

## Accepted Manuscript

Title: Genome-scale reconstruction of the *Streptococcus pyogenes* M49 metabolic network reveals growth requirements and indicates potential drug targets

Author: Jennifer Levering Tomas Fiedler Antje Sieg Koen W.A. van Grinsven Silvio Hering Nadine Veith Brett G. Olivier Lara Klett Jeroen Hugenholtz Bas Teusink Bernd Kreikemeyer Ursula Kummer



PII: S0168-1656(16)30007-4  
DOI: <http://dx.doi.org/doi:10.1016/j.jbiotec.2016.01.035>  
Reference: BIOTEC 7402

To appear in: *Journal of Biotechnology*

Received date: 26-6-2015  
Revised date: 3-1-2016  
Accepted date: 12-1-2016

Please cite this article as: Levering, Jennifer, Fiedler, Tomas, Sieg, Antje, van Grinsven, Koen W.A., Hering, Silvio, Veith, Nadine, Olivier, Brett G., Klett, Lara, Hugenholtz, Jeroen, Teusink, Bas, Kreikemeyer, Bernd, Kummer, Ursula, Genome-scale reconstruction of the *Streptococcus pyogenes* M49 metabolic network reveals growth requirements and indicates potential drug targets. *Journal of Biotechnology* <http://dx.doi.org/10.1016/j.jbiotec.2016.01.035>

This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

**Genome-scale reconstruction of the *Streptococcus pyogenes* M49 metabolic network reveals growth requirements and indicates potential drug targets**

Jennifer Levering<sup>a,\*#</sup>, Tomas Fiedler<sup>b#</sup>, Antje Sieg<sup>b</sup>, Koen W. A. van Grinsven<sup>c</sup>, Silvio Hering<sup>b</sup>, Nadine Veith<sup>a</sup>, Brett G. Olivier<sup>d</sup>, Lara Klett<sup>a</sup>, Jeroen Hugenholtz<sup>c</sup>, Bas Teusink<sup>d</sup>, Bernd Kreikemeyer<sup>b</sup>, Ursula Kummer<sup>a</sup>

Department of Modeling of Biological Processes, COS Heidelberg/BIOQUANT, Heidelberg University, Heidelberg, Germany<sup>a</sup>; Institute of Medical Microbiology, Virology and Hygiene, Rostock University Medical Centre, Rostock, Germany<sup>b</sup>; Laboratory for Microbiology, Swammerdam Institute for Life Sciences, Amsterdam, The Netherlands<sup>c</sup>; Amsterdam Institute for Molecules, Medicines and Systems, VU Amsterdam, The Netherlands<sup>d</sup>

#Address correspondence to Tomas Fiedler, [tomas.fiedler@med.uni-rostock.de](mailto:tomas.fiedler@med.uni-rostock.de) or

Jennifer Levering, [jlevering@ucsd.edu](mailto:jlevering@ucsd.edu).

\*

Current address: Department of Bioengineering, University of California, San Diego, 9500 Gilman Drive MC0412, La Jolla, 92093-0412, CA, USA

J.L. and T.F. contributed equally to this work.

## Highlights

- We present the first genome-scale model of the metabolism of *Streptococcus pyogenes*
- We experimentally tested the model using diverse data sets, e.g. auxotrophy experiments
- The model allows to understand and predict growth requirements of *S. pyogenes* in detail.
- The mechanism of pH adaptation in this organism is studied in detail.

## Abstract

Genome-scale metabolic models comprise stoichiometric relations between metabolites, as well as associations between genes and metabolic reactions and facilitate the analysis of metabolism. We computationally reconstructed the metabolic network of the lactic acid bacterium *Streptococcus pyogenes* M49. Initially, we based the reconstruction on genome annotations and already existing and curated metabolic networks of *Bacillus subtilis*, *Escherichia coli*, *Lactobacillus plantarum* and *Lactococcus lactis*. This initial draft was manually curated with the final reconstruction accounting for 480 genes associated with 576 reactions and 558 metabolites. In order to constrain the model further, we performed growth experiments of wild type and *arcA* deletion strains of *Streptococcus pyogenes* M49 in a chemically defined medium and calculated nutrient uptake and production fluxes. We additionally performed amino acid auxotrophy experiments to test the consistency of the model. The established genome-scale model can be used to understand the growth requirements of the human pathogen *S. pyogenes* and define optimal and suboptimal conditions, but also to describe differences and similarities between *S. pyogenes* and related lactic acid bacteria such as *L. lactis* in order to find strategies to reduce the growth of the pathogen and propose drug targets.

**Keywords:** Genome-scale metabolic model; metabolism; *Streptococcus pyogenes*; amino acid auxotrophies; lactic acid bacteria

## 1. Introduction

*Streptococcus pyogenes* (group A streptococcus [GAS]) belongs to the group of lactic acid bacteria (LAB) which are characterised by their capability to ferment glucose to lactic acid. LAB colonize multiple biotopes, including foods, plants or the human body. There is a great biodiversity amongst lactic acid-producing bacteria with respect to their genetics and consequent biochemical details, reflected in differences in acidification and pathogenicity. Some LAB play essential roles in food and beverage industry (e.g. for fermentation), while others, such as GAS possess pathogenic features (Levering et al., 2012).

*S. pyogenes* is one of the most widespread human pathogens and colonises the skin, tonsils, mucous membrane and deeper tissues. It causes many different infections such as pharyngitis, scarlet fever, impetigo, necrotizing fasciitis and streptococcal toxic shock syndrome. *S. pyogenes* can also cause immune-mediated post-infectious sequelae like acute rheumatic fever and acute glomerulonephritis (Walker et al., 2014).

In general, lactic acid bacteria have evolved in environments that are rich in amino acids, vitamins, purines and pyrimidines. As a consequence, they have complex nutritional requirements. Species differ in their ability to ferment individual carbohydrates and in their preferred carbon source (Gunnewijk et al., 2001). The most commonly used sugar for their cultivation is glucose. Growth of LAB requires supply of vitamins and related growth factors like p-aminobenzoic acid, biotin, riboflavin, thiamine, vitamin B<sub>6</sub> and vitamin B<sub>12</sub>, whereby the amount of required growth factors differ among the organisms. The amount and combination of amino

acids required for growth is characteristic for each LAB and depends upon the medium composition, e.g. again on the supplied vitamins. However, even if all vitamins are supplied in excess, lactic acid bacteria still require many amino acids for growth. Additionally, purine and pyrimidine bases are required which are precursors for nucleic acid synthesis. Besides its function as a buffer, acetate stimulates growth of most LAB. Furthermore, inorganic salts of potassium, manganese, magnesium and phosphoric acid are required by lactic acid bacteria. These complex nutritional requirements indicate that many biochemical pathways present in other organisms are absent in lactic acid bacteria (Tittsler et al., 1952).

In order to explore and understand the growth requirements of the human pathogen *S. pyogenes* we reconstructed the genome-scale metabolic network of this bacterium. The reconstructed metabolic network aids the identification of similarities and differences between GAS and related lactic acid bacteria. Understanding *S. pyogenes*' metabolism and responses to different growth environments will also aid the development of strategies to reduce its growth and to identify possible drug targets in the future.

Genome-scale models comprise a list of associations between reactions, enzymes, substrates and products, i.e. the stoichiometry of all metabolic reactions. They can be used e.g. to explore the response of an organism's metabolism to changes in its environment, gain insights into the genotype-phenotype relationship, identify the physiological states which are achievable by a given metabolic network or analyse perturbations like gene deletions or drug applications (Durot et al., 2009). The construction of a genome-scale model is based on the organism's genome and

requires a large amount of knowledge about the organism's metabolism when done in a careful and curated manner.

Genome-scale metabolic networks have been developed for many organisms including LAB such as *Lactococcus lactis* (Flahaut et al., 2013; Oliveira et al., 2005; Verouden et al., 2009), *Lactobacillus plantarum* (Teusink et al., 2006), *Streptococcus thermophilus* (Pastink et al., 2009) and very recently *Enterococcus faecalis* (Veith et al., 2015).

In this paper we present the reconstruction of the metabolic network of *S. pyogenes* strain 591, a serotype M49 strain. We concentrated on all metabolic reactions essential for cell growth, i.e. reactions involved in nucleotide biosynthesis and production of DNA and RNA, protein biosynthesis, the synthesis of membranes, cell wall and capsule components, primary and (poly)saccharide metabolism, amino acid metabolism, pathways for the synthesis of fatty acids and the production of vitamins and cofactors.

To speed up the development of the first genome-scale model of GAS the AUTOGRAPH method (Notebaart et al., 2006) was used. AUTOGRAPH is a semi-automatic approach facilitating the generation of draft reconstructions. Based on orthology search against genomes of related organisms with available metabolic reconstructions metabolic genes in the organism of interest's genome and their corresponding metabolic functions are predicted. Here, we exploited the manually curated metabolic networks of *Bacillus subtilis* 168 (Oh et al., 2007), *Escherichia coli* K12 (Feist and Palsson, 2010), *Lactobacillus plantarum* WCFS1 (Teusink et al., 2006) and *Lactococcus lactis* subsp. *cremoris* MG1363 (Verouden et al., 2009). This step was followed by a manual curation of the initial reconstruction which included a

consistency check and gap filling. In order to analyse the metabolic capabilities of the reconstruction PySCeS CBMPy (Olivier et al., 2005) was used and flux balance analysis (FBA) (Price et al., 2004; Varma and Palsson, 1994) was applied to find a feasible and optimal solution.

The whole genome-scale metabolic model was used to simulate our fermentation data of *S. pyogenes* M49 wild type and a corresponding *arcA* deletion strain. Furthermore, the model was used to explore the organism's reaction to perturbations in its environment, such as amino acid omissions, and to find strategies to reduce the growth of *S. pyogenes* and to propose drug targets by identifying essential genes.

## 2. Materials and Methods

### 2.1 Bacterial strains and growth conditions

*S. pyogenes* M49 591 wild type and corresponding *arcA* deletion strains were grown in static batch cultures at 37°C in either Todd-Hewitt broth supplemented with 0.5% (wt/vol) yeast extract (Oxoid) or a chemically defined medium for Lactic Acid Bacteria (CDM-LAB). CDM-LAB was essentially composed as described previously (Fiedler et al., 2011; Jonsson et al., 2009) with two exceptions: the arginine concentration was 0.5 g/L and the medium additionally contained 30 mM NaHCO<sub>3</sub> (see **Table A1**). GAS *arcA* deletion strain was maintained in medium containing 60 mg/liter spectinomycin. *Escherichia coli* DH5α harboring pASK-IBA3c derivative was grown on lysogeny broth (LB) medium supplemented with 300 mg/liter erythromycin, 60 mg/liter spectinomycin, or 20 mg/liter chloramphenicol, respectively. All *E. coli* cultures were grown at 37°C under ambient air conditions.

### 2.2 Chemostat cultures

*S. pyogenes* M49 wild type and mutant strains were grown in anaerobic glucose-limited chemostat cultures in CDM-LAB as described previously (Fiedler et al., 2011; Levering et al., 2012). In brief, cultures were grown at 37°C in a Biostat Bplus fermentor unit with a total volume of 750 ml at a stirring rate of 150 rpm. The pH was maintained at the indicated value by titrating with sterile 2 M KOH. Growth rates were controlled by the medium dilution rate (D; 0.05 h<sup>-1</sup>). Culture volume was kept constant by removing culture liquid at the same rate that fresh medium was added. The cultures were considered to be in steady-state when no detectable glucose remained in the culture supernatant and the optical densities (ODs), dry weights

(DWs), and product concentrations of the cultures were constant on two consecutive days.

### **2.3 Analysis of carbon fluxes**

Glucose, pyruvate, lactate, formate, acetate, succinate and ethanol were determined by high-pressure liquid chromatography (HPLC, LKB) with a Rezex organic acid analysis column (Phenomenex) at a temperature of 45°C with 7.2 mM H<sub>2</sub>SO<sub>4</sub> as the eluent, using a RI 1530 refractive index detector (Jasco) and AZUR chromatography software for data integration. Aspartic acid, serine, glutamic acid, glycine, histidine, arginine, threonine, alanine, proline, cysteine, tyrosine, valine, methionine, lysine, isoleucine, leucine and phenylalanine were determined by HPLC (Agilent) by use of the Waters AccQ Tag method. Fluorescence was analysed using a Hitachi F-1080 fluorescence detector set to 250 nm excitation and emission was recorded at 395 nm (Fiedler et al., 2011).

### **2.4 Amino acid auxotrophy and minimal media experiments**

*S. pyogenes* M49 was precultured overnight in static batch cultures at 37°C using Todd-Hewitt broth supplemented with 0.5% (wt/vol) yeast extract (Oxoid). After washing in PBS cells were inoculated in CDM-LAB. To determine the amino acid auxotrophies, single amino acids or combinations of amino acids were omitted from the CDM-LAB. Washed GAS cells were inoculated at an optical density (600 nm; OD<sub>600</sub>) of 0.05-0.075 and incubated statically for 24 h at 37°C using 3 ml of medium in 15 ml capped tubes. Growth was monitored by visual inspection and OD<sub>600</sub> measurements. In case of poor, but visible growth (i.e. final OD<sub>600</sub> below 1), cells were used for re-inoculation in the same medium to confirm prototrophy for the

respective amino acid(s). Omissions of other medium components such as nucleotide precursors or vitamins were done accordingly. All omissions have been tested using at least two independently prepared media and at least four separate inoculations.

## 2.5 Construction of recombinant vector and GAS strain

Chromosomal DNA of *S. pyogenes* M49 was extracted with the DNeasy Blood and Tissue Kit (Qiagen, Hilden, Germany) and used as template for PCR amplification of the upstream and downstream flanking regions of the *arcA* gene, respectively. Plasmid pFW5 served as a template for amplification of the spectinomycin resistance gene cassette *aad9*. All PCR amplifications were done with Phusion Taq Polymerase (Biozym). All PCR products were purified using the QIAquick PCR purification kit (Qiagen). Sizes of the PCR products and corresponding primers are listed in **Table A2**. For the construction of the *arcA* deletion plasmid, the PCR fragments of the *arcA* upstream and downstream flanking sequences were fused via EcoRV restriction sites and ligated into plasmid pASK-IBA3c via BamHI and HindIII restriction sites. The *aad9* cassette was ligated between the flanking regions via EcoRV restriction sites. *E. coli* DH5 $\alpha$  was used for sub-cloning of pASK-IBA3c derivatives.

The resulting recombinant vector, pASK-IBA3c\_arcA\_ko, was transformed into *S. pyogenes* M49 wild type cells. The resulting transformants were screened for spectinomycin resistance. Double crossover events resulting in the deletion of the respective genes were confirmed by appropriate PCR and Northern Blot analyses.

## 2.6 Metabolic network reconstruction and modelling

The AUTOGRAPH method (Notebaart et al., 2006) was applied to the Genbank NCBI (Bilofsky et al., 1988) annotation file of *S. pyogenes* M49 NZ131 (McShan et al., 2008) which is the only complete determined genome sequence of a GAS M49 serotype and four manually curated metabolic networks from *Bacillus subtilis* 168 (Oh et al., 2007), *Escherichia coli* K12 (Feist et al., 2007), *Lactobacillus plantarum* WCFS1 (Teusink et al., 2006) and *Lactococcus lactis* subsp. *cremoris* MG1363 (Verouden et al., 2009). The identification of genes having most likely an identical biological function in different organisms is based on orthology detection by INPARANOID (Remm et al., 2001). The output of AUTOGRAPH assigns each metabolic gene from the query genome a protein function. Additionally, orthologous genes from each reference organism and a reaction associated to the gene according to the reference organism's metabolic network are listed.

Subsequent manual curation of the initial metabolic network involved identification and resolving of inconsistencies and gaps, and introducing organism specific reactions as described in detail in (Thiele and Palsson, 2010). The network was completed by searching biochemical and metabolic databases such as UniProt (The Uniprot Consortium, 2010), BRENDA (Schomburg et al., 2002), KEGG (Kanehisa and Goto, 2000) and NCBI (Geer et al., 2010) as well as journal publications. Information regarding transport proteins was obtained from TransportDB (Ren et al., 2007) and TCDB (Saier et al., 2014).

To identify missing genes and proteins we performed homology searches using the blastp suite on the NCBI webserver with default settings. If no gene could be identified although experimental evidence about the presence of a certain metabolic reaction exists or the reaction is required to simulate growth, the reaction was

included as non-gene associated reaction. Reaction names and equations for the metabolic networks were mainly derived from the genome-scale models of *L. lactis* and *L. plantarum*. Species-specific reaction equations were manually created and named. Gene-protein-reaction assignments were based on the reference genome-scale models and the previously mentioned databases.

## 2.7 Flux balance analysis

The model was simulated and analysed using PySCeS CBMPy (<http://cbmpy.sourceforge.net>) (Olivier et al., 2005) which is a platform for constraint-based modelling and analysis. PySCeS CBMPy implements analyses such as flux balance analysis (FBA), flux variability analysis (FVA), element/charge balancing, and model editing (Olivier et al., 2005). We used the IBM ILOG CPLEX solver (IBM, 2012) which is free for academic uses.

Mathematically, the reconstructed metabolic network is represented by the stoichiometric matrix  $\mathbf{S}$ . In this matrix, each row represents a mass balance of a metabolite and each column represents a reaction. Elements of the matrix are the stoichiometric coefficients. Given that the simulated system is in steady-state it can be simulated by the equation  $\mathbf{S} \cdot \mathbf{v} = 0$  where the vector  $\mathbf{v}$  describes all fluxes through the network. FBA (Price et al., 2004; Varma and Palsson, 1994) was applied to analyse feasible and optimal flux distributions of the developed stoichiometric model. Constraints based on physiological aspects or experimental data were added to the individual metabolic fluxes as upper and lower boundaries to reduce the space of allowable flux distributions of the system. Measured concentrations were

transformed into fluxes of utilisation ( $q_i$ ) given in  $\text{mmol gDW}^{-1} \text{ h}^{-1}$  by

$$q_i = D \cdot \frac{C_{i,\text{supernatant}} - C_{i,\text{feed}}}{X_{\text{biomass}}}$$

with  $C_i$  is the concentration of metabolite  $i$  in  $\text{mmol per liter}$ ,  $X_{\text{biomass}}$  is the measured biomass concentration in  $\text{gDW per liter}$  and  $D$  is the dilution rate per hour. We subsequently defined the lower bound (LB) and upper boundary (UB) of the exchange fluxes in the model by the upper and lower boundaries of the two measurements as  $\text{LB} = \min(0.9 \cdot q_1, 0.9 \cdot q_2)$  and  $\text{UB} = \max(1.1 \cdot q_1, 1.1 \cdot q_2)$ . Thus, in addition to using the two different measurements as upper and lower bounds, we allow another 10% variation to account for the experimental uncertainty. We also repeated all calculations with even wider bounds (25%) and confirmed that all results are qualitatively the same. Constraining all reactions at once resulted in an infeasible problem indicating that some constraints were too tight. Therefore, we applied the constraints one at a time and adjusted the boundaries whenever necessary. For non-measured medium compounds the lower bound was calculated based on the concentration in CDM-LAB and the upper bound was set to 1000. For non-measured amino acids which are present in the medium (i.e., asparagine, cysteine, glutamine, tryptophan) we used -1 and 0 for the lower and upper boundaries. We selected -1 to resemble the fluxes of the measured amino acids. However, applying -1000, 1000 would give the same results. The applied constraints are given in Table **A3**.

As objective function we have chosen the biomass production. It comprises biomass components and growth-associated ATP consumption and its stoichiometric coefficients represent the molar quantities that are required to produce one gram of dry weight. Due to lacking information about the biomass composition of GAS in

literature, the biomass composition of a recent *L. lactis* reconstruction (Flahaut et al., 2013) was used. The biomass objective function includes vitamins and vitamin-derived cofactors to model vitamin requirements, namely co-enzyme A, NAD, tetrahydrofolate, thiamin pyrophosphate and undecaprenol (lipid II). Although vitamins do not make a quantitative impact, their inclusion in the biomass equation ensures that deletions that interfere with metabolism of essential vitamins are also lethal in the model. Additionally, reactions involved in the metabolic pathways for the following vitamins and cofactors were included in the reconstruction: NADP, molybdenum cofactor, riboflavin, thioredoxine, glutathione, biotin, vitamin C, vitamin B6, pyridoxal 5-phosphate. By constraining the model's nutrient uptake rates with experimental measurement this objective can be used to simulate microbial growth rates. From the resulting set of allowable flux distributions the optimal metabolic flux distribution was determined using linear programming under the steady-state criteria. Simulation results were compared with our experimental data and in case of discrepancies the model was adjusted accordingly based on information found in literature or databases.

Within the model, the growth rate is represented by the flux through the biomass reaction when constraining nutrient uptake rates with experimental measurements. Under steady-state conditions the growth rate equals the dilution rate. Therefore, we compared the simulated flux through the biomass objective function with the dilution rate of  $0.05 \text{ h}^{-1}$  used in our experiments.

## **2.8 Estimation of energetic parameters**

The genome-scale metabolic model distinguishes between growth associated maintenance (GAM,  $\text{mmol gDW}^{-1}$ ) and non-growth associated maintenance (NGAM,

mmol DW<sup>-1</sup> h<sup>-1</sup>). GAM defines the ATP required for biomass assembly and is accounted for in the biomass equation itself. This parameter was adapted from the *L. lactis* model and is set to 39.4 mmol ATP per gDW (Flahaut et al., 2013). The NGAM parameter was estimated from the model. After implementing the constraints according to our experimental data the NGAM was obtained by fixing the flux through the biomass reaction to the experimentally observed growth rate (since we constrain nutrient uptake fluxes in the model, biomass reaction flux represents growth rate in units h<sup>-1</sup>) and maximizing the flux through the ATP maintenance reaction. The observed objective value equals the NGAM parameter and was used to constrain the ATP maintenance reaction. The calculated NGAM is rather high compared to similar reconstructions, e.g. of *L. lactis*, and is probably an overestimate. This underlines that biomass composition for *S. pyogenes* needs to be better understood to be able to compute this more realistically.

## 2.9 Amino acid auxotrophies

To test the essentiality of single amino acids or combinations *in silico* we blocked the amino acid uptake from the medium, i.e. we set the lower bound for this amino acid's exchange reaction to zero. The organism's ability to grow in the absence of this amino acid was tested using FBA. Note that we used the genome-scale model wide flux boundaries (i.e. with boundaries set to +/-1000 or zero which is referred to as unconstrained genome-scale model in the following) since the corresponding experiments were performed in batch cultures.

## 2.10 Flux variability analysis

Often the FBA solution is not unique and different flux distributions exist satisfying the constraints and having the same quantitative objective value. To study these alternative optimal flux distributions flux variability analysis (FVA) (Mahadevan and Schilling, 2003) was performed. In a first step the objective value is calculated by solving the linear program. From this solution the range of flux variability for all reactions in the network is calculated by fixing the objective value and maximizing and subsequently minimizing the flux of each reaction in the network. FVA helps to identify robustness within the metabolic network. Reactions with low flux variability are likely to be more important to an organism compared to reactions allowing a higher variability of fluxes.

### **2.11 Gene essentiality analysis**

We studied the effect of a single gene deletion in the metabolic network of *S. pyogenes* by subsequently knocking out each gene within the network and setting the associated reactions to carrying no flux. FBA was used to predict the growth of the mutant strain. If the objective function value is lower than a certain threshold (here 5%) of the objective value without gene knock-out, the knock-out is considered to be lethal and the respective gene to be essential, otherwise the gene is considered to be not essential.

### **2.12 Minimal medium prediction**

Using the constructed genome-scale model we predicted a minimal medium by omitting one CDM-LAB component at a time and testing if this compound is essential by performing FBA (Borenstein et al., 2008; Handorf et al., 2008). The predicted minimal medium consists of all essential CDM-LAB compounds.

### 3. Results

#### 3.1 Setting up the genome-scale model

The reconstruction process is divided into four main steps (Thiele and Palsson, 2010) and starts with the reconstruction of a draft network based on *S. pyogenes*' genome annotation and protein homology to related organisms with available reconstructions (see Materials and Methods for more details on the reconstruction process). Second, the draft reconstruction was manually curated using additional resources such as primary literature and external databases. In the third step the curated reconstruction was converted into a mathematical model. We defined system boundaries, i.e., nutrient uptake and product secretion, according to the experimental set-up. The fourth and final step consists in network verification and evaluation. We verified that all biomass components could be synthesized and identified missing functions by comparison with known phenotypes and refined the model in an iterative manner if necessary.

##### *3.1.1 Metabolic network reconstruction and initial analysis*

We reconstructed the metabolic network of *S. pyogenes* M49 based on AUTOGRAPH, a semi-automatic approach which takes advantage of already existing and manually curated models (Notebaart et al., 2006). The draft reconstruction was manually curated using different databases and primary literature (see Materials and Methods). The final reconstruction comprised all reactions that are required to simulate growth in CDM-LAB (Fiedler et al., 2011; Jonsson et al., 2009) and contains 480 genes associated with 576 reactions and 558 metabolites.

This final version significantly differs from all individual models originally used as data basis. Thus, compared to the model of *L. lactis*, there are 403 reactions in common, but also 102 reactions in our model that are not present in *L. lactis* and 239 reactions in *L. lactis* that do not occur in our model. Similar numbers hold true for *L. plantarum* (see Figure S1).

According to the experimental set-up, the growing cell was simulated by a model in steady-state with continuous in- and outflow of CDM-LAB and metabolic products from the reaction vessel. Exchange reactions were defined depending on the simulated growth experiment whereas transport reactions were given by the abilities of the cultivated organism to take up or secrete substances. Since transporters in *S. pyogenes* are poorly studied, required reactions were incorporated based on experimental findings, e.g. measurement of pyruvate in the supernatant, or adopted from the *L. lactis* model (Verouden et al., 2009). An overview of the model features and the list of reactions is given in **Table 1** and **Table A4**, respectively.

The network reconstruction facilitated gaining insight into the metabolic capabilities of *S. pyogenes*. In the following we will discuss some points of special interest that occurred during metabolic reconstruction specifically for *S. pyogenes*.

First of all, the model reflects that GAS lacks the adenylate cyclase like all other Gram-positive bacteria (Stewart, 1993), and, thus, no cyclic adenosine monophosphate is produced. Furthermore, the model reflects that no glutaminyl-tRNA ligase (EC 6.1.1.18) exists in *S. pyogenes* as previously observed for *L. plantarum* (Teusink et al., 2006). Instead, a glutamyl-tRNA amidotransferase (EC 6.3.5.-) is present which allows the formation of correctly charged glutaminyl-tRNA through the transamidation of misacylated Glu-tRNA(Gln). Interestingly, and in

contrast to *S. pyogenes*, *L. lactis* possesses tRNA(Gln) and a glutaminyl-tRNA ligase. Since we used the *L. lactis* specific biomass objective function (see Materials and Methods), the respective contribution had to be changed to not rely on tRNA(Gln), but rather incorporate more tRNA(Glu).

Phospholipid biosynthesis is poorly studied for *S. pyogenes*. However, compounds synthesized in this pathway are essential for cell wall and lipoteichoic acid (LTA) production. Therefore, phospholipid biosynthesis was assumed to be present in *S. pyogenes* and the essential enzymes in this pathway were copied from the *L. lactis* reconstruction. Since GAS teichoic acid is free of ribitol (Matsuno and Slade, 1970) and is present in the form of LTA in *S. pyogenes* (Slabyj and Panos, 1976), reactions including ribitol teichoic acid or wall teichoic acids were removed from the model.

Like all other LAB, *S. pyogenes* has an incomplete tricarboxylic acid (TCA) cycle with the consequence that no succinyl-CoA is produced. Therefore, reactions that require succinyl-CoA in other organisms were changed to use acetyl-CoA in the model. Hereby, we follow Francke *et al.* (Francke *et al.*, 2005) who suggested this for *L. lactis*. Thus, this bacterium relies on glycolysis and pyruvate metabolism for energy production. Due to the defective TCA cycle *S. pyogenes* is not able to synthesize precursors for most amino acids. As described in more detail in the model testing section, *S. pyogenes* is auxotroph for 11 amino acids. Although all genes necessary for the conversion of histidine into glutamate are annotated in *S. pyogenes* (see KEGG histidine pathway, soz00340), our experimental data showed that *S. pyogenes* does not grow in the absence of glutamine and glutamate (also see below). This indicates that the conversion of histidine to glutamine is not working

*in vivo*. A reason for this might be the accumulation of one of the intermediates, formamide. We could not find any information about a formamide consuming reactions in any streptococcus or lactococcus.

Furthermore, growth of *S. pyogenes* requires the supply of vitamins, e.g. biotin (Mickelson, 1964). This vitamin is involved in fatty acid synthesis and amino acid metabolism. Another required vitamin is panthotenate which is needed for coenzyme A production. Due to the fact that many thiamine synthesis enzymes are missing in *S. pyogenes* we hypothesize that this compound is essential and needs to be taken up from the medium. Riboflavin is also taken up from the medium and is converted into flavin adenine dinucleotide, a redox cofactor. The flavin reductase is lacking in *S. pyogenes* and therefore no reduced riboflavin is produced. Furthermore, *S. pyogenes* cannot produce molybdopterin which serves as a cofactor for some enzymes *in vivo*.

Most enzymes participating in the folate and C1-THF pool synthesis are present except for the enzyme catalysing the transformation of dihydropteridine triphosphate into 7,8-dihydropteridine, which is one of the first steps. In *L. plantarum*, two enzymes, dihydroneopterin triphosphate pyrophosphatase (DNTPPA, EC 3.6.1.-) and dihydroneopterin monophosphate dephosphorylase (DNMPPA, EC 3.6.1.-), are catalysing this conversion. In the *L. lactis* reconstruction this step is catalysed by a reaction named unkFol with a so far unknown gene association. In *Streptococcus pneumoniae* this step is catalysed by a membrane-bound alkaline phosphatase (EC 3.1.3.1). A protein BLAST search of this phosphatase against the *S. pyogenes* NZ131 genome results in the hypothetical protein Spy49\_1023c as the sequence producing the most significant alignment (maximal score 155, total score 155, query

coverage 91%, E-value  $3e-48$ , maximal identity 41%). Since all other enzymes are present in this pathway, we included the alkaline phosphatase in the model which, interestingly, catalyses the same chemical reaction as unkFol in the *L. lactis* reconstruction.

Furthermore, the NADPH-dependent enzyme methylene-tetrahydrofolate reductase (EC 1.5.1.20) catalysing the conversion of 5,10-methylenetetrahydrofolate into 5-methyltetrahydrofolate, i.e. the active form of folate, is missing in *S. pyogenes*. Usually, the latter compound is used to recycle homocysteine back to methionine by methionine synthase (EC 2.1.1.13). Instead, *S. pyogenes* possesses a homocysteine S-methyltransferase which catalyzes the methionine production from homocysteine while converting S-Adenosyl-L-methionine into S-Adenosyl-L-homocysteine (EC 2.1.1.10).

### 3.1.2 Elemental and charge balance

The curated reconstruction was transformed into a computational model using PySCeS CMBPy (Olivier et al., 2005) and all reactions were balanced in terms of chemical elements and charge. There are a few exceptions in genome-scale models where the reactions in the end are necessarily unbalanced. Model-based input and output reactions allow the transfer of mass across the system boundaries and do not need to be balanced. In contrast, internal reactions need to be balanced due to the steady-state assumption. We used a function implemented in PySCeS CBMPy to check any problems with respect to this. In the end, the model contains three reactions out of 576 that are unbalanced (see **Table A5**). One out of these three reactions is unbalanced with respect to the chemical composition due to the involvement of proteins and the other two reactions are charge unbalanced.

Reactions that explicitly or implicitly include proteins cannot be balanced properly in a metabolic network since protein synthesis is not included, except as a general term for the biomass equation. The charge-imbalanced reactions are involved in LTA production. LTAs are large polymers with small fluxes. These three imbalanced reactions do not lead to any artificial gain or loss of matter which could have resulted in changes in other network fluxes.

### *3.1.3 Identification of type III extreme pathways*

Before constraining the model with our measured fluxes we tested the metabolic network for the presence of type III extreme pathways. Type III extreme pathways are internal cycles that lead to no net conversion of any metabolite and are artefacts of metabolic reconstructions (Price et al., 2002; Schilling et al., 2000; Thiele and Palsson, 2010). The presence of these cycles was tested by performing a flux variability analysis (FVA) and closing all exchange reactions, i.e. setting lower boundaries to zero.

We identified 13 reactions that are involved in these stoichiometrically balanced cycles (see **Table 3**). The ATP- and dATP-dependent guanylate kinases (GK1 and GK2, EC 2.7.4.8) and the nucleoside-diphosphate kinase NDPK8 (EC 2.7.4.6) are involved in the purine metabolism. By setting the lower boundaries to zero this thermodynamically infeasible cycle could be prevented. The second circulation group is driven by O<sub>2</sub> production through dihydroorotic acid dehydrogenase (DHORD1, EC 1.3.3.1) and could be eliminated by constraining the reversibility of DHORD1. This inhibits also the circulation of L-lactate oxidase (LOXL, EC 1.13.12.4) and L-lactate dehydrogenase (LDH\_L, EC 1.1.1.27) as well as glycerol 3-phosphate oxidase (G3PO, EC 1.1.3.21) and glycerol-3-phosphate dehydrogenase (G3PD1,

EC 1.1.1.94). The last stoichiometrically balanced cycle between the two phosphoribosylglycinamide formyltransferases GARFT<sub>met</sub> and GARFT (EC 2.1.2.2) and methenyltetrahydrofolate cyclohydrolase (MTHFC, EC 3.5.4.9) could be circumvented by constraining the lower bound of MTHFC to zero. With these changes in the model constraints we were able to remove all type III extreme pathways.

### 3.1.4 Constraining the model with fermentation data

Flux balance analysis (FBA) (Price et al., 2004; Varma and Palsson, 1994) finds a flux distribution through the metabolic network that optimizes an objective function given a set of constraints. We used experimentally derived flux values to constrain the solution space. **Figure 1** shows the experimentally determined fermentation pattern of *S. pyogenes* wild type and *arcA* deletion strain at two different pHs, 6.5 and 7.5, and a dilution rate of 0.05 h<sup>-1</sup>. As a lactic acid bacterium GAS is characterised by its capability to ferment glucose primarily to lactate. Under glucose-limited conditions as studied here, *S. pyogenes* carries out mixed acid fermentation and produces also formate, acetate and ethanol as shown in **Figure 1A**. At pH 7.5, *S. pyogenes* consumes all measured amino acids except for tyrosine and at pH 6.5 all amino acids except for aspartate and glutamate (**Figure 1B**). Typical pH adaptations, like an increased uptake of arginine and serine can be seen in the data. Knocking out arginine deiminase (*ArcA*) does not alter the organic acid end-product pattern but has a strong impact on amino acid metabolism as shown in **Figure 1C** and **Figure 1D**. Under the studied conditions,  $\Delta arcA$  mutants produced alanine, aspartate, glutamate, methionine, phenylalanine, proline and threonine. Interestingly,  $\Delta arcA$  shows a similar metabolic pattern as the wild type at lower pH, with the

obvious exception of arginine uptake. Thus, lactate production as well as serine uptake is increased.

We used the experimentally determined uptake and production fluxes to restrict exchange and transport reactions in the model as described in detail in Materials and Methods.

### *3.1.5 Maintenance and growth-associated energy coefficients*

Energy requirements for growth and maintenance are essential parameters in a genome-scale model as described in (Teusink et al., 2006) and need to be specified before the model can be used to predict growth rates. The growth associated energy parameter is specified in the biomass equation and was set to 39.4 mM/gDW as done for *L. lactis* before (Flahaut et al., 2013). We estimated the non-growth associated ATP maintenance parameter (NGAM) from the genome-scale model by fixing the biomass flux to  $0.05 \text{ h}^{-1}$  and maximizing the flux through the ATP maintenance reaction. Our model predicts the NGAM parameter to be i) 6.18 mM  $\text{gDW}^{-1} \text{ h}^{-1}$  for WT pH 6.5,  $D=0.05 \text{ h}^{-1}$ , ii) 5.07 mM  $\text{gDW}^{-1} \text{ h}^{-1}$  for WT pH 7.5,  $D=0.05 \text{ h}^{-1}$  and iii) 2.93 mM  $\text{gDW}^{-1} \text{ h}^{-1}$  for the  $\Delta arcA$  mutant strain at pH 7.5,  $D=0.05 \text{ h}^{-1}$ . These values are higher than the non-growth associated maintenance values applied in models of other lactic acid bacteria (Flahaut et al., 2013) but in the same range as experimentally determined for *L. lactis* NZ9000, *Enterococcus faecalis* V583 and *S. pyogenes* M49 grown in carbon limited continuous cultures (Fiedler et al., 2011).

## **3.2 Model analysis and testing**

### *3.2.1 Wild type and $\Delta arcA$ mutant show different fermentation patterns*

After applying our measured flux boundaries as constraints and specifying maintenance and growth associated energy constants we used FBA and FVA to analyse the constructed genome-scale metabolic model. As shown in **Table 2** we characterized reactions in the model as blocked or essential. Blocked reactions cannot carry any flux and indicate gaps or annotation errors in the genome-scale model (Ponce-de-León et al., 2013). Alternatively, gaps could be also due to reactions that possess a regulatory role which is not impacting a stoichiometric model and therefore does not carry flux when calculating feasible flux distributions. A reaction was classified as essential if it has to carry flux to allow biomass production activity. Additionally, we identified gap metabolites which only take part in blocked reactions as done in (Ponce-de-León et al., 2013). Since the respective numbers are dependent on the applied constraints they differ between the unconstrained complete network and the models simulating our three studied experimental conditions. The complete network obviously has the smallest number of essential reactions, since there is more flexibility with respect to fluxes and routes. By applying more constraints to the model the number of essential reactions required for growth increases. However, the ratio between reactions carrying flux (active) and reactions not carrying any flux (inactive) does not change between the unconstrained model and the ones simulating growth in CDM-LAB.

Our model was able to fit the growth rate of  $0.05 \text{ h}^{-1}$  for the wild type grown at pH 6.5 and 7.5 and for the *arcA* knock-out mutant grown at pH 7.5. We used the genome-scale model to study the different fermentation patterns of the wild type grown at different pH values and of the wild type compared with the  $\Delta arcA$  mutant grown at the same experimental conditions. Within the model, the flux distributions for the

different pH values result from the distinct experimental constraints. **Table A6** and **Table A7** depict the different fluxes. We used FVA to calculate the flux ranges and defined a flux as different, if the absolute difference of the optimal, minimal or maximal value is greater than  $10^{-4}$  between the two conditions. For the wild type, 134 fluxes were found to be different between pH 6.5 and pH 7.5. For wild type and *ΔarcA* mutant 112 fluxes differed.

Interestingly, the production of ammonium by the deamination of arginine and serine which are taken up in higher amounts at lower pH does not directly influence intracellular pH in the model, since the deamination reaction at the respective pH do not result in proton consumption. It is rather the subsequent production of ATP from the products of the deamination reaction that allows to pump more protons out of the cell and therefore helps to maintain a higher pH.

The *ΔarcA* mutant is obviously unable to generate ATP from arginine. Thus the quite similar shift in fermentation pattern which can be reproduced in the model can be explained by an increased ATP demand compared to the wild type under the chosen experimental conditions.

### 3.2.2 Amino acid auxotrophies

In order to determine the amino acid auxotrophies and prototrophies of *S. pyogenes* we performed growth experiments in CDM-LAB omitting sequentially each of the 20 proteinogenic amino acids and cystine from the medium, also in combinations. After 24 h of growth the optical densities (ODs) were determined and compared to growth in full CDM-LAB. The omission of alanine, asparagine, aspartic acid, cysteine, cystine, glutamine, glutamic acid and proline did not affect the OD compared to the

OD the *S. pyogenes* cultures reached after 24 h with complete CDM-LAB (**Figure 2A**). The omissions of all other amino acids led to a strong decrease in the final ODs. For these amino acids repeated re-inoculations in CDM-LAB lacking the respective amino acid were performed to account for adaptation in gene expression. The experiments revealed that *S. pyogenes* is also able to grow in the absence of serine and glycine or both when adaptation is allowed (**Figure 2B**). However, the combined omission of glutamine and glutamate, as well as the omission of the pair cysteine/cystine or their combination with serine does not allow growth of *S. pyogenes* as shown in **Figure 2A**. In summary, in CDM-LAB *S. pyogenes* is auxotrophic for 11 out of the 20 proteinogenic amino acids.

We used the collected amino acid auxotrophy information to assess the consistency of the reconstructed metabolic network. We tested *S. pyogenes*' ability to grow in the absence of single amino acids or combinations *in silico* by blocking the uptake of the corresponding amino acid(s) and using FBA. Since the experiments were performed in batch cultures we exploited the genome-scale model with default boundaries ( $\pm 1000$ ) to qualitatively predict growth or non-growth (see Materials and Methods). The model was able to correctly predict the outcome of almost all leave out experiments. However, there were two points of contradiction that we needed to resolve during the iterative model refinement step, namely predicting the growth in the absence of alanine and in the absence of glycine and serine. Initially, the computational results predicted no growth in the absence of alanine in contrast to the experimental data. According to KEGG and the genome annotation, *S. pyogenes* can convert glutamate into alanine and asparagine, which can be further transformed into aspartate. Although the model included the reaction producing

alanine from glutamate (alanine transaminase, EC 2.6.1.2) the reaction was not running due to problems with the  $\alpha$ -ketoglutarate (AKG) balance. The latter compound is produced concomitant with alanine and accumulated in the model. We were unable to find any *S. pyogenes* specific transport protein catalyzing AKG efflux or any other AKG consuming reaction resolving this problem. However, since AKG is an important precursor for many biomolecules, it is not unlikely that such an unknown reaction exists. Therefore, we introduced an artificial demand for  $\alpha$ -ketoglutarate symbolizing a knowledge gap while allowing the model to reproduce growth without alanine.

Furthermore, the model predicted no growth in the absence of glycine. Glycine and serine can be converted into one another by glycine hydroxymethyltransferase (EC 2.1.2.1). To produce glycine from serine, tetrahydrofolate is converted into methenyltetrahydrofolate. The latter two compounds are involved in the folate biosynthesis and in the one-carbon-pool of folate. There were several reactions missing in the latter pathway. After the incorporation of these reactions the model predicted growth in the absence of glycine, but not in CDM-LAB with the double leave-out of serine and glycine. According to KEGG serine and pyruvate can be converted into one another (L-serine deaminase, EC 4.3.1.17). This reaction was indicated to be irreversible in the KEGG database. However, in previous reconstructions (Becker and Palsson, 2005; Nogales et al., 2008), the reaction is assumed to be reversible. This assumption is without doubt questionable since according to the eQuilibrator database (<http://equilibrator.weizmann.ac.il>) the  $K_{eq}$  of the reaction is in the range of  $10^5$ . Allowing also the reverse reaction or assuming an

alternative yet unknown path interconverting the respective amino acids resolved this final problem.

An overview of the experimentally determined auxotrophies in comparison with previously published data (Slade et al., 1951) and with the behaviour of the refined genome-scale model is given in **Table 4**. All in all the experimental findings are in good agreement with the computational predictions. We should point out that according to literature cysteine, cystine, glycine, proline and serine are essential which was neither observed in our experiments nor in the model before (for cysteine, cystine, proline and serine) or after refinement (for glycine) (Slade et al., 1951). **Figure 3** summarizes the possible conversions of amino acids for *S. pyogenes* in growth in CDM-LAB predicted by the genome-scale model and verified by our experimental data.

### 3.2.3 Predicting a minimal medium

The reconstructed network was used to predict a minimal medium composition which is summarized in **Table 5**. Here, we want to highlight some central components: All essential amino acids and additionally glutamine or glutamate as well as cysteine or cystine are required for *in silico* growth of *S. pyogenes*. Furthermore, guanine or xanthine, which are interconvertible, and adenine are essential. Uracil is not required for growth but UTP which can be produced from CTP. Inosine can be produced from adenine. Similar to uracil, thymidine is not a biomass component but TTP which is produced from UTP.

If citrate is not supplied, aspartate or asparagine has to be supplied to ensure growth of *S. pyogenes* since the production of asparagine from glutamate requires

oxaloacetate which is synthesized from citrate. Additionally, the model requires ammonium or ammonia, sulfate, water and protons.

Thus, there is not one single possible composition of the minimal medium, but different possible combinations and the model can be used to verify each of these.

We experimentally tested the model predictions by growing *S. pyogenes* in CDM-LAB leaving out individual or combinations of the predicted non-essential medium compounds. Our experiments verify growth in CDM-LAB without the nucleotide precursors guanine, uracil, xanthine and combinations thereof as shown in **Figure 4**.

We were also able to validate growth of *S. pyogenes* in CDM-LAB without predicted non-essential vitamins, namely biotin, inosine, orotic acid, pyridoxamine, pyridoxine, riboflavin and thymidine. Omission of these vitamins lead to about 80 % biomass reduction compared to full medium, but still allowed significant growth of the bacteria (**Figure 4**).

Finally, we evaluated growth of *S. pyogenes* in a minimal medium leaving out all predicted non-essential compounds with and without glucose. However, in both cases GAS grew very poorly. Model predictions were based on excess abundance of all essential compounds. Thus, the discrepancy between the model and the experimental findings in the last test might be caused by limiting medium components.

#### *3.2.4 Essential gene analysis*

To analyse essential genes within the metabolic network of *S. pyogenes* M49 we knocked-out one gene at a time by setting the fluxes through all reactions associated with this gene to zero and performing FBA. We defined the effect of each gene on

the metabolism based on the objective function value after gene knock-out compared to the biomass production flux before knock-out and categorized the genes into four categories, namely unaffected, improved, affected and lethal as shown in **Table 6**. Obviously, the presented results strongly depend on the model, especially on the chosen objective function and on the accuracy of the implemented gene-reaction associations. For most of the genes (289) a knock-out did not significantly affect the growth of *S. pyogenes*. We identified 12 out of 480 genes whose deletion inhibited the biomass production rate and 179 gene knock-outs completely blocked growth. The gene essentiality analysis results are summarized in **Table A8**. Not surprisingly, among the genes that inhibit biomass production the majority of genes is coding for enzymes of the central metabolism like GAPDH, PYK, TPI and PGK. However, the fact that these are not essential points to an unexpected versatility of the central metabolism. This is not so much the case in amino acid metabolism, fatty acid metabolism and nucleotide metabolism, as well as protein synthesis in which the majority of gene products of essential genes participate.

Recently, Le Breton and co-workers performed a screen for genes essential for growth under optimal condition in *S. pyogenes* M1T1 5448 and M49 NZ131 using transposon mutagenesis (Le Breton et al., 2015). Based on all 1698 annotated genes in the GAS M49 genome, the authors identified 241 essential genes many of which are involved in key cellular processes and metabolic pathways such as central carbon metabolism and fatty acid synthesis. Here, we predict essential genes in metabolic pathways and thus, obviously, these are a subset of the genes determined by Le Breton and co-authors. Our analysis yields 179 essential genes mainly involved in amino acid metabolism, fatty acid metabolism, nucleotide metabolism

and protein synthesis. We found nine genes in central carbon metabolism to be essential, for example fructose-bisphosphate aldolase (EC 4.1.2.13), phosphofructokinase (EC 2.7.1.11) and enolase (EC 4.2.1.11) which were also identified by transposon mutagenesis. Le Breton *et al.* identified additional essential genes in glycolysis such as pyruvate kinase which, according to our model predictions, is not essential but slightly compromises growth.

We compared our essential gene analysis with their results to get a more general overview of similarities and discrepancies using Euler diagrams (see **Figure 5**). 103 out of the 178 essential genes predicted by the model are also identified as essential by Le Breton *et al.* (**Figure 5A**), 49 were classified as not conclusive by Le Breton *et al.*, and 26 are not-essential according to the transposon mutagenesis results but predicted to be essential by the model. We also noted an overlap between the genes classified as essential by Le Breton and non-essential genes of the genome-scale model (25 genes, see **Figure 5B**). Discrepancies between predicted and experimental determined essential and non-essential genes and vice-versa may indicate incorrect gene-reaction rules in the model.

## 4. Discussion

In this study we presented the first genome-scale metabolic network reconstruction for *S. pyogenes* M49. To guarantee a high-quality of the reconstructed metabolic network and the corresponding model we used established networks of related organisms as reference networks to obtain an initial reconstruction followed by an extensive manual curation phase on our side. We have measured input and output fluxes at two differential pH values, 6.5 and 7.5, for the wild type and also created an *arcA* deletion strain and determined its fermentation pattern at pH 7.5. We used the developed genome-scale model to analyse the data set.

As pointed out above, the shift in fermentation and amino acid metabolism is quite similar between the wild type at lower pH and the *arcA* deletion strain. When analysing this fact in the model, it became apparent that the increased uptake of arginine and serine at lower pH leads to an increased NGAM in the model, since both amino acids can be metabolized to yield ATP. We speculate that this increased ATP production seen in the model with the applied experimental constraints is used to pump out protons when subject to a more acidic environment. Thus, the lower pH is – in contrast to literature (LaSarre and Federle, 2011; Liu et al., 2003; Marquis et al., 1987; Novák and Loubiere, 2000) – not counterbalanced by direct consumption of protons during deamination, since at this pH ammonium is produced during deamination without proton consumption. This now readily explains why the *arcA* deletion strain is reacting in a similar way. Due to the lack of arginine metabolism, this strain has a lower NGAM which it tries to counterbalance by reacting in a similar way than the wildtype at lower pH which has a higher ATP demand.

To assess the consistency of the genome-scale model we performed amino acid leave-out experiments. The curated model predicted 9 out of 11 essentialities correctly. The existing discrepancies led to model refinement and are likely due to false annotation and missing annotations in the respective gene databases. Thus, folate metabolism was incomplete in the model and had to be completed to allow growth on glycine. Growth on alanine was prevented due to AKG accumulation and could be enabled by adding an AKG demand reaction. We also compared our findings to published amino acid auxotrophy data (Slade et al., 1951). According to Slade *et al.*, cysteine, cystine, glycine, proline and serine are essential but our experimental data show that *S. pyogenes* is able to grow without supplying these amino acids (see Table 4). In the cases of cysteine, cystine, proline and serine the results were very clear whereas the leave out of glycine required repeated inoculation (see Figure 2). Therefore, in the latter case, we cannot exclude the occurrence of mutations that enabled *S. pyogenes* to grow on glycine. However, the fact that the model also suggests growth in the absence of glycine – at least in the presence of folate metabolism – provides further evidence for glycine being non-essential. The striking discrepancies between literature and our findings might be caused by the usage of different media for the omission experiments. As mentioned before, the amount and combination of amino acids required for growth is characteristic for each LAB and depends upon the medium composition, i.e. upon the supplied vitamins.

The effect of the medium composition has been also analysed when predicting minimal media. These can have different compositions since what is essential obviously depends again strongly on what else is supplied. A striking example for

this fact is the early study by Mickelson *et al.* (Mickelson, 1964) showing improved growth of *S. pyogenes* in the absence of ASP/ASN due to their interference with CO<sub>2</sub> utilization.

We experimentally tested the model predictions regarding the essentiality of individual components by monitoring growth in leave-out experiments with predicted non-essential components. Growth occurred in all predicted examples. However, when setting up an absolute minimal medium leaving out every predicted non-essential component at the same time, only very poor growth was observed. This could be true to the limited availability of individual components in the media compared to abundant availability in the model.

Finally, the analysis of gene essentialities can help in the identification of novel drug targets. However, it is important to note that due to methodological restrictions only stoichiometric effects can be considered with the employed computational approach. Kinetic effects are not tractable and potential vulnerable targets which exist due to kinetic effects as the previously suggested GAPN (Levering *et al.*, 2012) will not be visible. As indicated by the essentiality analysis central metabolism in *S. pyogenes* seems to be very versatile and able to circumvent knock-outs of enzymes whereas this does not hold true for amino acid and nucleotide metabolism.

We compared the results of our essential gene analysis to a recent screen for genes essential for growth under optimal condition in *S. pyogenes* M1T1 5448 and M49 NZ131 using transposon mutagenesis (Le Breton *et al.*, 2015). As shown in Figure 5, most model predictions coincide with the results from the screen. However, there are some discrepancies. A false positive result (non-essential gene predicted although it is truly essential; 25 genes, Figure 5B) indicates that the reaction in the

model is either associated to isozymes but, actually, there are none present in *S. pyogenes*. or the model contains a reaction circumventing the reaction blocked by the gene knock-out. A false negative result (essential gene predicted although it is truly non-essential; 26 genes, Figure 5A) indicates that the gene-reaction rule misses an isozyme. The genome-scale model presents the current knowledge of *S. pyogenes* and, thus, contains knowledge gaps. Consequently, gene-reaction rules lack information or contain wrongly listed isozyme information. Although the model allows the identification of essential genes the results of such an analysis should be carefully examined given that the model does not take kinetic effects into account and focuses on a subset of the genome, here 26.8% of the annotated genes.

All in all, we provide here a manually curated and consistency checked genome-scale model of the pathogen *S. pyogenes* which allows the prediction of growth under different conditions and a directed search for essential genes.

## **Acknowledgement**

This work was financially supported by SysMO-LAB (Systems Biology of Microorganisms - Lactic acid bacteria). The authors would like to thank Michiel Wels for supplying the output of the AUTOGRAPH method and Erika Tsingos for literature and database search on the amino acid metabolism of *S. pyogenes*.

## References

- Becker, S.A., Palsson, B.Ø., 2005. Genome-scale reconstruction of the metabolic network in *Staphylococcus aureus* N315: An initial draft to the two-dimensional annotation. *BMC Microbiol.* 5, 8.
- Bilofsky, H.S., Burks, C., Fickett, J.W., Goad, W.B., Lewitter, F.I., Rindone, W.P., Swindell, C.D., Tung, C.S., 1988. The GenBank genetic sequence data bank. *Nucleic Acids Res* 14, 1–4.
- Borenstein, E., Kupiec, M., Feldman, M.W., Ruppin, E., 2008. Large-scale reconstruction and phylogenetic analysis of metabolic environments. *Proc. Natl. Acad. Sci. U. S. A.* 105, 14482–14487. doi:10.1073/pnas.0806162105
- Durot, M., Bourguignon, P.-Y., Schachter, V., 2009. Genome-scale models of bacterial metabolism: Reconstruction and applications. *FEMS Microbiol. Rev.* 33, 164–190.
- Feist, A.M., Henry, C.S., Reed, J.L., Krummenacker, M., Joyce, A.R., Karp, P.D., Broadbelt, L.J., Hatzimanikatis, V., Palsson, B.Ø., 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. doi:10.1038/msb4100155
- Feist, A.M., Palsson, B.O., 2010. The biomass objective function. *Curr. Opin. Microbiol.* In Press, , 344–349. doi:DOI: 10.1016/j.mib.2010.03.003
- Fiedler, T., Bekker, M., Jonsson, M., Mehmeti, I., Pritschke, A., Siemens, N., Nes, I., Hugenholtz, J., Kreikemeyer, B., 2011. Characterization of three lactic acid bacteria and their isogenic *ldh* deletion mutants shows optimization for YATP

- (cell mass produced per mole of ATP) at their physiological pHs. *Appl. Environ. Microbiol.* 77, 612–617.
- Flahaut, N.A.L., Wiersma, A., van de Bunt, B., Martens, D.E., Schaap, P.J., Sijtsma, L., Dos Santos, V.A.M., de Vos, W.M., 2013. Genome-scale metabolic model for *Lactococcus lactis* MG1363 and its application to the analysis of flavor formation. *Appl. Microbiol. Biotechnol.* 97, 8729–8739.
- Francke, C., Siezen, R.J., Teusink, B., 2005. Reconstructing the metabolic network of a bacterium from its genome. *Trends Microbiol.* 13, 550–558.
- Geer, L.Y., Marchler-Bauer, A., Geer, R.C., Han, L., He, J., He, S., Liu, C., Shi, W., Bryant, S.H., 2010. The NCBI BioSystems database. *Nucleic Acids Res.* 38, D492–D496.
- Gunnewijk, M.G., van den Bogaard, P.T., Veenhoff, L.M., Heuberger, E.H., de Vos, W.M., Kleerebezem, M., Kuipers, O.P., Poolman, B., 2001. Hierarchical control versus autoregulation of carbohydrate utilization in bacteria. *J. Mol. Microbiol. Biotechnol.* 3, 401–413.
- Handorf, T., Christian, N., Ebenhöf, O., Kahn, D., 2008. An environmental perspective on metabolism. *J. Theor. Biol.* 530–537.
- IBM, 2012. CPLEX Optimization Studio, 12.4.
- Jonsson, M., Saleihan, Z., Nes, I.F., Holo, H., 2009. Construction and characterization of three lactate dehydrogenase-negative *Enterococcus faecalis* V583 mutants. *Appl. Environ. Microbiol.* 75, 4901–4903.

- Kanehisa, M., Goto, S., 2000. KEGG: Kyoto encyclopedia of genes and genomes. *Nucleic Acids Res.* 28, 27–30.
- LaSarre, B., Federle, M.J., 2011. Regulation and consequence of serine catabolism in *Streptococcus pyogenes*. *J. Bacteriol.* 193, 2002–2012. doi:10.1128/JB.01516-10
- Le Breton, Y., Belew, A.T., Valdes, K.M., Islam, E., Curry, P., Tettelin, H., Shirtliff, M.E., El-Sayed, N.M., McIver, K.S., 2015. Essential genes in the core genome of the human pathogen *Streptococcus pyogenes*. *Sci. Rep.* 5, 9838.
- Levering, J., Musters, M.W.J.M., Bekker, M., Fiedler, T., Bellomo, D., de Vos, W.M., Hugenholtz, J., Kreikemeyer, B., Kummer, U., Teusink, B., 2012. Role of phosphate in the central metabolism of two lactic acid bacteria - a comparative systems biology approach. *FEBS J.* 279, 1274–1290.
- Liu, S.Q., Holland, R., McJarrow, P., Crow, V.L., 2003. Serine metabolism in *Lactobacillus plantarum*. *Int. J. Food Microbiol.* 89, 265–273. doi:10.1016/S0168-1605(03)00157-0
- Mahadevan, R., Schilling, C.H., 2003. The effects of alternate optimal solutions in constraint-based genome-scale metabolic models. *Metab. Eng.* 5, 264–276.
- Marquis, R.E., Bender, G.R., Murray, D.R., Wong, A., 1987. Arginine deiminase system and bacterial adaptation to acid environments. *Appl. Environ. Microbiol.* 53, 198.
- Matsuno, T., Slade, H., 1970. Composition and properties of a group A streptococcal teichoic acid. *J. Bacteriol.* 102, 747–752.

- McShan, W.M., Ferretti, J.J., Karasawa, T., Suvorov, A.N., Lin, S., Qin, B., Jia, H., Kenton, S., Najjar, F., Wu, H., Scott, J., Roe, B.A., Savic, D.J., 2008. Genome sequence of a nephritogenic and highly transformable M49 strain of *Streptococcus pyogenes*. *J. Bacteriol.* 190, 7773–7785.
- Mickelson, M.N., 1964. Chemically defined medium for growth *Streptococcus pyogenes*. *J. Bacteriol.* 88, 158–164.
- Nogales, J., Palsson, B.Ø., Thiele, I., 2008. A genome-scale metabolic reconstruction of *Pseudomonas putida* KT2440: iJN746 as a cell factory. *BMC Syst. Biol.* 2, 79.
- Notebaart, R.A., van Enkevort, F.H.J., Francke, C., Siezen, R.J., Teusink, B., 2006. Accelerating the reconstruction of genome-scale metabolic networks. *BMC Bioinformatics* 7, 296. doi:10.1186/1471-2105-7-296
- Novák, L., Loubiere, P., 2000. The metabolic network of *Lactococcus lactis*: Distribution of 14 C-labeled substrates between catabolic and anabolic pathways. *J. Bacteriol.* 182, 1136. doi:10.1128/JB.182.4.1136-1143.2000.Updated
- Oh, Y.-K., Palsson, B.O., Park, S.M., Schilling, C.H., Mahadevan, R., 2007. Genome-scale reconstruction of metabolic network in *Bacillus subtilis* based on high-throughput phenotyping and gene essentiality data. *J. Biol. Chem.* 282, 28791–28799.
- Oliveira, A.P., Nielsen, J., Förster, J., 2005. Modeling *Lactococcus lactis* using a genome-scale flux model. *BMC Microbiol.* 5, 39.

- Olivier, B.G., Rohwer, J.M., Hofmeyr, J.-H.S., 2005. Modelling cellular systems with PySCeS. *Bioinformatics* 21, 560–561.
- Pastink, M.I., Teusink, B., Hols, P., Visser, S., de Vos, W.M., Hugenholtz, J., 2009. Genome-scale model of *Streptococcus thermophilus* LMG18311 for metabolic comparison of lactic acid bacteria. *Appl. Environ. Microbiol.* 75, 3627–3633.
- Ponce-de-León, M., Montero, F., Peretó, J., 2013. Solving gap metabolites and blocked reactions in genome-scale models: application to the metabolic network of *Blattabacterium cuenoti*. *BMC Syst. Biol.* 7, 114.
- Price, N.D., Famili, I., Beard, D.A., Palsson, B.Ø., 2002. Extreme pathways and Kirchhoff's second law. *Biophys. J.* 83, 2879–2882. doi:10.1016/S0006-3495(02)75297-1
- Price, N.D., Reed, J.L., Palsson, B., 2004. Genome-scale models of microbial cells: Evaluating the consequences of constraints. *Nat. Rev. Microbiol.* 2, 886–897.
- Remm, M., Storm, C.E., Sonnhammer, E.L., 2001. Automatic clustering of orthologs and in-paralogs from pairwise species comparisons. *J. Mol. Biol.* 314, 1041–1052.
- Ren, Q., Chen, K., Paulsen, I.T., 2007. TransportDB: a comprehensive database resource for cytoplasmic membrane transport systems and outer membrane channels. *Nucleic Acids Res.* 35, D274–D279.
- Saier, M.H., Reddy, V.S., Tamang, D.G., Västermark, Å., 2014. The transporter classification database. *Nucleic Acids Res.* 42, D251–D258. doi:10.1093/nar/gkt1097

- Schilling, C.H., Letscher, D., Palsson, B.Ø., 2000. Theory for the systemic definition of metabolic pathways and their use in interpreting metabolic function from a pathway-oriented perspective. *J. Theor. Biol.* 203, 229–248. doi:10.1006/jtbi.2000.1073
- Schomburg, I., Chang, A., Schomburg, D., 2002. BRENDA, enzyme data and metabolic information. *Nucleic Acids Res.* 30, 47–49.
- Slabyj, B.M., Panos, C., 1976. Membrane lipoteichoic acid of *Streptococcus pyogenes* and its stabilized L-form and the effect of two antibiotics upon its cellular content. *J. Bacteriol.* 127, 855–862.
- Slade, H.D., Knox, G.A., Slamp, W.C., 1951. The amino acid nutrition of group A hemolytic *Streptococci*, with reference to the effect of glutathione on the cystine requirement. *J. Bacteriol.* 62, 669–675.
- Stewart, G.C., 1993. Catabolite repression in the gram-positive bacteria: Generation of negative regulators of transcription. *J. Cell. Biochem.* 51, 25–28.
- Teusink, B., Wiersma, A., Molenaar, D., Francke, C., De Vos, W.M., Siezen, R.J., Smid, E.J., 2006. Analysis of growth of *Lactobacillus plantarum* WCFS1 on a complex medium using a genome-scale metabolic model. *J. Biol. Chem.* 281, 40041–40048. doi:10.1074/jbc.M606263200
- The Uniprot Consortium, 2010. The Universal Protein Resource (UniProt) in 2010. *Nucleic Acids Res.* 38, D142–D148.
- Thiele, I., Palsson, B.Ø., 2010. A protocol for generating a high-quality genome-scale metabolic reconstruction. *Nat. Protoc.* 5, 93–121. doi:10.1038/nprot.2009.203.A

- Tittsler, R.P., Pederson, C.S., Snell, E.E., Hendlin, D., Niven, C.F., 1952. Symposium on the lactic acid bacteria. *Bacteriol. Rev.* 16, 227–260.
- Varma, A., Palsson, B.O., 1994. Stoichiometric flux balance models quantitatively predict growth and metabolic by-product secretion in wild-type *Escherichia coli* W3110. *Appl. Environ. Microbiol.* 60, 3724–3731.
- Veith, N., Solheim, M., van Grinsven, K.W.A., Olivier, B.G., Levering, J., Grosseholz, R., Hugenholtz, J., Holo, H., Nes, I., Teusink, B., Kummer, U., 2015. Using a genome-scale metabolic model of *Enterococcus faecalis* V583 to assess amino acid uptake and its impact on central metabolism. *Appl. Environ. Microbiol.* 81, 1622–33.
- Verouden, M.P.H., Notebaart, R.A., Westerhuis, J.A., van der Werf, M.J., Teusink, B., Smilde, A.K., 2009. Multi-way analysis of flux distributions across multiple conditions. *J. Chemom.* 23, 406–420.
- Walker, M.J., Barnett, T.C., McArthur, J.D., Cole, J.N., Gillen, C.M., Henningham, A., Sriprakash, K.S., Sanderson-Smith, M.L., Nizet, V., 2014. Disease manifestations and pathogenic mechanisms of group A *Streptococcus*. *Clin. Microbiol. Rev.* 27, 264–301.



## Appendix

**Table A1:** CDM-LAB medium. Composition of the chemically defined medium for lactic acid bacteria per liter.

**Table A2:** Oligonucleotide primers used for the construction of recombinant vector in this study.

**Table A3:** Experimentally determined metabolite fluxes and corresponding calculated flux boundaries.

**Table A4:** List of model reactions and metabolites.

**Table A5:** List of imbalanced reactions in the *S. pyogenes* metabolic network reconstruction.

**Table A6:** Comparison fermentation patterns wild type grown at pH 6.5 and at pH 7.5. We used FVA to calculate the flux ranges and defined a flux as different if the absolute difference of the optimal (optval), minimal (min) or maximal (max) value is greater than  $10^{-4}$  between the two conditions.

**Table A7:** Comparison fermentation patterns wild type and *arcA* knock-out strain. We used FVA to calculate the flux ranges and defined a flux as different if the absolute difference of the optimal (optval), minimal (min) or maximal (max) value is greater than  $10^{-4}$  between the two conditions.

**Table A8:** Essential gene analysis results.

Other supplementary files: SBML files containing the genome-scale metabolic models of *S. pyogenes* WT and *arcA* deletion strain, applied constraints, computational code and a readme file.

## Figure Legends

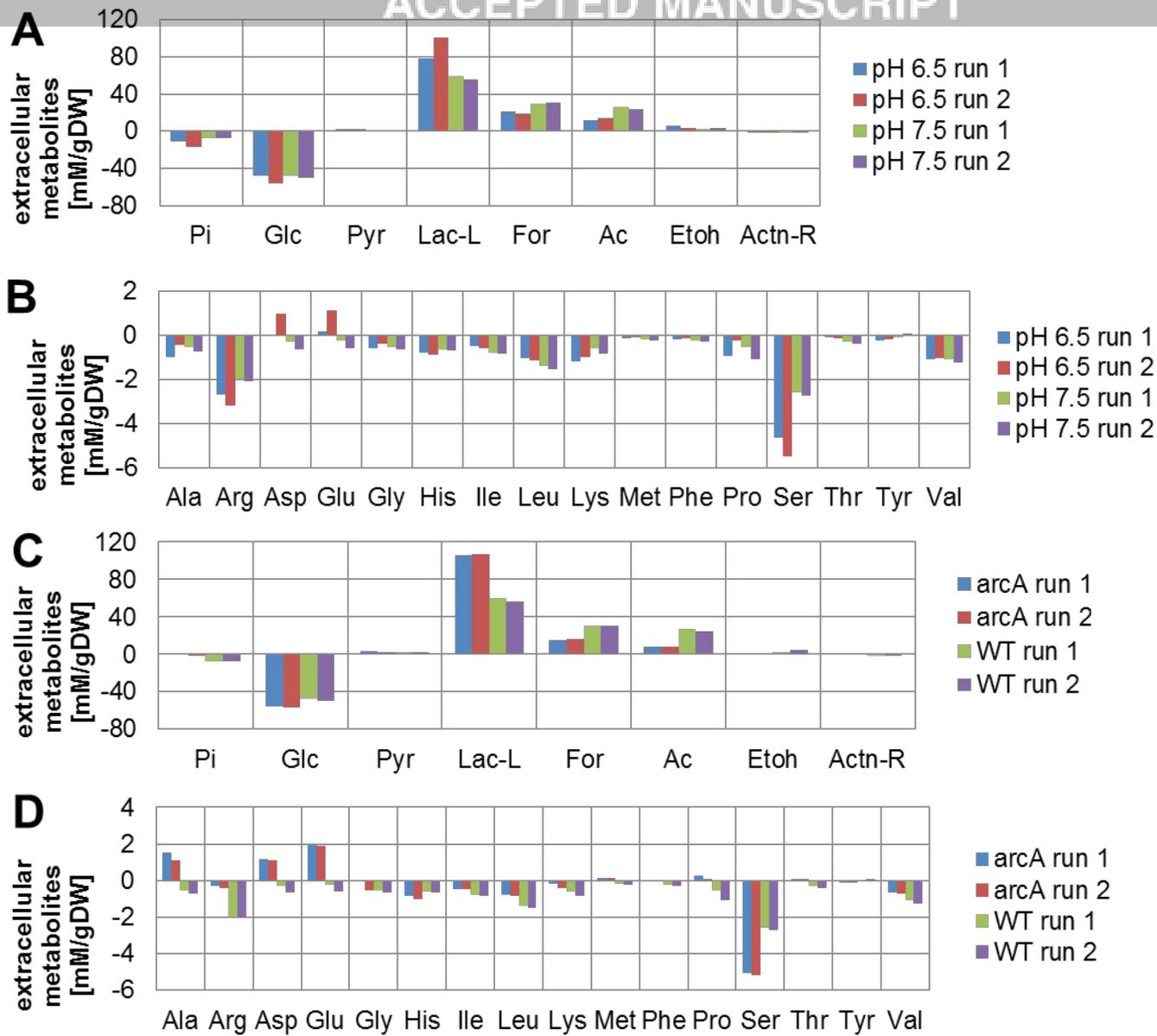
**Figure 1:** Experimentally measured metabolites consumed and produced by *S. pyogenes* wild type (A, B) and *arcA* knock-out (C, D) in duplicate. The cells were anaerobically grown in CDM-LAB in a bioreactor at two different pHs and a dilution rate of 0.05 per hour. The metabolite concentrations were normalized by the measured culture dry weight. Consumed metabolites have negative values, produced ones are represented by positive values.

**Figure 2:** Experimental determination of amino acid auxotrophies. Growth of *S. pyogenes* M49 in the absence of several single amino acids or amino acid combinations in CDM-LAB. (A) Optical densities after 24 h growth are shown. The y-axis gives information about the omitted amino acid(s). (B) Repeated re-inoculation of cultures grown in the absence of glycine or serine alone or in combination, respectively, lead to adaptation to the omissions and significant growth of the bacteria in the absence of these amino acids.

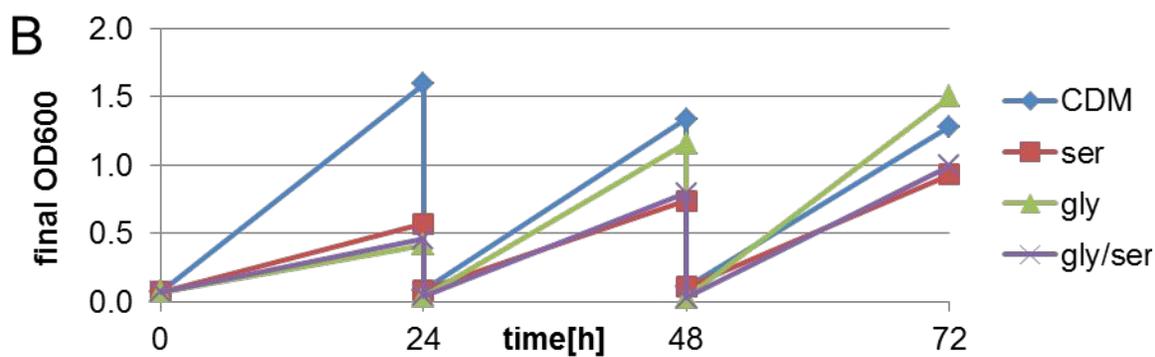
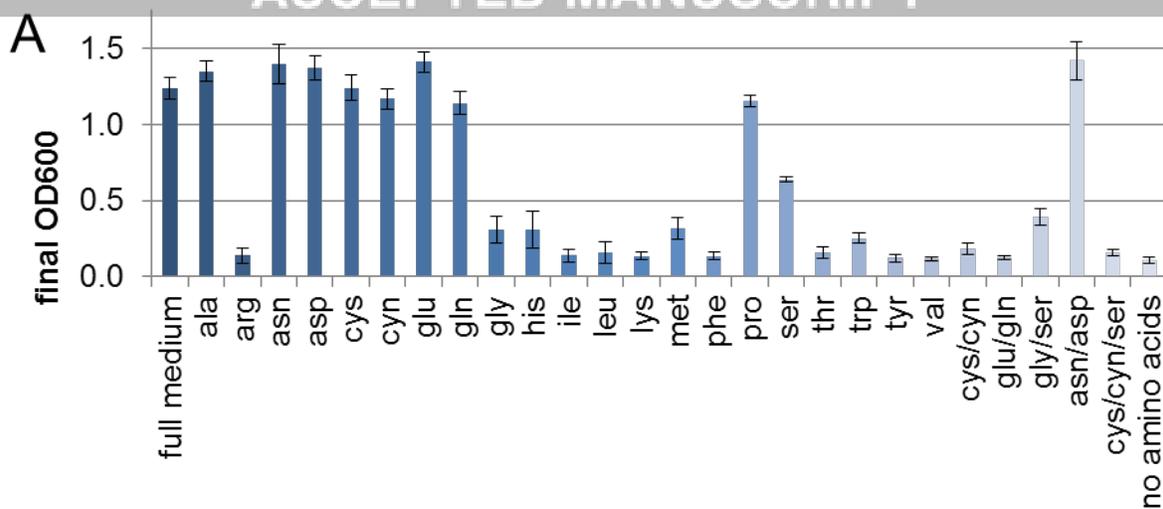
**Figure 3:** Amino acid conversions as predicted by the model and confirmed by our experimental data. Although the reaction transforming serine into cysteine (see dashed line) is capable of carrying flux under the simulated conditions, it does not in the absence of cysteine and cystine.

**Figure 4:** Growth of *S. pyogenes* M49 in CDM-LAB minimal variants. Optical density after 24 h growth in full CDM-LAB was set to 100% and all other values were related accordingly. Experimental data are shown in dark grey, model predictions are shown in light grey. Full=CDM-LAB; G=guanine; U=uracil; X=xanthine; non-essential vitamins=biotin, inosine, orotic acid, pyridoxamine, pyridoxine, riboflavin; Minimal=CDM-LAB w/o alanine, asparagine, aspartate, glutamate, glycine, proline, serine, acetate, thymidine, xanthine, uracil, bicarbonate, biotin, inosine, orotic acid, pyridoxamine, pyridoxine, riboflavin. Experimental data represent mean values and standard deviations of at least four independent experiments.

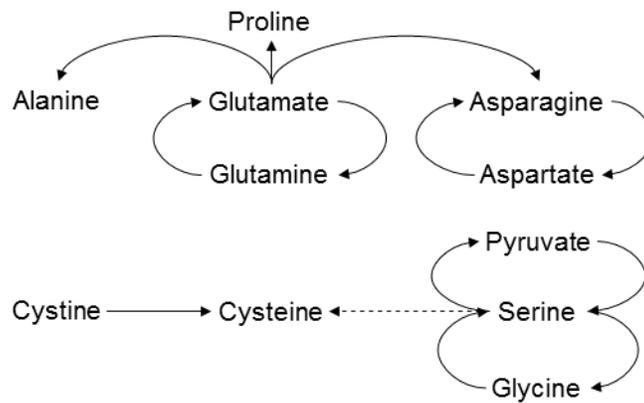
**Figure 5:** Comparison essential genes identified by Le Breton and coworkers (Le Breton et al., 2015) and by our computational approach. Panel A shows an Euler diagram of Le Breton's gene classification and the genes classified as essential by the genome-scale model. Panel B displays an Euler diagram of Le Breton's gene classification and the genes classified as non-essential by the genome-scale model. For each set, we give the number of genes intersecting with other sets and the number of genes without an overlap to any other set. Abbreviations: C, critical genes; E, non-essential genes; NC, non-conclusive genes; NE, non-essential genes; Model E/NE, essential/non-essential genes predicted by the genome-scale mode



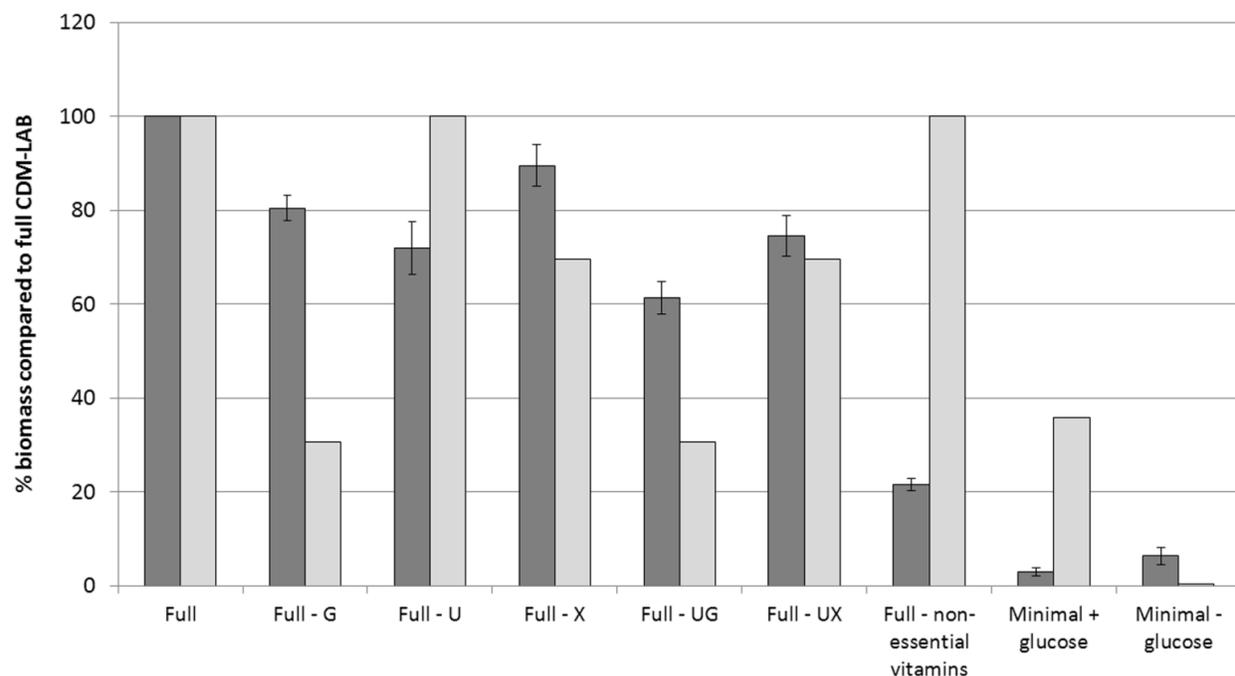
**Figure 1:** Experimentally measured metabolites consumed and produced by *S. pyogenes* wild-type at pH 6.5 and 7.5 (A, B) and *arcA* knock-out at pH 7.5 in comparison to wild-type at pH 7.5 (C, D) in duplicate. The cells were anaerobically grown in CDM-LAB medium in a bioreactor at two different pHs and a dilution rate of 0.05 per hour. The metabolite concentrations were normalized by the measured culture dry weight. Consumed metabolites have negative values, produced ones are represented by positive values.



**Figure 2:** Experimental determination of amino acid auxotrophies. Growth of *S. pyogenes* M49 591 in the absence of several single amino acids or amino acid combinations in CDM-LAB medium. (A) Optical densities after 24 h growth are shown. The y-axis gives information about the omitted amino acid(s). All ODs were averaged over three data sets, and the standard deviations are displayed. (B) Repeated re-inoculation of cultures grown in the absence of glycine or serine alone or in combination, respectively, lead to adaptation to the omissions and significant growth of the bacteria in the absence of these amino acids.

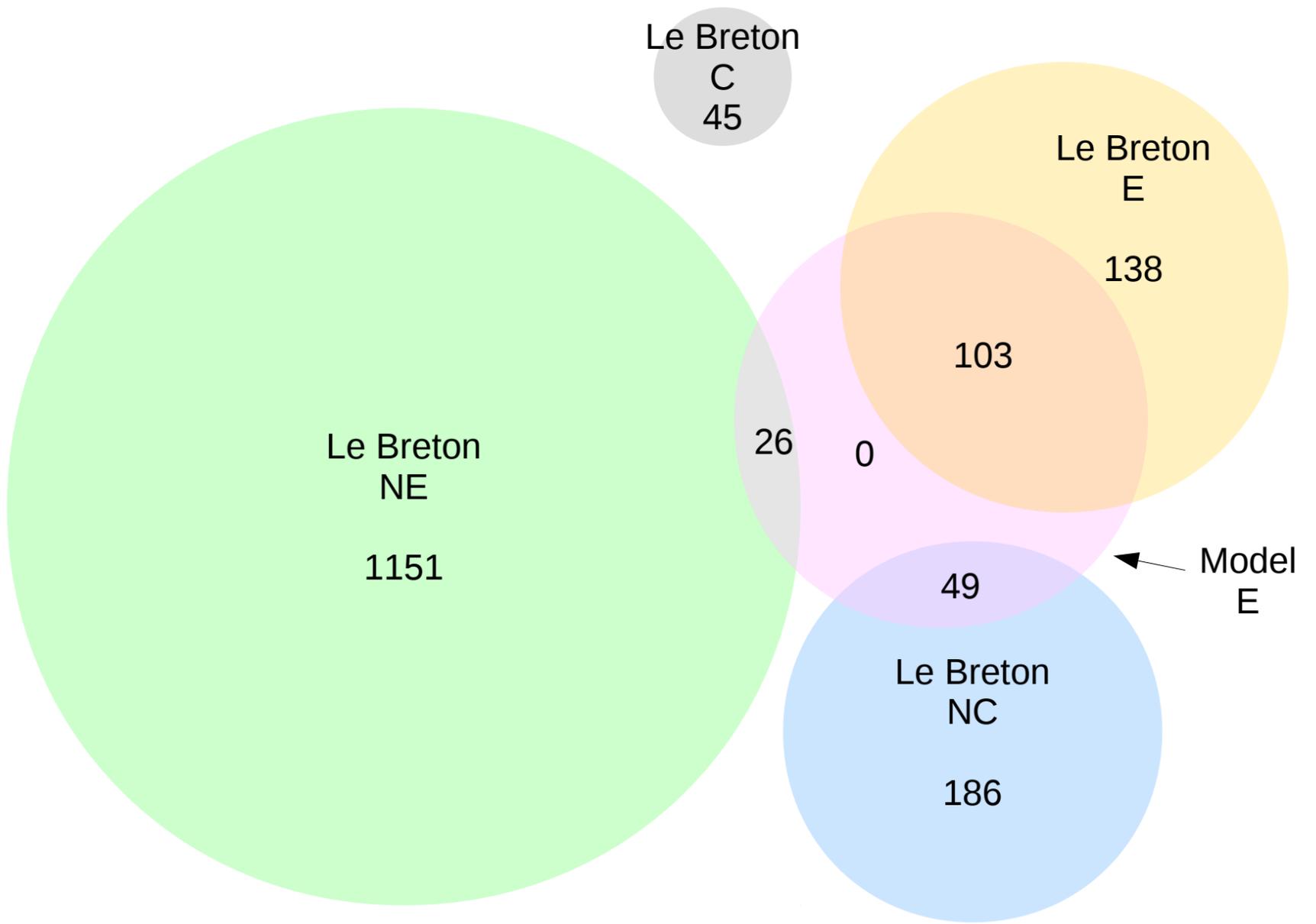


**Figure 3:** Amino acid conversions as predicted by the model and confirmed by our experimental data. Although the reaction transforming serine into cysteine (see dashed line) is capable of carrying flux under the simulated conditions, it does not in the absence of cysteine and cystine.

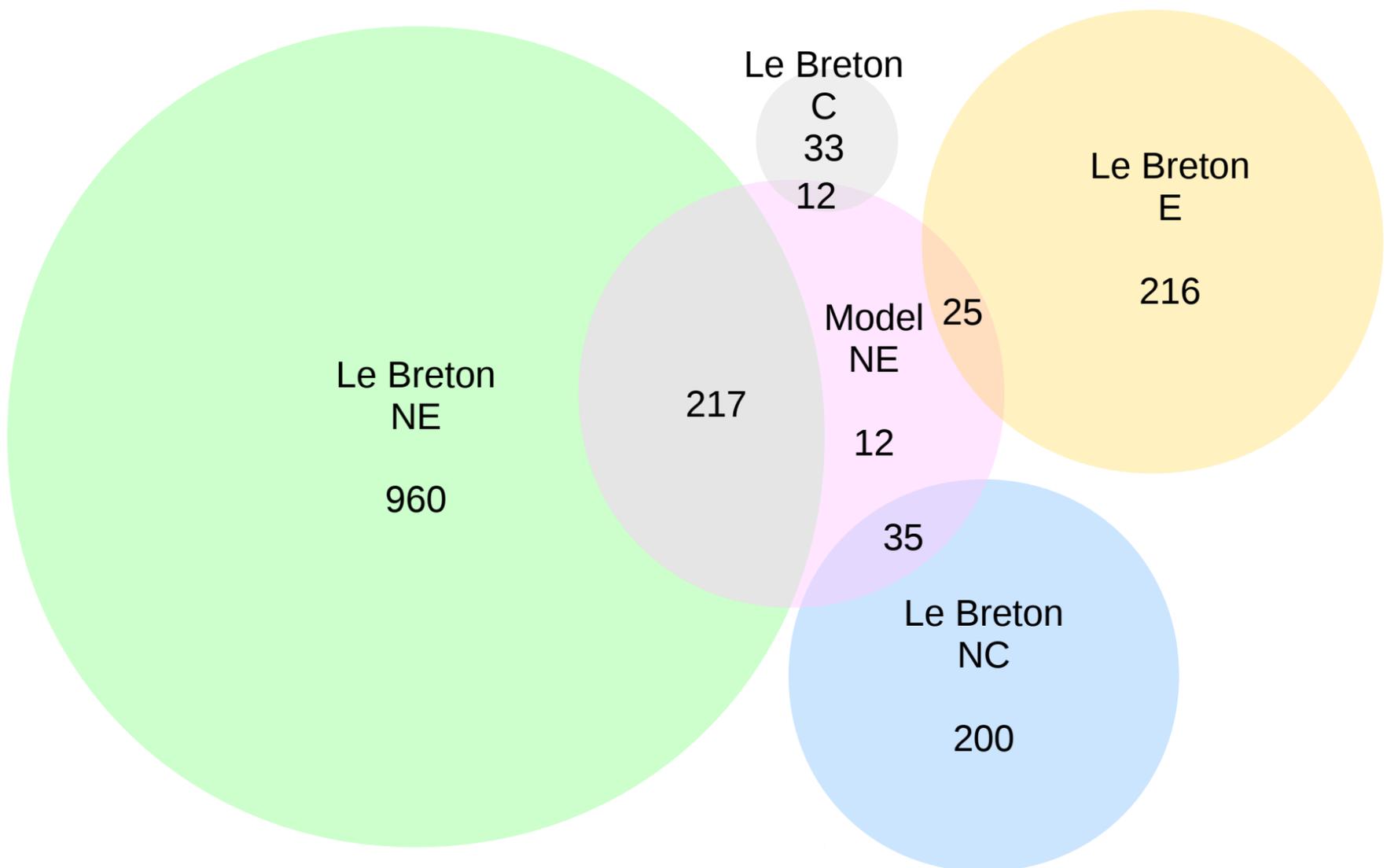


**Figure 4: Growth of *S. pyogenes* M49 in CDM-LAB minimal variants.** Optical density after 24 h growth in full CDM-LAB was set to 100% and all other values were related accordingly. Experimental data are shown in dark grey, model predictions are shown in light grey. Full=CDM-LAB; G=guanine; U=uracil; X=xanthine; non-essential vitamins=biotin, inosine, orotic acid, pyridoxamine, pyridoxine, riboflavin; Minimal=CDM-LAB w/o alanine, asparagine, aspartate, glutamate, glycine, proline, serine, acetate, thymidine, xanthine, uracil, bicarbonate, biotin, inosine, orotic acid, pyridoxamine, pyridoxine, riboflavin. Experimental data represent mean values and standard deviations of at least four independent experiments.

**A**



**B**



## Tables

**Table 1:** Characteristics of the *S. pyogenes* M49 genome-scale metabolic model in terms of genes, reactions and metabolites.

<b>Genes</b>	Genome (NCBI genome)	1788
	Model	480 (26.8%)
<b>Reactions</b>	Total	576
	Non-gene associated	61
	Exchange	69
	Transport	103
	Balanced	501
<b>Metabolites</b>	Total	558
	Intracellular	450
	Extracellular	108

**Table 2:** Reaction and metabolite characteristics of the *S. pyogenes* M49 genome-scale metabolic model for different constraints. We analysed the complete model without any flux constraints and the model used to simulate growth of wild type and *arcA* knock-out mutant on CDM-LAB at pH 6.5 and 7.5.

		<b>Complete network</b>	<b>WT pH 6.5</b>	<b>WT pH 7.5</b>	<b><math>\Delta arcA</math> pH 7.5</b>
<b>Reactions</b>	Blocked	182	199	190	201
	Essential	188	221	214	216
	Active	273	291	281	275
	Inactive	295	285	295	301
<b>Metabolites</b>	Gap	138	139	138	139

**Table 3:** Reactions involved in circulations. For each reaction involved in loops the model equation, the lower and upper boundaries, and the minimum and maximum possible flux through the reaction are given. Additionally, the last column gives the solution applied to the model constraints to prevent the circulation.

Reaction	Equation	Boundaries		FVA		Solution
		Lower	upper	min	max	
GK1	atp + gmp $\rightleftharpoons$ adp + gdp	-1000	1000	-1000	1000	LB = 0
GK2	datp + gmp $\rightleftharpoons$ dadp + gdp	-1000	1000	-1000	1000	LB = 0
NDPK8	atp + dadp $\rightleftharpoons$ adp + datp	-1000	1000	-1000	1000	LB = 0
DHORD1	dhor-S + o2 $\rightleftharpoons$ h2o2 + orot	-1000	1000	-1000	0	LB = 0
DHORD6	dhor-S + nad $\rightleftharpoons$ h + nadh + orot	-1000	1000	0	1000	
NOX1	h + nadh + o2 $\rightarrow$ h2o2 + nad	0	1000	0	1000	
LOXL	lac-L + o2 $\rightarrow$ h2o2 + pyr	0	1000	0	1000	
LDH_L	lac-L + nad $\rightleftharpoons$ h + nadh + pyr	-1000	1000	-1000	0	
G3PO	glyc3p + o2 $\rightarrow$ dhap + h2o2	0	1000	0	1000	
G3PD1	glyc3p + nad $\rightleftharpoons$ dhap + h + nadh	-1000	1000	-1000	0	
GARFT_met	methf + h2o + gar $\rightleftharpoons$ fgam + h + thf	0	1000	0	1000	
MTHFC	h2o + methf $\rightleftharpoons$ 10fthf + h	-1000	1000	-1000	0	LB = 0
GARFT	10fthf + gar $\rightleftharpoons$ fgam + h + thf	-1000	1000	-1000	0	

**Table 4:** Comparison of essential (A) and non-essential (P) amino acids for *S. pyogenes* between literature data (Slade, (Slade et al., 1951), our experimental data (Exp) and the genome-scale model predictions (GSM). Slade et al. performed single leave-out experiments but did not omit combination of amino acids at one time indicated by a blank field.

	ALA	ARG	ASN	ASP	CYS	CYN	GLN	GLU	GLY	HIS	ILE	LEU	LYS	MET	PHE	PRO	SER	THR	TRP	TYR	VAL	CYS/CYN	GLU/GLN	GLY/SER	ASN/ASP	CYN/CYS/SER
<b>Slade</b>	P	A	P	P	A	A	P	P	A	A	A	A	A	A	A	A	A	A	A	A	A					
<b>Exp</b>	P	A	P	P	P	P	P	P	P	A	A	A	A	A	A	P	P	A	A	A	A	A	A	P	P	A
<b>GSM</b>	P	A	P	P	P	P	P	P	P	A	A	A	A	A	A	P	P	A	A	A	A	A	A	P	P	A

**Table 5:** Minimal medium composition. Using FBA we categorized the CDM-LAB compounds as essential or non-essential. Some components are interconvertible, e.g. cystine and cysteine, and only one of these are required for *in silico* growth.

	<b>Essential compounds</b>	<b>Non-essential compounds</b>
<b>Amino acid mix</b>	Arginine	Alanine
	Cystine or cysteine	Asparagine
	Glutamine or glutamate	Aspartate
	Histidine	Cysteine or cystine
	Isoleucine	Glutamate or glutamine
	Leucine	Glycine
	Lysine	Proline
	Methionine	Serine
	Phenylalanine	
	Threonine	
	Tryptophan	
	Tyrosine	
	Valine	
<b>AGU mix</b>	Adenine	Uracil
	Guanine or xanthine	Xanthine or guanine
<b>Vitamins</b>	Aminobenzoate	Biotin
	Ascorbic acid	Inosine
	Nicotinic acid	Orotic acid
	Pantothenate	Pyridoxamine
	Thiamine	Pyridoxine
		Riboflavin
		Thymidine
<b>Other</b>	Citrate	Acetate
	Phosphate	Bicarbonate
	Ammonium or ammonia	Glucose
	Sulfate	
	Water	
	Protons	

**Table 6:** Essential gene analysis. To analyse essential genes within the metabolic network of *S. pyogenes* M49 we knocked-out each gene sequentially. We defined the effect of each gene on the metabolism based on the objective function value after gene knock-out compared to the biomass production flux before knock-out and categorized the genes into categories.

<b>Objective value in % of original value</b>	<b>Category</b>	<b>Quantity</b>
0 – 5 %	Lethal	179
5 – 95 %	Affected	12
95% – 105 %	Unaffected	289
> 105 %	Improved	0