Pyruvate

## Introduction:

*Mycoplasma* are gram positive bacteria adapted to an obligatory parasitic lifestyle able to infect a broad range of hosts<sup>1</sup>. It is estimated by the CDC that 2 million infections with *M. pneumoniae* (*MPN*) occur in the US alone on a yearly basis<sup>2</sup>. These infections lead to conditions ranging from mild to severe respiratory illness including life threatening conditions such as auto-immune diseases<sup>3</sup>. Therefore, it is imperative to improve our understanding of *MPN*.

*MPN* has been established as a model organism for systems biology and large dataset collections are available informing on its genome, transcriptome<sup>4</sup>, proteome<sup>5</sup>, metabolome<sup>6</sup>, and transcriptional adaptations<sup>7</sup>. The metabolism and energetic expenditure of *MPN* have been thoroughly studied by combining a genome-scale constraint based model of metabolism with detailed experimental characterizations<sup>8</sup>. Neither energy production nor uptake of protein building blocks appear to limit growth of *MPN* only protein synthesis was found to be growth limiting<sup>6,9</sup>. Maintenance requirements are high for *MPN*, so most of the energy is devoted to maintenance instead of growth<sup>8</sup>.

*MPN* like other *Mycoplasma*'s has adopted to its pathogenic lifestyle leading to a severely reduced genome. Despite their small genome size and limited number of enzymes and relatively low number of regulators<sup>10</sup>, *Mycoplasma*'s are still able to adapt to a large number of conditions and still genes can be removed, as they have been seen not to be essential<sup>11</sup>. Many essential genes are however only present in some *Mycoplasma* species, suggesting alternatives to a minimal genome exist<sup>12</sup>. Since there are only few regulatory elements in the genome<sup>10</sup> we hypothesized that a lot of the adaptability of *MPN* to adapt to changing growth conditions must be due to network structure and to allosteric control of its metabolism.

In this study we investigated the metabolism of *MPN* with a focus on central carbon metabolism and its allosteric control. Dynamic models were successfully used to investigate regulation and adaptation of central carbon metabolism to changing environmental conditions in other organism<sup>13–20</sup>.

Therefore, in this study we will combine analysis of metabolomics a large number of samples taken from varying environmental conditions, OE and KO mutants with a dynamic model of glycolysis and pyruvate metabolism, to identify key controlling metabolites and enzymes in central carbon metabolism. We tested single or combined addition of 1) an ATPase reaction, 2) O<sub>2</sub> inhibition of Lactate Dehydrogenase and 3) NAD regeneration by NoxE using O<sub>2</sub> to this model as potential mechanisms for *MPN*'s adaptability to various conditions.

We trained and tested the model's ability to predict a wide range of environmental

conditions, single and double overexpression mutants as well as mutants with single gene deletions. The model was able to predict these mutants with reasonable accuracy.

In a recent study, local sensitivity analysis on a dynamic model of *E. coli* central carbon metabolism identified robustness as one of properties of central carbon metabolism of *E. coli*<sup>16</sup>. This robustness is a system property resulting from the many feed-forward and feed-backward interactions in metabolism, such as allosteric control of glucose uptake as well as lactate and acetate metabolism. Another study in *E. coli* revealed that only three metabolites (FBP, F1P and cAMP) account for about 70% of the expression variability of central carbon metabolism enzymes through control of two transcription factors<sup>21</sup>. Similarly, our model predicts the central carbon metabolism of *MPN* to be inherently robust to changing conditions and identifies two main hubs of metabolic control. Clustering of samples of FBA OE and LDH KO mutants corroborate assumed allosteric control of LDH FBP. Additionally, the analysis of metabolomics data of *MPN* indicated that glycolipid metabolism might be linked to the high energy metabolites needed for growth of *MPN*. Our findings are in agreement with recent findings where some key lipids were identified to be needed for *MPN* growth on serum free medium<sup>22</sup>.

## Materials and methods

### Bacterial strains and culture conditions

*M. pneumoniae* strain M129 (passage 33-34) was grown in modified Hayflick medium and transformed by electroporation with the pMT85 transposon as previously described<sup>9</sup>. Briefly, cells were split 1:10, and washed twice with 10 mL and collected in 300 µl Electroporation buffer (8 mM Hepes·HCl, 272 mM sucrose, pH 7.4) three days later. Cells (50 µl) were electroporated with 5 µg plasmid in 1 mm gapped cuvettes at 1.25 kV, 100 Ω, 25 µF (Gene Pulser Xcel Electroporator, Bio-Rad). Cells were recovered in Hayflick for 2 h at 37 °C, diluted 1:5 in Hayflick with 200 µg mL<sup>-1</sup> gentamycin, selected for three days and then maintained with 80 µg mL<sup>-1</sup> gentamycin. The cell lines used are detailed in **Error!**

**Reference source not found..**

### Transposon insertion mutants obtained by haystack mutagenesis

For the isolation of *M. pneumoniae* mutants, we used a collection of strains carrying insertions of transposon Tn4001<sup>23</sup>. The presence of the desired mutant was assayed by PCR

using one primer that hybridizes to the transposon (directed outwards), and a second primer specific for the gene of interest. Mass spectroscopy is used to verify absence of the corresponding protein.

## Over Expression mutant construction

Genes to be overexpressed were cloned in the transposon Tn4001<sup>24</sup> control of the promoter of the EF-tu gene<sup>24</sup>.

## Growth curves

To obtain equal amounts of each sample, initial inocula for the growth curves were quantified. Briefly, cells were grown for 3 days in 25-cm<sup>2</sup> flasks, collected in 1 mL medium and 100 µl was used for quantification using the BCA (bicinchoninic acid) protein assay kit (Pierce, see below). Same amounts of total protein (1 µg) were aliquoted per well in a 96-multiwell plate in duplicates. Two hundred µl of Hayflick medium was added per well and the cells were incubated in a Tecan Infinite plate reader at 37°C. The “growth index” (absorbance 430/560 nm, settle time at 300 msec and number of flashes equal to 25) was obtained every hour for 5 days as published<sup>9</sup>. To quantify growth, we determined two slopes of the growth curve. The first one is based on the time interval from 10 to 30 h (“early slope”) and the second one on the whole growth curve (“late”). The early slope was determined by considering the maximum median of the slope between two time points (eq. 1) separated by three time measurements over successive periods of 30 time points. The late slope was determined by considering the maximum median value of the slope between two time points separated by four time measurements (eq. 2) over successive periods of 30 time points.

$$\text{Early Slope} = (\text{value (time [i])} - \text{value (time[i+3])}) / (\text{time[i]} - \text{time[i+3]}) \text{ (eq. 1)}$$

$$\text{Late Slope} = (\text{value (time [i])} - \text{value (time[i+4])}) / (\text{time[i]} - \text{time[i+4]}) \text{ (eq. 2)}$$

The early slope is more representative of growth, while the late slope reflects the metabolic activity.

On the other hand, biomass was quantified at 48 h (early stationary phase) by inoculating a twin 96-well plates, in the same conditions as above. After incubation for two days at 37°C under, medium was sucked out, cells were carefully washed twice with 200 µl PBS and lysed with 100 µl lysis buffer (10 mM Tris·HCl, 6 mM MgCl<sub>2</sub>, 1 mM EDTA, 100 mM NaCl, 0.1% Tx-100, pH 8, and 1× Protease Inhibitor Cocktail, Roche) at 4°C. In the same first 96-well plate, cell lysates were kept on ice and extracted protein was quantified by BCA Protein Assay Kit (Pierce, see below).

The protein concentrations at 48 h and early slope are more representatives of growth, while the late slope and the value of A<sub>430/560</sub> at midpoint reflect the metabolic activity. These four parameters of growth and metabolism were analysed for each batch of experiments. Outliers (larger than quartile 3, Q3) by at least 1.5 times the interquartile range (IQR), or smaller than Q1 by at least 1.5 times the IQR) were removed to calculate the mean and the standard deviation of each of the parameters for each batch. Values larger or smaller than the mean by at least 2 times the standard deviation of each parameter were considered to determine fast- and slow-growing/metabolizing clones, respectively.

### Strain cultivation and growth conditions of mutant and perturbation samples

A 300 cm<sup>2</sup> flask was inoculated 1:10 with the lab stock and 100 mL of Hayflick and grown for 3-4 days at 37°C. Then, medium was removed, and cells were scrapped and resuspended in 12 mL medium. From this inoculum, 75 cm<sup>2</sup> flasks were seeded with 1 mL of inoculum in 20 mL of Hayflick. After 6 hours of incubation (i.e. when cells reached stationary growth phase) the cells were treated as follows, before the standard extraction protocol:

- Glucose starvation: remove medium and add new Hayflick medium without glucose. Incubate sample for 5 h at 37°C. Long incubation time is required to deplete glucose from Hayflick medium.
- Amino acid starvation: take half of the medium, add 200 mg of DL-serine hydroxamate (10 mg/mL), mix and add again to the cells; incubate cells with for 15 min at 37°C.

- Fe<sup>2+</sup> depletion: Add directly to the flask the iron chelator 2,2'-Bipyridine at a final concentration of 3 mM, incubate for 30 min at 37° C.
- Oxidative stress: Add directly to the flask H<sub>2</sub>O<sub>2</sub>, to 0.5%, incubate 15 min at 37C.
- Glycerol: Add directly to the flask glycerol to 1% -v/v, incubate 30 min at 37C.

## Sample preparation for metabolomics

*M. pneumoniae* cells were grown in 6-well culture dishes as described above until reaching 80-90% confluency. Culture medium was aspirated, and cells were rapidly washed twice at 37° C with 1 mL of buffer (75 mM ammonium carbonate at pH 7.4 and 0.1% glucose). After aspiration of washing buffer, plates were immersed in liquid nitrogen to quench metabolism and stored at -80° C for less than 4 days until further processing. After aspiration of washing buffer, plates were immersed in liquid nitrogen to quench metabolism and stored at -80° C for less than 4 days until further processing.

To extract metabolites, plates were placed on a 75° C heating block and 700 µL of extraction solution (70%-v/v ethanol in water at 75° C) were added to each well. After incubating for 3 min, the supernatant was collected and transferred to ice, and the extraction was repeated once. Pooled extracts were dried under vacuum and stored at -80° C prior to metabolomics analyses.

## Nontargeted metabolomics

All samples were measured in triplicate. Metabolomics samples were analysed by flow-injection time-of-flight MS with an Agilent 6550 iFunnel QToF instrument (Agilent, Santa Clara, CA, U.S.A.) operated in negative ionization mode at 4 GHz high-resolution in a range from 50-1,000 *m/z* using published settings<sup>25</sup>. The mobile phase was 60:40 isopropanol:water (v/v) and 1 mM NH<sub>4</sub>F at pH 9.0 supplemented with 10 nM hexakis(1H-, 1H-, 3H-tetrafluoropropoxy)phosphazine and 80 nM taurocholic acid for online mass correction. Spectral processing and ion annotation based on accurate mass within 0.001 Da of metabolites in the *M. pneumoniae* MyMPN database<sup>26</sup>, allowing for [M-H]<sup>-</sup> and [M+F]<sup>-</sup> ions and [1x<sup>12</sup>C->1x<sup>13</sup>C] neutral gain and keeping for each metabolite only the ion with lowest *m/z*.



in case of multiple matching ions, was performed using Matlab R2015b (The Mathworks, Nattick, MA, U.S.A.) as described previously<sup>25</sup>. Metabolomics data were normalized to the summed abundance of a group of amino acids (Ser, Pro, Ala, Val, Thr, Leu/Ile, Met, Phe, Tyr) found to strongly correlate in each sample. In mycoplasma amino acids are not made but imported and they are fairly constant. Therefore, we could use the summed values for the less variable amino acids to normalize. A similar approach is used in free label quantitative proteomics. Subsequently, log<sub>2</sub>-transformed fold-changes and *P*-values (two-sided *t* tests, with q-values computed from raw p-values to enable false discovery rate adjustment<sup>27</sup>) were calculated to determine relative metabolite abundances compared to control samples and their statistical significance.

## Targeted metabolomics

Samples were injected into a Waters Acquity UPLC with a Waters T3 column (150 mm x 2.1 mm x 1.8 mm; Waters Corporation, Milford, MA) coupled to a Thermo TSQ Quantum Ultra triple quadrupole instrument (Thermo Fisher Scientific, Waltham, MA) with electrospray ionization. Compound separation was achieved by a gradient of two mobile phases (i) 10 mM tributylamine, 15 mM acetic acid, 5% (v/v) methanol and (ii) 2-propanol. In total, 138 metabolites covering carbohydrate and energy metabolism, amino acid metabolism, nucleotide metabolism and other pathways were targeted. Further details are published elsewhere<sup>28</sup>.

## Proteomics

Cells were grown in a 25-cm<sup>2</sup> flask for 3 days as above, washed with PBS and lysed/collected in 4% SDS, and 0.1 M Hepes·HCl pH 7.5. Samples were reduced with dithiothreitol (15 μM, 30 min, 56°C), alkylated in the dark with iodoacetamide (180 nmols, 30 min, 25°C) and digested with 3 μg LysC (Wako) O/N at 37°C and then with 3 μg of trypsin (Promega) for eight hours at 37°C following FASP procedure (Filter-aided sample preparation 48). After digestion, the peptide mix was acidified with formic acid and desalted with a MicroSpin C18 column (The Nest Group, Inc) prior to LC-MS/MS analysis. The peptide mixes were analysed using a LTQ-Orbitrap Velos Pro mass spectrometer (Thermo

Fisher Scientific) coupled to an EasyLC (Thermo Fisher Scientific). Peptides were loaded onto the 2-cm Nano Trap column with an inner diameter of 100  $\mu\text{m}$  packed with C18 particles of 5  $\mu\text{m}$  particle size (Thermo Fisher Scientific) and were separated by reversed-phase chromatography using a 25-cm column with an inner diameter of 75  $\mu\text{m}$ , packed with 1.9  $\mu\text{m}$  C18 particles (Nikkyo Technos). Chromatographic gradients started at 93% buffer A and 7% buffer B with a flow rate of 250  $\text{nl min}^{-1}$  for 5 minutes and gradually increased 65% buffer A and 35% buffer B in 120 min. After each analysis, the column was washed for 15 min with 10% buffer A and 90% buffer B. Buffer A: 0.1% formic acid in water. Buffer B: 0.1% formic acid in acetonitrile.

The mass spectrometer was operated in DDA mode and full MS scans with 1 micro scans at resolution of 60,000 were used over a mass range of  $m/z$  350-2,000 with detection in the Orbitrap. Auto gain control (AGC) was set to 1 E6, dynamic exclusion (60 seconds) and charge state filtering disqualifying singly charged peptides was activated. In each cycle of DDA analysis, following each survey scan the top twenty most intense ions with multiple charged ions above a threshold ion count of 5,000 were selected for fragmentation at normalized collision energy of 35%. Fragment ion spectra produced via collision-induced dissociation (CID) were acquired in the Ion Trap, AGC was set to 5e4, isolation window of 2  $m/z$ , activation time of 0.1 ms and maximum injection time of 100 ms was used. All data were acquired with Xcalibur software v2.2.

Proteome Discoverer software suite (v2.0, Thermo Fisher Scientific) and the Mascot search engine (v2.5, Matrix Science) were used for peptide identification. Samples were searched against a *M. pneumoniae* database with a list of common contaminants and all the corresponding decoy entries (87,059 entries). Trypsin was chosen as enzyme and a maximum of three mis-cleavages were allowed. Carbamidomethylation (C) was set as a fixed modification, whereas oxidation (M) and acetylation (N-terminal) were used as variable modifications. Searches were performed using a peptide tolerance of 7 ppm, a product ion tolerance of 0.5 Da. Resulting data files were filtered for FDR < 5 %. Protein Top 3 areas were calculated with unique peptides per protein.

## Data and model management

All omics data, modelling files as well as a backup of modelling pipeline and simulation outputs are available via the Seek data and model management platform for maximum reproducibility (<http://doi.org/10.15490/FAIRDOMHUB.1.INVESTIGATION.133.3>)<sup>29,30</sup>. SBtab<sup>31</sup>, a tabular exchange format was used to add minimal information compliant with the Minimal Information Requirements In the Annotation of Models (MIRIAM)<sup>32</sup> compliant annotation and to add Systems Biology Ontology (SBO) identifiers<sup>33</sup> for metabolites, reactions and parameters in the model.

## Data analysis

Relative metabolite measurements were log10-transformed following the recommendation of Jauhiainen et al<sup>34</sup>. Pearson correlations between metabolites were computed using R v3.4.2<sup>35</sup>. To remove batch effects in the metabolites measurement, values were normalized by dividing them by the median measured metabolite value per batch<sup>36</sup>. Fold Change (FC) metabolite measurements for the 40 independent samples were analysed through Principal Component analysis using the prcomp package. Pearson correlation between samples and metabolites were calculated and used to generate heatmaps of sample correlations and metabolite correlations. We used the metabolite correlation matrix for [M+F]-ion data and filtered on correlations with at least p-value cut-off of 0.001. We calculated the Euclidean distance using the complete linkage method and used Hierarchical clustering and cut tree to identify 6 clusters in the metabolite correlation data. Absolute measurements needed for simulations with the model were obtained by multiplying relative metabolite values from [M-H]<sup>-</sup> measurements with quantitative metabolite measurements at 24 h. Similarly, enzyme concentrations of samples with overexpression (OE) of enzymes were obtained by multiplying relative measurements for these mutants with absolute measurements of the wild type at their respective time point. These computations were performed using Python. Additionally, we calculated Pearson correlation between metabolite concentration and estimated growth of 24 time-series samples (P-value <0.05) applying Benjamin Hocheman Multiple testing correction.

## Model construction and numerical implementation

A dynamic model was build of glycolysis an pyruvate metabolism. A base model containing all reactions in glycolysis, pyruvate metabolism from the MyMPN database<sup>26</sup>. Different additions to this model were tested such as individual and combined additions of 1) an ATPase reaction, 2) LDH inhibition by oxygen and 3) a NoxE reaction for NAD regeneration using oxygen. The tested models include i) the base model ii) the base model and the ATPase reaction iii) base model with NoxE reaction iv) base model with both ATPase and NoxE reaction v) based model with NoxE reaction and LDH inibhition by oxygen and vi) the base model with all three modifications. In case intermediate metabolites were not measurable, reactions were lumped in a single reaction. This was the case for Phosphoglycerate kinase (PGK), Glycerate phosphomutase (GMP) and enolase (ENO). These three reactions were combined in reaction re07 lumping the enzymatic reactions of PGK&GMP&ENO. Similarly, phosphotransacetylase (PTA) and acetate kinase (ACK) were combined in reaction re10 which lumps the enzymatic reactions of PTA&ACK.

Allosteric control was assumed to be similar to allosteric control in *Lactococcus lactis* as presented in the model by Costa *et al*<sup>37</sup> due to the lack of *MPN* specific information on allosteric control. Allosteric control includes three activator and five inhibitor effects. Reactions were modelled using modular rate laws except for transport reactions for which Hill type kinetics were used. Enzyme concentrations were included as reaction parameter to allow model predictions at varying protein concentrations. The base model contains 10 equations and 72 parameters of which 10 represent experimentally determined enzyme concentrations, 5 represent equilibrium constants ( $K_{eq}$ ) and 1 is a Hill coefficient. The remaining 56 parameters represent Michaelis-Menten constants, activation constants and inhibition constants which are not known for *MPN*. The model was built using COPASI<sup>38</sup>.

## Initial parameter values

Proteomics measurements for 6, 24 and 48-hour timepoints were used as estimates for enzyme concentrations at respective time points. In case of multi-subunit enzymes, the most abundant single copy subunit was chosen to represent the enzyme concentration. Many lower abundance subunits are only expressed under specific conditions and as such are not representative of the abundance of these glycolysis enzymes. We also tested using the

average, but no major differences were found. Equilibrium constants were gathered from [www.equilibrator.org](http://www.equilibrator.org) assuming an ionic strength of 0.1 and a pH of 7<sup>39</sup>. Initial values for Monod constants and allosteric control constants were randomly selected between 0.01 and 100 fold of the observed metabolite concentrations at 24h. An overview of the six models reactions and equations can be found in Supplementary file 1 A.

## Model selection and parameter estimation

The base model has 72 parameters of which 56 are unknown while model that includes ATPase, NoxE and O<sub>2</sub> inhibition of LDH has 80 parameters of which 63 are unknown. Parameter estimation was performed training the models on metabolomics steady state data obtained from growth curve samples for 6h, 24h and 48h time points grown on medium containing 60 mM of glucose. These samples were selected as training data due to the completeness of the data available for these three times points. Only for these three samples, measurements for all 17 metabolites present in the model were available. In addition, protein copy number, glucose uptake, lactate secretion, and acetate secretion measurements were available for these three samples. Steady state concentration for 6h, 24h and 48h grown on a lower glucose concentration of 10 mmol were used as internal validation data. Internal validation data is used by COPASI to stop the parameter estimation algorithm from overfitting parameters to the training data. The large time interval between the samples means that metabolite concentrations in each sample can be assumed to be independent from the concentrations of the other samples. Therefore, each sample was treated as an independent steady state.

COPASI's<sup>38</sup> build in Genetic programming algorithm was used to estimate parameters using a maximum of 1000 generations with a population size of 500 models with normalized sum of squares as weights. 100 independent parameter estimations were run per model. Optimal parameters were searched within a range of 10<sup>-2</sup>-10<sup>2</sup>-fold of the observed metabolite concentration at 24h for Monod constants and allosteric control constants while maximum reaction velocity values were searched within a range of 10<sup>-2</sup>-10<sup>3</sup>. The performance of the six models were compared based on the distribution of the mean square error values for each of the 100 parameter estimations.

## Simulations, local and global sensitivity analysis

The model was used to predict steady state concentrations for 40 independent samples comprised of OE and knock out (KO) mutants, perturbations, and time-series measurements in different growth conditions measured in triplicate. In these simulations, the input for the model was concentration data of 11 metabolites: acetyl coenzyme A (AcCoA), acetate (ACE), adenosine diphosphate (ADP), adenosine triphosphate (ATP), coenzyme A (CoA), diacylglycerol phosphate (DGP), fructose 6-phosphate (F6P), fructose-1,6-bisphosphatase (FBP), glucose 6-phosphate (G6P), glyceraldehyde-3-phosphate (GAP), lactate (LAC), nicotinamide adenine dinucleotide (NAD), reduced nicotinamide adenine dinucleotide (NADH), phosphoenolpyruvate (PEP), orthophosphate (Pi\_Int), pyruvate (PYR), external glucose (GLC\_Ext). Measurements for NAD and NADH are approximate. For each sample, 1000 steady state simulations were performed while sampling from the log normal distribution of metabolite measurements. By comparing sampled measurements and sampled simulation values, measurement error and its propagation are incorporated in model predictions.

Not all metabolites present in the model were measured for all independent samples. Reference values from measurements taken at 6h, 24h and 48h of growth on high glucose concentrations used to train the model were used to set the initial concentration of NAD, NADH and orthophosphate. Reference values were also used for CoA, Acetyl-CoA, and Lactic Acid (LAC) for some of the independent samples (Supplementary Material 2).

To compare the error between simulated and measured metabolite concentrations in a consistent manner, we used the symmetric Mean Absolute Percentage Error (sMAPE). sMAPE is a measure of prediction accuracy used for forecasting methods. This method has the advantage of providing an equal error to positive and negative errors for log normal distributed data such as metabolite measurements and predictions<sup>40</sup>.

We performed global sensitivity analysis for all kcat, Monod constants, activation and inhibition constants using a 100,000 Latin Hypercube sampling<sup>41,42</sup>. Samples were constructed by sampling from the log linear distribution of each parameter's respective search range.

The above described operations were performed using Python 3.6.5 with the Tellurium 2.0.18 and Roadrunner 1.4.24 high performance SBML simulation and analysis libraries<sup>43,44</sup>. The

pyDOE package was used for Latin Hypercube sampling. Conda version 4.3.21 was used for package management.

## Modelling oxygen diffusion

Oxygen concentrations were calculated based on the initial oxygen concentration in the culture flasks and the acetate production rate which requires NAD to be regenerated from NADH by the oxygen dependent reaction catalysed by NoxE. The initial oxygen concentration was calculated with the ideal gas law, using the temperature used in cultivation (37 degrees Celsius) and atmospheric pressure. Volumes, surface area height of the medium were calculated based on the medium and inoculant volume and the specifications of the T300 cell culture flask<sup>45</sup>.

Diffusion of oxygen from the head space into the medium was calculated using the Wilke and Chang correlation<sup>46</sup> while Fick's law<sup>47</sup> was used to calculate the diffusion of oxygen to the bottom of the flask at 6h, 24h, 48h and 96 hours of growth. The calculated oxygen concentrations were added to the metabolomics measurements for the 95 independent samples.

## Results

Datasets were collected growing *MPN* in a large number of conditions.

*M. pneumoniae* was grown in suspension until sedimenting after 6 h of incubation in rich medium in non-aerated, non-stirred conditions mimicking its host environment. At several time points during the growth of wild-type *MPN*, samples were taken for metabolomics, biomass, pH and acetate concentration measurements. In addition, relative metabolite concentrations were measured by untargeted metabolomics for 40 different samples from environmental perturbations, genetic mutations and at 6h, 24h, 48h and 96 hours of growth. The metabolomics data and targeted proteomics data for these samples is available in Supplementary file 2.

Among these 40 datasets, there were data corresponding to OE mutants for all glycolysis and pyruvate metabolism enzymes except for pyruvate dehydrogenase (PDH) of which the



complex is large to clone and OE, as well as for the KO of LDH (Mpn674). The fold change in mRNA and or protein concentrations for genes targeted in each mutant were measured. Of the 40 datasets, 17 are mutants that target enzymes for which a reaction is present in the model. Of these mutants 2 are KO mutants and 14 are OE mutants and 1 is a combined KO and OE mutant (Table 1). Additionally, there are 6 mutants targeting enzymes in the pentose phosphate pathway which is connected to the glycolysis via F6P.

Changes in mRNA and protein concentration of enzymes targeted in over expression (OE) and knock out (KO) mutants were also measured. Table 1 gives an overview of these conditions and mutants.

Unless stated otherwise, relative metabolite concentrations were measured at steady state in non-aerated conditions. In cases where different conditions were used, or where additional omics data were measured, this is indicated in Table 1. Targeted metabolomics were used to measure protein concentration for all OE mutants targeting central carbon metabolism enzymes.

Experiment	Type	Enzyme in model
24h_timecourse *	Absolute Metabolomics, Proteomics, Glucose uptake, acetate lactate secretion rate and lactate secretion rate	NA
48h_timecourse *	Absolute Metabolomics, Proteomics, Glucose uptake, acetate lactate secretion rate and lactate secretion rate	NA
6h_timecourse *	Absolute Metabolomics, Proteomics, Glucose uptake, acetate lactate secretion rate and lactate secretion rate	NA
24h_timecourse_4 **	Glucose uptake, acetate lactate secretion rate and lactate secretion rate	NA
48h_timecourse_4 **	Glucose uptake, acetate lactate secretion rate and lactate secretion rate	NA
6h_timecourse_4 **	Glucose uptake, acetate and lactate secretion rate	NA
blank_control_7	Control	NA
water_control_7	Control	NA
WT_5	Control	NA
WT_perturbation_7	Control	NA
KO51_mutant_6	mutant, glpD KO	NA
MPN025-OE_6	mutant, tsr OE	FBA
MPN025-OE_7	mutant, tsr OE	FBA



MPN051-KO_5	mutant, glpD KO	NA
MPN051-OE_7	mutant, glpD KO	NA
MPN250-OE_5	mutant, pgi OE	PGI
MPN250-OE_7	mutant, pgi OE	PGI
MPN302-OE_6	mutant, pfkA OE	PFK
MPN302-OE_7	mutant, pfkA OE	PFK
MPN303-OE_5	mutant, pyk OE	PYK
MPN303-OE_6	mutant, pyk OE	PYK
MPN303-OE_7	mutant, py OE	PYK
MPN430-OE_6	mutant, gap OE	GAP
MPN606-OE_6	mutant, eno OE	ENO
MPN627-OE_7	mutant, ptsI OE	NA
MPN674-KO, NoxE OE_5	mutant, ldh KO, noxE OE	LDH, NOXE
MPN674-KO_5	mutant, ldh KO	LDH
MPN674-KO_6	mutant, ldh KO	LDH
MPN674-OE_5	mutant, ldh OE	LDH
MPN674-OE_7	mutant, ldh OE	LDH
Tn674_mutant_7	mutant, ldh KO	LDH
Tn051_Gly_perturbation_7	mutant, perturbation glpD KO	NA
AA_perturbation_6	Perturbation	NA
Fe_perturbation_6	Perturbation	NA
Glu_perturbation_6	Perturbation	NA
Glucose_starv_perturbation_7	Perturbation	NA
Gly_ctrl_perturbation_7	Perturbation	NA
Gly_perturbation_6	Perturbation	NA
Ox_perturbation_6	Perturbation	NA
WT_noGluc_perturbation_7	Perturbation	NA
M129_timecourse_24h_3 ***	time course, perturbation	NA
M129_timecourse_48h_3 ***	time course, perturbation	NA
M129_timecourse_6h_3 ***	time course, perturbation	NA
M129_timecourse_96h_3 ***	time course, perturbation	NA

\* Used to train the model.

\*\* Used to validate the model.

\*\*\* *M. pneumoniae* M129 grown in aerated conditions.

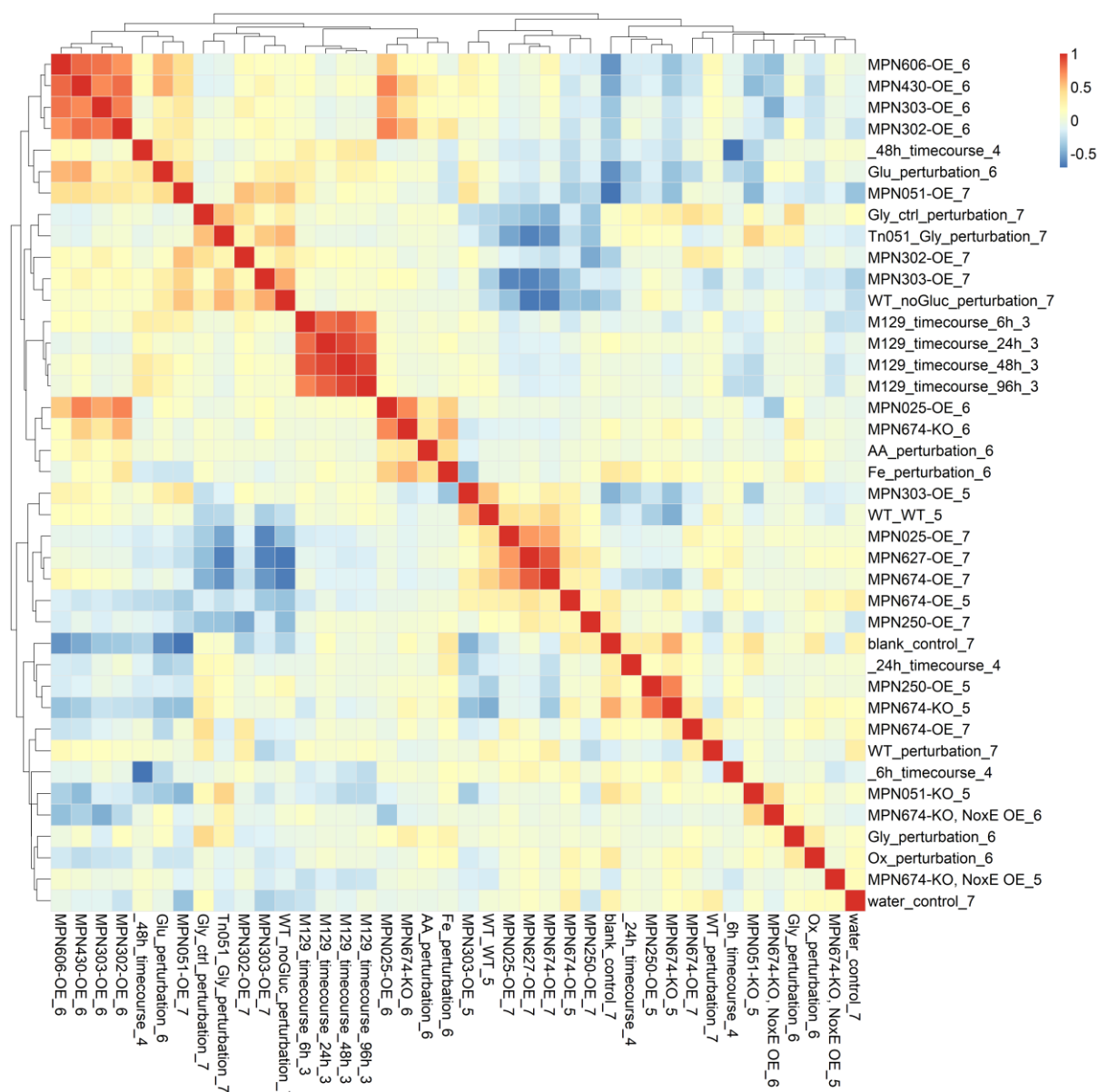
We explored the measurements of metabolite concentrations for the various samples shown in Table 1. Similar conditions as well as KO of genes in the same pathway cluster together.

An example of this is the clustering of all M129 samples which are the only samples grown in aerated conditions (Figure 1).

Two clusters exist of OE mutants targeting glycolysis. The first cluster contains an OE mutant of FBP, phosphotransferase MPN627 involved in mannitol and mannose uptake<sup>8</sup> and LDH. This clustering corroborates the assumed allosteric activation of LDH by FBP. The second cluster contains OE mutants of ENO, GAP, PYK and PFKA.

Another interesting cluster contains *MPN* perturbation, growth without oxygen and growth without amino acids which cluster together with both an LDH KO and FBA OE. These four samples have in common that the conditions are growth inhibiting. The clustering of FBA OE which degrades FBP together with LDH KO can be explained by the positive allosteric control of FBP on LDH.

Two annotation techniques were used to measure metabolite abundance,  $[M+F]^-$  ion and  $[M-H]^-$  ion detection<sup>48</sup>. Some differences are present in the correlations between individual metabolites in the  $[M+F]^-$  ion and  $[M-H]^-$  data, however, clustering of samples for both detection techniques is highly similar. A heatmap that compares both  $[M+F]^-$  ion and  $[M-H]^-$  data can be found in the Supplementary files 1B.



**Figure 1 Heatmap of mutant and perturbation sample clustering based on their metabolite profile**

In addition to clustering of samples based on their relative metabolite concentrations we studied the clustering of metabolites of these samples. We use the Pearson correlations between metabolites to build a network of metabolite-metabolite interactions (Figure 2). We identified a neatly defined cluster structure. The largest cluster contains sn-glycero-3-phosphocholine, CDP-Choline, folate and methionine cycle metabolites, orthophosphate, all nucleotide three phosphates (ATP, CTP, UTP) and pentose phosphate metabolism (PPP) metabolites. This cluster suggest a link between sn-glycero-3-phosphocholine catabolism, energy production, and nucleobase salvaging by phosphorylation and de-oxidation. Higher concentrations of PRRP associated with these metabolites are needed to convert nucleobases



guanine, GMP, dATP and dCTP. Pyruvate is positively correlated with dATP and dCTP which are needed for DNA synthesis. This cluster is strongly associated to the sub cluster of sn-Glycero-3-phosphocholine and tetrahydrofolate metabolites from the first cluster.

Based on the observed cluster, there are three main lessons to be learned. Firstly, clustering of sn-Glycero-3-phosphocholine with ATP, CTP and UTP metabolite correlation profiles supports the theory<sup>50</sup> that glycerol-3-phosphate derived from sn-Glycero-3-phosphocholine functions as a carbon and energy source for *MPN*. Addition of phosphatidylcholine to a defined minimal medium for *MPN* indeed optimizes growth<sup>22</sup>. Secondly, sn-Glycero-3-phosphocholine clusters together with folate and methionine cycle one carbon metabolites and as such is likely the main one carbon donor in *MPN* metabolism. Thirdly, positive correlation between CDP-choline, 3-phospho-D-glyceroyl-phosphate and sedoheptulose-7 phosphate, as well as between sn-Glycero-3-phosphocholine and the cluster containing PRPP indicate a link between sn-Glycero-3-phosphocholine and pentose phosphate metabolism.

## Over Expression of glycolytic enzymes

To further study the control of different glycolytic enzymes on central carbon metabolism we, analysed the fold change of enzymes in glycolysis and pyruvate metabolism when OE single as well as some combinations of glycolytic enzymes (Table 1)

**Table 1 Log2 Fold change expression values of OE mutants. I: Mutant PTA ACK did not show any OE of ACK. II: OE are significantly different from the wild type. III: OE values are significant, but the changes are noisy.**

		<i>PTA</i> <i>ACK</i> <sup>I</sup>		<i>PFK</i>	<i>PFK</i> <i>PTA</i>	<i>PFK</i> <i>LDH</i>	<i>LDH</i>
		MPN428; MPN533	MPN428	MPN302;	MPN302; MPN428	MPN302; MPN674	MPN674
		fold change	fold change	fold change	fold change	fold change	fold change
PTS	MPN207	0.07	0.18	-0.04	-0.18	-0.07	0.10
PFK	MPN302	-0.08	-0.10	2.73 <sup>II</sup>	2.66 <sup>II</sup>	2.71 <sup>II</sup>	-0.20
PGI	MPN250	-0.12	0.09	-0.16	-0.05	-0.04	0.04
FBA	MPN025	0.28	0.54	0.04	0.47	0.75	0.62
GAPD	MPN430	-0.22	-0.14	-0.27	-0.31	0.19	0.07
H							

PGK	MPN429	-0.01	0.04	0.07	0.20	0.05	-0.01
PGM	MPN628	0.06	0.21	-0.20	0.03	0.51	0.25
ENO	MPN606	0.24	0.30	0.13	0.32	0.41	0.35
PYK	MPN303	-0.20	-0.13	-0.11	0.12	0.01	-0.15
LDH	MPN674	0.37	0.54	0.01	0.34	1.50 <sup>II</sup>	1.69 <sup>II</sup>
lplA	MPN389	0.29	0.18	0.17	0.27	0.35	0.19
pdhD	MPN390	0.03	-0.04	0.08	0.19	-0.08	-0.11
pdhC	MPN391	0.24	0.12	0.10	0.23	0.15	0.15
pdhB	MPN392	0.34	0.19	0.27	0.29	0.37	0.36
pdhA	MPN393	0.07	-0.01	-0.11	0.06	-0.08	-0.09
nox	MPN394	0.71	0.50	0.12	0.43	0.80	0.52
pta	MPN428	2.11 <sup>II</sup>	2.22 <sup>II</sup>	0.28	1.93 <sup>II</sup>	0.26	0.33
ack	MPN533	0.36	0.34	0.28	0.43	0.40	0.30
GlpD	MPN051	0.41	0.44	0.19	0.27	0.60	0.49
atpC	MPN597	-0.16	0.21	-0.21	-0.11	0.27	0.31
atpD	MPN598	0.15	0.29	0.02	0.34	0.53	0.42
atpG	MPN599	0.20	0.30	0.20	0.11	0.71	0.32
atpA	MPN600	0.00	0.16	-0.01	0.09	0.16	0.27
atpH	MPN601	-0.15	0.18	-0.09	0.05	0.43	0.12
atpF	MPN602	-0.08	0.06	-0.30	-0.49	0.18	0.23
atpE	MPN603	-1.53 <sup>III</sup>	-0.52	-0.16	1.23 <sup>III</sup>	0.30	-1.76
atpB	MPN604	0.20	-0.15	-0.54	-0.60	-0.10	0.19
tkl	MPN082	0.02	-0.01	0.03	0.18	0.20	0.26
tim	MPN629	0.21	0.31	0.13	0.13	0.37	0.47

What we see is that when OE these other enzymes in the pathway don not widely change. These results suggests that allosteric regulation and circuit topology might play a great role on the control of central carbon metabolism of *MPN*.

## Exploration of metabolite's concentration at steady state: study of associations.

We build a dynamic model of central carbon metabolism including glycolysis and pyruvate metabolism. We trained this model with a limited subset of data and use the model to identify key regulatory elements in glycolysis.

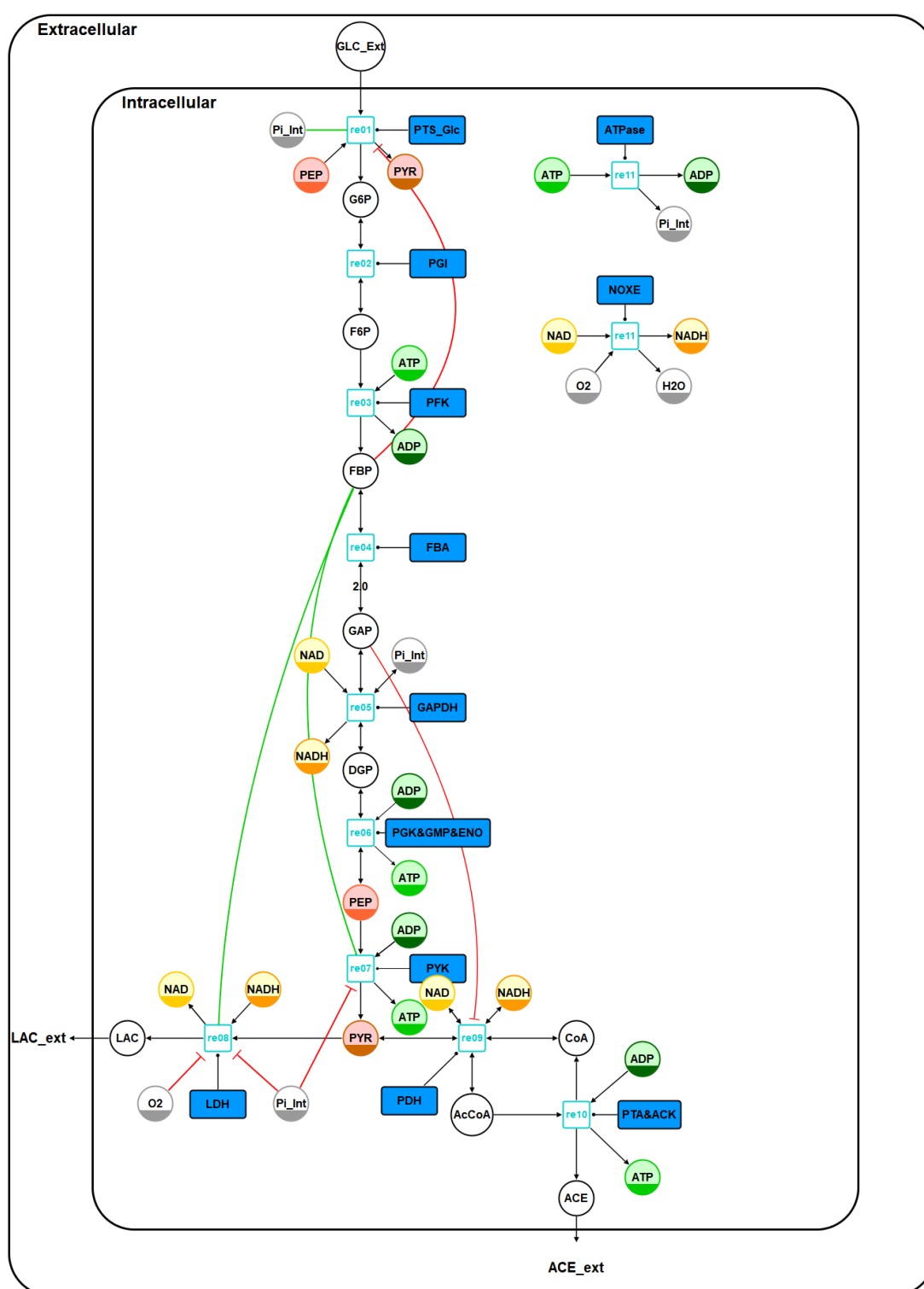
Different additions to this model were tested such as individual and combined additions of 1) an ATPase reaction to account for varying ATP demand, 2) LDH inhibition by oxygen and 3) a NoxE reaction for NAD regeneration using oxygen. The tested models include i) the base model ii) the base model and the ATPase reaction iii) base model with NoxE reaction iv) base model with both ATPase and NoxE reaction v) based model with NoxE reaction and LDH inhibition by oxygen and vi) the base model with all three modifications.

We found the model with addition of NoxE to have the best fitting in multiple parameter estimations, therefore we kept the addition of the NoxE to the model we used for further simulations/

In case intermediate metabolites were not measurable, reactions were lumped in a single reaction. This was the case for Phosphoglycerate kinase (PGK), Glycerate phosphomutase (GMP) and enolase (ENO). These three reactions were combined in reaction re07 lumping the enzymatic reactions of PGK&GMP&ENO. Similarly, phosphotransacetylase (PTA) and acetate kinase (ACK) were combined in reaction re10 which lumps the enzymatic reactions of PTA&ACK.

Allosteric control was assumed to be similar to allosteric control in *Lactococcus lactis* as presented in the model by Costa *et al*<sup>37</sup> due to the lack of *MPN* specific information on allosteric control. Allosteric control includes three activator and five inhibitor effects. Reactions were modelled using modular rate laws except for transport reactions for which Hill type kinetics were used. Enzyme concentrations were included as reaction parameter to allow model predictions at varying protein concentrations. The base model contains 10 equations and 72 parameters of which 10 represent experimentally determined enzyme concentrations, 5 represent equilibrium constants ( $K_{eq}$ ) and 1 is a Hill coefficient. The remaining 56 parameters represent Michaelis-Menten constants, activation constants and inhibition constants which are not known for *MPN*. The model was built using COPASI<sup>38</sup>.

An overview of the model's reactions and equations and the additions tested can be found in in Supplementary material 1A.



**Figure 3: Schema of the model.** The diagram meets the Systems Biology Graphical Notation (SBGN) standard<sup>53</sup> with the exception of the LAC and ACE export arrows which in the model are present as syncs for LAC and ACE: Arrowheads represent reactions (black arrowhead end), catalysis (open circle end), activation (green) and inhibition (red). Circles indicate metabolites. Half-filled circles are clone makers to indicate the metabolites appears multiple



times in the diagram (green=adenine nucleotides, yellow=redox equivalents, red=PEP/PYR). Blue filled rectangles describe macromolecules (enzymes, transporters) that catalyse a reaction. In case the reaction stoichiometry is different to 1, the reaction stoichiometry is given as text at the reaction arrow. Blue empty rectangles indicate reaction identifiers in the model. Metabolite abbreviations: AcCoA = acetyl coenzyme A, ACE = acetate, ADP = adenosine diphosphate, ATP = adenosine triphosphate, CoA = coenzyme A, DGP = diacylglycerol phosphate, F6P = fructose 6-phosphate, FBP = fructose-1,6-bisphosphatase, G6P = glucose 6-phosphate, GAP = glyceraldehyde-3-phosphate, LAC = lactate, NAD = nicotinamide adenine dinucleotide, NADH = reduced nicotinamide adenine dinucleotide, PEP = phosphoenolpyruvate, Pi\_Int = orthophosphate, PYR = pyruvate (PYR), GLC\_Ext = external glucose.

We compared the mean square error, for all parameter sets (Figure 4). Models that include ATPase (models 2, 4 and 6) have on average the largest error and took the most iterations to reach a stable solution. The addition of NoxE was shown to reduce the error in model predictions (see Figure 4). A zoomed in version of Figure 4 as well as correlation analysis of parameter sets is available in Supplementary file 1B. Since the model with addition of NoxE performed best, further analyses were continued using this model. This model loaded with the best performing parameter set was deposited in BioModels<sup>54</sup> and assigned the identifier MODEL1911200003.

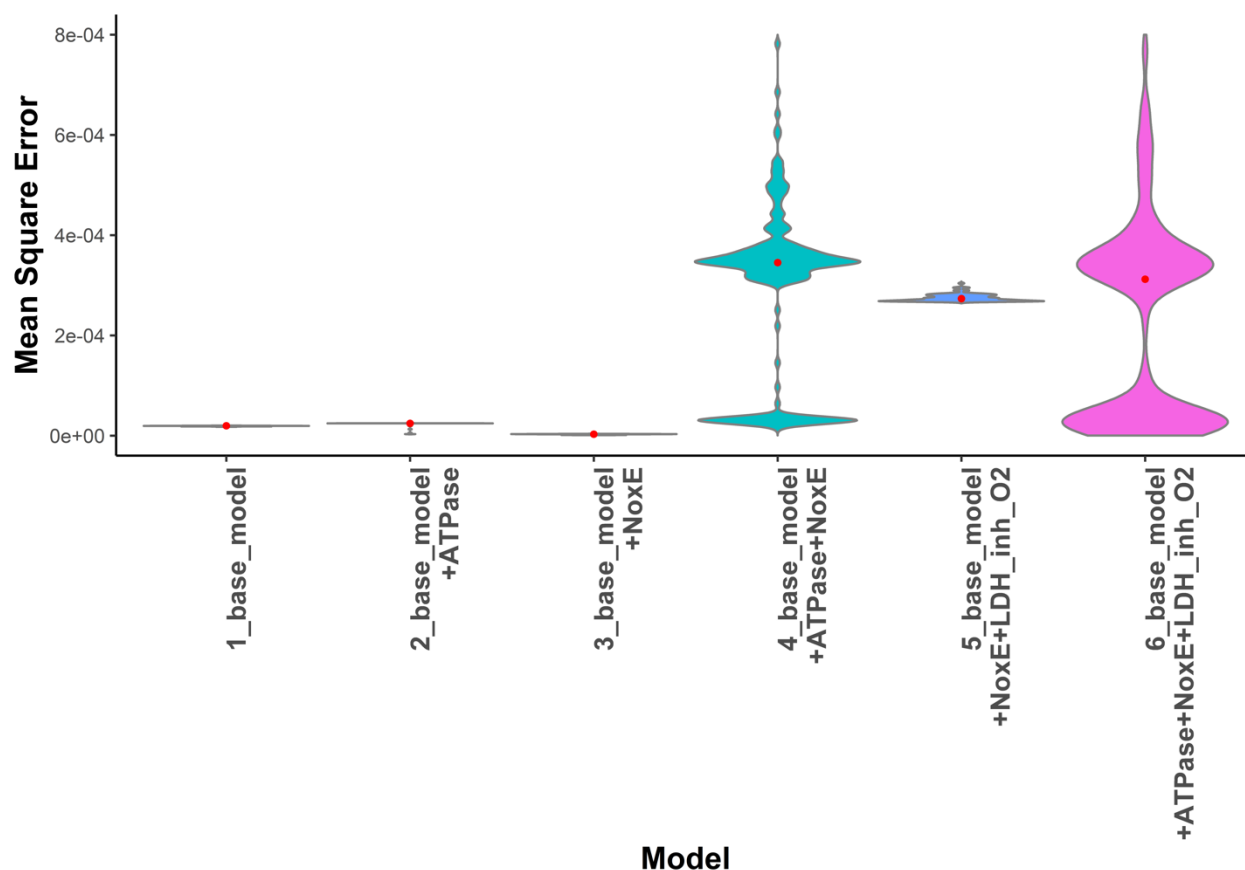
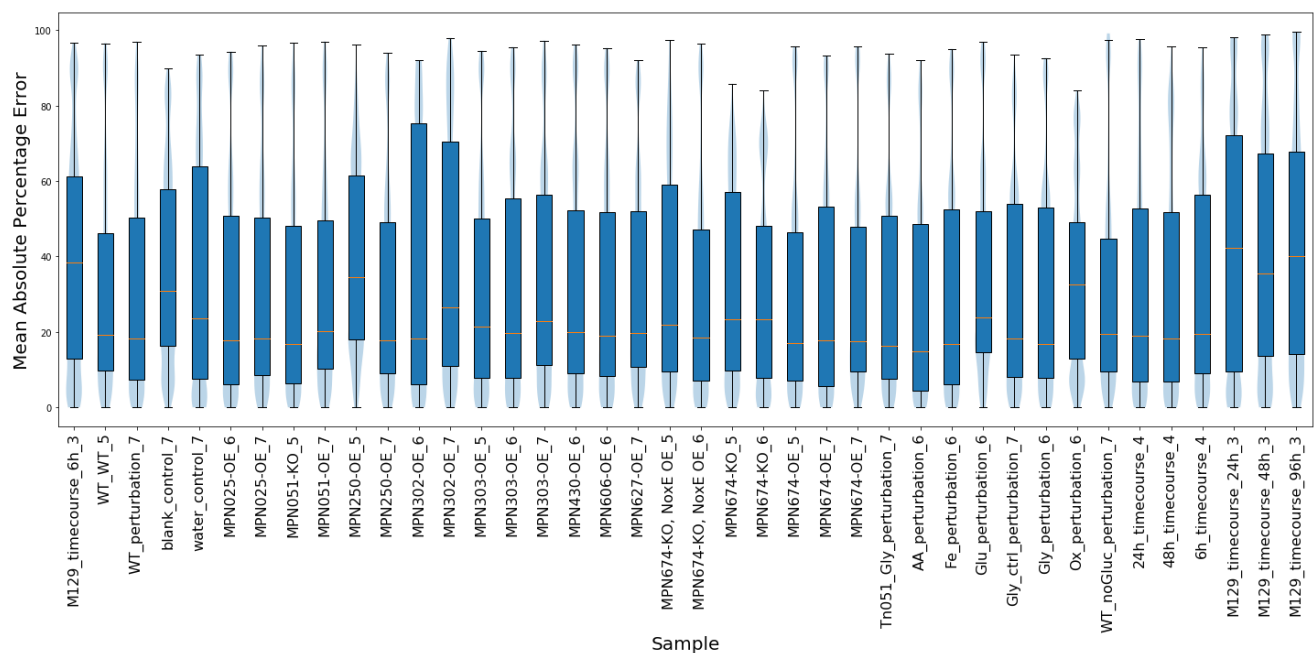


Figure 4 Violin plot comparing the mean square error, for all 100 parameter sets for six models. The red dot indicates the median value.

### Simulating perturbations, KO and OE mutants

We used the trained model to predict steady state metabolite concentrations for 40 independent samples. Sample's metabolite concentration mean and standard deviation values were determined by measuring samples in triplicate. In addition to being measured in triplicates, biological replicates were available for all OE targeting glycolysis and pyruvate metabolism.

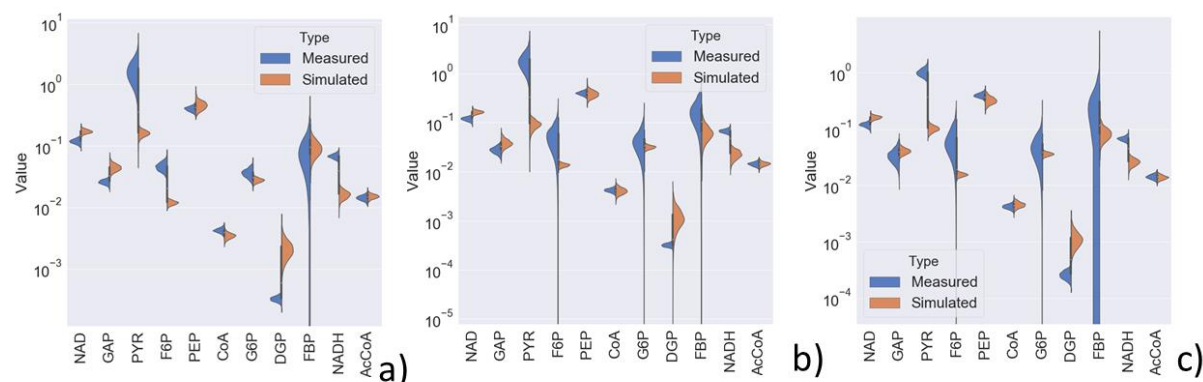
For each of the 40 samples, 1000 independent simulations were performed with slightly different initial concentrations to explore the impact of biological variability and uncertainty associated to measurement error rates. On average the model predicted these samples with reasonable accuracy (Figure 4).



**Figure 5 Symmetric Mean Absolute Percentage Error between simulated and measured values using a 1000x times sampling. sMAPE for all metabolites per sample are combined.**

The largest error in predicted metabolite concentrations occur for the samples of MPN129 growth in aerated conditions. These results were to be expected as these samples are clearly forming an outlier since the growth conditions are so vastly different from the other samples. From samples corresponding to perturbations in growth condition, the oxidative stress perturbation (0.15%  $\text{H}_2\text{O}_2$ ) had the largest prediction error.

The relatively simple model here presented reproduces the states attained under a broad range of perturbations such as glucose concentrations up to a factor 10 lower as compared to the training data. The use of proteomics data is most likely one of the main reasons the model simulates OE and KO mutants of enzymes in glycolysis relatively well. The calculated metabolite concentrations are, on average within a factor 3 of measured values for all mutants and perturbation samples, which is comparable to the accuracy of the training set. An example of measured and simulated values of sampled simulations can be seen in Figure 6.



**Figure 6 Measured and simulated steady state concentrations at a) 6h of growth, b) 24h of growth and c) 48h of growth on low glucose concentration. Measurement for DGP, NAD and NADH are reference values from samples grown on high glucose concentration.**

Additionally, concentrations of for most metabolites are predicted within the 95% confidence interval of the *in-vitro* data (Supplementary material 1C). Exceptions to this rule are PYR and ATP and NADH. Simulated concentrations for ATP and NADH are systematically lower than measured values while PYR is systematically predicted to have a higher than measured concentration. Pyruvate is hard to quantify since it easily is degraded causing variation in its measurement due to for example different times between measurement and sampling. As such we cannot quantify the accuracy of the predictions for pyruvate. For the metabolites such as DGP, NAD and NADH, reference values from high glucose conditions were used to set initial concentrations. As such deviation from these reference values in the simulations in steady state was to be expected.

## Metabolic control

We used two types of metabolic control analysis to understand which enzymes and metabolites effort the greatest control on glycolysis: Local Sensitivity analysis and Global Sensitivity analysis. Control coefficients are unit less measures of the relative steady state change in a system variable, in our case the flux through PFK, in response to a relative change in a parameters value. Global sensitivity provides information on parameters that exert control independent of a specific parameter value. We achieved this by sampling parameter sets uniformly from the parameter search space. These parameters sets are not specific for *MPN* since they are not fitted to any *MPN* data. Local sensitivity analysis on the other hand is based on control coefficients derived at the steady state using parameter sets

fitted to *MPN* specific data. As such the approaches are fundamentally different and complementary in the information they provide. For our local sensitivity analysis, we use the best performing 10 parameter sets to calculate metabolic control coefficients (**Error! Reference source not found.**). Since these parameter sets are independent of one another, if metabolic control is higher for certain parameters based on multiple parameter sets, we can conclude the control to be relevant since it is a result of the fitting to *MPN* data. Additionally, we also performed local sensitivity analysis while sampling from the measurement distribution of metabolites to investigate the effect of measurement error on control coefficients. The steady state changes for each sampled simulation, as such we can see how the uncertainty in metabolite concentrations propagates and creates uncertainty in the control coefficients calculated at these steady states. An overview of our metabolic control analyses can be found in Supplementary file 1D. We found two main control hubs PTS\_Glc + PFK and LDH+PDH+PYK. The first control hub consists of parameters associated to PTS\_Glc and PFK and represents metabolism in the upper part of glycolysis, the second hub consists of parameter associated to LDH, PDH and PYK and part of pyruvate metabolism.

### Simulating combined OE and KO mutants

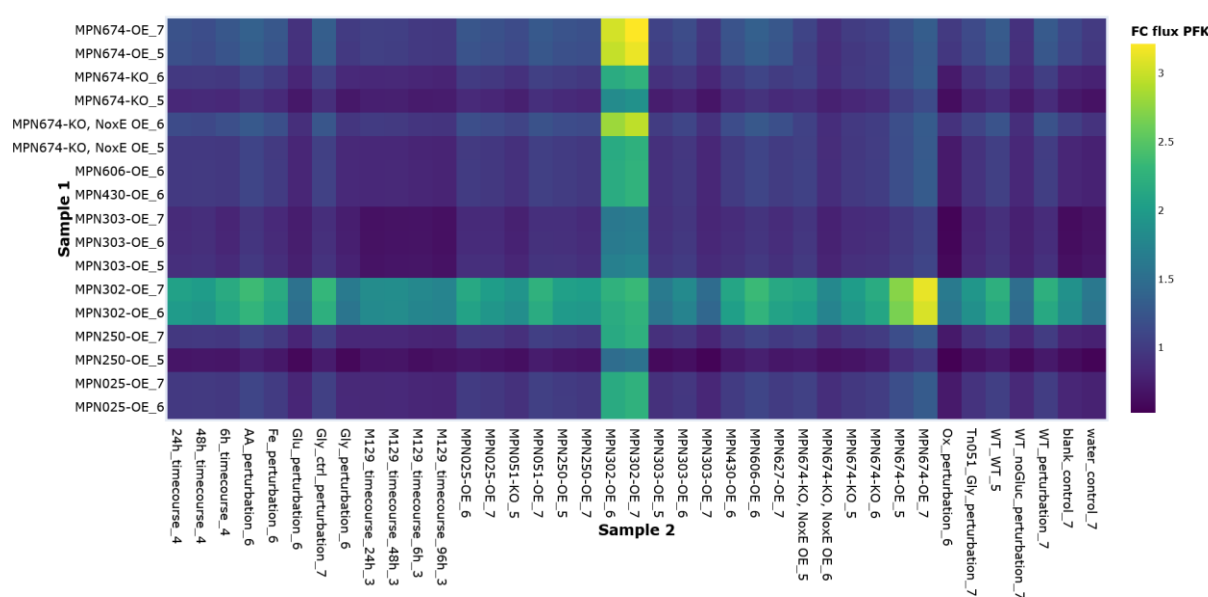
We used the model to simulate the combined effect of genetic perturbations targeting glycolysis enzymes (OE, KO) combined with a second perturbation, either genetic or environmental. This analysis can identify bottlenecks in central carbon metabolism consisting of combinations of enzymes. Such bottlenecks cannot be identified through local sensitivity analysis or when simulating single over expressions. The expected variations in the flux through glycolysis is shown in Figure 7. For most of these combined perturbations, only minor changes were observed. However, simulations of OE of PFK show greatly increased flux through glycolysis while oxygen stress, iron limitation and growth of MPN129 in aerated conditions lead to greatly reduced flux.

However, combination of PFK OE with OE of lactate dehydrogenase (LDH) or phosphotransacetylase (PTA) with acetate kinase (ACK), increase the flux nearly as much and is realistic to obtain *in vivo* since OE mutants for each of these enzymes individually are available.

Based on these simulations, combined OE mutants were suggested for lab validation: PFK+LDH, PFK+NOXE. Growth curves, protein and metabolite concentrations were measured for the PFK+LDH OE mutant but not for combined PFK+NOXE OE. Combined PFK+PDH OE did not increase flux through glycolysis. However, some positive epistasis on

the growth for the combined OE of PFK+LDH was observed, with higher biomass and higher acidification than the individual mutants. None of the combined OE mutants increased the growth rate of *MPN* with respect to the wild type as could be expected since energy metabolism is not growth limiting in *MPN*<sup>9</sup>.

The simulation results of the double OE mutants show that glycolysis is robust. Meaning that the network structure that includes feed forward and feed backward allosteric control, makes the glycolysis of *MPN* inherently robust.



**Figure 7** Fold change flux through PFK for combinatorial mutants and perturbations in steady state, relative to the wild type at 24h. The primary sample, of which only the protein OE values are used, are shown on the y-axis while the secondary sample, of which both changes to parameters and metabolite concentrations are used, are shown on the x-axis. The colour represents the fold change in flux through glycolysis compared to the wild type.

## Discussion:

Metabolomics data for a large number of perturbations, OE and KO mutants were collected. We observe that metabolite concentration of samples in general do not vary that much and that samples of similar conditions as well samples of enzymes OE and KO mutants of enzymes close to other in the metabolic network cluster together. In glycolysis we observed no clear clustering between metabolites. We also see that there is relatively little difference in the expression of glycolytic enzymes when OE enzymes in glycolysis and pyruvate metabolism. These results suggest that central carbon metabolism of *MPN* is robust against

perturbation. Literature research revealed that glycolysis and pyruvate metabolism in many species is observed to be robust against perturbations<sup>16,55,56</sup>. Arguably robustness is an even more important property for a minimal organism such as *MPN* where biological noise can be expected to have much larger effect than for many organisms with a larger cell volume. The property of robustness agrees with model results that show only minor changes occur even when OE multiples glycolysis enzymes in silico. We tested the effect of adding a reaction for additions of 1) ATPase, 2) O<sub>2</sub> inhibition by LDH and 3) a NoxE reaction. Only the addition of a NoxE reaction resulted in a much better fit to the training data. Arguably, addition of NoxE improves parameter identifiability since the models without this reaction use a fixed reference values for NAD and NADH. The models that include NoxE allow NAD and NADH concentrations to change, this added flexibility and improves parameter identifiability since NAD and NADH associated parameter can now be used to account for differences in metabolite concentration opposed to being fixed values based on reference values. Additionally, NAD/NADH are known to have control over central carbon metabolism in other organisms such as *L. lactis*<sup>57,58</sup> therefore we argue it is likely they also have control over central carbon metabolism in *MPN*.

Steady state simulations showed the model to be flexible since it can predict metabolite concentrations well for all mutant samples. Part of this flexibility is the result of including parameters representing the enzyme concentration, therefore genetic perturbations are accounted for in simulations. Similarly, by using measurements of cofactors from these conditions, the model can approximate the effects of these simulations on central carbon metabolism. Additionally, we argue the flexibility of the model might partly be the result of the inherent robustness of central carbon metabolism in *MPN*.

## Conclusion

In this study we integrated experimental data and model simulations and analysis to explore the robustness of central carbon metabolism of *MPN* in steady state.

Firstly, we analyzed metabolomics data. We observed that samples from similar conditions as well as samples of OE and KO mutants of enzymes close to each other in the metabolic network cluster together. Samples from vastly varying conditions, such as aerated conditions, do not cluster with the other samples. Secondly, we build a model to simulate samples of various single or combined perturbations. The simple model presented in this study can predict metabolite concentration with reasonable accuracy for a wide range of conditions and

OE and KO mutants.

Two control hubs were identified using our dynamic model a) upper glycolysis (PTS\_Glc + PFK) and b) lower pyruvate metabolism (LDH+PDH+PYK). No single or combined OE mutant of glycolysis and pyruvate metabolism enzymes resulted in a higher growth rate although OE of PFK and LDH resulted in somewhat higher acidification indicating there might be higher flux through glycolysis. These results are in agreement with studies that show that glycolysis and pyruvate metabolism in *MPN* is not growth limiting. Both the results from the analysis of our samples as well as the model results, suggest robustness to be a central property of *MPN* glycolysis and pyruvate metabolism.

## Acknowledgement

We acknowledge support of the Spanish Ministry of Economy, Industry and Competitiveness (MEIC) to the EMBL partnership, the Spanish Ministry of Economy and Competitiveness, ‘Centro de Excelencia Severo Ochoa 2013-2017’, the European Research Council (ERC) under the European Union’s Horizon 2020 research and innovation program under agreement No 670216 (MYCOCHASSIS), the European Union’s Horizon 2020 research and innovation programme under grant agreement No 634942 (MycoSynVac), the CERCA Programme / Generalitat de Catalunya, FEDER project from Instituto Carlos III (ISCIII, Acción Estratégica en Salud 2016) (reference CP16/00094) and “Secretaria d’Universitats i Recerca del Departament d’Economia i Coneixement de la Generalitat de Catalunya” (2014SGR678). We would like to thank dr.ir. A. Rinzema for his expert advice on calculating the various oxygen diffusion rates in our experimental setup.

## Contributions

LS, VMdS and MS-D conceived the study. NZ developed the model and performed the simulations with guidance and input from PS, VMdS, LS and MS-D. EY, DS, SM and CG performed the experiments with guidance and input from ML-S and LS. NZ wrote the manuscript with input from VMdS, PS, ML-S, LS and MS-D.

## Competing interests

The authors declare no Competing Financial or Non-Financial Interests.



## Corresponding authors

Correspondence to Luis Serrano or Maria Suarez Diez.

## References

1. Timenetsky, J., Santos, L. M., Buzinhani, M. & Mettifogo, E. Detection of multiple mycoplasma infection in cell cultures by PCR. *Brazilian J. Med. Biol. Res.* **39**, 907–914 (2006).
2. Mycoplasma pneumoniae Surveillance and Reporting | CDC. <https://www.cdc.gov/pneumonia/atypical/mycoplasma/surv-reporting.html> (2020).
3. Khan, F. Y. & A.yassin, M. Mycoplasma pneumoniae associated with severe autoimmune hemolytic anemia: case report and literature review. *Brazilian J. Infect. Dis.* **13**, 77–79 (2009).
4. Güell, M. *et al.* Transcriptome Complexity in a Genome-Reduced Bacterium. *Science* (80-. ). **326**, 1268–1271 (2009).
5. Maier, T. *et al.* Quantification of mRNA and protein and integration with protein turnover in a bacterium. *Mol. Syst. Biol.* **7**, 511 (2011).
6. Maier, T. *et al.* Large-scale metabolome analysis and quantitative integration with genomics and proteomics data in Mycoplasma pneumoniae. *Mol. Biosyst.* **9**, 1743 (2013).
7. Junier, I., Unal, E. B., Yus, E., Lloréns-Rico, V. & Serrano, L. Insights into the Mechanisms of Basal Coordination of Transcription Using a Genome-Reduced Bacterium. *Cell Syst.* **2**, 391–401 (2016).
8. Wodke, J. a H. *et al.* Dissecting the energy metabolism in Mycoplasma pneumoniae through genome-scale metabolic modeling. *Mol. Syst. Biol.* **9**, 653 (2013).
9. Yus, E. *et al.* Impact of Genome Reduction on Bacterial Metabolism and Its Regulation. *Science* (80-. ). **326**, 1263–1268 (2009).
10. Yus, E. *et al.* Determination of the Gene Regulatory Network of a Genome-Reduced Bacterium Highlights Alternative Regulation Independent of Transcription Factors. *Cell Syst.* **9**, 143–158.e13 (2019).
11. Glass, J. I. *et al.* Essential genes of a minimal bacterium. *Proc. Natl. Acad. Sci.* **103**, 425–430 (2006).
12. Kamminga, T. *et al.* Persistence of Functional Protein Domains in Mycoplasma Species and their Role in Host Specificity and Synthetic Minimal Life. *Front. Cell. Infect. Microbiol.* **7**, 31 (2017).
13. Kurata, H. & Sugimoto, Y. Improved kinetic model of Escherichia coli central carbon metabolism in batch and continuous cultures. *J. Biosci. Bioeng.* **125**, 251–257 (2018).
14. Machado, D., Herrgård, M. J. & Rocha, I. Modeling the Contribution of Allosteric Regulation for Flux Control in the Central Carbon Metabolism of E. coli. *Front. Bioeng. Biotechnol.* **3**, 1–11 (2015).
15. Oddone, G. M., Mills, D. a. & Block, D. E. A dynamic, genome-scale flux model of Lactococcus lactis to increase specific recombinant protein expression. *Metab. Eng.* **11**, 367–381 (2009).
16. Millard, P., Smallbone, K. & Mendes, P. Metabolic regulation is sufficient for global and robust coordination of glucose uptake, catabolism, energy production and growth in Escherichia coli. *PLOS Comput. Biol.* **13**, e1005396 (2017).
17. Smallbone, K. *et al.* A model of yeast glycolysis based on a consistent kinetic characterisation of all its enzymes. *FEBS Lett.* **587**, 2832–2841 (2013).
18. Murabito, E. *et al.* Monte-Carlo Modeling of the Central Carbon Metabolism of Lactococcus lactis: Insights into Metabolic Regulation. *PLoS One* **9**, e106453 (2014).

19. Kamminga, T. *et al.* Metabolic modeling of energy balances in *Mycoplasma hyopneumoniae* shows that pyruvate addition increases growth rate. *Biotechnol. Bioeng.* **114**, 2339–2347 (2017).
20. Kesten, D., Kummer, U., Sahle, S. & Hübner, K. A new model for the aerobic metabolism of yeast allows the detailed analysis of the metabolic regulation during glucose pulse. *Biophys. Chem.* **206**, 40–57 (2015).
21. Kochanowski, K. *et al.* Few regulatory metabolites coordinate expression of central metabolic genes in *Escherichia coli*. *Mol. Syst. Biol.* **13**, 903 (2017).
22. Gaspari, E. *et al.* Model-driven design allows growth of *Mycoplasma pneumoniae* on serum-free media. *npj Syst. Biol. Appl.* **6**, 33 (2020).
23. Halbedel, S. & Stülke, J. Probing In Vivo Promoter Activities in *Mycoplasma pneumoniae* : a System for Generation of Single-Copy Reporter Constructs. *Appl. Environ. Microbiol.* **72**, 1696–1699 (2006).
24. Halbedel, S., Busse, J., Schmidl, S. R. & Stülke, J. Regulatory Protein Phosphorylation in *Mycoplasma pneumoniae*. *J. Biol. Chem.* **281**, 26253–26259 (2006).
25. Fuhrer, T., Heer, D., Begemann, B. & Zamboni, N. High-Throughput, Accurate Mass Metabolome Profiling of Cellular Extracts by Flow Injection–Time-of-Flight Mass Spectrometry. *Anal. Chem.* **83**, 7074–7080 (2011).
26. Wodke, J. a. H. *et al.* MyMpn: a database for the systems biology model organism *Mycoplasma pneumoniae*. *Nucleic Acids Res.* **43**, D618–D623 (2015).
27. Storey, J. D. & Tibshirani, R. Statistical significance for genomewide studies. *Proc. Natl. Acad. Sci.* **100**, 9440–9445 (2003).
28. Buescher, J. M., Moco, S., Sauer, U. & Zamboni, N. Ultrahigh Performance Liquid Chromatography–Tandem Mass Spectrometry Method for Fast and Robust Quantification of Anionic and Aromatic Metabolites. *Anal. Chem.* **82**, 4403–4412 (2010).
29. Wolstencroft, K. *et al.* The SEEK. in *Methods in Enzymology* vol. 500 629–655 (2011).
30. Wilkinson, M. D. *et al.* The FAIR Guiding Principles for scientific data management and stewardship. *Sci. Data* **3**, 160018 (2016).
31. Lubitz, T. *et al.* SBTAB: a flexible table format for data exchange in systems biology. *Bioinformatics* **32**, 2559–2561 (2016).
32. Juty, N., Le Novère, N. & Laibe, C. Identifiers.org and MIRIAM Registry: community resources to provide persistent identification. *Nucleic Acids Res.* **40**, D580–D586 (2012).
33. Courtot, M. *et al.* Controlled vocabularies and semantics in systems biology. *Mol. Syst. Biol.* **7**, 543 (2011).
34. Jauhiainen, A. *et al.* Normalization of metabolomics data with applications to correlation maps. *Bioinformatics* **30**, 2155–2161 (2014).
35. R Core Team. R: A Language and Environment for Statistical Computing. (2017).
36. Wu, Y. & Li, L. Sample normalization methods in quantitative metabolomics. *J. Chromatogr. A* **1430**, 80–95 (2016).
37. Costa, R. S., Hartmann, A., Gaspar, P., Neves, A. R. & Vinga, S. An extended dynamic model of *Lactococcus lactis* metabolism for mannitol and 2,3-butanediol production. *Mol. Biosyst.* **10**, 628 (2014).
38. Hoops, S. *et al.* COPASI--a COMplex PATHway SIMulator. *Bioinformatics* **22**, 3067–3074 (2006).
39. Flamholz, A., Noor, E., Bar-Even, A. & Milo, R. eQuilibrator--the biochemical thermodynamics calculator. *Nucleic Acids Res.* **40**, D770–D775 (2012).
40. Chen, C., Twycross, J. & Garibaldi, J. M. A new accuracy measure based on bounded relative error for time series forecasting. *PLoS One* **12**, e0174202 (2017).

41. Iman, R. L., Helton, J. C. & Campbell, J. E. An Approach to Sensitivity Analysis of Computer Models: Part I—Introduction, Input Variable Selection and Preliminary Variable Assessment. *J. Qual. Technol.* **13**, 174–183 (1981).
42. Iman, R. L., Helton, J. C. & Campbell, J. E. An Approach to Sensitivity Analysis of Computer Models: Part II—Ranking of Input Variables, Response Surface Validation, Distribution Effect and Technique Synopsis. *J. Qual. Technol.* **13**, 232–240 (1981).
43. Choi, K., Medley, J. K., Cannistra, C. & Matthias, K. Tellurium : A Python Based Modeling and Reproducibility Platform for Systems Biology. *bioRxiv* (2016) doi:<https://doi.org/10.1101/054601>.
44. Somogyi, E. T. *et al.* libRoadRunner: a high performance SBML simulation and analysis library: Table 1. *Bioinformatics* **31**, 3315–3321 (2015).
45. Merk TTP T300 tissue culture flask specifications. *TPP® tissue culture flasks / Sigma-Aldrich* [www.sigmaaldrich.com/catalog/product/sigma/z707562?lang=en&region=NL](http://www.sigmaaldrich.com/catalog/product/sigma/z707562?lang=en&region=NL).
46. Wilke, C. R. & Chang, P. Correlation of diffusion coefficients in dilute solutions. *AIChE J.* **1**, 264–270 (1955).
47. Tyrrell, H. J. V. The origin and present status of Fick’s diffusion law. *J. Chem. Educ.* **41**, 397 (1964).
48. Tiernan, T., Chang, C. & Cheng, C. C. Formation and Reactions of Negative Ions Relevant to Chemical Ionization Mass Spectrometry . I . Cl Mass Spectra of Organic Compounds Produced by F- Reactions. *Environ. Health Perspect.* **36**, 47–62 (1980).
49. Yang, M. & Vousden, K. H. Serine and one-carbon metabolism in cancer. *Nat. Rev. Cancer* **16**, 650–662 (2016).
50. Schmidl, S. R. *et al.* A Trigger Enzyme in Mycoplasma pneumoniae: Impact of the Glycerophosphodiesterase GlpQ on Virulence and Gene Expression. *PLoS Pathog.* **7**, e1002263 (2011).
51. Joseph, J. & Loscalzo, J. Methoxistasis: Integrating the Roles of Homocysteine and Folic Acid in Cardiovascular Pathobiology. *Nutrients* **5**, 3235–3256 (2013).
52. Großhennig, S., Schmidl, S. R., Schmeisky, G., Busse, J. & Stülke, J. Implication of Glycerol and Phospholipid Transporters in Mycoplasma pneumoniae Growth and Virulence. *Infect. Immun.* **81**, 896–904 (2013).
53. Novère, N. Le *et al.* The Systems Biology Graphical Notation. *Nat. Biotechnol.* **27**, 735–741 (2009).
54. Chelliah, V. *et al.* BioModels: Ten-year anniversary. *Nucleic Acids Res.* **43**, D542–D548 (2015).
55. Ghorbaniaghdam, A., Henry, O. & Jolicoeur, M. An in-silico study of the regulation of CHO cells glycolysis. *J. Theor. Biol.* **357**, 112–122 (2014).
56. Gehrmann, E. *et al.* Robustness of glycolysis in yeast to internal and external noise. *Phys. Rev. E* **84**, 021913 (2011).
57. Neves, A. R. *et al.* Is the Glycolytic Flux in Lactococcus lactis Primarily Controlled by the Redox Charge? *J. Biol. Chem.* **277**, 28088–28098 (2002).
58. NEVES, A., POOL, W., KOK, J., KUIPERS, O. & SANTOS, H. Overview on sugar metabolism and its control in – The input from in vivo NMR. *FEMS Microbiol. Rev.* **29**, 531–554 (2005).